From sequence to ecology: siderophore-receptor coevolution algorithm predicts bacterial interactions in complex communities

Shaohua Gu[1,2,3], Zhengying Shao[3], Yuanzhe Shao[2], Shenyue Zhu[3], Di Zhang[1], Richard Allen[4], Ruolin He[1], Jiqi Shao[1], Guanyue Xiong[1], Zeyang Qu[1], Alexandre Jousset[3], Ville-Petri Friman[5], Zhong Wei[3*], Rolf Kümmerli[4*], Zhiyuan Li[1,2*]

[1] Center for Quantitative Biology, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, 100871, China

[2] Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, 100871, China

[3] Jiangsu Provincial Key Lab for Organic Solid Waste Utilization, Key lab of organic-based fertilizers of China, Nanjing Agricultural University, Nanjing, P R China

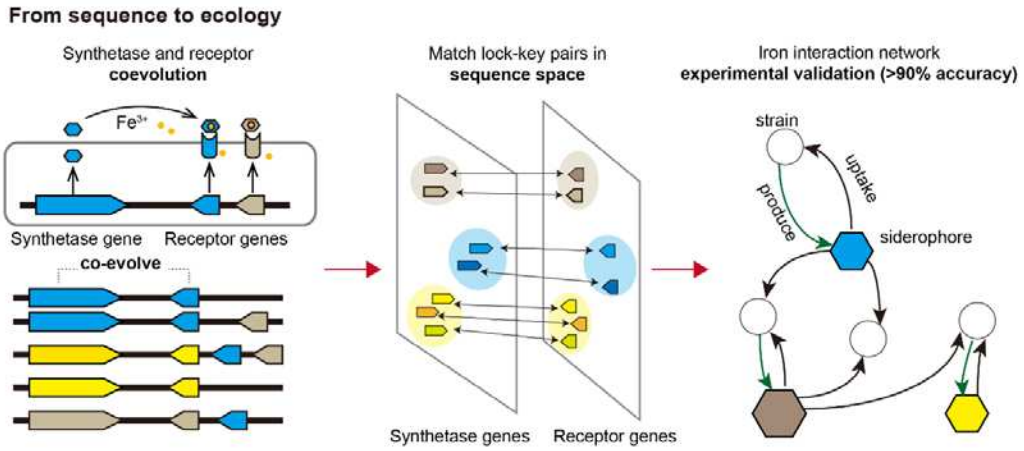
[4] University of Zurich, Department of Quantitative Biomedicine, Winterthurerstr. 190, 8057 Zurich, Switzerland

[5] University of Helsinki, Department of Microbiology, 00014, Helsinki, Finland

[*]Corresponding authors (email: weizhong@njau.edu.cn; rolf.kuemmerli@uzh.ch; zhiyuanli@pku.edu.cn)

Abstract

Unlocking the secrets of microbial interactions through genomics is pivotal for advancing microbial ecology. In most ecosystems, the scarcity of iron makes iron-mediated interactions a central theme in shaping microbial communities. Bacteria have evolved diverse strategies, including the production of siderophores—diverse secondary metabolites—to scavenge iron from their surroundings. Here, we use bioinformatic tools to predict siderophore iron-interaction networks among 1928 *Pseudomonas* strains from sequence data. Our approach uses coevolution analysis to group siderophore synthetase clusters and receptors used for uptake into key-lock pairs. Through a mix of computational analyses and experimental validation, we reconstruct *Pseudomonas* iron-interaction networks across a spectrum of habitats, from soil to water, plants, and human-related environments and reveal substantial differences in network structure and connectivity across habitats. Altogether, our sequence-to-interaction mapping tool empowers researchers to decode bacterial ecology in complex microbiomes, setting the stage for novel interventions to engineer microbiome functionality.



Introduction

Microbial communities populate all ecosystems on earth from terrestrial to aquatic environments, impacting human health, agriculture, and industry [1-3]. The dynamics and functioning of these communities are shaped by complex interactions between microorganisms [4,5]. As the number of sequenced microbial genomes continues to grow exponentially [6,7], there is a growing interest in predicting microbial interaction networks based on the genomic data. Such efforts potentiate major advances, as many microorganisms cannot be cultured in the laboratory [8], while their roles in natural communities can potentially be inferred through sequence-to-interaction mapping. Currently, sequence-to-interaction mapping approaches primarily focus on metabolic interactions, with Genome-scale Metabolic Models (GEMs) serving as the primary tool for establishing the pan-reactome of microbial communities [9,10]. These methods infer metabolic reactions from the genome annotation of enzymes, and then reconstruct a flux model to understand how microorganisms take up essential nutrients and release metabolic byproducts into the environment [11-13].

Despite the significance of primary nutrients, like carbon and nitrogen, there is increasing evidence that secondary metabolism also plays a major role in shaping microbial interactions [14,15]. Nearly all microbes actively synthesize compounds to fulfill a diverse set of functions, including resource scavenging, motility, attack of and defense against competitors, and communication [16,17]. These compounds, referred to as "secondary metabolites," were previously considered non-essential for microbial growth in a laboratory setting, but have since been shown to be critical for competitiveness in natural environments [14,15,18,19]. However,

genome data-based sequence-to-interaction mapping have rarely been applied to secondary metabolism, although this could provide fundamentally new insights into microbial community assembly rules.

Here, we developed a secondary metabolite sequence-to-interaction approach focusing on iron-scavenging siderophores, one of the most prevalent classes of microbial secondary metabolites [20]. Iron is a critical nutrient for microbial survival, as it is used as a catalytic group in enzymes guiding key biological processes such as respiration and replication [21].

However, the concentration of bioavailable iron is typically below the level required for microbial survival in most habitats [21-23]. In response to iron limitation, nearly all bacteria produce siderophores, a chemically diverse class of low molecular weight compounds that efficiently chelate iron from the environment [24,25]. Siderophores are typically diffusible and able to chelate iron over a broad physical range [26]. Once iron is bound, the complex is recognized and taken up by specific receptors in microbial cell membranes [25]. Given their diffusible nature, siderophores mediate a range of social interactions. For bacteria possessing receptors capable of recognizing the iron-siderophore complex, siderophores act as a public good promoting cooperation between individuals [25,27]. When bacteria possess receptors for siderophore uptake but do not produce siderophores themselves they can act as cheaters by exploiting the public goods secreted by others [24,25]. Finally, for bacteria unable to recognize and uptake a specific siderophore, the iron-siderophore complex restricts access to iron and intensifies iron competition [25,28]. Consequently, siderophore-mediated interactions can have an important impact on microbial community composition and dynamics [29-32], yet the prediction of such interactions from sequence data is an unsolved challenge.

We have previously developed bioinformatic pipelines to predict the chemical structure of siderophores and to identify receptors from sequence data. Here, we apply our data mining approach to infer how receptor and siderophore had co-evolved and to find matching siderophore-receptor pairs to predict interaction networks in bacterial communities. We focus on *Pseudomonas* spp. (1928 strains) featuring 188 predicted variants of pyoverdine (their main siderophore) and 94 groups of FpvA receptors (siderophore receptors for pyoverdines)[33]. We developed a Co-evolution Pairing Algorithm that revealed nearly 50 unique lock-key groups, where groups of FpvA receptors emerge as specific “locks” that recognize corresponding pyoverdines as the “key”. Validation experiments yielded prediction accuracies of over 90%. Based on the predicted lock-key pairs, we reconstructed the iron-interaction networks among the 1928 *Pseudomonas* strains. We noticed that network topologies differ fundamentally between different ecological habitats (soil, plant, water, human). Taken together, our work provides a robust sequence-to-interaction mapping tool to predict social interaction networks mediated by secondary metabolites in complex microbial communities.

Results

Section 1: Three classes of pyoverdine strategies in *Pseudomonas* and the lock-key (receptor-synthetase) principle of co-evolution

In our recent work, we developed bioinformatic pipelines to predict pyoverdine molecule structures and receptors based on the genome sequence data. Among 1928 nonredundant *Pseudomonas* strains, we identified 188 chemically different pyoverdine types and 94 different

receptor groups[33]. In this work, we seek to understand how different *Pseudomonas* strains interact through these pyoverdines and receptors.

We first compared all the strains according to phylogeny, ecological habitat, and pyoverdine function. At the phylogenetic level, our data set features a large diversity of *Pseudomonas* species, whereby *P. aeruginosa* (28.7%), *P. fluorescens* (7.0%), *P. syringae* (6.0%), and *P. putida* (2.2%) were the most abundant ones (Figure 1a). The strains originated from a diverse set of habitats, including humans (21.2%), soils (13.6%), plants (12.1%), and water (6.4%), although the origin of many strains (39.5%) is unknown (Figure 1a).

