Higher-order interactions shape microbial interactions as microbial community complexity increases

Manon A. Morin[1]*, Anneliese J. Morrison[2,3], Michael J. Harms[2,3], and Rachel J. Dutton[1,4*]

[1] University of California San Diego, La Jolla, 92093, USA

[2] Department of Chemistry and Biochemistry, University of Oregon, Eugene, OR, United States

[3] Institute of Molecular Biology, University of Oregon, Eugene, OR, United States

[4] Center for Microbiome Innovation, Jacobs School of Engineering, University of California San Diego, La Jolla, 92093, USA

[*] mamorin@ucsd.edu ; rjdutton@ucsd.edu

ABSTRACT

Non-pairwise interactions, or higher-order interactions (HOIs), in microbial communities have been claimed to explain the emergent features in microbiomes. Yet, the re-organization of microbial interactions between pairwise cultures and larger communities remains largely unexplored from a molecular perspective but is central to our understanding and further manipulation of microbial communities. Here, we used a bottom-up approach to investigate microbial interaction mechanisms from pairwise cultures up to 4-species communities from a simple microbiome (*Hafnia alvei*, *Geotrichum candidum*, *Pencillium camemberti* and *Escherichia coli*). Specifically, we characterized the interaction landscape for each species combination involving *E. coli* by identifying *E. coli's* interaction-associated genes using an RB-TnSeq-based interaction assay. We observed a deep reorganization of the interaction-associated genes, with very few 2-species interactions conserved all the way up to a 4-species community and the emergence of multiple HOIs. We further used a quantitative genetics strategy to decipher how 2-species interactions were quantitatively conserved in higher community compositions. Epistasis-

based analysis revealed that, of the interactions that are conserved at all levels of complexity, 82% follow an additive pattern. Altogether, we demonstrate the complex architecture of microbial interactions even within a simple microbiome, and provide a mechanistic and molecular explanation of HOIs.

Introduction

Microbiomes are multidimensional systems containing up to thousands of interacting species. In part due to this complexity, a common strategy to investigate their biology has been to use bottom-up or reductionist approaches. Using *in vitro* microbial communities or monocultures, the features of microbial communities are measured in low-dimensional settings, with the underlying objective to understand whether high-dimension phenotypes can be predicted from lower-level observations. The lack of predictability of microbial community features from monoculture observations has demonstrated the critical importance of microbial interactions in shaping microbial communities. For instance, community assembly[1,2], community function[3], community resistance to invasion[4] or its effects on its host[5] are highly different from the simple combination of individual species effects. Some of these studies have also shown that, while 2-species, or pairwise culture, observations could partially predict what happens in higher communities[1,2,6], context-specific phenotypes emerge as the microbial community becomes more complex, limiting the use of pairwise information to obtain a descriptive picture of larger communities. For instance, the function (amylolytic activity) of a soil-associated microbial community of 6 species has been shown to differ from the simple linear combination (*i.e*. the addition) of the observed function of pairwise cultures[3]. Similarly, 13 to 43% of measured *Drosophila melanogaster* traits (lifespan, reproduction and development) in flies carrying a microbiome of 5 species were not predictable from the associated traits from flies inoculated with pairwise combinations of the same 5 species[5]. In the zebrafish gut microbiome, while strong negative pairwise interactions were measured, the assembly of a more complex microbiome of 5 species strikingly diverged from any assembly

prediction based on pairwise interactions, and all 5 species actually co-occur in the final microbiome[7].

Reorganization of the interaction profile with varying community complexity and the presence of higher-order interactions (HOIs) are the most likely explanation for the lack of predictability of complex community phenotypes from pairwise observations. In ecology, HOIs are described as key to stabilizing communities and promoting biodiversity in biological ecosystems [8–10] They are commonly described as the modification of pairwise interactions when another species is introduced, or any interaction that cannot be described by a pairwise model [11,12]. For instance, while the bacterium *Escherichia coli* can successfully invade cultures of *Chlamydomonas reinhardtii* as well as cultures of *Tetrahymena thermophila*, it is unable to invade a coculture of the alga and the ciliate. This is due to *C. reinhardtii* inhibition of *E. coli* aggregation specifically in the presence of *T. thermophila*, which renders the bacterium vulnerable to predation[13]. Changes in microbial interactions and HOIs are undoubtedly a significant feature of microbiome ecology and functioning. Thus, investigating HOIs and deciphering how microbial interactions are rearranged when microbial systems increase in complexity are essential to our understanding of microbiomes and our ability to manipulate them.

In this work, we investigate how interactions are reorganized when the complexity of a microbial community increases from two to four species. We use a simplified cheese rind microbiome composed of the two gamma-proteobacteria *E. coli* and *Hafnia alvei*, the yeast *Geotrichum candidum* and the filamentous fungus *Penicillium camemberti*. Our previous work comparing the pairwise interaction patterns of this model microbiome to the interaction pattern in the full community highlighted the prevalence of HOIs and the lack of conservation of pairwise interactions[14]. However, we couldn't precisely resolve the origin of these HOIs or specific rules underlying the emergence of these HOIs. For instance, we couldn't determine whether 2-species interactions are alleviated by the introduction of a specific species or by the introduction of all the other species. Similarly, we couldn't conclude whether community-specific interactions are

77 actually specific to the community or whether they arise at an intermediate level and
78 are maintained in the whole community. Relying on the ability to deconstruct and reconstruct this
79 model system and on the RBTnSeq-based interaction assay we have previously optimized to
80 compare gene fitness values across multiple conditions[15], we aim to identify quantitative changes
81 in gene fitness values for *E. coli* in different interactive conditions to identify the genetic basis of
82 interactions at every level of community complexity.
83 Comparing interaction-associated genes across conditions, we can identify pairwise interaction-
84 associated genes that are also observed at higher levels of complexity as well as genes
85 associated with HOIs. Here, HOIs are defined as interaction-associated genes that emerge in the
86 3 or 4 species cultures and 2-species interaction-associated genes that are no longer observed
87 in the 3 or 4 species cultures. Analysis of the genes associated with HOIs and their functions
88 allowed us to further characterize the deep reorganization of the interaction landscape and
89 highlights that it is mostly associated with the reprogramming of metabolic interactions and the
90 introduction of a fungal partner. We further focus on the genetic basis of 2-species interactions
91 that are conserved in higher levels of complexity to elucidate principles behind interaction
92 conservation. We use an epistasis and quantitative genomics approach[16,17] to understand
93 whether interactions that are conserved follow a linear, or additive, pattern. The evaluation of
94 interaction effects as quantitative traits allows us to define another form of HOIs as cases in which
95 there is a lack of additivity in interactions that are conserved from simpler to more complex
96 community composition. This form of HOIs is consistent with a more quantitative definition of
97 HOIs, highly similar to the definition of epistasis in population genetics[18], that identify HOIs (or
98 epistasis) as any deviation, for a given quantitative trait, from the prediction of a linear model
99 where only pairwise interactions are included. Carrying out this analysis, we observe that 82% of
100 the conserved interactions follow an additive pattern of conservation from 2-species to 4-species,
101 and that 18% of the conserved 2-species interaction is associated with non-linear models of
102 conservation.

103 Overall, our work provides a unique illustration of the highly complex reorganization of interaction
104 mechanisms when microbial community complexity changes. This provides a mechanistic
105 explanation of HOIs in microbial communities that is essential for the further global understanding
106 of microbial communities.
107
108 **Results**
109
110 **Sets of interaction-associated genes change across interactive conditions**
111 To investigate how microbial interactions are reorganized in a microbial community with
112 increasing complexity, we reconstructed *in vitro* a modified bloomy rind cheese-associated
113 microbiome on Cheese Curd Agar plates (CCA plates) as described in our previous work[14]. The
114 original community is composed of the gamma-proteobacterium *H. alvei*, the yeast *G. candidum*
115 and the mold *P. camemberti.* Using a barcoded transposon library of the model bacterium *E. coli*
116 as a probe to identify interactions, we investigated microbial interactions in 2-species cultures (*E.*
117 *coli* + 1 community member), in 3-species cultures (*E. coli* + 2 community members) and in 4-
118 species cultures (or whole community: *E. coli* + 3 community members) (Figure 1A).
119 Quantification of species' final CFUs after 3 days of growth highlighted consistent growth for *H.*
120 *alvei* and *G. candidum* independent of the culture condition and slightly reduced growth for *E. coli*
121 in interactive conditions compared to growth alone except for following growth with *P. camemberti*
122 (Supplementary Figure 1). Although we were unable to quantify spores of *P. camemberti* after
123 three days, growth of *P. camemberti* was visually evident in all of the expected samples.
124 Previously, we developed an assay and a pipeline to identify microbial genes associated with
125 interactions by adapting the original RB-TnSeq approach[19] to allow for consistent implementation
126 of biological replicates as well as for direct quantitative comparison of fitness values between
127 different culture conditions[15]. More specifically, the original RBTnSeq assay relies on the use of
128 a pooled library of randomly barcoded transposon mutants of a given microorganism (RBTnSeq

129 library)[19]. Measuring the variation of the abundance of each transposon mutant before and after
130 growth, the pipeline allows the calculation of a fitness value for each mutant. A negative fitness
131 indicates decreased growth of the mutant relative to a wild type strain, whereas a positive fitness
132 value indicates increased growth in the studied condition. Then, to identify genes associated with
133 interactions, we measure and compare gene fitness across the different studied conditions, for
134 example, comparing growth alone to growth in the presence of another species. Then, any gene
135 whose fitness values significantly change between such conditions is identified as an interaction-
136 associated gene.
137 In this work, we used the *E. coli* RBTnSeq Keio_ML9 library[19] and grew it for 3 days alone or in
138 the seven different interactive conditions studied here (Figure 1A). For each interactive condition,
139 we calculated the Interaction Fitness Effect (IFE) for 3699 *E. coli* genes as the difference between
140 the gene fitness in the studied interactive condition and the gene fitness in growth alone
141 (Supplementary Data 1). To identify genes associated with interactions, we then tested for all the
142 IFEs that are significantly different from 0 (adjusted p-value ≤ 0.1; two-sided t-test and Benjamini-
143 Hochberg correction for multiple comparison[20]). Negative IFE occurs when gene fitness
144 decreases in the interactive condition, and positive IFE occurs when gene fitness improves in the
145 interactive condition. Here, we identified between 6 (with *P. camemberti*) and 71 (with *H. alvei* +
146 *P. camemberti*) significant IFEs per condition (Figure 1B). Both negative IFEs and positive IFEs
147 were found in each interactive condition except for the 2-species culture with *P. camemberti*,
148 where only negative interactions were identified. A total of 330 significant IFEs associated with
149 218 unique genes were identified (as the same gene can have a significant IFE in multiple
150 conditions) including 125 genes associated with negative IFE and 120 genes associated with
151 positive IFE (Supplementary Figures 2 and 3).
152
153 To gain insight into the interaction mechanisms among microbes, we next analyzed the functions
154 associated with IFEs. Here, the vast majority of genes associated with significant IFEs are part of

