Trait-trait relationships and tradeoffs vary with genome size in prokaryotes

Sara Beier[1,2*], Johannes Werner[1,3], Thierry Bouvier[4], Nicolas Mouquet[4,5], Cyrille Violle[6]

[1] Leibniz Institute for Baltic Sea Research Warnemünde (IOW), Department of Biological Oceanography, Rostock, Germany
[2] UMR 7621 Laboratoire d'Océanographie Microbienne, Observatoire Océanologique de Banyuls-sur-Mer, Sorbonne Université, Banyuls-sur-Mer, France
[3] High Performance and Cloud Computing Group, Zentrum für Datenverarbeitung (ZDV), Eberhard Karls University of Tübingen, 72074 Tübingen, Germany
[4] MARBEC, Université de Montpellier, CNRS, Ifremer, IRD, Montpellier, France.
[5] FRB – CESAB, Montpellier, France.
[6] CEFE, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France.
Corresponding author*

Abstract

We report genomic traits that have been associated with the life history of prokaryotes and highlight conflicting findings concerning earlier observed trait correlations and tradeoffs. In order to address possible explanations for these contradictions we examined trait-trait variations of 11 genomic traits from ~ 18,000 sequenced genomes. The studied trait-trait variations suggested: (i) the predominance of two resistance and resilience-related orthogonal axes and (ii) at least in free living species with large effective population sizes whose evolution is little affected by genetic drift an overlap between a resilience axis and an axis of resource usage efficiency. These findings imply that resistance associated traits of prokaryotes are globally decoupled from resilience related traits and in the case of free-living communities also from resource use efficiencies associated traits. However, further inspection of pairwise scatterplots showed that resistance and resilience traits tended to be positively related for genomes up to roughly five million base pairs and negatively for larger genomes. This in turn may preclude a globally consistent assignment of prokaryote genomic traits to the competitor - stress-tolerator - ruderal (CSR) schema that sorts species depending on their location along disturbance and productivity gradients into three ecological strategies and may serve as an explanation for conflicting findings from earlier studies. All reviewed genomic traits featured significant phylogenetic signals and we propose that our trait table can be applied to extrapolate genomic traits from taxonomic marker genes. This will enable to empirically evaluate the assembly of these genomic traits in prokaryotic communities from different habitats and under different productivity and disturbance scenarios as predicted via the resistance-resilience framework formulated here.

BOX 1: Glossary

Species: The species concept developed for macroorganisms cannot be directly transferred to asexually proliferating microorganisms. However, sequenced genomes sharing >94% of their average nucleotide identity (Konstantinidis and Tiedje, 2005) or operational taxonomic units (OTUs) delineated e.g. from amplicon sequence variants (ASVs) of taxonomic marker genes approximate natural units that reflect ecological species and will be referred to as species.

Generalist/specialist: The characterization of species along the generalist specialist continuum is based on the organisms' niche breadth, where ecological specialization indicates a limited niche breadth and niche breadth has been defined as the variety of resources, habitats, or environments used by a given species (Sexton et al., 2017). It has recently been pointed out that the unambiguous characterization of species along the specialist generalist continuum is challenging as niche breadth estimations usually refer to a specific range of measured conditions, and for instance, a resource generalists may simultaneously be a temperature specialist (Bell and Bell, 2021). Still, even though not practically measurable, the increasing number of biotic and abiotic conditions in a multivariate niche space under which a species can proliferate can be considered as an increasingly large multidimensional niche breadth of this species.

Colonizer/competitor: Competition-colonization tradeoff models assume that species can occupy a niche as colonizer by efficiently colonizing empty habitat patches or a niche as competitor by outcompeting species within sites (Tilman, 1994; Mouquet et al., 2005).

r/K selection: The theory of r- and K-selection postulates that r-strategists allocate an increased fraction of resources to reproduction under conditions of high density independent mortality (Gadgil and Solbrig, 1972).

Oligotrophic/copiotrophic: Oligotrophic species are able to grow in nutrient poor environments while copiotrophic species require high concentrations of inorganic and organic compounds (Poindexter, 1981).

CSR terminology: The CSR schema sorts species into competitors (C), stress-tolerators (S) and ruderals (R). Ruderals have an advantage in habitats with high disturbance levels, stress tolerators dominate habitats with high stress levels such as low nutrient conditions or extreme temperatures, while competitors profit from their competitive advantage over other organisms in resource-rich habitats (Grime, 1977).

Genomic trait: The term trait has previously been defined as 'any morphological, physiological or phenological feature measurable at the individual level, from the cell to the whole-organism level, without reference to the environment or any other level of organization' (Violle et al., 2007). Here, in order to adapt the use of "traits" to genomic data, genomic traits will be defined as variables that characterize prokaryotic species and which can be delineated from their genome sequence data, such as its genome size. We particularly focus on genomic traits that had been associated in earlier studies with one of the above listed life history traits.

1. Introduction

Prokaryotes contribute largely to the global organic carbon budget (Whitman et al., 1998), are main drivers of major element cycling (Konopka et al., 2015), and are therefore key compounds of earth functioning. However, natural microbial communities are typically extremely diverse and complex, and it remains challenging to predict prokaryote ecosystem functioning and community dynamics in response to environmental changes (Shade et al., 2012; Bardgett and Caruso, 2020). A large body of research highlights the impact of structural community properties such as diversity and species interaction patterns on community functioning (Poisot et al., 2013; Duffy et al., 2017). The effect of structural community properties, however, depends on the characteristics of individual species in a community mediated via their traits and the distributions of these traits have

been shown to influence community functioning and dynamics (Enquist *et al.*, 2015). Microbes harbor an enormous functional versality regarding the number of metabolic functions and pathways they are capable of, which poses a challenge in selecting meaningful functional descriptors to infer overall community functioning. To simplify the assignment of functional attributes to microbial communities it has been suggested to characterize communities based on the distribution of overall life histories rather than focusing on the potentially large number of traits related to specific metabolic pathways (e.g. Bardgett and Caruso, 2020; Malik *et al.*, 2020). Life history traits determine how species allocate available energy among survival, growth and reproduction and are therefore decisive for the overall production and stability of an ecosystem.

Several traditional life history classifications allocate species into binary categories either concerning their response to environmental change or in dependence of the resource availability required for growth (Box 1). For instance, the characterization of species along the specialist-generalist continuum is related to their tolerance against environmental change (Kassen, 2002). It has been pointed out that traits that encompass the capability of species to tolerate or to adapt to changing conditions is related to *resistance*, i.e. the ability of organisms to withstand disturbances (Nimmo et al., 2015).