To assess diversity at the pyoverdine functional level, we checked for the absence or presence of pyoverdine synthesis clusters and counted the number of FpvA receptors per strain. We found three basic types of pyoverdine-utilization strategies (Figure 1b). "Single-receptor producers" are the most common type (985 strains, 51.1%) and refer to strains with one pyoverdine synthesis locus and one FpvA receptor gene. "Multi-receptor producers" are the second most common type (679 strains, 35.2%) and refer to strains with one pyoverdine synthesis cluster and multiple receptor genes. "Non-producers" are the least common type (264 strains, 13.7%) and refer to strains that lack the pyoverdine synthesis cluster but contain at least on receptor gene. While each strain possesses on average two to three FpvA receptor genes, no strain carries more than one pyoverdine synthesis cluster. This observation is in line with the expected high costs of pyoverdine synthetase, which is based on a series of gigantic modular enzymes known as NPRS (non-ribosomal peptide synthetases)[34].

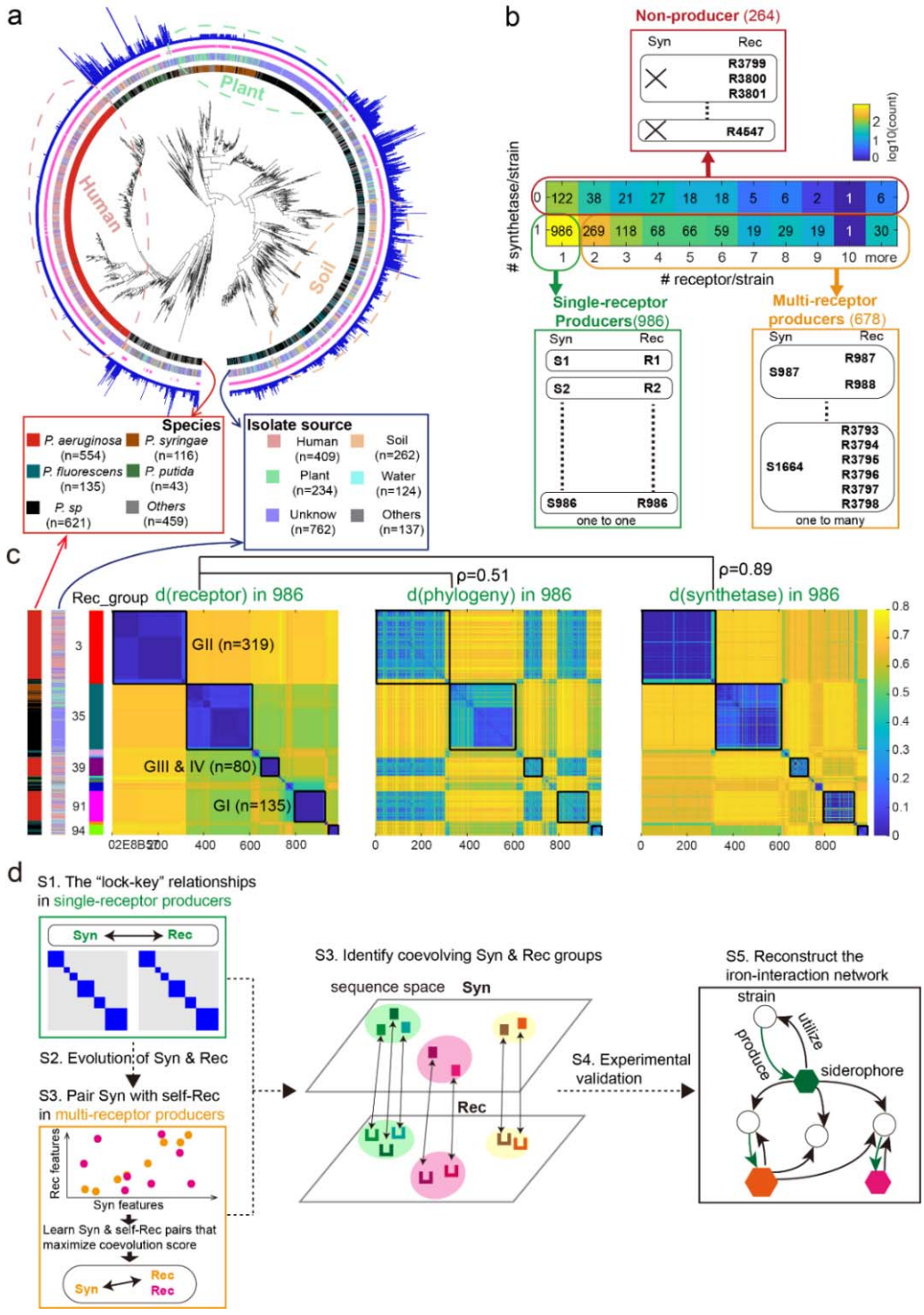


Figure 1 Classification of *Pseudomonas* strains and elucidation of the co-evolution between pyoverdine synthetase and receptors. **a.** Phylogenetic relationship among the 1928 *Pseudomonas* strain based on the concatenated alignment of 400 single-copy conserved genes. Starting

126 from inside, colors in the first ring distinguish the five most prevalent species, with "Others" representing
127 the remaining less abundant species. Colors in the second ring distinguish the four most prevalent
128 sources of isolation. In the third ring, claret and blank regions cover strains with complete pyoverdine
129 synthetase clusters and strains without synthetase gene clusters, respectively. In the fourth blue ring, the
130 bar height indicates the number of FpvA receptors present in each strain. **b.** Strains can be classified into
131 three types by scoring the presence/absence of a synthetase cluster and counting the number receptors
132 in each genome: (i) single-receptor producers containing one pyoverdine synthetase cluster and one
133 FpvA receptor gene; (ii) multi-receptor producers containing one pyoverdine synthetase cluster and
134 several FpvA receptor genes; and (iii) Non-producers lacking synthetase gene but containing at least one
135 receptor gene. **c.** Heatmap visualizing distances between feature sequences of the FpvA receptors and
136 the pyoverdine synthetase clusters and between FpvA features sequences and phylogenetic genes
137 among the 986 single-receptor producers. The hierarchical clustering of the strains is based on the FpvA
138 feature sequences for all three heatmaps. The black squares on the heatmaps denote the five major
139 FpvA groups. Three of these groups correspond to the receptors found among *P. aeruginosa* strains and
140 are labelled with black text. **d.** Scheme of our approach to predict lock-key interactions between
141 pyoverdines and receptors from sequence data. S-labels refer to the respective results section of our
142 work.

143 Based on these findings, we hypothesize that in each single-receptor producer, the sole
144 receptor should recognize its self-produced pyoverdine. Consequently, when the synthetase
145 structure is altered, the receptor should correspondingly change sequences to preserve their
146 matching relationship. This implies that synthetase and receptor pairs engage in molecular
147 co-evolution. To test this hypothesis, we focused on the 986 single-receptor producers and

calculated the degree of covariation between sequence distances matrices of the receptor, the synthesis cluster, and the 400 conserved phylogenetic genes. For the receptors (FpvA) and the synthesis cluster, we used the feature sequences that are most predictive of receptor specificity and pyoverdine molecular structure[33].

In support of the co-evolution hypothesis, we found a strong correlation between the distance matrixes of the receptors and the synthesis clusters (Pearson's r=0.89), a correlation that is much stronger than between the receptor and the phylogeny matrix (Pearson's r=0.51) (Figure 1c). Notably, we observed strong clustering patterns in the sequence space of the receptors, forming distinctive blocks that closely match the clustering patterns of their corresponding synthesis clusters. Using our receptor clustering pipeline[33], we identified 17 receptor groups among the 986 single-receptor producers. Importantly, three out of the 17 receptor groups represent the FpvA receptors characteristic of the human pathogen *P. aeruginosa* (text marked in Figure 1c left panel, as type I-IV FpvAs), and their associated synthetase groups were known to produce pyoverdines that these receptors could selectively uptake[24,35]. These analyses support the hypothesis that cognate receptors and synthesis genes have co-evolved in single-receptor producers, resulting in one-to-one “lock-key” relationships: Each group of receptors, characterized by similar sequence features, acts as a "lock" specifically recognizes pyoverdines (the "key") produced a corresponding group of synthetases.