155 an interaction network, highlighting the presence of genes with connected functions and from
156 similar pathways (Figure 1C). A significant fraction of the genes associated with a negative IFE
157 are part of amino acid biosynthesis and transport (17% - Figure 1C and Supplementary Figures
158 2 and 4), and more specifically with histidine, tryptophan and arginine biosynthesis. This points to
159 competition for these nutrients between *E. coli* and the other species. Another large set of genes
160 is related to nucleotide metabolism and transport (14% - Figure 1C and Supplementary Figures
161 2 and 5), highlighting competitive interactions for nucleotides and/or their precursors. The majority
162 of these genes relate to purine nucleotides and more specifically to the initial steps of their *de*
163 *novo* biosynthesis associated with the biosynthesis of 5-aminoimidazole monophosphate (IMP)
164 ribonucleotide. Of the genes with a positive IFE, 15% are related to amino acid biosynthesis and
165 transport (Figure 1C and Supplementary Figures 3 and 4), suggesting cross feeding of amino
166 acids between *E. coli* and the other species. More specifically, this includes phosphoserine,
167 serine, homoserine, threonine, proline and arginine. The presence of amino acid biosynthetic
168 genes among both negative and positive IFEs indicate that trophic interactions (competition
169 versus cross-feeding) depend on the type of amino-acid and/or the species interacting with *E.*
170 *coli*. For both negative and positive IFEs, numerous genes were annotated as transcriptional
171 regulators (Figure 1C and Supplementary Figures 2 and 3) emphasizing the importance of
172 transcriptional reprogramming in response to interactions. These transcriptional regulators
173 include metabolism regulators as well as regulators of growth, cell cycle and response to stress.
174 Finally, these interaction-associated genes and these interaction mechanisms are consistent with
175 previous findings in this microbiome[14] as well as in a study of bacterial-fungal interactions
176 involving *E. coli* and cheese rind isolated fungal species[15].
177
178 **Introduction of a third-interacting species deeply reshapes microbial interactions**
179 The differences in the number and sign of significant IFEs observed among the different
180 interactive conditions, with different numbers of interaction species, suggest that the number and

181 type of interacting partners influence interaction mechanisms. To characterize how the
182 interactions are reorganized with community complexity, we then investigated if and how the
183 genetic basis of interactions changes when the number of interacting partners increases by
184 comparing the genes associated with significant IFE in 2-species cultures, in 3-species cultures
185 and then in 4-species cultures.
186 First, we have identified 104 IFEs associated with 98 genes in 2-species cultures as well as 168
187 IFEs associated with 136 unique genes in 3-species conditions (Supplementary Figure 6 and
188 Supplementary Data 2). Comparing these gene sets, we can identify how the interaction-
189 associated genes change when a third-species is added to a 2-species culture. We identified 45
190 2-species interaction-associated genes maintained in at least one 3-species condition
191 (maintained interaction-genes), 55 2-species interaction-associated genes no longer associated
192 with interaction in any 3-species condition (dropped interaction-genes) and 100 3-species
193 interaction-associated that aren't associated with any 2-species interaction-associated genes
194 (emergent interaction-genes) (Figure 2A, Supplementary Figure 6 and Supplementary Data 3).
195 Both dropped and emerging interaction-associated genes represent 3-species HOIs; the third
196 species either removes an existing interaction or brings about a new one.
197 We further carried out functional analysis of maintained, dropped and emerging interaction-genes
198 to elucidate whether maintained and HOIs interaction-genes would be associated with specific
199 functions and thus interaction mechanisms (Figure 2B). For each set of genes, we calculated the
200 fraction of genes of that set associated with a given COG ontology category. Metabolism and
201 transport is the most observed COG group (Figure 2B - teal dots). For maintained interaction-
202 genes, this indicates that some trophic interactions can be maintained from 2-species to 3-species
203 conditions. For instance, serine biosynthetic genes *serA*, *serB* and *serC* as well as threonine
204 biosynthetic genes *thrA*, *thrB* and *thrC* are associated with positive IFEs in the 2-species condition
205 with *G. candidum* as well as in the 3-species conditions involving *G. candidum* (Supplementary
206 Figure 4). This suggests that, (i) *G. candidum* facilitates serine and threonine cross feeding and

207 (ii) this cross-feeding is still observed when another species is introduced. However, metabolism-
208 related genes identified among the dropped and emerging interaction-agenes indicate that many
209 trophic interactions are also rearranged through HOIs. Genes associated with lactate catabolism
210 (*lldP* and *lldD*) and lactate metabolism regulation (*lldR*) have a negative IFE in the 2-species
211 culture with *H. alvei*, suggesting competition for lactate between *E. coli* and *H. alvei*. Yet, these
212 genes are no longer associated with a significant IFE when at least another partner is introduced
213 (Supplementary Figure 7). Histidine biosynthesis genes *hisA*, *hisB*, *hisD*, *hisH* and *hisI* have a
214 negative IFE in the 2-species culture *with H. alvei* and sometimes in the 3 species culture with *H.*
215 *alvei* + *P. camemberti*. However, the negative IFE is alleviated whenever *G. candidum* is present,
216 suggesting that potential competition for histidine between *E. coli* and *H. alvei* is alleviated by this
217 fungal species (Supplementary Figure 4). Also, genes related to the COG section "Information
218 storage and processing" are mostly found among HOIs genes, suggesting a fine-tuning of specific
219 cellular activity depending on the interacting condition. For instance, we identified many
220 transcriptional regulators of central metabolism among the dropped genes (*rbsR* and *lldR*) and
221 the emerging genes (*purR*, *puuR*, *gcvR* and *mngR*), highlighting again the reorganization of
222 trophic interactions associated with HOIs. Also, many transcriptional regulators broadly
223 associated with growth control, cell cycle and response to stress were found among the emerging
224 interaction-genes with 3-species (*hyfR*, *chpS*, *sdiA*, *slyA* and *rssB*), underlining a noticeable
225 modification of *E. coli's* growth environment with 3-species compare to with 2-species.
226 Finally, we further aimed to understand whether HOIs are associated with the introduction of any
227 specific species (Figure 2C and Supplementary Figure 8). We observe that interaction-associated
228 genes with *H. alvei* are more likely to be dropped, as 65% of them are alleviated by the
229 introduction of a fungal species (Figure 2C). This can be seen, for instance, with the
230 reorganization of *E. coli* and *H. alvei* trophic interactions following the introduction of *G. candidum*
231 (alleviation of lactate and histine competition for instance). Also, we observe that 76% of the
232 interactions in the 3-species cultures with *H. alvei* + *P. camemberti* and 65% in the 3-species

233 culture with *H. alvei* + *G. candidum* are emerging genes (compared to 38% of emerging
234 interaction-associated genes in the 3-species condition with *G. candidum* + *P. camemberti*)
235 (Figure 2C). For the 3-species with *H. alvei* + *P. camemberti*, they include for instance the genes
236 associated with purine *de novo* biosynthesis (*purR*, *purF*, *purN*, *purE*, *purC*) and the genes
237 associated with pyrimidine *de novo* biosynthesis (*pyrD*, *pyrF*, *pyrC*, *carA* and *ulaD*), suggesting
238 important trophic HOIs. For the 3-species with *H. alvei* + *G. candidum*, emerging interaction genes
239 include for example the transcriptional regulators *chpS*, *sdiA* and *slyA*, indicating the presence of
240 a stress inducing environment. Together, these observations suggest that the introduction of a
241 fungal partner may introduce multiple 3-species HOIs by both canceling existing interactions and
242 introducing new ones.
243
244 **HOIs are prevalent in a 4-species community**
245 To further decipher whether microbial interactions continue to change with increasing community
246 complexity, we investigated the changes in the genetic basis of interactions going from 3-species
247 to 4-species experiments. We identified 58 interaction-genes in the 4-species condition (*E. coli*
248 with *H. alvei* + *G. candidum* + *P. camemberti*), compared with 145 genes associated with
249 interactions in any 3-species conditions. Comparing these two sets of genes we identify: 26 3-
250 species interaction-genes that are maintained in the 4-species condition (including 16 directly
251 from 2-species interactions), 115 3-species interaction-genes that are no longer associated with
252 interactions in the 4-species condition (dropped interaction-genes) and 32 interaction-genes that
253 are observed solely in the 4-species condition (emerging interaction-genes) (Figure 3A,
254 Supplementary Figure 6 and Supplementary Data 3). Both dropped and emerging interaction-
255 genes represent 4-species HOIs. Here, HOIs are remarkably abundant when introducing a single
256 new species and moving up from 3-species interactions to 4-species interactions. Functional
257 analysis of maintained and HOI genes reveals the presence of many metabolism related genes
258 in every gene set (Figure 3), suggesting that some trophic interactions can be maintained from 3-

259 species to 4-species interactions while some other trophic interactions are rearranged with HOIs.
260 For instance, most of the genes of the initial steps of *de novo* purine biosynthesis have been
261 found to have a negative IFE in the 3 species condition with *H. alvei* + *P. camemberti* (*purC*, *purE*,
262 *purF*, *purL* and *purN*) as well as in the pairwise condition with *H. alvei* for *purH* and *purK*
263 (Supplementary Figure 5), suggesting competition for purine initial precursor IMP in these
264 conditions. Yet, the introduction of the yeast *G. candidum* as a fourth species cancels the negative
265 IFE value, suggesting that the competition is no longer happening in its presence. Altogether, the
266 observation of noticeable trophic HOIs moving up from 2 to 3 species and then from 3 to 4-species
267 interaction highlights a consistent reorganization of trophic interactions along with community
268 complexity. Also, genes related to Cell wall/membrane/envelope biogenesis are found abundantly
269 among the 4-species emerging genes (Figure 3B) and they represent the largest functional
270 fraction of this gene set. These genes have a negative IFE and are related to Enterobacterial
271 Common Antigen (ECA) biosynthetic processes (*wecG*, *wecB* and *wecA*) (Supplementary Figure
272 9). While the roles of ECA can be multiple but are not well defined[21], they have been shown to be
273 important for response to different toxic stress, suggesting the development of a specific stress in
274 the presence of the four species.
275 As for the 2 to 3 species comparison, we investigated whether the introduction of a specific fourth
276 species would be most likely associated with HOIs. The 3-species culture that appears to be the
277 least affected by the introduction of a fourth member is with *G. candidum* + *P. camemberti* where
278 34% of the observed interactions are still conserved in the 4-species condition after the
279 introduction of *H. alvei* (versus 22% for with *H. alvei* + *G. candidum* when *P. camemberti* is added
280 and 21% for with *H. alvei* + *P. camemberti* when *G. candidum* is added) (Figure 3C and
281 Supplementary Figure 10). Together, these observations suggest that, again, the introduction of
282 a fungal partner may introduce multiple 4-species HOIs.
283 Finally, by increasing the number of interacting species in our system and investigating interaction
284 gene maintenance and modification with every increment of community complexity, we are able