The r/K selection theory refers to the fraction of resources allocated to reproduction (Gadgil and Solbrig, 1972). R-strategists can be described as opportunists that proliferate fast in response to opportune environmental changes while K-strategists are typically strong competitors that allocate more resources in efficient resource usage rather than growth. The competitor/colonizer classification (Tilman, 1994) differentiates between organisms with competitive advantage in either empty or densely populated habitats: as a consequence the best competitor will outcompete all other species at low disturbance rates, while species with high colonization efficacy will dominate ecosystems at high disturbance rates (Hastings, 1980). Both, r-strategists as well as species with high colonization efficacy should be characterized by short lag-phases and fast growth rates enabling them to respond fast to changing environmental conditions or rapidly colonize empty habitats. It has been proposed that traits that indicate rapid reproduction or strong re-colonization capabilities are linked to an organism's *resilience*, defined as its capacity to recover after a disturbance (Nimmo et al., 2015).

Resistance and resilience related traits represent two distinct components of ecological units (i.e. populations, communities or ecosystems) that determine their response to disturbances (Shade et al., 2012; Nimmo et al., 2015). High levels of resistance and resilience are both associated with costs: Dall and Cuthill (1997) suggested costs that are inherent with a generalist's life-style can include performance reductions from having more ecological variables to monitor. High potential growth rates in microbes often come at the expense of low resource usage efficiency (Fierer et al., 2007; Roller et al., 2016). In agreement with this, tradeoffs between resilience and resistance have been observed in microbes, where species can allocate transcriptional resources in the expression of stress resistance genes at the cost of a reduced expression of genes that promote growth (Ferenci,

2016). Such tradeoffs between resistance and resilience have been described not only at the species but also at the community level (e.g. de Vries and Shade, 2013; Garcia *et al.*, 2020; Piton *et al.*, 2021).

Resource availability can also shape the distribution of resistance or resilience related life histories due to the above mentioned potential costs that are linked to the response to disturbance. Traditionally, in microbial ecology, large emphasis has been put in sorting species according to the nutrient levels required for their growth, in either oligotrophic or copiotrophic species (Koch, 2001; Fierer, 2017). It has been suggested that copiotrophic species that require high nutrient levels for growth can commonly be characterized as generalists (Christie-Oleza et al., 2012) and / or r-strategists that react fast to nutrient pulses (Fierer, 2017). However, although some binary life history classifications may overlap, these terms should not be used interchangeably as they are defined differently (Box 1).

The more complex CSR theoretical schema, that goes beyond a binary classification of life-history traits, sorts species into three classes of ecological strategies (competitor vs. stress-tolerator vs. ruderal species) depending on their location along two major environmental gradients: disturbance and productivity (Grime, 1977). This framework was originally developed in plant ecology and integrates different dimensions in the interplay of resistance, resilience and resource availability. Several studies have suggested applying the CSR framework in the field of microbial ecology (e.g. Krause *et al.*, 2014; Fierer, 2017). However, it has recently been claimed that the reliance of heterotroph microbial organisms on external carbon and energy sources distinguishes them from autotroph plants and complicates the application of the CSR framework to microbial communities dominated by heterotrophs. The CSR framework has therefore been adapted to microbial ecology by classifying high yield (Y) – resource acquisition (A) – stress tolerator (S) categories (Malik et al., 2020). Here Y refers to species with high carbon use efficiency, A replaces the plant competitor strategy because microbial competition is mainly over resources and the S strategy refers to species that are adapted to stress exposure due to deviations from ambient.

Although a number of prokaryote representatives from naturally abundant and relevant lineages have recently brought into culture (e.g. Marshall and Morris, 2013; Neuenschwander *et al.*, 2018) the majority of prokaryotic species remains uncultured (Steen et al., 2019). Accordingly, the physiologically measured species characteristics that inform about the life-histories of prokaryotes are only available for a small minority of cultured representatives. However, a growing number of genome sequences from so far uncultured prokaryotic strains are available. In prokaryotes the myriad of responses to environmental change, such as changes in resource availability, abiotic stressors or biological interactions, is coded in their genomic material. Thus, the DNA of a population contains the information of its performance under all possible conditions or its coverage of the n-dimensional niche space. However, an incomplete understanding of the complexity of gene regulation from genome sequence data currently limits our capacity to predict gene expression patterns and consequently the specific phenotype of a strain in a given environment.

Still, it has been demonstrated via machine learning approaches that the presence of functional genes within genomes was tightly linked to the ecological niche occupied by the corresponding prokaryotes, explaining ~50% of their niche variability in a high-dimensional niche space (Alneberg et al., 2020). A more simplified possibility to characterize prokaryotes via their genomic information is to extract and evaluate simple parameters, so-called genomic traits, from their genomes, such as the presence or copy number of a specific gene or genome size. Although this option may lack accuracy due to the high degree of simplification, various studies found empirical and statistically validated evidence for correlations between certain genomic traits and the binary life history strategies listed in BOX1.

Hereafter, we summarize literature about genomic traits that were associated with statistical support to life history strategies with a focus on binary classifications and earlier reported correlation among these traits. In order to resolve some conflicting findings concerning earlier proposed genomic trait assignments to life histories, their pairwise correlations and tradeoffs, we inspected the covariation patterns among 11 genomic traits that are considered as life history proxies, from ~ 18,000 sequenced genomes representing ~9,000 species in a multivariable trait space. As a result, we categorized the studied genomic traits within a resistance-resilience based framework. This framework resolved conflicts and may be used in future studies to address community dynamics under different disturbance and productivity scenarios in variable habitats.

We demonstrated that all analyzed genomic traits featured significant phylogenetic signals and proposed a simple tool to empirically evaluate the theoretical predictions made here via the extrapolation of genomic traits based on taxonomic marker genes from uncultured taxa.

2. Genomic traits as life history proxies in prokaryotes

A number of earlier studies has addressed prokaryotic genome features that give indirect evidence for physiological characteristics and the life-history strategies of the corresponding organisms (Figure 1). Particularly the classification as generalist, measured as an organism's physiological versatility or its ability to colonize multiple habitat types, has been related to a range of different genomic characteristics. This includes features such as a large genome size (Barberan et al., 2014; Bentkowski et al., 2015; Cobo-Simon and Tamames, 2017; Sriswasdi et al., 2017) and a high fraction of regulatory genes (Parter et al., 2007; Kostadinov et al., 2011). It indeed seems reasonable that organisms that are able to live and proliferate under variable conditions need more genes for sensing or for coping with a range of different growth conditions. They further need a larger number of transcription factors (%TF, i.e. regulatory genes) to regulate genes that are alternatively expressed depending on the prevailing environmental conditions. Based on the same considerations it has been suggested that a high frequency of genes acquired via horizontal gene transfer (%HGT) would increase the versatility of prokaryotes and allow them to grow in a larger number of environments (Takemoto, 2013).

It has further been argued that the life style as a generalist requires enhanced growth rates. In agreement with this, high growth rates (Freilich et al., 2009) as well as elevated codon usage biases (CUB, Botzman and Margalit, 2011) have been associated with a generalist life style. The CUB describes the phenomenon that synonymous codons are used unevenly among genes, where genes coding for highly expressed proteins are enriched in codons that reflect the taxon-specific tRNA pool. The CUB has been shown to be specifically pronounced in fast-growing organisms and was therefore interpreted as a genomic feature that correlates with maximal growth rates of prokaryotes (Vieira-Silva and Rocha, 2010).