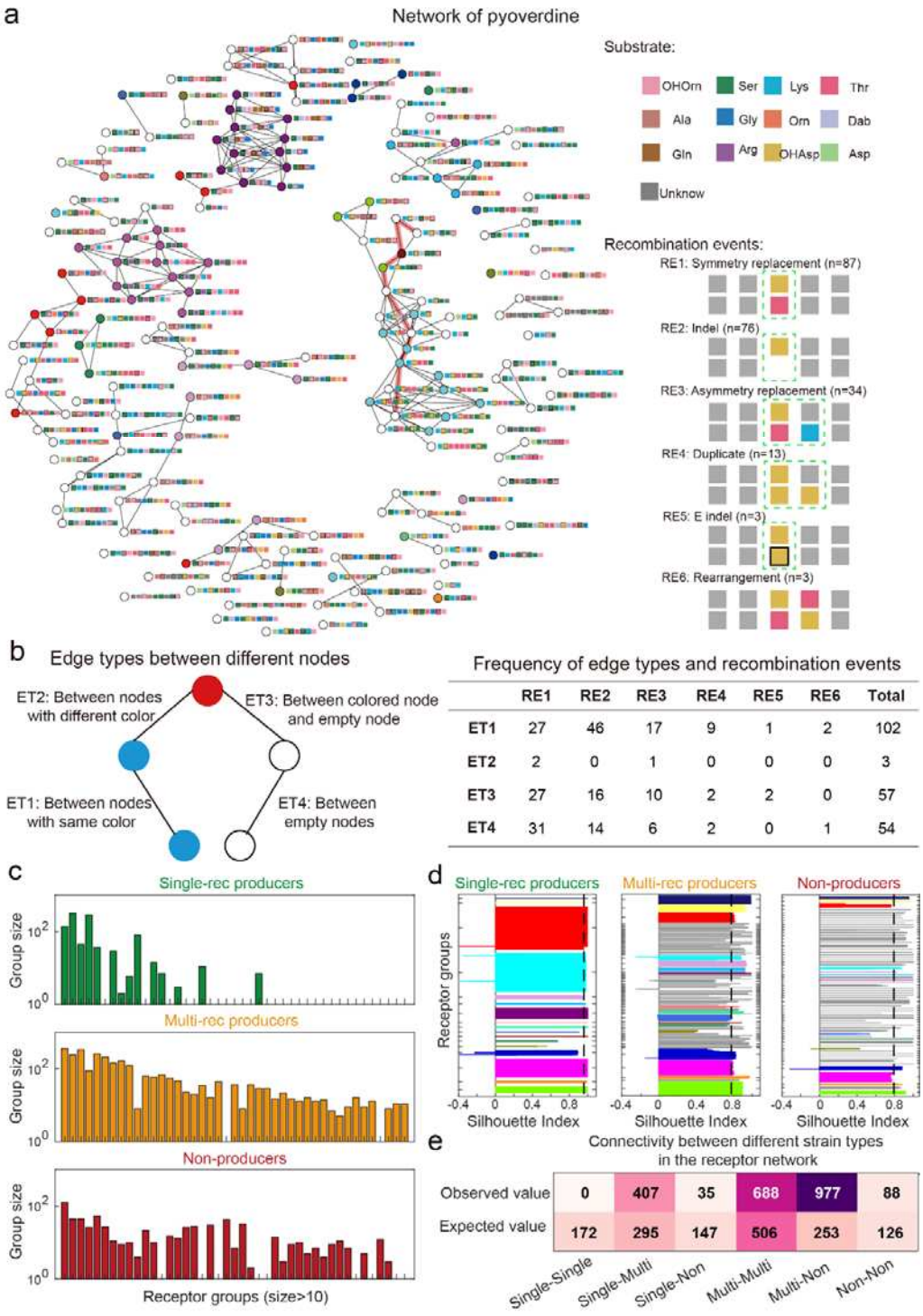
To be able to reconstruct the iron interaction network in *Pseudomonas* communities, we now need to uncover the lock-key groups in multi-receptor producers and to match the receptors of any strain to the synthetase groups of all producers in a community. To achieve

this, we followed a five-step approach, which is summarized in Figure 1d. In section 2, we systematically studied the evolution and diversification of the pyoverdine synthesis clusters and receptors in 1928 *Pseudomonas* strains. In section 3, we used this knowledge to develop an unsupervised algorithm based on coevolutionary principles, in order to allocate synthesis clusters and receptors into lock-key groups. This analysis yields a complete prediction map regarding the groups of pyoverdines each strain is capable of producing and utilizing. In section 4, we use two experimental systems to validate the predicted siderophore-mediated interactions in model *Pseudomonas* communities. In section 5, we finally construct the predicted pyoverdine-mediated interaction networks for soil, plant, water, and human-associated *Pseudomonas* communities.

Section 2: Concurrent diversification patterns of pyoverdine synthetases and receptors reveal multi-receptor producers as evolutionary hubs.

We observed that the structural diversity of pyoverdines was much higher when strains carried multiple receptors (108 unique structures among 678 strains) compared to when they carried a single receptor (47 unique structures among 986 strains). Only 33 structures were predicted to be produced by both receptor type strains. This indicates that pyoverdine diversity is tightly linked with the apparently different iron-acquisition strategies of multi-receptor and single-receptor producers. Here, we explore this diversity in detail and ask how do pyoverdine NRPSs evolve. The 188 pyoverdine structures varied in their length, chirality, and substrate composition, agreeing with the notion that recombination dominates NRPS evolution [36]. We developed the "NRPS tracer" algorithm, which calculates the minimal number of editing events required to transform one pyoverdine product into another. Our algorithm builds on the double

193 cut and join (DCJ)-indel model from Bohnenkampers et al. [37].



194

195 **Figure 2 Diversifications of pyoverdine synthetases and receptors reveal**

196 **multi-receptor producers as evolutionary hubs. a.** Network of all the 188 known pyoverdine

structures connected by one-step evolutionary events. Stripes of colored squares show the amino-acid sequence of pyoverdines with black-framed boxes representing D-type amino acids. Colored circles (nodes) indicate the different FpvA receptor groups found among single-receptor producers. Empty nodes depict pyoverdines exclusively found among multi-receptor producers. The bold orange line shows a representative path of pyoverdine diversification. The positioning of pyoverdine stripes and nodes has no specific meaning. The lower right panel show the scheme of six different recombination events in the evolution of pyoverdine NRPS clusters. Grey and colored boxes indicate conserved and recombination events, respectively. Black-framed boxes indicate recombination involving E domain alterations (e.g. affecting L- vs. D-conformation). **b.** Frequency of edge types and recombination events in the network (right panel). There are four edge types (left panel): ET1 – edges connecting pyoverdine structures from single-receptor producers sharing the same receptor group, ET2 – edges connecting pyoverdine structures from single-receptor producers featuring different receptor groups, ET3 – edges connecting pyoverdine structures that only appeared in multi-receptor producers and these appeared at least one time in single-receptor producers, ET4 – edges connecting pyoverdine structures exclusively occurring in multi-receptor producers. **c.** The 43 largest FpvA receptor groups with more than 10 members (sorted by group size) and their frequency among single-receptor producers, multi-receptor producers, and non-producers. **d.** Silhouette index analysis on the compactness of all receptor groups in single-receptor producers (left panel), multi-receptor producers (middle panel) and non-producers (right panel). Colors represent all the 17 receptor groups found among single-receptor producers. All other receptor groups are shown in black. The dashed vertical lines represent the average of the Silhouette index across all the receptor groups within each strain class. **e.** The observed and expected connectivity value between different strains types in the sequence similarity network of all 4547 FpvAs receptors (Figure S4).

To keep analysis tractable, we focused on pairs of pyoverdine structures that can transform from one to the other by a single recombination event and for which the number of NRPS modules involved in the transformation is ≤ 3 (corresponding to the maximum number of amino acids a NRPS enzyme can incorporate into the pyoverdine backbone). We identified 216 such single-event transformations. They connect 148 of the 188 pyoverdine structures into multiple sub-networks (Figure 2a). While many sub-networks are small the five largest connect 70 out of 188 structures and 1016 out of 1664 strains (Figure S1). The most frequent genetic recombination events were domain and subdomain symmetry replacement (40%), insertions and deletions (indel, 35%) and asymmetry replacement (16%), while duplication events, E-domain indels and rearrangements (total 9%) were much rarer (Figure 2a, the lower right panel).

We then asked whether pyoverdine structures from the same sub-network tend to share similar receptors. To address this question, we allocated nodes to each structure and filled them with the colors of the corresponding receptor groups found in single-receptor producers (17 receptor groups, Figure 1c). Thus, we had 80 colored nodes for the 47 structures exclusively found in single-receptor producers and the 33 structures occurring in both producer types. The 108 nodes of structures that only occurred in multi-receptor producers were left empty (Figure 2a). We then counted the number of cases in which two nodes were connected by the same vs. different colored receptors (Figure 2b, split according to recombination type). We found that the great majority of edges connecting nodes of single-receptor producers were of the same color (97%), while different-color connections were rare (3%). This result strongly supports the molecular co-evolution hypothesis. By contrast, structures that only occur in

multi-receptor producers (empty node) were as often connected to colored nodes (57 events) than to other empty nodes (54 events). This indicates that multi-receptor producers can connect sub-networks of different structures from single-receptor producers (Figure 2a, bold orange line and Figure S2). Altogether, our analysis reveals that multi-receptor producers follow different iron-uptake strategies and have undertaken different evolutionary trajectories in terms of pyoverdine diversification (Figure 2). The latter implies that multi-receptor producers (and probably also non-producers) should also differ in the FpvA receptors they possess, a question we address in the subsequent section.