285 to build our understanding of the architecture of interactions in a microbial community. Altogether,
286 we have observed a total of 218 individual genes associated with interactions in any experiment.
287 Only 16 of them (7%) were conserved across all levels of community complexity (Figure 3D).
288 Starting from 2-species interaction genes, 48% of them were maintained with 3-species and only
289 15% (16 out of 104) were still maintained with 4-species. Thus, we demonstrate here a
290 progressive loss and replacement of 2-species interactions as community complexity increases
291 and the prevalent apparition of HOIs. Tracking back the origins of the genetic basis of interactions
292 in the 4-species experiment that represents the full community of our model, we identify that 28%
293 of the full community interactions can be traced back to 2-species interactions, 18% are from 3-
294 species interaction and 54% are specific to the 4-species interaction (Figure 3D and 3E). Most of
295 the maintained interaction-genes from 2-species as well as from 3-species are associated with
296 metabolism (Figure 3D and Supplementary Figure 11) while Signal transduction and cell
297 membrane biosynthesis genes are most abundant among the 4-species interaction-genes as
298 previously mentioned. To conclude, this shows that the genetic basis of interactions and thus the
299 sets of microbial interaction are deeply reprogrammed at every level of community complexity
300 and illustrates the prevalence of higher order interactions (HOIs) even in simple communities.
301
302 **The majority of maintained 2-species interaction-genes in the 4-species culture follows an**
303 **additive conservation behavior**
304 While HOIs are abundant in the 4-species condition, our data yet suggest that up to 28% of the
305 interactions are maintained from 2-species interactions. However, we don't know whether and
306 how 2-species interactions are quantitatively affected by the introduction of other species and
307 whether they would follow specific quantitative models of conservation. For instance, we can
308 wonder how the strength of a given 2-species interaction is modified by the introduction of one or
309 two other species, or how two 2-species interactions associated with the same gene will combine
310 when all the species are present. In other words, can we treat species interactions as additive

311 when we add multiple species? Such information would generate a deeper mechanistic
312 understanding of the architecture of microbial interactions while allowing us to potentially predict
313 some whole community interactions from 2-species interactions. Here, two main hypothetical
314 scenarios can be anticipated. First, the conservation of 2-species interactions follows a linear or
315 additive behavior, where the introduction of other species either doesn't affect the strength of the
316 conserved 2-species interaction or two similar 2-species interactions combine additively. The
317 second scenario identifies non-linear or non-additive conservation of 2-species interactions,
318 where the strength of the conserved 2-species interaction is modified by the introduction of other
319 species or two similar 2-species interactions are not additive. The second scenario would
320 encompass for instance synergistic effects or inhibitory effects following the introduction of more
321 species. We next use an epistasis and quantitative genomics approach to understand whether
322 interactions that are conserved follow a linear, or additive, pattern. For the 16 genes that are
323 associated with interaction in 2-species cultures, in associated 3-species cultures and in the 4-
324 species condition, we use epistasis analysis to test the linear behavior of IFE when the number
325 of interacting species increases, as IFEs are quantitative traits related to the interaction strength.
326 In multi-dimensional systems, an epistasis analysis quantifies the additive (or linear) behavior of
327 conserved quantitative traits. In quantitative genetics, for instance, epistasis measures the
328 quantitative difference in the effects of mutations introduced individually versus together[18,22,23].
329 Using a similar rationale, we can use IFEs as a quantitative proxy for interaction strength and test
330 whether the IFEs of the maintained interaction genes in 3-species and in 4-species conditions
331 result from the linear combination of associated 2-species IFEs (Figure 4A). Nonlinear
332 combination, or non-additivity of 2-species IFEs in higher community level also highlights higher-
333 order interactions.
334 We adapted the pipeline Epistasis[17], originally designed for quantitative genetics investigation.
335 We implemented the linear model with the gene fitness values for growth alone, for each of the
336 2-species conditions, for each of the 3-species cultures and for the 4-species condition. For each

337 gene, the software finds the simplest mathematical model that reproduces the observed IFEs
338 across all levels of community complexity. In the simplest case, the model will have a term
339 describing the effects for adding each species individually to the *E. coli* alone culture; that term
340 corresponds to the 2-species IFE. Then, if the IFE for two *E. coli*'s partners combined (3-species
341 IFE) differs from the sum of their individual effects (corresponding 2-species IFE), the software
342 adds a term capturing this epistasis (Figure 4A). Here, we call that term 3-species epistatic
343 coefficient or $\varepsilon_{i,j}$. Finally, if the IFE for the combined community (*E. coli* plus all three species; 4-
344 species condition) differs from the prediction based on the 2-species and 3-species terms, the
345 software will add a high-order interaction term to the model (Figure 4B). Here, we name that term
346 4-species epistatic coefficient or ε_{ijk}.
347
348
349 We performed this analysis on the 16 genes that are associated with interactions at every level
350 of community complexity. To identify real additive behavior of IFE from non-additivity, we screen
351 for 3-species epistatic coefficients and 4-species epistatic coefficients that are significantly
352 different from 0 (adjusted p-value ≤ 0.01, correction for multiple testing Benjamin-Hochberg). We
353 found that 13 genes behaved additively from 2-species to 4-species culture, with no epistatic
354 contributions in the 3-species conditions nor in the 4-species condition (Figure 4C, *(i)*). One gene
355 (*gadW*) exhibited nonlinear conservation of IFE only in the 4-species condition, but additive IFE
356 conservation from 2-species to 3-species (Figure 4C, *(ii)*). Another gene (*lsrG*) showed epistasis
357 in one 3-species condition but no epistasis in the 4-species condition (Figure 4C, *(iii)*) Finally, one
358 gene (*gltB*) displayed both non-additivity in 3-species and 4-species conditions (Figure 4C, *(iv)*).
359 If we look more closely at the additive genes, we find genes (*betA*, *betT*, *purD* and *purH*) are
360 associated with the conservation of negative IFEs (Supplementary Figure 12). While *betA* and
361 *betT* are associated with choline transport (*betT*) and glycine betaine biosynthesis from choline
362 (*betA*)[24], *purD* and *purH* are associated with *de novo* purine biosynthesis[25]. This suggests that

363 requirements for glycine betaine biosynthesis from choline and for purine biosynthesis caused by
364 microbial interactions, possibly due to competition for the nutrients used as precursors, are
365 additively conserved from individual 2-species interactions requirements. Also, 5 genes
366 associated with amino acid biosynthesis (*serA*, *thrC, cysG, argG* and *proA*) are associated with
367 the additive conservation of positive IFE (Supplementary Figure 12), suggesting that cross
368 feeding can be additive when the community complexity increases. Altogether, this highlights the
369 existence of 2-species interactions, including trophic ones, conserved in an additive fashion in the
370 highest-level of complexity.
371 This leaves 3 genes (18%) of the maintained 2-species interaction-associated genes that are
372 associated with non-additive behavior, and thus HOIs, at at least one higher level of community
373 complexity (Figure 4C - *(ii)*, *(iii)* and *(iv)*). The gene *gadW* is associated with non-additivity at the
374 4-species level, suggesting that while IFEs are additive in 3-species cultures, the introduction of
375 a fourth species introduces HOI. Moreover, the observed 4-species IFE is greater than the IFE
376 predicted by a linear model (Figure 4D), highlighting a potential synergistic effect when the 4
377 species are together. The gene *lsrg* is associated with non-additivity only at the 3-species culture
378 w *G.c* + *P.c*. More specifically, this indicates that HOI arise when these 2 fungal species are
379 interacting together with *E. coli,* but that no more HOI emerge when *H. alvei* is introduced (ie, the
380 4-species IFE can be predicted by the linear combination of the lower levels IFEs). As the
381 observed IFE for the 3-species condition w *G.c* + *P.c* is greater than the predicted IFE (Figure
382 4C), this suggests a synergistic effect between the 2 fungal species. Finally, the gene *gltB* is
383 associated with non-additivity at both the 3-species and 4-species levels. For this gene, the
384 conservation of IFE is never associated with an additive model. Here, the observed 4-species IFE
385 is not as negative as it would be as the result of the linear combination of the associated lower
386 IFE (Figure 4D), suggesting the existence of a possible IFE threshold, or plateau effect.
387 Altogether, this indicates that maintained 2-species-interactions can follow nonlinear behaviors
388 that could involve synergistic effects, inhibitory effects or constraints.

389

390 **Discussion**

391 Interactions between microbes are responsible for the specific and multiple phenotypes observed
392 in microbial communities compared to monocultures. Knowledge of microbial interactions could
393 be the key to controlling microbial communities, but deciphering these interactions is challenging
394 in complex microbiomes. Moreover, pairwise culture interactions are often insufficient to predict
395 what happens in more complex microbiomes, suggesting an important reprogramming of
396 interactions as community complexity increases in terms of the number of species present[3,5,7,13,14].
397 Understanding the restructuring of these interactions in complex communities is thus essential to
398 comprehending the biology of microbial communities. Using an *in vitro* multi-kingdom community,
399 we performed a molecular investigation of the reorganization of interaction profiles as community
400 complexity changed. Relying on the tractability of our model system and an RBTnSeq-based
401 interaction assay, we tracked interactions in 2-species, 3-species and 4-species cultures. In this
402 work, the combination of a qualitative and quantitative comparison of interaction profiles at the
403 molecular level underlines the complex dynamics of interaction reorganization with community
404 complexity and the existence of multiple forms of higher-order interactions.

405 This work offers an example of the different forms that HOIs can take in biological systems. We
406 report multiple mechanistic HOIs as defined in ecological studies and represented here by any
407 pairwise interaction-associated genes that are not observed in 3 and more species conditions as
408 well as any interaction-associated genes observed in higher-levels than pairwise cultures[12]. We
409 also report another form of HOIs, as defined in quantitative genetics, which are associated with
410 the non-additive behavior of conserved 2-species interactions in 3 or 4 species communities[22].
411 Here, each HOI level refers to different biological phenomena occurring in the same biological
412 system. Yet, they are essential and complementary to decipher the extremely convoluted
413 architecture and dynamics of microbial interactions and microbiome biology.

414 As the number of interaction-associated genes strongly decreased in the 4-species culture
415 compared to 2 or 3 species setups, our work points out a strong reduction of the interaction
416 landscape with 4 species that is associated with the loss of many lower-level interactions and the
417 emergence of context-specific interactions. While 43% of the 2-species interaction-associated
418 genes are still found with 3 species, only 15% are still found with 4 species, representing less
419 than a third (28%) of the total of 4-species interactions. This highlights the increasing dilution and
420 replacement of original 2-species interactions. To summarize, the more complex a community
421 gets, the more mechanistic HOIs emerge. While our work was limited to 4 species, it remains
422 necessary to verify whether this statement will be true for more complex communities: whether
423 more 2-species interactions will be lost and more HOIs will keep emerging at each level or whether
424 the interaction landscape will stabilize. Indeed, Friedman *et al*., have highlighted that growth
425 observation for 2- and 3-species combination could predict the assembly of 7- to 8-species[2]; this
426 would suggest that interactions, or at least key interactions driving community assembly, could
427 change less dramatically in higher complexity microbiomes.