	Generalist /Specialist	r/K-strategist	Colonizer/ competitor	Copiotrophic/ oligotrophic
Genome size	*Barberan et al. 2014 (soil) *Bentkowski et al. 2015 *Cobo-Simon and Tamames 2017 *Sriswasdi, et al. 2017			*Lauro et al. 2009 (marine) *Okie et al. 2020 (freshwater) *Chen et al. 2020 (soil) *Leff et al. 2015 (soil)
%TF	*Parter et al. 2007 *Kostadinov et al. 2011 (marine)			
RRN	*Cobo-Simon and Tamames 2017		*Nemergut et al. 2016 (soil, sediment) *Shrestha et al. 2007 (soil)	* Lauro et al. 2009 (marine) *Okie et al. 2020 (freshwater) *Chen et al. 2020 (soil)
% HGT	*Takemoto 2013			
Prophages				* Lauro et al. 2009 (marine)
%GC				*Hellweger et al. 2018 *Okie et al. 2020 (freshwater)
CUB	*Botzman and Margalit 2011			*Chen et al. 2020 (soil) *Okie et al. 2020 (freshwater)
Maximal growth rate (physiologically measured)	*Freilich et al. 2009	*Freilich et al. 2010 *Roller et al. 2016		

Figure 1: Overview of earlier literature that linked genomic traits or the physiologically measured maximal growth rates of prokaryotes to their life-history traits. If studies refer to organisms from a specific ecosystem, this is given in brackets. Green (or red) fields indicate studies that found positive (or negative) relationships ($p < 0.05$) between the listed genomic feature and a life style as generalist, r-strategist, colonizer or copiotroph. In one study no p-value was reported, but mathematical evidence for increased environmental tolerance of species with large genomes was provided via the computational simulation of evolutionary processes (Bentkowski et al., 2015). Gray fields indicates studies where neither a significant relationship and not even a positive or negative trend was detected ($p>0.5$).

In contrast, the number of rRNA gene copies (RRN) in prokaryote genomes could not be associated significantly to the classification as either generalist or specialist (Cobo-Simon and Tamames, 2017). Instead, the RRN decreased significantly during the community succession after environmental disturbances (Shrestha et al., 2007; Nemergut et al., 2016). A high RRN at early successional stages can be interpreted as evidence for a life style as colonizer and corroborates observations that associated high RRN with short lag phases and high growth rates (Klappenbach et al., 2000; Stevenson and Schmidt, 2004; Vieira-Silva and Rocha, 2010; Roller et al., 2016).

A significant correlation between growth rate and competitive ability (Klappenbach et al., 2000; Stevenson and Schmidt, 2004; Vieira-Silva and Rocha, 2010) or carbon use efficiency (Roller et al., 2016) demonstrated that prokaryotes can generally be well classified along the r/K-strategist continuum. Organisms with short duplication times and accordingly high CUB are therefore likely to be r-strategists.

It had been hypothesized that the evolution towards GC depleted genomes is an adaption to nutrient poor conditions because GC pairs contain one more nitrogen atom compared to AT pairs (Giovannoni et al., 2005; Grzymski and Dussaq, 2012). Hellweger and colleagues (2018) demonstrated that beside other mechanisms, such as mutation biases, particularly N limitation but also C-limitation impacted the evolution of genomes towards GC depletion. Accordingly, low GC content can be interpreted as a genomic trait that is indicative of an oligotrophic life style in prokaryotes. Alternatively, may a high frequency of GC pairs that have three hydrogen bounds compared to only two hydrogen bounds in AT pairs enhance the resistance of cells with high GC content against some specific stressors, such as heat or desiccation stress. Probably for this reason, a higher GC content was found in arid, nutrient-poor soils exposed to heat and dehydration stress than in soils with higher nutrient contents but lower stress exposure (Chen et al., 2020).

Lauro et al. (2009) suggested that high RRN, an elevated number of prophages and a large genome size are more common in marine copiotrophic than in marine oligotrophic bacteria. A recent study demonstrated in agreement with this that nutrient additions to oligotroph lake water induced a significant increase of community mean genome size, RRN, CUB and GC content (Okie et al., 2020). In contrast, nutrient additions to soil environments resulted in the selection of prokaryotes with smaller genomes (Leff et al., 2015) and a recent comparison of microbial communities in nutrient rich versus nutrient depleted soils did not reveal a significant difference in their average genome sizes (Chen et al., 2020). In addition, enhanced CUBs were found associated with copiotrophic microbial communities in soil environments (Chen et al., 2020).

3. Trait-trait correlations in the light of physiological constraints

A number of earlier studies explored pairwise correlations among the above-mentioned genomic traits or with the organism's physiologically measured maximal growth rate. Most of these studies found significant positive correlations e.g. between genome size and %TF, RRN and growth rates (Figure 2). As a consequence, if large genome sizes and a high %TF are indicative of a generalist life history and a high RRN and elevated growth rates indicate r-strategists and/or colonizers, a generalist's life style should be associated with a simultaneous life style as r-strategist and colonizer. This inference however contradicts the often observed and above outlined tradeoff between resistance and resilience in microbial communities (e.g. Ferenci, 2016; Garcia et al., 2020; Piton et al., 2021). A resistance-resilience tradeoff would instead imply a negative correlation between resistance associated traits, such as genome size or transcription factors versus resilience associated traits, such as rRNA gene copy numbers or growth rates. We want to point out that some studies did not find significant positive relationships between genome size and growth rates and although the observed correlation coefficients were positive, they were very weak (Vieira-Silva and Rocha, 2010; Westoby et al., 2021b). Still, also in these cases no negative correlations were detected as one would expect from a tradeoff between resistance and resilience.

	Genome size	%TF	Maximal growth rate (physiologically measured value)
%HGT	*Cordero and Hogweg 2009b		
RRN	*Klappenbach et al. 2000		*Stevenson and Smith 2004 *Vieira-Silva and Rocha 2010 *Roller et al. 2016
Prophages			*Touchon et al. 2016
Genome size		*Van Nimwegen 2003 *Molina and van Nimwegen 2009 *Charoensawan et al 2010 *Grilli et al. 2012 *Cordero and Hogweg 2009a	*Touchon and Rocha 2007 *Freilich et al. 2009 *Vieira-Silva and Rocha 2010
%TF			*Freilich et al. 2009
CUB			*Vieira-Silva and Rocha 2010 *Long et al. 2021 *Weissman et al. 2021

Figure 2: Earlier reported pairwise correlations among genomic traits and /or physiologically measured maximum growth rates. Dark green fields indicate positive significant correlations ($p < 0.05$) and light green field indicates a positive trend below the significance level ($0.5 > p > 0.05$).