The previously identified 4547 FpvAs genes cluster into 94 distinct groups based on their feature sequences. For our analysis, we focus on the 43 largest groups comprising more than 10 members (Figure 2c). We found that FpvA receptors of the three strain types were unequally distributed across the 43 groups (Figure 2c). Receptors of single-receptor producers were restricted to 14 out of the 43 groups (32.6%). Conversely, receptors of multi-receptor producers occurred in almost all groups (95.3%). Similarly, receptors from non-producers were also found in many groups (83.7%) with two distinct groups containing exclusively non-producer receptors. This simple frequency analysis shows that lock-key relationships from single-receptor producers are not sufficient to recover the whole iron interaction network because multi-receptor producers and non-producers possess a much more diverse FpvA receptor repertoire than single-receptor producers.

This notion receives further support when comparing the compactness of receptor groups across the three strain types (Figure 2d). We observed that receptors from single-receptor producers tend to connect more compactly (mean silhouette index = 0.96±0.16), while

receptors from multi-receptor producers (mean silhouette index = 0.78±0.19) and non-producers (mean silhouette index = 0.79±0.20) were more dispersed in the sequence space (Figure S3). This observation indicates that receptors from single-receptor producers are more conserved, whereas receptors from multi-receptor producers harbor greater variation. The distinct and conserved nature of receptors from single-receptor producers also emerges when conducting a network analysis with all 4547 FpvAs using their feature sequences (Figure S4). When focusing on the shortest distances between receptor groups, we noticed that none of the 17 receptor groups from single-receptor producers are connected (Figure 2e). In contrast, connections are disproportionately enriched among receptor groups of multi-receptor producers and non-producers, covering 94% and 50% of all the detected 2195 shortest between-group distances, respectively (Figure 2e). These results indicate that evolutionary trajectories of receptors differ between the three strain types. While single-receptor producers harbor conserved and evolutionarily distinct receptors, receptors of multi-receptor producers and non-producers are much more diverse and evolutionarily connected.

Taken together, our results suggest that multi-receptor producers are the main reservoir for siderophore and receptor diversification. These strain types should thus be able to take up several pyoverdine types and are thus expected to form denser interaction networks at the ecological level.

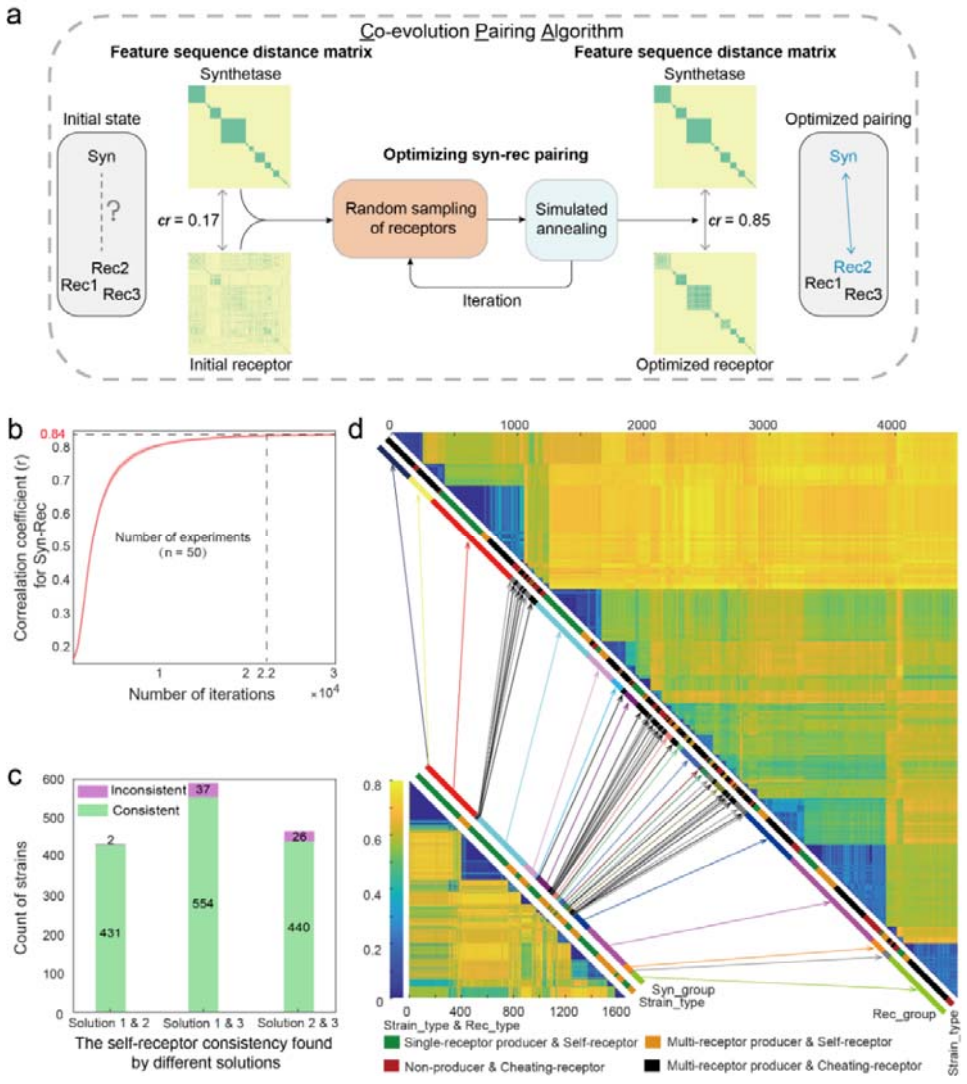
Section 3: Matching synthetases and receptors in sequence space results in 47 lock-key groups.

Next, we aim is to establish a lock-key receptor-pyoverdine interaction map across all three strain types. A first task in this process is to identify receptors in multi-receptor producers that

are used to take up the self-produced pyoverdine. The first intuitive solution is to check for receptors proximate to the pyoverdine synthetase. We can refer to single-receptor producers for the proximity threshold, where 95.6% (943 of 986) of FpvA receptors locate within 20 kb distance from the pyoverdine synthetase cluster. When applying this proximity threshold to multi-receptor producers we can identify putative self-receptors in 87.1% (591 out of 678) of the strains (Solution 1). An alternative approach is to use the lock-key pairs identified for single-receptor producers and check whether similar pairs occur in multi-receptor strains (Solution 2). However, as shown in Figure 2c, this approach only works for the 17 receptor groups found in single-receptor producers and could be applied to 68.7% (466 of 678) multi-receptor strains. Even when combining the Solutions 1 and 2, more than half of the receptor groups could not be paired to any pyoverdine synthetase.

We thus developed an unsupervised learning algorithm, termed Co-evolution Pairing Algorithm (Solution 3), which matches the feature sequence of the synthetase cluster in each strain with its receptors by searching for the set of synthetase-receptor combinations that maximizes co-evolutionary association. Among all multi-receptor producers, there are 678 synthetases and 2812 receptors in total. First, considering that NRPS pathways mainly evolve by large genetic rearrangement like recombination, we used the synthetase feature sequences (concatenated Amotif4-5 regions with consideration of recombination, See Method for details) to build the 678x678 synthetase distance matrix (Figure 3a). We then picked a random receptor as putative self-receptor for each multi-receptor producer and used the receptor features sequences (168 Pro to 295 Ala) to calculate the corresponding 678x678 receptor distance matrix. Subsequently, we calculated co-evolution coefficient *cr*, defined as the

309 Pearson's correlation coefficient between the two matrices (see Method for details). The initial
310 random self-receptor assignment resulted in poor co-evolution coefficients. We thus
311 introduced an iterative optimization process, during which putative self-receptors were shuffled
312 within each multi-receptor producer. We discarded iterations that decreased *cr* values and
313 continued with those that increased *cr* values until an optimization plateau was reached
314 (Figure 3b, *cr* = 0.84). We predicted the self-receptor of all multi-receptor producers based on
315 the final assignment.



316

317 **Figure 3 Developing an unsupervised algorithm to identify self-receptor in**

multi-receptor producers and establishing the lock-key pairs of synthetase and receptor subgroups. **a.** The flowchart of the Co-evolution Pairing Algorithm (Solution 3) that matches the synthetase in each strain to its "self-receptor", by an unsupervised learning scheme that optimizes co-evolutionary strength between the sequence distance matrices of pyoverdine synthetase and matched receptors. The mean correlation coefficients (r) between the two matrices before and after the optimization were shown. **b.** The correlation coefficient (r) and stability of the algorithm were examined by multiple rounds of learning (exp 1 to 50, with final r recorded in the brackets). **c.** The consistency of self-receptors identified by different solutions. The consistency is calculated by comparing the self-receptor found by the multi-receptor producer based on the two supervised solutions (Solution 1 and 2) and one unsupervised algorithm (Solution 3). **d.** Lock-key pairs connecting the sequence spaces of synthases with their self-receptors in both single-receptor producers and multi-receptor producers. The colored (use the same color code as Figure 2a) and black lines represent groups with single-receptor producer and without single-receptor producer, respectively.