428 Our molecular approach enabled us to identify that most of the reorganization of the
429 interaction profile is associated with the reprogramming of metabolic or trophic interactions
430 including both competition for nutrients and cross-feeding. As more species are introduced it
431 appears that the dynamics of nutrient consumption is rearranged. For instance, we observed that
432 some 2-species competition for amino acids and for lactate between *E. coli* and *H. alvei* can be
433 alleviated by the introduction of *G. candidum*. While the competition for amino acids is likely
434 alleviated by amino acid cross-feeding from *G. candidum*, as this species is known to release
435 amino acids into the environment through digestion of proteins and peptides[26–28], the mechanism
436 relieving competition for lactate is unclear. Possibly, *G. candidum* provides other nutrients sources
437 like amino acids that alleviate the need for *E. coli* to rely on lactate. Trophic interactions are
438 described as core determinants in community assembly[29,30] and simple rules of metabolic

interdependencies and metabolic specialization can be sufficient to predict the assembly of rather complex communities[31–33]. Yet in complex systems, such as the gut, with complex nutrient composition, the important reorganization of metabolic interactions as community complexity changes could likely explain the difficulty in predicting community assembly and composition in some studies[7]. As previously suggested[34], ecological and metabolic factors associated with the present species such as niche overlap, degree of metabolic specialization and species similarities are likely to be drivers of these metabolic HOIs. Indeed, microorganisms display incredible metabolic abilities, from nutrient usage to rapid metabolic switching, and in the presence of multiple nutrient sources and/or other microorganisms they are likely to readjust the sequence of nutrient uptake. Niche occupation and nutrient access are also two crucial aspects in the success or failure of microbial invasion[4,35]. The reorganization of metabolic interactions in different community composition could likely explain the poor predictability of invasion resistance from simple species-combinations and the emergence resistance-specific phenotype.

To some extent, our work also highlights the importance of fungal species as major actors in reshaping interaction networks. Here, more 3-species HOIs and 4-species HOIs were observed when one fungal species was added either as a third species with two bacteria or as a fourth species with another fungus and two bacteria. In microbial communities, fungi are known to impact community structure and access to nutrients through the formation of hyphal highways[36,37], to impact community assembly through environmental modification[38–40] as well as to impact community protection through the production of multiple secondary metabolites with antibiotic or antimicrobial properties[15,39,41]. In this specific context, fungi-associated HOIs seem to be related to their metabolic characteristics, whether through cross-feeding or competition for nutrients. Yet, the emergence of specific requirements for ECA biosynthetic genes in *E. coli* when all species are present, suggesting a potential toxic stress, may reveal other fungal context-specific properties, while it could also inform about a more precise role of the ECA in *E. coli*. We believe

464 this work contributes, along with other recent studies, to advocate for the need to include fungal
465 species more frequently into microbiome work studies.
466
467 Finally, to understand principles behind the conservation of pairwise interactions, we used
468 epistasis analysis to quantify the additive or non-additive behavior of conserved 2-species
469 interactions in 3 and 4 species conditions. Epistasis analysis offers an adaptable approach to test
470 the linear behavior of quantitative traits in systems with increasing complexity, whether it is the
471 effects of accumulation of mutations in a genome[18,42], the effect of drug combinations[43] or the
472 dynamics of predators-prey ecosystems[44]. Recent studies in the field of microbiome research
473 have also relied on epistasis analysis to elucidate fundamentals of microbial communities such
474 as how the nutrient composition of the environment determines the assembly and the diversity of
475 a microbial community[34,45], how the functional landscape of microbial community is built[3] or how
476 the commensal microbiome determines the development, lifespan and reproduction of its host[5].
477 In this work, using the quantitative metric IFE for the strength of interactions, we used an epistatic
478 model to characterize the additivity or non-linearity of IFE of conserved 2-species interactions
479 when *E. coli* is growing with 2 and 3 other species (3-species and 4-species conditions
480 respectively). We observed that most of the conserved 2-species interaction genes followed an
481 additive model of conservation, including trophic-interaction related genes. While our study was
482 limited to a small number of interacting species, it remains to be investigated if this linear behavior
483 is propagated if the community complexity increases again. While we have observed both
484 synergistic effects and limitations effects for nonlinear conservation of interaction strength, we
485 believe that such behavior are likely to arise in more complex contexts and that additivity will
486 eventually stop. For instance, additivity of pairwise interactions is likely to saturate and to stop at
487 higher-levels of complexity due to environmental constraints (nutrient supply for instance) and/or
488 the metabolic capacities of the different species as highlighted in [34]. While this theory can apply
489 to conserved trophic interactions, like the ones highlighted in our work, other scenarios could

490 include molecule-driven or environment-mediated interactions such as exchange of metabolites,
491 quorum sensing, antibiosis and modification of the physiochemical properties of the environment.
492 While trait additivity is fairly easy to quantify in high-dimensional systems, our study, along with
493 other work using epistasis to study microbial communities[3,5,34,45] suggest that additivity of
494 quantitative microbial features is not the only scenario in nature, and strongly emphasizes the
495 need to develop mathematical approaches to accurately understand the dynamics and
496 functionalities of microbiomes.
497
498 Overall, our work identifies the reorganization of microbial interactions along with community
499 complexity at the molecular level. While more advanced modeling is still required, this knowledge
500 of the mechanistic reorganization of interactions and patterns of HOIs is essential for bottom-up
501 approaches to be able predict, from minimal information, the biology of complex microbiomes and
502 will offer novel opportunities to design synthetic microbial communities or control natural ones.
503

504 **Methods**

505 Strains and media

506 *Strains*

507 The bloomy rind cheese community was reconstructed with the same strains from [14]: *H. alvei*
508 JB232 isolated from cheese[46] and two industrial cheese strains: *G. candidum* (Geotrichum
509 candidum GEO13 LYO 2D, Danisco – CHOOZITTM, Copenhagen, Denmark) and *P. camemberti*
510 (PC SAM 3 LYO 10D, Danisco - CHOOZITTM).
511

512 *Medium*

513 All assays have been carried out on 10% cheese curd agar, pH7 (CCA) (10% freeze-dried Bayley
514 Hazen Blue cheese curd (Jasper Hill Farm, VT), 3% NaCl, 0.5% xanthan gum and 1.7% agar).
515 The pH of the CCA was buffered from 5.5 to 7 using 10M NaOH.

516

517 **Competition assay – RBTnSeq Assays**

518 The *E. coli* barcoded transposon library Keio_ML9[19] was used for all RBTnSeq assays on CCA
519 during a 3-day growth in eight different culture conditions: alone, 2-species conditions (with *H.*
520 *alvei*; with *G. candidum;* and with *P.camemberti*), 3-species conditions (with *H. alvei* + *G.*
521 *candidum*; with *H. alvei* + *P. camemberti*; and with *G. candidum* + *P. camemberti*), and in the
522 single 4-species condition (with *H. alvei* + *G. candidum* + *P. camemberti*) (Figure 1).

523 From the species and library inoculation to cell harvest after 3 days of growth, we followed the
524 same procedure as described in[14]. We amplified the *E. coli* barcoded transposon library Keio_ML9
525 into 25 mL of liquid LB-kanamycin (50 mg/mL) up to an OD of 0.6-0.8. A 5mL sample of this
526 preculture was spun down and kept at -80C as the T0 sample required for fitness calculation. The
527 remaining cells were then washed in PBS1x-Tween0.05% and used to inoculate the competition
528 assays with $7*10^6$ cells of the library on each 100 mm petri dish plate. When necessary, the other
529 community members were then inoculated at the following densities: for *H. alvei*: $7*10^6$ cells; for
530 *G. candidum*: $7*10^6$ cells; for *P. camemberti*: $7*10^5$ cells.

531 After 3 days, cells were harvested by flooding the plates with 1.5mL of PBS1X-Tween0.05%,
532 gentle scraping and then transferring the resuspended cells into 1.5mL tubes. Cells were
533 centrifuged for 3 min at RT at 10 KRPM and stored at -80C until gDNA extraction.

534 All assays have been performed in triplicate.

535

536 **gDNA extraction, library preparation and sequencing**

537 gDNA from the samples of the competition assays was extracted by phenol-chloroform extraction
538 (pH 8) as described in [14]. For each sample, we resuspended the cells into 500 mL of buffer B (200
539 mM NaCl, 20 mM EDTA) and then transferred them into a 2mL screw-capped tube previously
540 filled with 125 mL of 425-600 mm acid-washed beads and 125 mL of 150-212 mm acid-washed
541 beads. Then, 210mL of SDS 20% and 500mL of Phenol:Chloroform (pH 8) were added to each

542 sample before mechanical lysis by vortexing the tubes for 2 minutes at maximum speed. Tubes
543 were then centrifuged for 3 min at 8 KRPM at 4C and 450mL of aqueous phase were recovered
544 for each sample. 45 mL of sodium acetate 3M and 450 mL of ice-cold isopropanol were added
545 and tubes were incubated for 10 min at -80C. Tubes were then centrifuged for 5 min at 4C at 13
546 KRPM, supernatant was removed and the pellet was washed in 750 mL of 70% ice-cold ethanol
547 before being resuspended in 50mL of DNAse/RNAse free water.
548 For library preparation, the 98C BarSeq PCR described in [19] was used to amplify the barcoded
549 region of the transposons and PCR was performed in a final volume of 50 mL: 25 mL of Q5
550 polymerase master mix (New England Biolab), 10 mL of GC enhancer buffer (New England
551 Biolab), 2.5 mL of the common reverse primer (BarSeq_P1[19]) at 10 mM, 2.5 mL of a forward
552 primer from the 96 forward primers (BarSeq_P2_ITXXX[19]) at 10 mM and 50 ng to 2 mg of gDNA.
553 The following PCR program was used: (i) 98℃ - 4 min, (ii) 30 cycles of: 98℃ – 30 s; 55℃ – 30 s;
554 72℃ – 30 s, (iii) 72℃ – 5 min. After the PCR, 10 mL of each of the PCR products were pooled
555 together to create the BarSeq library. 200 mL of the pooled library were purified using the MinElute
556 purification kit (Qiagen) and final elution of the BarSeq library was performed in 30 mL in DNAse
557 and RNAse free water.
558 The BarSeq library was quantified using Qubit dsDNA HS assay kit (Invitrogen) and then
559 sequenced on HiSeq4000 (50 bp, single-end reads), by the IGM Genomics Center at the
560 University of California San Diego.
561
562 **Data processing and interaction fitness effects (IFE) analysis**
563 For each library, BarSeq reads were first processed using the Perl script BarSeqTest.pl from[19] to
564 obtain the count file (all.poolcount) containing the number of reads per barcode for each
565 sample. This pipeline requires a table where each barcode is mapped to a location in the genome.
566 The Arkin lab (Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley,

567 California, USA) kindly provided the TnSeq table for the *E. coli* library. The original script used for
568 this analysis originates from [19] is publicly available on https://bitbucket.org/berkeleylab/feba.
569 The generated all.poolcount information was then implemented into custom R scripts from[15] to
570 determine the average fitness scores for each gene across three RBTnSeq assay replicates for
571 each of the eight cultures (https://github.com/DuttonLab/RB-TnSeq-Microbial-interactions).
572 Insertion mutants that did not have a sufficient T0 count in each condition or that were not centrally
573 inserted (10–90% of gene) were removed from the analysis and counts were then normalized
574 using a set of five reference genes (*glgP*, *acnA*, *modE*, *leuA* - average of 52 strains
575 each). Detailed explanation of the fitness calculation strategy can be found in the Readme
576 document as well as in[15]. In each condition, to assess the possible variability between replicates,
577 we measured the correlation between the three replicates (Supplementary Figure 13). In this set
578 of experiments, Pearson coefficient between replicates varied from 0.79 to 0.86 suggesting low
579 technical noise.
580 Gene fitness values were then compared between interactive conditions and *E. coli* growth alone
581 conditions using two-sided *t*-tests (when the equality of variance was verified by Fisher test) and
582 correction for multiple comparison (Benjamini–Hochberg method[20]). Comparisons associated
583 with an adjusted *P* value lower than 10% were considered a significant interaction fitness effect.
584