In order to elucidating conflicting relationships between these earlier observations and the resistance-resilience tradeoff, we here examined covariations among multiple genomic traits. These traits were extracted from the genomic material of sequenced prokaryotic genomes available via the JGI/IMG platform (https://img.jgi.doe.gov/, Chen et al., 2021; Mukherjee et al., 2021). For downstream analyses we considered those genomes that are integrated into the default reference database in the PICRUSt2 software (Douglas et al., 2020) and which infers the genomic content of uncultured prokaryotes from closely related genomes via taxonomic marker genes. In total 17,856 of the 20,000 genomes that are integrated in the PICRUSt2 default reference phylogenetic tree could be downloaded via the JGI/IMG database (date of download: 24.08.2021). These genomes cover a broad phylogeny (73 phyla; 172 classes; 382 orders; 762 families; 2,669 genera; 8,847 species) and include genomes from yet uncultured candidate phyla (Table S1).

Additionally to the genomic traits presented above (Figure 1), we also considered genome level parameters for gene richness (i.e. the number of different genes in a genome) and the gene duplication level (i.e. the average number of gene copies per gene in a genome). This is because the genome size can increase due to the integration of new genes or due to the duplication of already existing genes and either one of these two mechanisms may have different consequences for the resistance level of prokaryotes: on the one hand may different copies of the same gene be expressed alternatively in response to environmental change, for instance when they differ in their pH optimum (Miyashita et al., 1991) and therefore contribute to resistance levels analogously to two different genes that are can be expressed alternatively. On the other hand may multi copy genes have an adaptive effect to stressful environment due to an enhanced dosage effect (Kondrashov, 2012). Some genomic traits are provided directly by the JGI genome statistics, while others were computed from the genome sequences. An inspection of the JGI/IMG provided RRN data indicated inaccuracies for this specific genomic trait (Supplementary Figures S1/S2). We

therefore extrapolated RRNs for the JGI/IMG reference genomes affiliating with the same genus of genomes stored in the Ribosomal RNA Operon Copy Number Database (rrnDB, https://rrndb.umms.med.umich.edu/) which provides curated RRN data (Stoddard et al., 2015). In order to remove phylogenetic redundancy from the dataset we aggregated mean trait values for all genomes at the species level. The affiliation of genomes at the species level was determined via the GTDB taxonomy (Chaumeil et al., 2020). In the case of genomes that could not be annotated at the species level to the GTDB taxonomy, we considered genomes sharing >94% of their average nucleotide identity as affiliating to the same species (Konstantinidis and Tiedje, 2005). Scripts that were used for determining the here presented genomic traits and the assignment of genomes to species are available via GitHub (https://github.com/sarabeier/genomic.traits) and a trait table is provided in the supplementary material (Supplementary Table S1).

Covariation patterns among the above introduced genomic traits in the multivariate trait space were illustrated via a principal component analysis (PCA, Figure 3). The first two principal components explained > 65% of total variability, while the remaining principal components each contributed < 10% to total variability (Supplementary Figure S3). A random removal of 1, 2, 3 or 4 variables from the PCA demonstrated that the covariation patterns among the remaining genomic traits patterns stayed robust (Supplementary Figure S4). An inspection of individual pairwise correlation strengths (Figure 4) revealed that the absolute spearman rank correlation index rho ranged from 0.04 to 0.92. All correlations were significant due to the large number of included genomes (8,847 species, p-value adjusted via Bonferroni correction for multiple comparisons: Padj < 0.05). Below we will highlight some of the 35 pairwise correlations whose strength or shape displayed via local fitting can be interpreted well in the light of possible physiological constraints.

Genome size, gene duplication level, gene richness and the %TF aligned approximately along the first principal component (PC1) while covarying positively with each other (Figure 3). An inspection of the pairwise scatter plots revealed a pronounced linear relationship between the genome size and the gene duplication level (Figure 4). In contrast, the increase of gene richness along with genome size rather followed a saturation curve. Apparently, the enrichment of genomes with new genes occurred only until a certain genome size threshold (~ five million base pairs), after which a further genome size increase was mostly due to the duplication of already present genes. A similar saturation pattern was observed for the pairwise correlation of the fraction transcription factors against genome size (Figure 4): a steep positive relationship was apparent approximately up to the genome size until which gene richness increased, while after that threshold a less pronounced increase was observed. Obviously, acquiring new genes requires a stronger enrichment in regulatory genes, than does the duplication of genes. This observation implies that multicopy genes are often, although not explicitly, under the control of the same regulatory operon and can in this case not be expressed alternatively in response to changing conditions. An increasing tolerance to environmental changes due to a genome size increase should therefore, in the case of larger genomes, be primarily due to the dosage effect of replicated genes. The shape of

the above reported pairwise relationships accordingly illustrates the physiological mechanisms that link the variables for %TF and genome size depending on whether genome size increases due to the acquisition of new genes or gene duplication. The positive covariation among all four traits underlines their common association with species classifications along the specialist-generalist continuum suggested in literature for genome size and the %TF (Figure 1).