We then checked for consistency in self-receptor identification across the three solutions (Figure 3c). Solution 1 and Solution 2 can be classified as supervised machine learning, and they yield high levels of consistency (99.5% across 433 strains). The unsupervised Solution 3 also shares high consistency with Solution 1 (93.7%, for 591 strains) and Solution 2 (94.4%, for 466 strain). These high degrees of consistency legitimate all three solutions, with Solution 3 having the advantage of being applicable to all strains.

With the help of our co-evolution pairing algorithm, we could predict and allocate a self-receptor to each of the 1664 pyoverdine-producing strains, segregating into 47 distinct lock-key groups (Figure 3d and Figure S5). All single-receptor producers and 84.4%

multi-receptor producers belong to 17 receptor groups, while the remaining 15.6% self-receptors found among multi-receptor producers segregate into 30 additional receptor groups (Figure S5). This leaves us with 2883 FpvA receptors among multi-receptor producers and non-producers that are not self-receptors and could be used to uptake pyoverdines produced by other strains, representing cheating. Indeed, when mapping these 2883 non-self-receptors to the co-evolution matrix, we found that 2703 receptors associated with one of the 47 lock-key groups, suggesting that they can take up the corresponding non-self-produced pyoverdine. There were only 180 FpvA receptors (4.0%) that could not be linked to any of the 47 lock-key groups. They belong to rare receptor groups (Figure S5) that are either erroneously identified as FpvA receptors or match rare pyoverdine structures not covered by the 1664 producers in our dataset. Taken together, we can now predict iron-interactions between strains based on the siderophores they produce and the receptors they possess for siderophore utilization.

Section 4: The receptor-synthetase lock-key pairs successfully predict pyoverdine-mediated interactions in experiments.

We conducted two validation experiments to show that our bioinformatic lock-key approach reliably predicts pyoverdine-mediated interactions with accuracy > 90%. The first validation was conducted with a *Pseudomonas* community from the Nanjing (China) collection originally isolated from the tomato rhizosphere [31]. We chose the genomes of 24 independent strains and subjected them to our previously developed bioinformatic pipelines to predict pyoverdine molecular structures and to identify all FpvA receptors[33]. These 24 strains included 4 single-receptor producers, 16 multi-receptor producers and 4 non-producers (Figure S6). We

applied our Co-evolution Pairing Algorithm to identify the self-receptors of all 20 producers (Solution 3) and verified that all the predicted self-receptors were within the 20 KB genome proximity threshold to the synthetase genes (Solution 1). Subsequently, we assigned the self-receptors to the previously identified 47 lock-key paring groups in our full database, and found that 13 pairing groups occur among the 20 producer strains (Figure S7). Combining this information with the knowledge on non-self-receptors occurring in each strain, we predicted the pyoverdine-mediated interactions between the 24 strains in our community (Figure 4a).

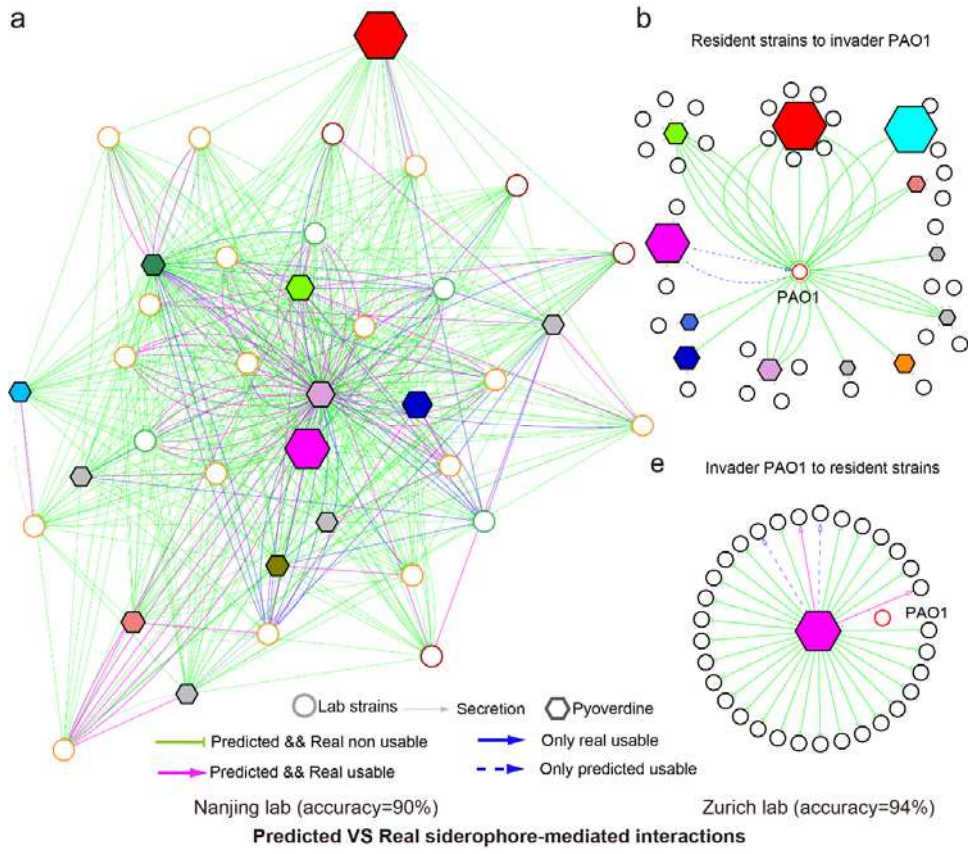


Figure 4 The receptor-synthetase lock-key pairings inferred from sequence data reliably predict experimentally observed pyoverdine-mediated iron interactions. a. Predicted vs. observed iron-interaction network among the 24 experimental strains. Each circular node represents an experimental strain. Green, yellow, and red circular nodes represent single-receptor producers,

multi-receptor producers and non-producers, respectively. Hexagons represent the predicted 13 lock-key receptor-pyoverdine groups. Edges from strain nodes to lock-key nodes represent pyoverdine production, while edges from lock-key nodes to strain nodes represent utilization. Green (pyoverdine non-usable) and pink (pyoverdine usable) edges depict cases in which experimental observations match bioinformatically predicted interactions. Blue edges depict incorrectly predicted pyoverdine interactions. The pyoverdine groups that appeared at least once in single-receptor producers are shown as colored hexagones with the color of the respective receptor group, whereas the pyoverdine groups exclusively secreted by the multi-receptor producer are represented by grey hexagons. **b-c.** Predicted vs. observed iron-interaction networks based on data from a previous study carried out in Zurich lab. The predicted interactions were inferred by the algorithms presented in this study, while the experimental data is taken from Table S2 of Figueiredo et al. (DOI: 10.1111/ele.13912).

For the experimentally validation, we first confirmed that the 20 producers can indeed produce pyoverdine under iron-limited conditions, while the 4 non-producers cannot (Figure S6). We then followed a modified version of our previously established protocols to calculate the net effect pyoverdine has on the growth of other strains (GE_{Pyo}), while controlling for the effects of other metabolites in the supernatant [31]. In principle, $GE_{Pyo} > 0$ indicates pyoverdine-mediated facilitation. However, because there is substantial experimental variation between experimental replicates, we increased a threshold value of $GE_{Pyo} > 0.05$ and classified values above this threshold as positive interactions, where the receiving strain can use the respective pyoverdine for iron acquisition (interaction type 1). Conversely, $GE_{Pyo} \leq 0.05$ values were classified as neutral or negative interactions, where the receiving strain cannot use the respective pyoverdine for iron acquisition (interaction type 0). This approach

397 allowed us to infer an experimental pyoverdine-mediated interaction network (Figure 4a and
398 Figure S8), in which 90% of the observed interactions matched (based on the sign) the
399 predicted interactions from sequence data.

400 The second experimental validation involved strains from the Zurich (Switzerland)
401 collection, isolated from soil and freshwater habitats [25]. In this case, we used published
402 experimental data from the literature [38]. The focus of this earlier study was to test whether the
403 opportunistic human pathogen *P. aeruginosa* PAO1 can invade natural soil and pond
404 communities based on its ability to use pyoverdine from the natural isolates. We used data
405 from all the strains for which genome sequences were available (PAO1 and 33 natural
406 isolates), to establish pyoverdine-mediated interaction networks (Figure 4b-c). We then
407 applied our bioinformatic pipelines as explained for the Nanjing collection and found a high
408 level of consistency (94%) between the predicted and observed pyoverdine-mediated
409 interaction in pairwise cultures (Figure 4b-c).

410 The high consistency between observed and predicted interactions among both the
411 Nanjing and the Zurich strain collection demonstrates that siderophore-mediated microbial
412 interactions can be predicted based on genome-sequence analysis alone using the lock-key
413 relationship between receptor and synthetase genes.