585 **Epistasis model analysis**

586 We used the Python package Epistasis[17] to run the epistasis analysis on *E. coli*'s genes. We
587 implemented the model with the average fitness values across the three biological replicates
588 along with the corresponding variance, for each gene and for each of the culture conditions
589 (Alone, w *H.a*, w. *G.c* , w *P.c*, w *H.a* + *G.c*, w *H.a* + *P.c*, w *G.c* + *P.c*, w *H.a* + *G.c* + *P.c*). The
590 model is run for each gene individually. The 'genotypes' expected in the model correspond to the
591 culture conditions and are binary coded based on the presence or not of *E. coli*'s partners *G.*
592 *candidum*, *H. alvei* and *P. camemberti*. For instance, the Alone condition corresponds to

phenotype 000, the w *G.c* condition is coded 100, the w *H.a* + *P.c* condition is coded 011, the w *H.a* + *G.c* + *P.c* is coded 111 and so on. The 'phenotype' implemented in the model corresponds to the average fitness value across three biological replicates. After generating the genotype-phenotype map for each gene, the model was run a first time using the 'local' parameter to set-up the center of the map, and returns the set of epistatic coefficients for each gene calculated as follows:

$$IFE_{obs,Ha,Gc} = IFE_{Ha} + IFE_{Gc} + \varepsilon_{Ha,Gc}$$

$$IFE_{obs,Ha,Pc} = IFE_{Ha} + IFE_{Pc} + \varepsilon_{Ha,Pc}$$

$$IFE_{obs,Gc,Pc} = IFE_{Gc} + IFE_{Pc} + \varepsilon_{Gc,Pc}$$

$$IFE_{obs,Ha,Gc,Pc} = IFE_{Ha} + IFE_{Gc} + IFE_{Pc} + \varepsilon_{Ha,Gc} + \varepsilon_{Ha,Pc} + \varepsilon_{Gc,Pc} + \varepsilon_{Ha,Pc,Gc}$$

To determine whether each coefficient was different from zero, we generated 7,000 pseudoreplicates from our experimental variance from each gene and then calculated the probability each coefficient was above or below zero from this distribution. P-values were then adjusted for multiple comparison testing using Benjamini Hochberg Correction[20]. As it has been shown previously that epistatic analyses can fail if the effects of mutations combine non-linearly instead of linearly (e.g. mutational effects multiply rather than add), we used the epistasis package to look for such nonlinearity. However, no nonlinearity was observed (Supplementary Figure 14).

References

1. Ortiz A, Vega NM, Ratzke C, Gore J. Interspecies bacterial competition regulates community assembly in the C. elegans intestine. ISME J 2021;Available from: http://dx.doi.org/10.1038/s41396-021-00910-4

2. Friedman J, Higgins LM, Gore J. Community structure follows simple assembly rules in microbial microcosms. Nat Ecol Evol 2017;1(5):109.

621 3. Sanchez-Gorostiaga A, Bajić D, Osborne ML, Poyatos JF, Sanchez A. High-order interactions distort
622 the functional landscape of microbial consortia. PLoS Biol 2019;17(12):e3000550.

623 4. Stressmann FA, Bernal-Bayard J, Perez-Pascual D, et al. Mining zebrafish microbiota reveals key
624 community-level resistance against fish pathogen infection. ISME J 2021;15(3):702–19.

625 5. Gould AL, Zhang V, Lamberti L, et al. Microbiome interactions shape host fitness. Proc Natl Acad Sci
626 U S A 2018;115(51):E11951–60.

627 6. Vandermeer JH. The competitive structure of communities: An experimental approach with protozoa.
628 Ecology 1969;50(3):362–71.

629 7. Sundarraman D, Hay EA, Martins DM, Shields DS, Pettinari NL, Parthasarathy R. Higher-Order
630 Interactions Dampen Pairwise Competition in the Zebrafish Gut Microbiome. MBio 2020;11(5).
631 Available from: http://dx.doi.org/10.1128/mBio.01667-20

632 8. Grilli J, Barabás G, Michalska-Smith MJ, Allesina S. Higher-order interactions stabilize dynamics in
633 competitive network models. Nature 2017;548(7666):210–3.

634 9. Mayfield MM, Stouffer DB. Higher-order interactions capture unexplained complexity in diverse
635 communities. Nat Ecol Evol 2017;1(3):62.

636 10. Levine JM, Bascompte J, Adler PB, Allesina S. Beyond pairwise mechanisms of species coexistence
637 in complex communities. Nature 2017;546(7656):56–64.

638 11. Bairey E, Kelsic ED, Kishony R. High-order species interactions shape ecosystem diversity. Nat
639 Commun 2016;7:12285.

640 12. Billick I, Case TJ. Higher Order Interactions in Ecological Communities: What Are They and How Can
641 They be Detected? Ecology 1994;75(6):1529–43.

642 13. Mickalide H, Kuehn S. Higher-Order Interaction between Species Inhibits Bacterial Invasion of a
643 Phototroph-Predator Microbial Community. Cell Syst 2019;Available from:
644 http://dx.doi.org/10.1016/j.cels.2019.11.004

645 14. Morin M, Pierce EC, Dutton RJ. Changes in the genetic requirements for microbial interactions with
646 increasing community complexity. Elife 2018;7. Available from: http://dx.doi.org/10.7554/eLife.37072

647 15. Pierce EC, Morin M, Little JC, et al. Bacterial-fungal interactions revealed by genome-wide analysis
648 of bacterial mutant fitness. Nat Microbiol 2020;Available from: http://dx.doi.org/10.1038/s41564-020-
649 00800-z

650 16. Heckendorn RB, Whitley D. Predicting epistasis from mathematical models. Evol Comput
651 1999;7(1):69–101.

652 17. Sailer ZR, Harms MJ. Detecting High-Order Epistasis in Nonlinear Genotype-Phenotype Maps.
653 Genetics 2017;205(3):1079–88.

654 18. Weinreich DM, Lan Y, Wylie CS, Heckendorn RB. Should evolutionary geneticists worry about higher-
655 order epistasis? Curr Opin Genet Dev 2013;23(6):700–7.

656 19. Wetmore KM, Price MN, Waters RJ, et al. Rapid quantification of mutant fitness in diverse bacteria by
657 sequencing randomly bar-coded transposons. MBio 2015;6(3):e00306-15.

658 20. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach
659 to Multiple Testing. J R Stat Soc Series B Stat Methodol 1995;57(1):289–300.

660 21. Rai AK, Mitchell AM. Enterobacterial Common Antigen: Synthesis and Function of an Enigmatic
661 Molecule. MBio 2020;11(4). Available from: http://dx.doi.org/10.1128/mBio.01914-20

662 22. Sanchez A. Defining Higher-Order Interactions in Synthetic Ecology: Lessons from Physics and
663 Quantitative Genetics. Cell Syst. 2019;9(6):519–20.

664 23. Taylor MB, Ehrenreich IM. Higher-order genetic interactions and their contribution to complex traits.
665 Trends Genet 2015;31(1):34–40.

666 24. Rkenes TP, Lamark T, Strøm AR. DNA-binding properties of the BetI repressor protein of Escherichia
667 coli: the inducer choline stimulates BetI-DNA complex formation. J Bacteriol 1996;178(6):1663–70.

668 25. Jensen KF, Dandanell G, Hove-Jensen B, WillemoËs M. Nucleotides, Nucleosides, and Nucleobases.
669 EcoSal Plus 2008;3(1). Available from: http://dx.doi.org/10.1128/ecosalplus.3.6.2

670 26. Boutrou R, Kerriou L, Gassi J-Y. Contribution of Geotrichum candidum to the proteolysis of soft
671 cheese. Int Dairy J 2006;16(7):775–83.

672 27. Aziza M, Couriol C, Amrane A, Boutrou R. Evidences for synergistic effects of Geotrichum candidum
673 on Penicillium camembertii growing on cheese juice. Enzyme Microb Technol 2005;37(2):218–24.

674 28. Ozturkoglu-Budak S, Wiebenga A, Bron PA, de Vries RP. Protease and lipase activities of fungal and
675 bacterial strains derived from an artisanal raw ewe's milk cheese. Int J Food Microbiol 2016;237:17–
676 27.

677 29. Zengler K, Zaramela LS. The social network of microorganisms - how auxotrophies shape complex
678 communities. Nat Rev Microbiol 2018;16(6):383–90.

679 30. Zuñiga C, Li T, Guarnieri MT, et al. Synthetic microbial communities of heterotrophs and phototrophs
680 facilitate sustainable growth. Nat Commun 2020;11(1):3803.

681 31. Guo X, Boedicker JQ. The Contribution of High-Order Metabolic Interactions to the Global Activity of
682 a Four-Species Microbial Community. PLoS Comput Biol 2016;12(9):e1005079.

683 32. Enke TN, Datta MS, Schwartzman J, et al. Modular Assembly of Polysaccharide-Degrading Marine
684 Microbial Communities. Curr Biol 2019;29(9):1528-1535.e6.

685 33. Campbell EA, Korzheva N, Mustaev A, et al. Structural mechanism for rifampicin inhibition of bacterial
686 rna polymerase. Cell 2001;104(6):901–12.

687 34. Pacheco AR, Osborne ML, Segrè D. Non-additive microbial community responses to environmental
688 complexity. Nat Commun 2021;12(1):1–11.

689 35. Kurkjian HM, Akbari MJ, Momeni B. The impact of interactions on invasion and colonization resistance
690 in microbial communities. PLoS Comput Biol 2021;17(1):e1008643.

691 36. Zhang Y, Kastman EK, Guasto JS, Wolfe BE. Fungal networks shape dynamics of bacterial dispersal
692 and community assembly in cheese rind microbiomes. Nat Commun 2018;9(1):336.

693 37. Warmink JA, Nazir R, Corten B, van Elsas JD. Hitchhikers on the fungal highway: The helper effect
694 for bacterial migration via fungal hyphae. Soil Biol Biochem 2011;43(4):760–5.

38. Cosetta CM, Kfoury N, Robbat A, Wolfe BE. Fungal volatiles mediate cheese rind microbiome assembly. Environ Microbiol 2020;22(11):4745–60.

39. Frey-Klett P, Burlinson P, Deveau A, Barret M, Tarkka M, Sarniguet A. Bacterial-fungal interactions: hyphens between agricultural, clinical, environmental, and food microbiologists. Microbiol Mol Biol Rev 2011;75(4):583–609.