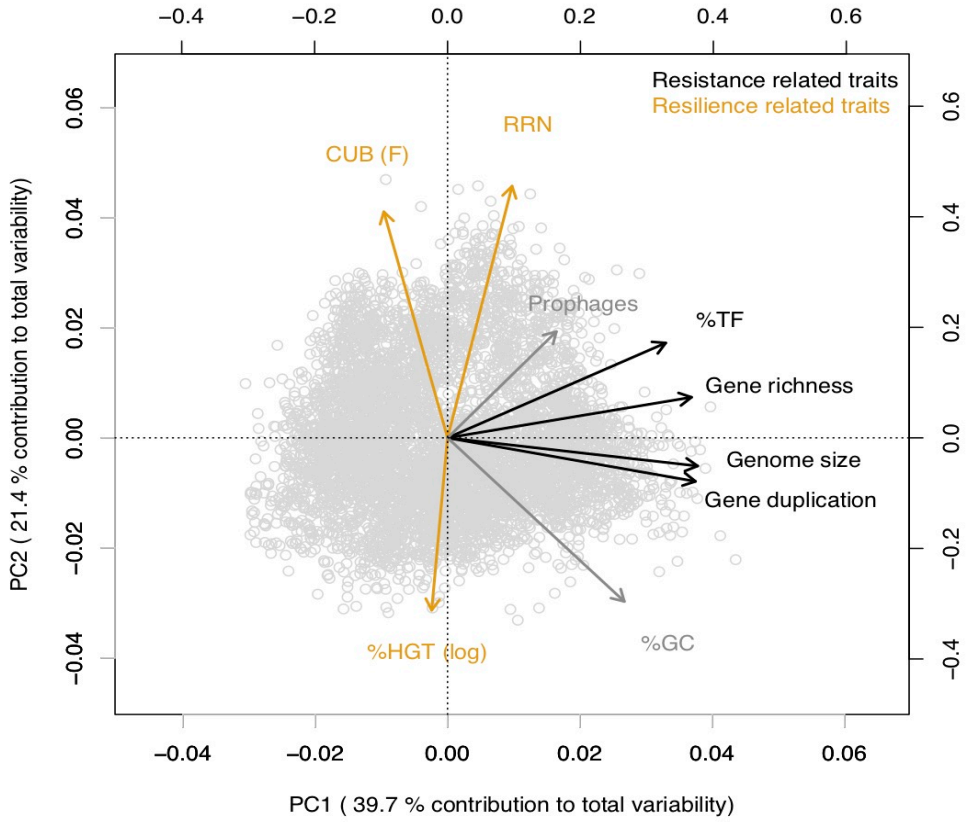

Figure 3: Principal component analysis illustrating covariations among genomic traits from 17,856 sequenced prokaryotic genomes available via the JGI/IMG platform (https://img.jgi.doe.gov/) that were averaged at the species level (Supplementary Table S1). 5,823 out 8,847 species shared values for all considered traits and were included in the analyses. The value for %HGT was log(x+0.001) transformed to approximate a normal distribution. We have chosen to display the CUB parameter F (Vieira-Silva and Rocha, 2010) instead of generation time estimations that can be delineated from the CUB, because F is unambiguously defined for all genomes. In contrast, generation time estimations have been suggested to be inaccurate for genomes with large CUB values.

Earlier studies suggested moreover that the % HGT, the CUB and the RRN were either linked to generalist-specialist classifications or correlated positively with genome size (Figure 1, 2), which would imply an alignment of these variables along PC1. Yet, these variables aligned rather along the second principal component (PC2) and were accordingly only weakly correlated to either of the four above described resistance related variables (Figure 3,4). Indeed, a weak correlation

between genome size and growth rates that had been observed in several earlier studies (Figure 2, Vieira-Silva and Rocha, 2010; Westoby et al., 2021b) was supported by the likewise weak overall correlation between genomes size and CUB (rho=-0.12) in our analysis.

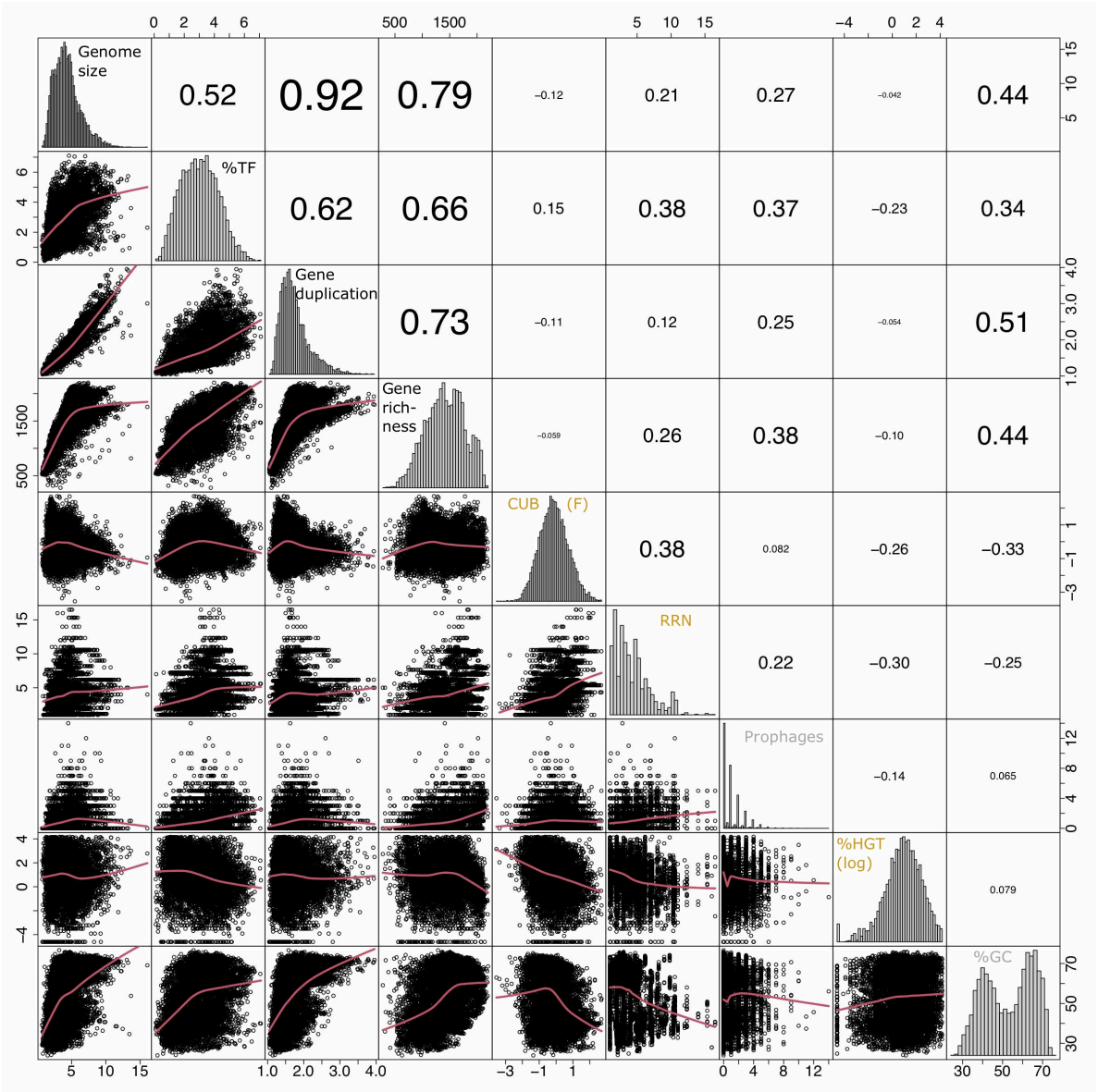

Figure 4: Overview of pairwise correlation patterns among genomic traits from 8,847 species delineated from sequenced prokaryotic genomes available via the JGI/IMG platform (https://img.jgi.doe.gov/) and created via the R command chart.Correlation in the R package PerformanceAnalytics (v2.0.4). The lower panels illustrate pairwise correlation plots fit via loess smoothing statistics with the smoothing parameter f set to 2/3 and the number of robustness iterations set to 3. The value in the upper panels indicate strength of the pairwise correlations (rho, Spearman rank correlation). The diagonal panels illustrate the distribution of the input variables. The CUB was estimated via the variable F as detailed elsewhere (Vieira-Silva and Rocha, 2010). The value for %HGT was log(x+0.001) transformed to approximate normal distribution. The variables RRN and prophages were not transformed as no transformation option improved the fit to normal distribution. The raw values of all data are given in Supplementary Table S1.