414

415 **Section 5: Pyoverdine interaction networks differ across habitats**

416 We then applied the lock-key pairing approach to our full data set to reconstruct the
417 pyoverdine-interaction network among all the 1928 *Pseudomonas* strains (Figure S9). To keep
418 traceability in such an enormous network, we allocated strains into behavioral groups sharing

the same “pyoverdine-interaction behavior”. Specifically, strains that produce the same pyoverdine type and can utilize the same repertoire of pyoverdines are allocated to the same behavior group and represented as (circular) nodes in the network. There were 407 such behavioral nodes. We included a second type of (hexagonal) nodes in the network and they represent the 47 different receptor-pyoverdine groups identified based on the lock-key algorithm (Figure 3). Edges from behavior nodes to lock-key nodes represent pyoverdine production, while edges from lock-key nodes to behavior nodes represent utilization. Overall, the network featured 307 production edges and 1788 utilization edges. This network analysis does not only reveal that certain lock-key nodes are much more densely connected to behavior nodes than others, it also reveals a high connectivity between behavior groups and the widespread ability of strains to use non-self-produced pyoverdines, indicative of potential cheating.

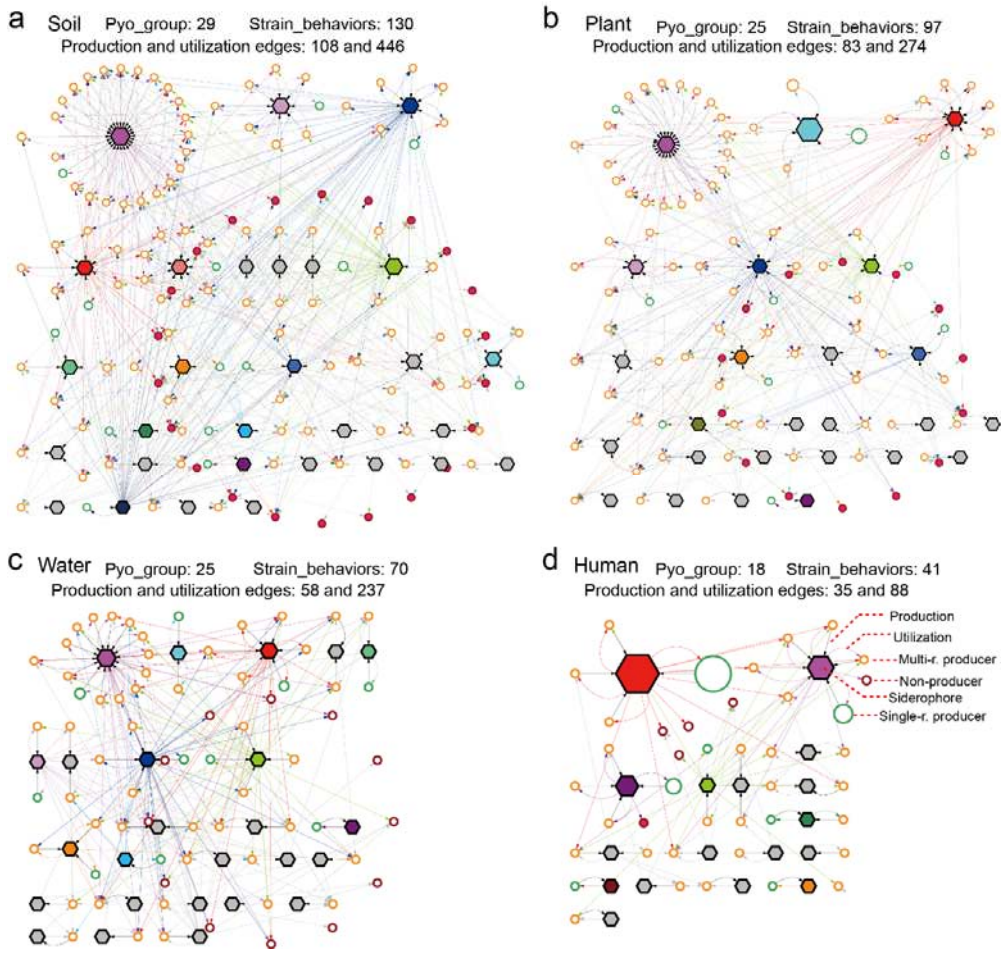


Figure 5 *Pseudomonas* iron-interaction networks vary across habitats. The predicted iron-interaction networks mediated by pyoverdines among *Pseudomonas* isolates from soil (**a**), plant (**b**), water (**c**) and human (**d**) habitats. The interaction networks were built using the Cytoscape's yFiles circular layout. Circular nodes represent behavioral groups (i.e., strains that produce the same pyoverdine type and utilize the same repertoire of pyoverdines) with node size being proportional to the number of strains that exhibit this behavior. Green, yellow, and red circular nodes represent single-receptor producers, multi-receptor producers and non-producers, respectively. Hexagonal nodes represent the lock-key receptor-pyoverdine groups with node size being proportional to the number of strains utilizing this siderophore. The pyoverdine groups once appeared in single-receptor producers were highlighted with the receptors' group colors (colored hexagons) and the pyoverdine groups only

442 secreted by the multi-receptor producer in the database represented by grey hexagons. Edges from
443 behavior nodes to lock-key nodes represent pyoverdine production, while edges from lock-key nodes to
444 behavior nodes represent utilization, with the same color of the pyoverdine node.
445
446 One biological limitation of the above analysis is that most of the strains would never meet
447 in nature as they were isolated from different environments. That is why we created separate
448 networks for strains isolated from soil (262 strains), plant (234), water (124), and
449 human-derived (409) habitats. We found that strain type frequencies varied fundamentally
450 between the four habitats (Figure 5). For example, in the soil-derived strains, there were 56.9%
451 multi-receptor producers, 27.5% single-receptor producers and 15.7% non-producers (Table
452 S1). In contrast, there were only 10.0% multi-receptor producers and 4.0% non-producers, but
453 86.1% single-receptor producers among human-derived strains. These differences in strategy
454 frequencies affected network topology and connectivity. Particularly, the number of behavior
455 groups was higher for soil (130, 0.50 = scaled relative to number of strains), plant (97, 0.41),
456 and water (70, 0.56) habitats than for human-related habitats (41, 0.10). Similarly, the
457 distribution of the 47 pyoverdine lock-key groups differed across habitats, with more groups
458 occurring in soil (29, 0.11 = scaled relative to number of strains), plant (25, 0.11), and water (25,
459 0.20) habitats compared human-derived habitats (18, 0.04). These differences affected the
460 number of utilization edges, which is higher in soil (446, 0.25= scaled relative to 1788
461 utilization edges), plant (274, 0.15) and water (237, 0.13) habitats than in human-derived
462 habitats (88, 0.05). Important to note is also that many behavior groups are unique in each
463 habitat: soil (80, 19.7%), plant (56, 13.0%), water (43, 10.6%), and human (26, 6.4%), and that

464 only 8 behavior groups (Figure S10a) and 11 lock-key groups are shared by all four habitats
465 (Figure S10b). This latter finding suggests divergent evolution of pyoverdine-interaction
466 networks across habitats. Taken together, these results show that our algorithms can predict
467 pyoverdine-interaction networks in natural communities and reveal key differences between
468 habitats.
469

470 **Discussion**

471 Predicting interactions between microbes from sequence data offers exciting opportunities for
472 understanding the ecology and evolution of microbiomes. While sequence-to-interaction
473 mapping has predominantly been carried out for primary metabolism involving resource
474 consumption, conversion, and cross-feeding, there are few approaches to reconstruct
475 microbial interactions based on secondary metabolites (antibiotics, toxins, siderophores,
476 surfactants) [39-41]. In our paper, we offer such an approach by developing a bioinformatic
477 approach to infer iron-interaction networks mediated by pyoverdines (a class of siderophores)
478 within communities of *Pseudomonas* bacteria. To achieve our goal, we analyzed patterns of
479 pyoverdine and pyoverdine-receptor evolution and their co-evolution from sequence data for
480 1928 strains and reconstructed strain interaction networks for soil, plant, water, and
481 human-derived habitats. Our experimentally validated approach provides a roadmap on how
482 to perform sequence-to-interaction mapping for secondary metabolism, and to obtain new
483 insights on the effects of these compounds on community assembly and dynamics.

484 Our approach underscores several challenges associated with secondary metabolites, to
485 which we offer solutions. The first challenge is that secondary metabolites are often built

through complicated synthesis machineries like NRPS and PKS, such that the chemical structure of the metabolite can hardly be directly inferred from sequencing data. We solved this challenge in our first paper[33], in which we developed an approach based on feature sequences that allowed us to infer the chemical structure of 188 pyoverdines produced by the strains in our data set. The second challenge is to identify pyoverdine receptor genes among the many different types of siderophore receptor genes each strain possesses. We also solved this challenge in our first paper[33], by again using feature sequences to identify 4547 pyoverdine receptor genes belonging to 94 groups. The third challenge is to pair pyoverdines to matching receptors within and across strains. For this purpose, we developed an unsupervised learning algorithm (called Co-evolution Paring Algorithm), which yielded 47 synthetase-receptor lock-key pairs and allowed us to reconstruct iron-interaction networks.