40. Lessard M-H, Viel C, Boyle B, St-Gelais D, Labrie S. Metatranscriptome analysis of fungal strains Penicillium camemberti and Geotrichum candidum reveal cheese matrix breakdown and potential development of sensory properties of ripened Camembert-type cheese. BMC Genomics 2014;15:235.

41. Boutrou R, Guéguen M. Interests in Geotrichum candidum for cheese technology. Int J Food Microbiol 2005;102(1):1–20.

42. Fisher R. A. The correlation between relatives on the supposition of Mendelian inheritance. Transaction of the Royal Society of Edinburgh 191; 52(2):399-433. Available from: http://digamoo.free.fr/fisher1919.pdf

43. Bliss CI. The toxicity of poisons applied jointly1. Ann Appl Biol 1939;26(3):585–615.

44. Sih A, Englund G, Wooster D. Emergent impacts of multiple predators on prey. Trends Ecol Evol 1998;13(9):350–5.

45. Estrela S, Sanchez-Gorostiaga A, Vila JC, Sanchez A. Nutrient dominance governs the assembly of microbial communities in mixed nutrient environments. Elife 2021;10. Available from: http://dx.doi.org/10.7554/eLife.65948

46. Wolfe BE, Button JE, Santarelli M, Dutton RJ. Cheese rind communities provide tractable systems for in situ and in vitro studies of microbial diversity. Cell 2014;158(2):422–33.

Acknowledgments

The authors would like to thank the following people and groups: the Arkin Lab and the Deutschbauer Lab at UC Berkeley for the *E. coli* Keio_ML9 RB-TnSeq library; K. Jepsen at the IGM Genomics Center at the University of California, San Diego for assistance with sequencing; and all members of the Dutton Lab for constructive comments on the manuscript. This work was supported by the National Institute of Health New Innovator grant #DP2 AT010401 (R.J.D), the NIH Institutional Training grant NIH-T32GM007413 (A.J.M), National Science Foundation CAREER Award DEB-1844963 (M.J.H),

Author Contributions

R.J.D., M.J.H. and M.A.M. conceptualized the study. M.A.M. performed the experiments. The Epistasis pipeline was written and adapted by M.J.H. and A.J.M. Data analyses were performed by M.A.M. and A.J.M.

The article was written by M.A.M. and revised with input from all authors. The figures were made by M.A.M with input from all authors. The study was supervised by R.J.D. and M.J.H.

Additional information / Competing Interest Statement

The authors declare that no financial and non-financial competing interests exist in relation to the work described.

Figure legends

Figure 1: Changes of *E. coli*'s interaction-associated genes in 2-species, 3-species and 4-species cultures. A. Experimental design for the identification of interaction-associated genes in 7 interactive conditions from the Brie community. The *E. coli* RBTnSeq Keio_ML9 (Wetmore *et al*., 2015) is either grown alone or in 2, 3 or 4 species cultures to calculate *E. coli* gene fitness in each condition (in triplicate). Interaction fitness effect (IFE) is calculated for each gene in each interactive culture as the difference of the gene fitness in the interactive condition and in growth alone. IFE that are significantly different from 0 (two-sided t-test, Benjamini-Hochberg correction for multiple comparisons) highlight the genes associated with interaction in an interactive condition. **B.** Volcanoplots of IFEs calculated for each interactive condition. Adjusted p-values lower than 0.1 highlight significant IFEs. Negative IFEs (blue) identify negative interactions and positive IFE (red) identify positive interactions. Numbers on each plot indicate the number of negative (blue) or positive (red) IFEs. **C**. Functional analysis of the interaction-associated genes (significant IFEs). Interaction-associated genes have been separated into two groups: negative IFE and positive IFE. For each group, we represent the STRING network of the genes (Nodes). Edges connecting the genes represent both functional and physical protein association and the thickness of the edges indicates the strength of data support (minimum required interaction score: 0.4 – medium confidence). Nodes are colored based on their COG annotation and the size of each node is proportional to the number of interactive conditions in which that given gene has been found associated with a significant IFE. Higher resolution of the networks with apparent gene names are found in Supplementary Figures 2 and 3.

Figure 2: Comparison of the genetic basis of interaction for 2-species and 3-species conditions. A. Venn Diagram of 2-species and 3-species interaction-gene sets. This Venn Diagram identifies 2-species interaction-genes that are dropped when a third species is introduced (Left side; Dropped interaction-genes = any 2-species gene that are not found in any 3-species condition), 2-species interaction-genes that are maintained in at least one associated 3-species condition (Intersection; Maintained interaction-genes) and interaction-genes that are specific to 3-species condition (Right side; Emerging interaction-genes). **B.** Functional analysis of the dropped, maintained and emerging interaction-genes from 2-species to 3-species. Each dot represents the fraction of genes of the studied gene set associated with a given COG category (Number of genes found in the category / Total number of genes in the gene set). The color of the dots indicates the general COG group of the COG category: Teal: Metabolism; Blue: Information storage and processing; Orange: Cellular Processes and Signaling ; Grey: Unknown or no COG category. **C.** Species-level analysis of 3-species HOIs: for each 2-species condition, we measure the fraction of interaction-genes that are dropped in associated 3-species cultures (Dropped in 3-species) or maintained in at least one of the 3-species cultures (Maintained in 3-species); for each 3-species condition, we measure the fraction of interaction-genes that have been conserved from at least one

associated 2-species condition (Maintained from 2-species) or that are emerging with 3-species (Emerging in 3-species).

Figure 3: Organization of the interactions in the 4-species community. A. Venn Diagram of 3-species and 4-species interaction-gene sets. This Venn Diagram identifies 3-species interaction-genes that are dropped when a fourth species is introduced (Left side; Dropped interaction-genes = any 3-species gene that are not found in the 4-species condition), 3-species interaction-genes that are maintained in the 4-species condition (Intersection; Maintained interaction-genes) and interaction-genes that are specific to 4-species condition (Right side; Emerging interaction-genes). **B.** Functional analysis of the dropped, maintained and emerging interaction-genes from 3-species to 3-species. Each dot represents the fraction of genes of the studied gene set associated with a given COG category (Number of genes found in the category / Total number of genes in the gene set). The color of the dots indicates the general COG group of the COG category: Teal: Metabolism; Blue: Information storage and processing; Orange: Cellular Processes and Signaling; Grey: Unknown or no COG category. **C.** Species-level analysis of 4-species HOIs: for each 3-species cultures we measure the fraction of interaction-genes that is conserved in the 4-species culture (Maintained in 4-species) and the fraction of interaction-genes that has been dropped (Dropped in 4-species). **D.** Alluvial plots of the interaction genes across community complexity levels. **E**. STRING network of the 4-species interaction genes (Nodes). Edges connecting the genes represent both functional and physical protein association and the thickness of the edges indicates the strength of data support (minimum required interaction score: 0.4 – medium confidence). Nodes are colored based on the level of community complexity the genes are conserved from.

Figure 4: Quantitative analysis of IFE conservation for the interaction associated genes conserved from 2-species to 4-species conditions. A. Schematized quantitative epistasis/non-linearity measured in 3-species conditions (with partner i and j). Epistasis (ε_{ij}) is the difference between the individual IFE of partner i and partner j (red and orange bars) versus placing them together (green). Mathematically, we need three terms (IFE_i, IFE_j, and ε_{ij}) to reproduce the observed IFE for the 3-species condition. **B.** This analysis can be extended to higher levels of community complexity: 4-species (*E. coli* with 3-partners i, j, and k). The model first accounts for epistasis between i/j, i/k, and j/k. In this example, i and j exhibit epistasis; i/k and j/k are additive (dark blue and purple). The predicted IFE for the 4-species community is the sum of the individual 2-species effects (red, orange, light blue) and the 3-species epistatic terms (green). The 4-species epistatic coefficient is the difference between this low-order prediction and the observed IFE for the i,j,k community (pink). **C.** Conservation profiles of the 16 2-species interaction-associated genes conserved up to 4-species. 2-species conditions: a colored square indicatea the 2-species condition(s) in which the interaction-associated gene was identified; a grey square indicates non-significant 2-species IFEs. 3-species conditions: a teal square indicates that the associated IFE is associated with additive behavior from associated 2-species IFE (no ε_{ij} epistatic coefficient), a red square indicates that the associated IFE display non-additivity from 2-species IFE and thus epistasis, a grey square corresponds to a 3-species condition that is not associated with significant 2-species IFE (no epistasis analysis performed); 4-species condition: a teal square indicates that the associated IFE is associated with additive behavior (no ε_{ijk} epistatic coefficient) , a red square indicates that the associated IFE is associated with non-additivity from lower-order IFE. **D.** Comparison of the observed and predicted IFE for the genes and condition associated with 3-species and 4-species non-additive IFE.

Supplementary material list

Supplementary Figures:

<u>Supplementary Figure 1</u>: Growth of the community species in the studied conditions

<u>Supplementary figure 2:</u> Functional analysis of interaction-associated genes with negative IFEs

<u>Supplementary figure 3</u>: Functional analysis of interaction-associated genes with positive IFEs

820 Supplementary figure 4: IFE profiles of Amino acid biosynthesis genes identified in this study

821 Supplementary figure 5: IFE profiles of Purine biosynthesis associated genes identified in this

822 study

823 Supplementary Figure 6: Comparison of interaction-associated genes across the different levels

824 of community complexity

825 Supplementary figure 7: IFE profiles of lactate metabolism genes

826 Supplementary figure 8: Condition specific comparison of interaction-associated genes for 2 and

827 3-species conditions

828 Supplementary figure 9: IFE profiles of Enterobacterial Common Antigen (EAC) genes

829 Supplementary figure 10: Condition specific comparison of interaction-associated genes for 3 and

830 4-species condition

831 Supplementary figure 11: Functional network of the 4-species interaction-genes and their origin

832 Supplementary figure 12: IFE profiles of the 16 2-species interaction genes maintained up to 4-

833 species

834 Supplementary figure 13: Pearson correlation of gene fitness across replicates

835 Supplementary figure 14: Non-linearity analysis of IFE in the Epistatis model

836

837 *Supplementary Datasets*:

838 Supplementary Data 1: RB-TnSeq based interaction analysis (Fitness values, Interaction Fitness

839 Effects and associated statistics)

840 Supplementary Data 2: Interaction-associated genes at each level of community complexity

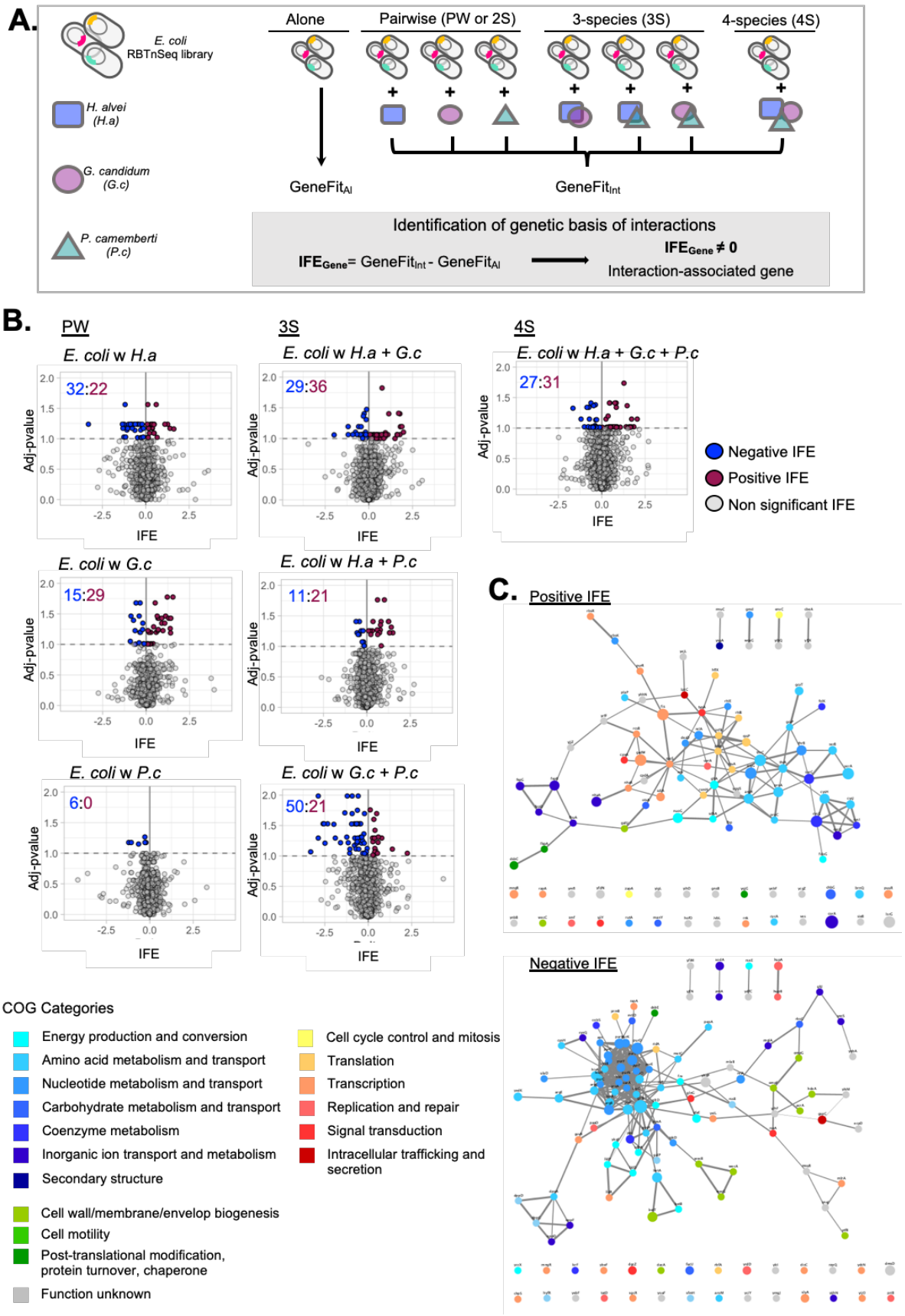
841 Supplementary Data 3: Comparison of interaction-associated genes across the different levels of

842 community complexity

843

844 **Figures**

845


A.
E. coli RBTnSeq library
H. alvei (H.a)
G. candidum (G.c)
P. camemberti (P.c)
Alone
Pairwise (PW or 2S)
3-species (3S)
4-species (4S)
GeneFitAl
GeneFitInt
Identification of genetic basis of interactions
IFEGene= GeneFitInt - GeneFitAl
IFEGene ≠ 0
Interaction-associated gene
B.
PW
3S
4S
E. coli w H.a
32:22
E. coli w H.a + G.c
29:36
E. coli w H.a + G.c + P.c
27:31
E. coli w G.c
15:29
E. coli w H.a + P.c
11:21
E. coli w P.c
6:0
E. coli w G.c + P.c
50:21
Adj-pvalue
IFE
Negative IFE
Positive IFE
Non significant IFE
C.
Positive IFE
Negative IFE
COG Categories
Energy production and conversion
Amino acid metabolism and transport
Nucleotide metabolism and transport
Carbohydrate metabolism and transport
Coenzyme metabolism
Inorganic ion transport and metabolism
Secondary structure
Cell wall/membrane/envelop biogenesis
Cell motility
Post-translational modification, protein turnover, chaperone
Function unknown
Cell cycle control and mitosis
Translation
Transcription
Replication and repair
Signal transduction
Intracellular trafficking and secretion

847 **Figure 1: Changes of *E. coli*'s interaction-associated genes in 2-species, 3-species and 4-species**
848 **cultures. A.** Experimental design for the identification of interaction-associated genes in 7 interactive
849 conditions from the Brie community. The *E. coli* RBTnSeq Keio_ML9 (Wetmore *et al*., 2015) is either grown
850 alone or in 2, 3 or 4 species cultures to calculate *E. coli* gene fitness in each condition (in triplicate).
851 Interaction fitness effect (IFE) is calculated for each gene in each interactive culture as the difference of the
852 gene fitness in the interactive condition and in growth alone. IFE that are significantly different from 0 (two-
853 sided t-test, Benjamini-Hochberg correction for multiple comparisons) highlight the genes associated with
854 interaction in an interactive condition. **B.** Volcanoplots of IFEs calculated for each interactive condition.
855 Adjusted p-values lower than 0.1 highlight significant IFEs. Negative IFEs (blue) identify negative
856 interactions and positive IFE (red) identify positive interactions. Numbers on each plot indicate the number
857 of negative (blue) or positive (red) IFEs. **C**. Functional analysis of the interaction-associated genes
858 (significant IFEs). Interaction-associated genes have been separated into two groups: negative IFE and
859 positive IFE. For each group, we represent the STRING network of the genes (Nodes). Edges connecting
860 the genes represent both functional and physical protein association and the thickness of the edges
861 indicates the strength of data support (minimum required interaction score: 0.4 – medium confidence).
862 Nodes are colored based on their COG annotation and the size of each node is proportional to the number
863 of interactive conditions in which that given gene has been found associated with a significant IFE. Higher
864 resolution of the networks with apparent gene names are found in Supplementary Figures 2 and 3.
865

866

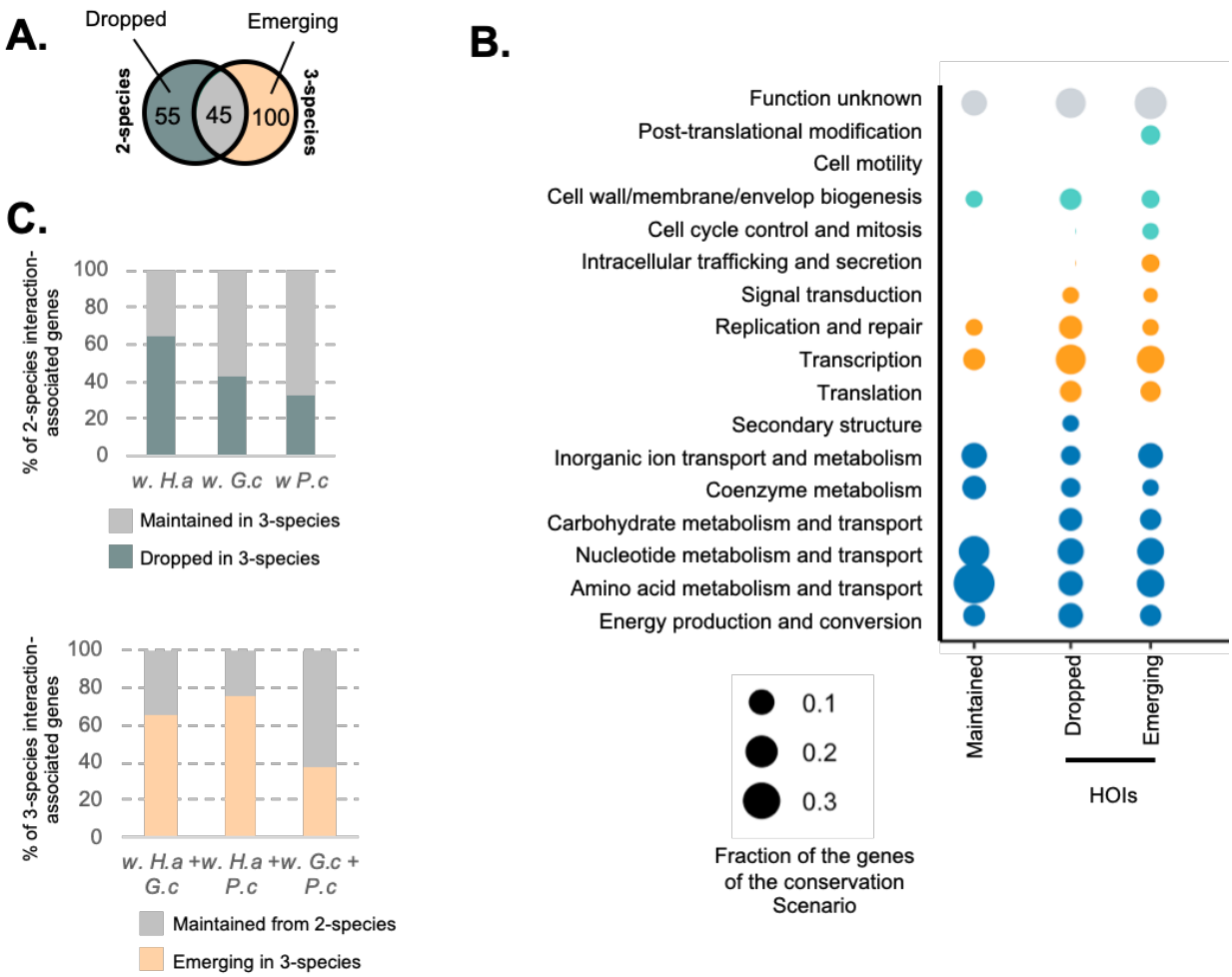


Figure 2: Comparison of the genetic basis of interaction for 2-species and 3-species conditions. A. Venn Diagram of 2-species and 3-species interaction-gene sets. This Venn Diagram identifies 2-species interaction-genes that are dropped when a third species is introduced (Left side; Dropped interaction-genes = any 2-species gene that are not found in any 3-species condition), 2-species interaction-genes that are maintained in at least one associated 3-species condition (Intersection; Maintained interaction-genes) and interaction-genes that are specific to 3-species condition (Right side; Emerging interaction-genes). **B.** Functional analysis of the dropped, maintained and emerging interaction-genes from 2-species to 3-species. Each dot represents the fraction of genes of the studied gene set associated with a given COG category (Number of genes found in the category / Total number of genes in the gene set). The color of the dots indicates the general COG group of the COG category: Teal: Metabolism; Blue: Information storage and processing; Orange: Cellular Processes and Signaling ; Grey: Unknown or no COG category. **C.** Species-level analysis of 3-species HOIs: for each 2-species condition, we measure the fraction of interaction-genes that are dropped in associated 3-species cultures (Dropped in 3-species) or maintained in at least one of the 3-species cultures (Maintained in 3-species); for each 3-species condition, we measure the fraction of interaction-genes that have been conserved from at least one associated 2-species condition (Maintained from 2-species) or that are emerging with 3-species (Emerging in 3-species).