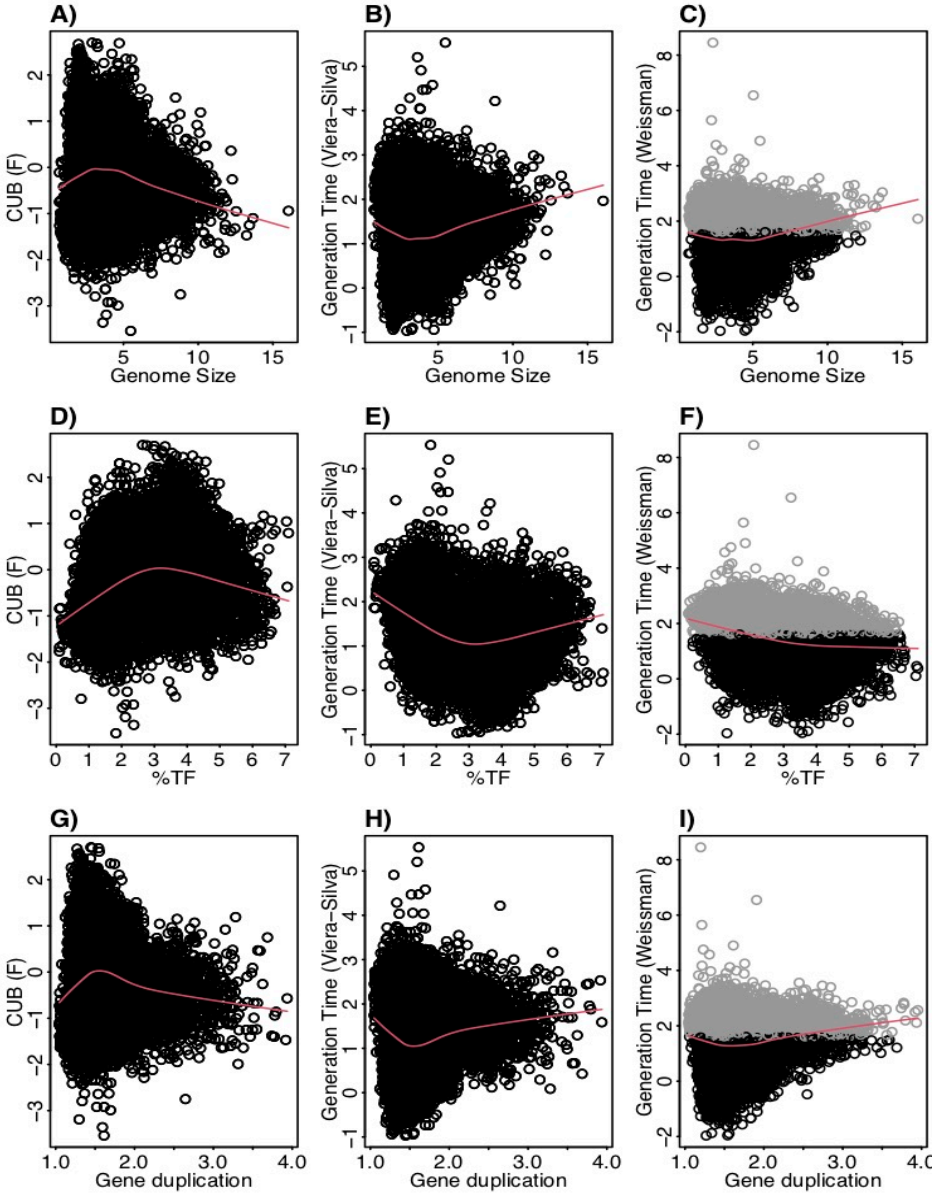

Figure 5: Scatterplots displaying pairwise relationships between the codon usage bias (F, Vieira-Silva and Rocha, 2010) the generation time (log-transformed) estimated as detailed in Vieira-Silva et al. (2010) and the generation time (log-transformed) estimated using the R package gRodon (v0.0.0.9000, Weissman et al., 2021) against genomes size, %TF and gene duplication. Genomes with low codon usage bias resulting in predicted generations times (gRodon) exceeding 5 h (highlighted in gray) represent species with generations times >5 h, while the exact value is inaccurate. The absence of genomes with elevated generation times at the extremes of the value ranges for the displayed resistance related traits is accordingly independent from these inaccuracies.

However, when inspecting the pairwise scatterplots in more details, local fitting suggested a hump-shaped relationship between CUB and genome size, that to our knowledge has not yet been described elsewhere: a positive trendline occurred until a genome size of roughly five million base pairs, after which the direction turned into a negative trendline (Figure 4). Obviously, the observed

hump-shaped trendline appeared mainly due to the absence of very small as well as very large genomes with high CUB values. In contrast, genomes with intermediate genome sizes were associated with almost the full range of possible CUB values, resulting in a kind of pyramid-shaped distribution of data points in the pairwise scatterplots. A recent study highlighted that the relationship between CUB values and minimal generation times is inaccurate for prokaryotes with minimal generation times > 5h (Weissman et al., 2021), which complicates the ecological interpretation of CUB values. We therefore want to point out that the observation of prokaryotes with very large or small being exclusively slow growers was not impacted by these inaccuracies in the relationship between CUB values and generation times (Figure 5).

Physiological reasons for a non-monotone relationship between growth rate and genome size could be due to opposing mechanisms that prevail under different genome sizes ranges: on the one hand, it had been proposed that an increasing proportion of genes involved in metabolic pathways, that was observed along with increasing genome sizes (Konstantinidis and Tiedje, 2004) causes higher metabolic rates. Consequently, an increased availability of energy supply via ATP should lead to enhanced growth rates (DeLong et al., 2010). In this scenario, the growth rate of very small genomes would be limited by the available energy. On the other hand, it has been argued that cells with small genomes feature higher growth rates than cells with large genomes, because they can initiate a new replication cycle before the previous rounds have been finished (Vieira-Silva and Rocha, 2010). Furthermore, the proportion of genes encoding the translation of mRNA into proteins, DNA replication or cell division is reduced in large genomes (Konstantinidis and Tiedje, 2004), which could lead to reduced growth rates. We propose that these latter two physiological constraints limit the growth rates of species with very large genomes. It has been discussed earlier that genome size reduction in prokaryotes can on the one hand be induced via genetic drift, a mechanism that should affect particularly intracellular parasites with small effective population sizes. Conversely, genome streamlining due to adaptive selection occur typically in response to nutrient limitation in aquatic systems (Giovannoni et al., 2014). Still, the above highlighted physiological consequences of genome size reduction should apply to all small genomes. Indeed, both, intracellular parasites as well free living oligotrophic organisms with small genome sizes are typically characterized as slow growing organisms that are sensitive to environmental change (Couturier and Rocha, 2006; Joseph and Goebel, 2007; Parter et al., 2007; Giovannoni et al., 2014). We argue based on these considerations that genome size itself causes the above described physiological constraints and characteristics, independent from the evolutionary forces selecting for prokaryotes with enlarged or reduced genome size.