Although co-evolution analyses are a widely used computational tool, employed in diverse areas ranging from ab initio protein structure to host-pathogen interaction predictions [42], we could not use existing algorithms, such as DCA, SCA, and Evoformer [43-45]. The reason is that these classical site-based co-evolution methods depend on paired sequences between which the degree of covariation is quantified. We faced two levels of complexity in this context. First, the existence of multi-receptor producer strains impeded direct assignments of synthetase-receptor pairs, and a major part of our algorithm is therefore devoted to the identification of the correct receptor for the self-produced pyoverdine of a strain. Second, due to the variation in the number of pyoverdine synthetase modules, site-based covariation quantification is not applicable, and whole-sequence alignment also performs poorly in characterizing co-evolution due the complexity of pyoverdine synthesis. Therefore, we had to

define reasonable distance metrics that capture the signatures of co-evolution and then apply the Co-evolution Pairing Algorithm to maximize coevolutionary strength between synthetases and receptors. Our new pipeline has the potential to be applied to many other microbial traits. For example, microbial membrane receptors co-evolve with phages[46,47], and pairing phages with the receptor they utilize for infection could provide insights into host-pathogen co-evolution.

While our work provides a novel framework for efficient sequence-based prediction of siderophore-mediated microbial interactions, there are a number of limitations that require careful consideration. First, our approach assumes that sequence similarity indicates functional similarity, i.e., similar gene sequences can be put into “behavioral groups” that produce and uptake the same pyoverdine. In our first paper, we show that this is a reasonable assumption when feature (but not when full) sequences are used. Second, we assume discrete lock-key relationships between paired groups of pyoverdine synthetases and receptors. While it is certainly true that self-receptors have high affinity for the pyoverdine the strain produces, it is also known that receptors can be promiscuous and take up other pyoverdine types although at lower efficiencies[35]. Receptor promiscuity is not covered by our approach. Third, we assume that each producer strain must have at least one self-receptor that recognizes its own pyoverdine. This assumption is reasonable and empirically well documented[48]. Fourth, our approach focusses entirely on the presence/absence of pyoverdine synthetases and receptors, and does not consider regulation. This comes with two shortcomings that affect our prediction accuracy: (i) there are strains that have an intact pyoverdine synthetase machinery but are functionally non-producers[24]. These non-producers

are not detected by our approach, which can lead to errors in interaction predictions. (ii) Strains can vary considerably in the amount of pyoverdine they produce with production levels influencing the strength of interactions, a parameter that cannot be measured with our approach. However, despite these limitations our empirical verification experiments reveal an accuracy of 90% and 94% of correctly predicted interactions.

While the primary goal of our work was to establish methods for sequence-to-interaction mapping, our results already yielded several new biological insights on iron-interaction networks in pseudomonads. First, we identified three different iron-acquisition strategies (single-receptor producers, multi-receptor producers, non-producers) that can co-exist and appear to follow different evolutionary trajectories. Specifically, single-receptor producers showed relatively low structural diversity in pyoverdines and their receptor groups, suggesting relatively conserved iron-uptake strategies. In contrast, pyoverdine and receptor diversity is much higher among multi-receptor producers, identifying them as the evolutionary hub for both pyoverdine and receptor diversification. Second, we found that the above findings have direct consequences for iron-network topologies because multi-strain producers (featuring high pyoverdine and receptor diversity) can connect sub-networks into larger interaction networks. Indeed, soil-, plant-, and water-associated habitats, dominated by multi-receptor producers, featured highly connected iron-interaction networks. Conversely, iron-interaction networks were small and fragmented in human-derived habitats dominated by single-receptor producers. This latter finding potentially stems from a (relatively recent) ecological expansion of *P. aeruginosa* from environmental habitats to humans and other hosts. This ecological expansion could have facilitated diversification, as evidenced by the diverse array of *P. aeruginosa* strains

552 in our data set. However, it appears that pyoverdine synthetase and receptors were little
553 affected by the diversification and remained conserved.

554 In conclusion, we succeeded to develop a sequence-to-interaction mapping approach for
555 siderophores that has high potential to deliver new insights into species and strain interaction
556 networks in bacterial communities. Given that iron is a key trace element that is limited in most
557 environments, siderophore-mediated interactions are an ideal entry point for secondary
558 metabolite analysis from sequence data. While we focused on *Pseudomonas* strains, we know
559 that siderophore-mediated interactions occur across the species boundaries. For example. *P.*
560 *aeruginosa* possesses receptors to take up enterobactin produced by *Enterobacteriaceae* spp.
561 and schizokinen produced by. *Ralstonia solanacearum* [49]. Thus, the next step would be to
562 apply our concepts to more diverse bacterial communities to derive microbiome-level
563 iron-interaction maps. Moreover, pyoverdines represent a particularly complex group of
564 secondary metabolites, such that our evolution-guided approach should be easily
565 customizable to simpler groups of secondary metabolites.

566 **Data Availability**

567 The source code and parameters used are available in the supplementary material.

568 **Acknowledgements**

569 We thank Xuedi Huang for the insight on pyoverdine receptor analysis.

570

571 **Funding**

572 This work was supported by the National Key Research and Development Program of China
573 (No. 2021YFF1200500, 2021YFA0910700), National Natural Science Foundation of China

(No. 42107140, No. 32071255, No.41922053), National Postdoctoral Program for Innovative Talents (No. BX2021012). R.K. is supported jointly by a grant from the Swiss National Science Foundation no. 310030_212266. V-P.F. is supported jointly by a grant from UKRI, Defra, and the Scottish Government, under the Strategic Priorities Fund Plant Bacterial Diseases program (BB/T010606/1), Research Council of Finland, and The Finnish Research Impact Foundation.

Author contributions

Shaohua Gu performed the majority of computational analysis in this research and drafted the manuscript. Zhengying Shao and Shenyue Zhu performed the experiment of testing pyoverdine-mediated interaction between *Pseudomonas* strains. Yuanzhe Shao built the NRPStracer pipeline to infer patterns of diversification in pyoverdine synthetase. Di Zhang developed unsupervised co-evolution pairing algorithm for identifying self-receptor. Ruolin He, Jiqi Shao, Guanyue Xiong, Zeyang Qu assisted in cleaning up the codes. Alexandre Jousset and Ville-Petri Friman offered insightful comments and assisted in revising and writing of the manuscript. Rolf Kümmerli and Zhong Wei oversaw the project, designed experiments and revised the manuscript. Zhiyuan Li conceptualized, oversaw the project and revised the manuscript.

Competing interests

The authors declare no competing interests.

References

1 Cavicchioli, R. *et al.* Scientists' warning to humanity: microorganisms and climate change. *Nature Reviews Microbiology* **17**, 569-586, doi:10.1038/s41579-019-0222-5 (2019).

2 Rani, N., Sangwan, P., Joshi, M., Sagar, A. & Bala, K. in *Microbial Wastewater Treatment* (eds Maulin P. Shah & Susana Rodriguez-Couto) 83-102 (Elsevier, 2019).

3 Wang, X. *et al.* Phage combination therapies for bacterial wilt disease in tomato. *Nature Biotechnology* **37**, 1513-1520, doi:10.1038/s41587-019-0328-3 (2019).

4 Kuramitsu Howard, K., He, X., Lux, R., Anderson Maxwell, H. & Shi, W. Interspecies Interactions within Oral Microbial Communities. *Microbiology and Molecular Biology Reviews* **71**, 653-670, doi:10.1128/MMBR.00024-07 (2007).

5 Konopka, A., Lindemann, S. & Fredrickson, J. Dynamics in microbial communities: unraveling mechanisms to identify principles. *The ISME Journal* **9**, 1488-1495, doi:10.1038/ismej.2014.251 (2015).

6 Handelsman, J. Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiology and Molecular Biology Reviews* **68**, 669-685, doi:10.1128/MMBR.68.4.669-685.2004 (2004).

7 Almeida, A. *et al.* A new genomic blueprint of the human gut microbiota. *Nature* **568**, 499-504, doi:10.1038/s41586-019-0965-1 (2019).

8 Schloss, P. D. & Handelsman, J. Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. *Genome Biology* **6**, 229,

doi:10.1186/gb-2005-6-8-229 (2005).

9 Gu, C., Kim, G. B., Kim, W. J., Kim, H. U. & Lee, S. Y. Current status and applications of genome-scale metabolic models. *Genome Biology* **20**, 121, doi:10.1186/s13059-019-1730-3 (2019).

10 Colarusso, A. V., Goodchild-Michelman, I., Rayle, M. & Zomorrodi, A. R. Computational modeling of metabolism in microbial communities on a genome-scale. *Current Opinion in Systems Biology* **26**, 46-57, doi:10.1016/j.coisb.2021.04.001 (2021).