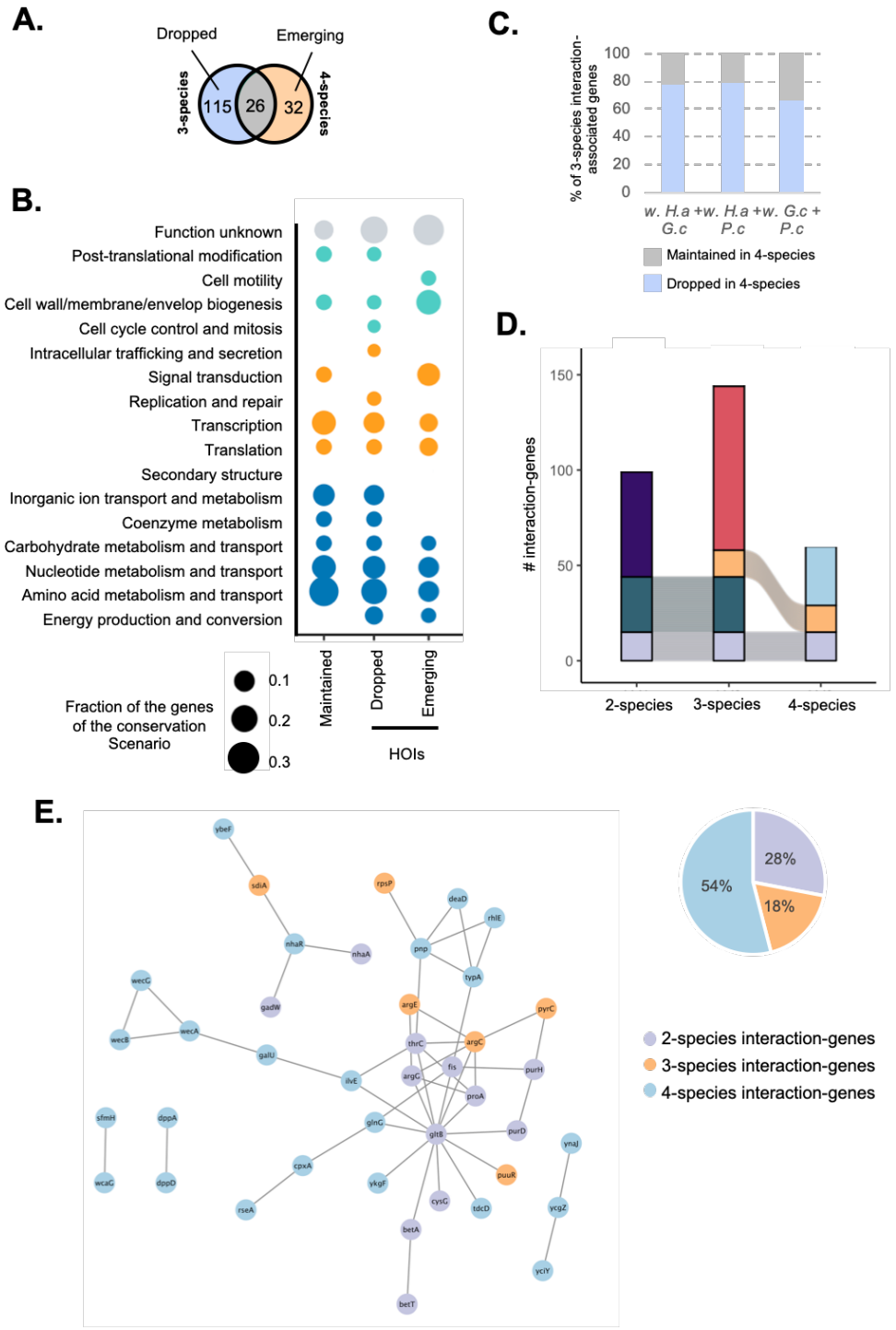


Figure 3: Organization of the interactions in the 4-species community. A. Venn Diagram of 3-species and 4-species interaction-gene sets. This Venn Diagram identifies 3-species interaction-genes that are dropped when a fourth species is introduced (Left side; Dropped interaction-genes = any 3-species gene that are not found in the 4-species condition), 3-species interaction-genes that are maintained in the 4-species condition (Intersection; Maintained interaction-genes) and interaction-genes that are specific to 4-species condition (Right side; Emerging interaction-genes). **B.** Functional analysis of the dropped, maintained and emerging interaction-genes from 3-species to 3-species. Each dot represents the fraction

894 of genes of the studied gene set associated with a given COG category (Number of genes found in the
895 category / Total number of genes in the gene set). The color of the dots indicates the general COG group
896 of the COG category: Teal: Metabolism; Blue: Information storage and processing; Orange: Cellular
897 Processes and Signaling ; Grey: Unknown or no COG category. **C.** Species-level analysis of 4-species
898 HOIs: for each 3-species cultures we measure the fraction of interaction-genes that is conserved in the 4-
899 species culture (Maintained in 4-species) and the fraction of interaction-genes that has been dropped
900 (Dropped in 4-species). **D.** Alluvial plots of the interaction genes across community complexity levels. **E**.
901 STRING network of the 4-species interaction genes (Nodes). Edges connecting the genes represent both
902 functional and physical protein association and the thickness of the edges indicates the strength of data
903 support (minimum required interaction score: 0.4 – medium confidence). Nodes are colored based on the
904 level of community complexity the genes are conserved from.
905

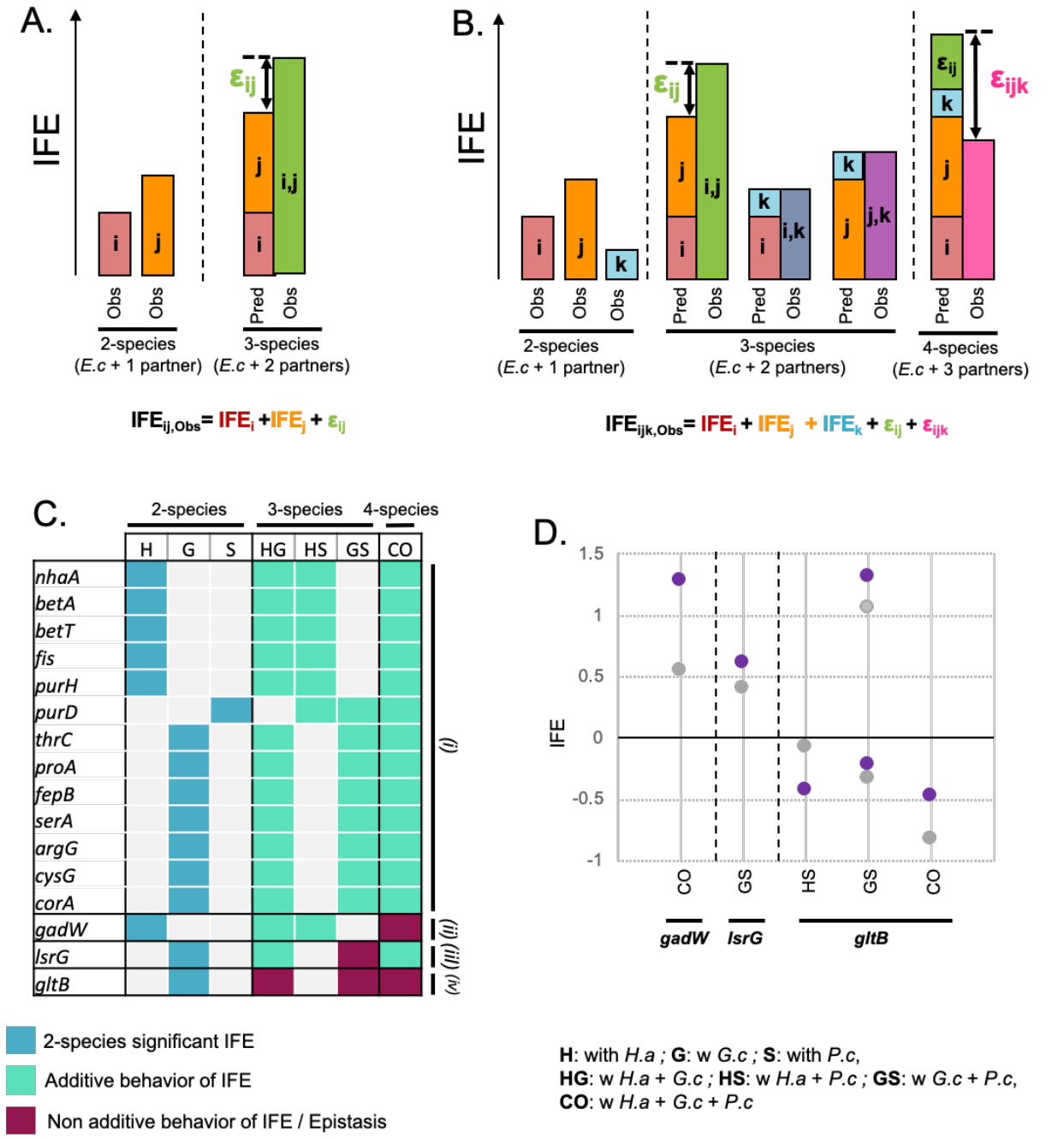


Figure 4: Quantitative analysis of IFE conservation for the interaction associated genes conserved from 2-species to 4-species conditions. A. Schematized quantitative epistasis/non-linearity measured in 3-species conditions (with partner i and j). Epistasis (ε_{ij}) is the difference between the individual IFE of partner i and partner j (red and orange bars) versus placing them together (green). Mathematically, we need three terms (IFE_i, IFE_j, and ε_{ij}) to reproduce the observed IFE for the 3-species condition. **B.** This analysis can be extended to higher levels of community complexity: 4-species (*E. coli* with 3-partners i, j, and k). The model first accounts for epistasis between i/j, i/k, and j/k. In this example, i and j exhibit epistasis; i/k and j/k are additive (dark blue and purple). The predicted IFE for the 4-species community is the sum of the individual 2-species effects (red, orange, light blue) and the 3-species epistatic terms (green). The 4-species epistatic coefficient is the difference between this low-order prediction and the observed IFE for the i,j,k community (pink). **C.** Conservation profiles of the 16 2-species interaction-associated genes conserved

up to 4-species. 2-species conditions: a colored square indicatea the 2-species condition(s) in which the interaction-associated gene was identified; a grey square indicates non-significant 2-species IFEs. 3-species conditions: a teal square indicates that the associated IFE is associated with additive behavior from associated 2-species IFE (no ε_{ij} epistatic coefficient), a red square indicates that the associated IFE display non-additivity from 2-species IFE and thus epistasis, a grey square corresponds to a 3-species condition that is not associated with significant 2-species IFE (no epistasis analysis performed); 4-species condition: a teal square indicates that the associated IFE is associated with additive behavior (no ε_{ijk} epistatic coefficient) , a red square indicates that the associated IFE is associated with non-additivity from lower-order IFE. **D.** Comparison of the observed and predicted IFE for the genes and condition associated with 3-species and 4-species non-additive IFE.