A trend for hump-shaped relationships and / or pyramid-shaped distribution of data points in the pairwise scatterplots was not only visible if plotting genome sizes against CUB values: several other pair-wise comparisons between traits associated with the specialist-generalist continuum (genome size, gene duplication level and the fraction of transcription factors) versus and traits that covaried positively with CUB (RRN, number of prophages) displayed similar profiles (Figure 4).

The positive covariation between the number of prophages with the RRN and, to a certain extent, with CUB and, accordingly, with maximum growth rate has already been outlined earlier (Touchon et al., 2016). It has been argued that a higher number of prophages in potentially fast growing bacteria is a consequence of their opportunist life style that provides more variable growth states and resources for the production of virions. This in turn was suggested to favor lysogeny and therefore the presence of prophages (Touchon et al., 2016). However, the number of prophages was placed in-between the PCA axes PC1 and PC2 and also covaried positively with %TF and gene richness. This seems reasonable as the integration of prophages into the genome adds more genes to the genome. The lateral transfer of genes via prophages is one of several mechanisms leading to HGT. It is therefore remarkable that we found a negative correlation between the number of prophages with the %HGT (Figure 3). We conclude from this observation that the overlap between the detected HGT events and prophages was low. This is possibly due to a high host-specificity of bacteriophages, which would lead to gene transfer only between closely related strains. We assume that the phylogeny-based method that was applied to detect the %HGT in the JGI genome statistics and that identifies genes with phylogenetic foreign origin in genomes, does not resolve well genes acquired from close relatives (Markowitz et al., 2010).

The %HGT covaried not only negatively with the number of prophages, but also with CUB (Figure 3). Such negative covariation has to our knowledge not yet been reported. However, the discovery that the %HGT events in genomes is connected to their optimal growth temperature (Gophna et al., 2015) may be linked to the above mentioned negative covariation, as species with high growth optimum tend to have comparably high maximal growth rate (i.e. low minimal generation time; Vieira-Silva and Rocha, 2010). A high degree of sub-species genome plasticity in the typically slow growing members of the SAR11 clade (Ward et al., 2017) is in agreement with this observation. It indeed seems reasonable that a fine tuned genetic replication machinery is necessary for achieving high maximal growth rate, which may easily be impeded by a high fraction of genes of foreign origin. Our data therefore point to a possible tradeoff between high growth rates and the ability of genomes to stably integrate foreign DNA from different phylogenetic origin. It needs though to be considered that the phylogeny-based methods to detect %HGT events may overestimate HGT events for genomes with few closely related genomes that belong to the same phylogenetic group. The above suggested tradeoff should therefore be confirmed in other datasets, possibly in combination with other approaches to detected HGT events.

The GC content did not exhibit a pronounced covariation with any other genomic trait. However, a certain level of covariation could be detected with genome size and the gene duplication level or with CUB and the fraction of horizontally transferred genes (Figure 3). The limited covariation of the GC content with any other trait could be due to the complex selection forces, including on the one hand nutrient availability and on the other hand exposure to heat or desiccation stress that in combination drive the GC content evolution (Hellweger et al., 2018; Chen et al., 2020).

402 ## 4. Synthesis

403 Based on the covariation patterns observed here in combination with findings from earlier studies,
404 we propose that the %TF, genome size, gene richness and gene duplication represent resistant
405 related traits and align approximately with the first principal component of the PCA (Figure 3). In
406 contrast, traits including CUB, RRN and %HGT that according to earlier studies or in agreement
407 with our analyses exhibited covariations with growth rates or lag phases and that hence could be
408 considered as resilience related traits aligned rather with the second principal component of the
409 PCA (Figure 3).

410 We did not assign %GC to resistance or resilience related traits as it was positioned in-between
411 PC1 and PC2 and did not align closely to any of the other traits. Although a high GC content can
412 protect cells from heat or desiccation stress resistance, it can furthermore not be considered as a
413 general resistance trait enabling for instance tolerance against changes in the resource type supply.
414 Likewise, we did not assign the number of prophages to either resistance nor resilience related
415 traits based on its position in-between PC1 and PC2. Apart from an earlier suggested link between
416 an elevated number of prophages with high growth rates and an opportunistic lifestyle (Touchon
417 et al., 2016) an increasing number of prophages will also increase the genome size and other
418 resistance related traits.

419 The not fully congruent overlap between the genomic traits that we assigned to either resistance
420 or resilient related traits suggests that these traits cover different aspects of resistance and
421 resilience, such as increased growth rates versus shorter lag phases in the case of resilience related
422 traits.

423 The roughly orthogonal position of several resistance and resilience traits in prokaryotes across
424 whole range genome sizes reflects the results of some previous studies, which suggested that
425 genome size and growth rate are not related (Vieira-Silva and Rocha, 2010; Westoby et al., 2021b).
426 However, the analyses presented here were based on a database that exceeded those used in these
427 former studies by one to two orders of magnitude. An inspection of pairwise trait correlations
428 indicated the presence of non-monotone relationships between several resistance and resilience
429 related genomic traits that to our knowledge were not yet reported in earlier studies: for instance,
430 genomes with a size up to roughly four million base pairs featured according to a local fitting
431 approach a positive relationship with CUB, while after approximately five million base pairs it
432 turned to a negative relationship (Figure 4, 5). Furthermore, genomes up to this threshold continued
433 to increase due to a combination of newly acquired genes and gene duplication events, while larger
434 genomes increased rather due to gene duplication events (Figure 4). Last but not least aquatic
435 environments are typically characterized by genomes smaller than four million base pairs, while
436 soil environments usually harbor genomes larger than five million base pairs (Giovannoni et al.,
437 2014). Correlation analyses of partial datasets exhibited in agreement with the trendlines obtained
438 by local fitting significant positive correlations between several resistance versus resilience related

traits, if considering genomes up to a size of four million base pairs. The opposite was true in most cases if considering genomes larger than five million base pairs (Table 1). Results from these partial correlation analyses (Table 1) highlight that non-monotone relationships are sensitive to the range of data points included. We argue that non-monotone relationships are the reason for contradicting findings e.g. the relationship between genome size and growth rates, that have been described as either positively related (Freilich et al., 2009; DeLong et al., 2010) or as largely unrelated dimensions (Vieira-Silva and Rocha, 2010; Westoby et al., 2021b). Furthermore, the previously reported superlinear positive correlation between genome size and HGT events (Cordero and Hogeweg, 2009b) or the assignment of generalist species to high %HGT, large CUBs or increased growth rates (Figure 2), which was not supported by our findings, could be due to the specific set of genomes that were used in the respective analyses.