11 Pacheco, A. R., Moel, M. & Segrè, D. Costless metabolic secretions as drivers of interspecies interactions in microbial ecosystems. *Nature Communications* **10**, 103, doi:10.1038/s41467-018-07946-9 (2019).

12 Gude, S. & Taga, M. E. Multi-faceted approaches to discovering and predicting microbial nutritional interactions. *Current Opinion in Biotechnology* **62**, 58-64, doi:10.1016/j.copbio.2019.08.005 (2020).

13 Louca, S. & Doebeli, M. Calibration and analysis of genome-based models for microbial ecology. *eLife* **4**, e08208, doi:10.7554/eLife.08208 (2015).

14 Penn, K. *et al.* Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria. *The ISME Journal* **3**, 1193-1203, doi:10.1038/ismej.2009.58 (2009).

15 Scherlach, K. & Hertweck, C. Mediators of mutualistic microbe–microbe interactions. *Natural Product Reports* **35**, 303-308, doi:10.1039/C7NP00035A (2018).

16 Netzker, T. *et al.* Microbial communication leading to the activation of silent fungal

secondary metabolite gene clusters. *Front Microbiol* **6**, 299, doi:10.3389/fmicb.2015.00299 (2015).

17 Demain, A. L. & Fang, A. The natural functions of secondary metabolites. *Adv Biochem Eng Biotechnol* **69**, 1-39, doi:10.1007/3-540-44964-7_1 (2000).

18 Gao, H., Weitao, T. & He, Q. Coping with the environment: how microbes survive environmental challenges. *Int J Microbiol* **2011**, 379519, doi:10.1155/2011/379519 (2011).

19 Wolf, D. M., Vazirani, V. V. & Arkin, A. P. Diversity in times of adversity: probabilistic strategies in microbial survival games. *Journal of Theoretical Biology* **234**, 227-253, doi:10.1016/j.jtbi.2004.11.020 (2005).

20 He, R. *et al.* Knowledge-guided data mining on the standardized architecture of NRPS: Subtypes, novel motifs, and sequence entanglements. *PLOS Computational Biology* **19**, e1011100, doi:10.1371/journal.pcbi.1011100 (2023).

21 Andrews, S. C., Robinson, A. K. & Rodríguez-Quiñones, F. Bacterial iron homeostasis. *FEMS Microbiology Reviews* **27**, 215–237, doi:10.1016/s0168-6445(03)00055-x (2003).

22 Boyd, P. W. & Ellwood, M. J. The biogeochemical cycle of iron in the ocean. *Nat. Geosci.* **3**, 675-682, doi:10.1038/ngeo964 (2010).

23 Emerson, D., Roden, E. & Twining, B. The microbial ferrous wheel: iron cycling in terrestrial, freshwater, and marine environments. *Frontiers in Microbiology* **3**, doi:10.3389/fmicb.2012.00383 (2012).

24 Butaitė, E., Baumgartner, M., Wyder, S. & Kümmerli, R. Siderophore cheating and

cheating resistance shape competition for iron in soil and freshwater Pseudomonas communities. *Nature Communications* **8**, 414, doi:10.1038/s41467-017-00509-4 (2017).

25 Kramer, J., Özkaya, Ö. & Kümmerli, R. Bacterial siderophores in community and host interactions. *Nature Reviews Microbiology* **18**, 152-163, doi:10.1038/s41579-019-0284-4 (2020).

26 Leventhal, G. E., Ackermann, M. & Schiessl, K. T. Why microbes secrete molecules to modify their environment: the case of iron-chelating siderophores. *Journal of The Royal Society Interface* **16**, 20180674, doi:10.1098/rsif.2018.0674 (2019).

27 West, S. A., Diggle, S. P., Buckling, A., Gardner, A. & Griffin, A. S. The Social Lives of Microbes. *Annual Review of Ecology, Evolution, and Systematics* **38**, 53-77, doi:10.1146/annurev.ecolsys.38.091206.095740 (2007).

28 Miethke, M. & Marahiel Mohamed, A. Siderophore-Based Iron Acquisition and Pathogen Control. *Microbiology and Molecular Biology Reviews* **71**, 413-451, doi:10.1128/MMBR.00012-07 (2007).

29 Seyoum, Y., Baye, K. & Humblot, C. Iron homeostasis in host and gut bacteria – a complex interrelationship. *Gut Microbes* **13**, 1874855, doi:10.1080/19490976.2021.1874855 (2021).

30 Schaible, U. E. & Kaufmann, S. H. E. Iron and microbial infection. *Nature Reviews Microbiology* **2**, 946-953, doi:10.1038/nrmicro1046 (2004).

31 Gu, S. *et al.* Competition for iron drives phytopathogen control by natural rhizosphere microbiomes. *Nature Microbiology* **5**, 1002–1010, doi:10.1038/s41564-020-0719-8

(2020).

32 Gu, S. *et al.* Siderophore-Mediated Interactions Determine the Disease Suppressiveness of Microbial Consortia. *mSystems* **5**, e00811-00819, doi:10.1128/mSystems.00811-19 (2020).

33 Shaohua, G. *et al.* From sequence to molecules: Feature sequence-based genome mining uncovers the hidden diversity of bacterial siderophore pathways. *bioRxiv*, 2023.2010.2030.564663, doi:10.1101/2023.10.30.564663 (2023).

34 Bozhüyük, K. A. J. *et al.* De novo design and engineering of non-ribosomal peptide synthetases. *Nature Chemistry* **10**, 275-281, doi:10.1038/nchem.2890 (2018).

35 Meyer, J. M. *et al.* Use of siderophores to type pseudomonads: the three Pseudomonas aeruginosa pyoverdine systems. *Microbiology (Reading)* **143 (Pt 1)**, 35-43, doi:10.1099/00221287-143-1-35 (1997).

36 Baunach, M., Chowdhury, S., Stallforth, P. & Dittmann, E. The Landscape of Recombination Events That Create Nonribosomal Peptide Diversity. *Mol Biol Evol* **38**, 2116-2130, doi:10.1093/molbev/msab015 (2021).

37 Bohnenkämper, L., Braga, M. D. V., Doerr, D. & Stoye, J. Computing the Rearrangement Distance of Natural Genomes. *Journal of Computational Biology* **28**, 410-431, doi:10.1089/cmb.2020.0434 (2020).

38 Figueiredo, A. R. T., Özkaya, Ö., Kümmerli, R. & Kramer, J. Siderophores drive invasion dynamics in bacterial communities through their dual role as public good versus public bad. *Ecology Letters* **25**, 138-150, doi:10.1111/ele.13912 (2022).

39 Tsilimigras, M. C. B. & Fodor, A. A. Compositional data analysis of the microbiome:

fundamentals, tools, and challenges. *Annals of Epidemiology* **26**, 330-335, doi:10.1016/j.annepidem.2016.03.002 (2016).

40 Weiss, S. *et al.* Correlation detection strategies in microbial data sets vary widely in sensitivity and precision. *ISME J* **10**, 1669-1681, doi:10.1038/ismej.2015.235 (2016).

41 Faust, K. & Raes, J. Microbial interactions: from networks to models. *Nature Reviews Microbiology* **10**, 538-550, doi:10.1038/nrmicro2832 (2012).

42 Chen, H. M., Guo, W., Shen, J., Wang, L. & Song, J. N. Structural Principles Analysis of Host-Pathogen Protein-Protein Interactions: A Structural Bioinformatics Survey. *Ieee Access* **6**, 11760-11771, doi:10.1109/Access.2018.2807881 (2018).

43 Greensmith, J., Feyereisl, J. & Aickelin, U. The DCA: SOMe comparison. *Evolutionary Intelligence* **1**, 85-112, doi:10.1007/s12065-008-0008-6 (2008).

44 Mirjalili, S. SCA: A Sine Cosine Algorithm for solving optimization problems. *Knowledge-Based Systems* **96**, 120-133, doi:10.1016/j.knosys.2015.12.022 (2016).

45 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589, doi:10.1038/s41586-021-03819-2 (2021).

46 Hampton, H. G., Watson, B. N. J. & Fineran, P. C. The arms race between bacteria and their phage foes. *Nature* **577**, 327-336, doi:10.1038/s41586-019-1894-8 (2020).

47 Kortright, K. E., Chan, B. K., Koff, J. L. & Turner, P. E. Phage Therapy: A Renewed Approach to Combat Antibiotic-Resistant Bacteria. *Cell Host & Microbe* **25**, 219-232, doi:10.1016/j.chom.2019.01.014 (2019).

48 Bodilis, J. *et al.* Distribution and evolution of ferripyoverdine receptors in Pseudomonas aeruginosa. *Environ Microbiol* **11**, 2123-2135,

doi:10.1111/j.1462-2920.2009.01932.x (2009).

49 Kümmerli, R. Iron acquisition strategies in pseudomonads: mechanisms, ecology, and evolution. *Biometals* **36**, 777-797, doi:10.1007/s10534-022-00480-8 (2023).