Table 1: Correlation coefficients between selected resistance and resilience related traits for species with small and large genomes

Comparison (resistance vs resilience)	Genomes < 4 Mbp[a] (e.g. aquatic communities)	Genomes > 5 Mbp[a] (e.g. soil communities)
Genome size vs CUB	0.05*	-0.28***
Genome size vs RRN	0.09***	-0.02
Gene duplication vs CUB	0.12***	-0.23***
Gene duplication vs RRN	0.07***	-0.16***
%TF vs CUB	0.44***	-0.12***
%TF vs RRN	0.83***	0.23***
Genome size vs CUB	0.05*	-0.28***
Genome size vs RRN	0.09***	-0.02
Gene duplication vs CUB	0.12***	-0.23***

[a]Correlation coefficients (rho, Spearman rank correlation). Bonferroni adjusted p-value: *Padj<0.05; ***Padj < 0.001.

Noticeably, not all partial pairwise correlations were strong or resulted in a differential correlation profile for larger and smaller genomes (Table 1). Still, we believe that as a consequence of the inconsistent correlation patterns reported here, positive relationships between resistance and resilience are more likely to occur in aquatic habitats: a positive covariation of genomic traits that we here classified as resistance and resilience related traits had been recently observed in an aquatic fertilization experiment (Okie et al., 2020). Instead, tradeoffs between resistance and resilience should be more likely in soil habitats. Indeed, the above outlined tradeoff between functional resistance and resilience has to our knowledge mainly been observed in soil environments (e.g., Garcia et al., 2020; Piton et al., 2021). This is in agreement with the negative correlation between several resistance and resilience related traits particularly among larger genomes. Habitat specific PCA patterns with genomes originating from soils, aquatic habitats or from the digestive tract furthermore supported our findings: the position of resistance versus resilience related traits shifted with increasing size of the included genomes from positive towards negative covariations along the first principal component (Figure 6). The here observed shifts may be even more extreme in natural habitats, as a considerable part of genomes in the reference

database originated from cultured strains. Particularly in aquatic habitats these are likely biased towards larger genomes while oligotroph large genomes in soil habitats might be underrepresented, thereby resulting in genome size distributions that are not typical for these habitat types.

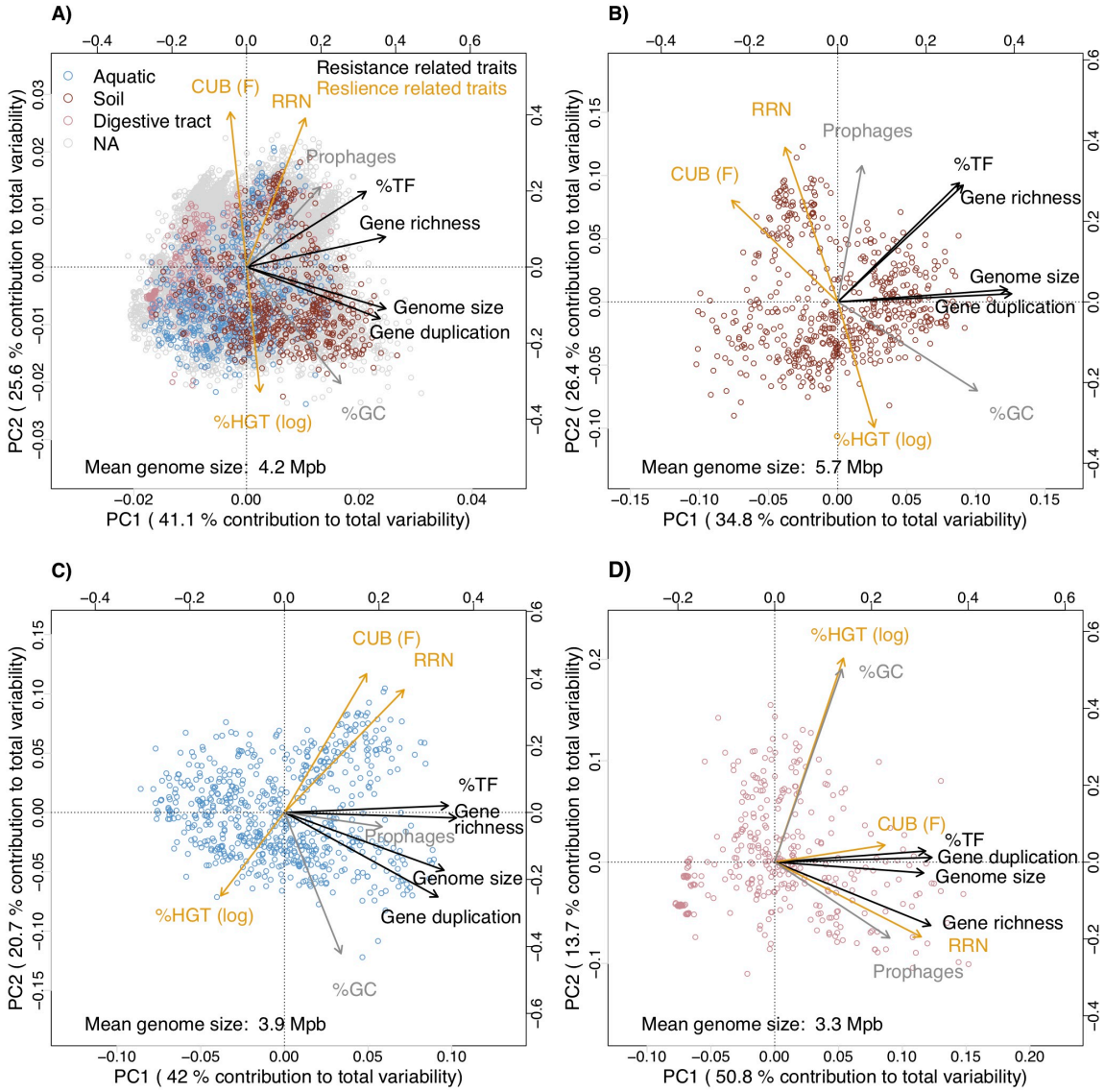

Figure 6: Principal component analyses illustrating covariations among genomic traits from 17,856 JGI/IMG prokaryotic genomes dependence on the habitat type. A) All genomes. B) Genomes originating from soil habitats. C) Genomes originating from aquatic habitats. D) Genomes originating from the digestive tract. In order to display three habitat types from which individual genomes originated, we did not aggregate the genomes at the species level. The three habitat types were determined via text search from the habitat information available via the JGI/IMG database (Supplementary Table S1; ignore.case=TRUE): all genomes containing the strings ‘soil’ or ‘rhizosphere’ in the habitat description where classified as originating from soil habitats; all genomes containing the strings ‘aquatic’ or ‘marine’ or ‘water’ where classified as originating from aquatic habitats; all genomes containing the strings ‘oral’ or ‘stomach’ or ‘gut’ or ‘intestinal’ or ‘feces’ were classified as originating from the intestinal tract. The remaining genomes were not further classified.

Literature suggests consistently that increasing resource availability leads to the selection of fast growing opportunists (i.e. r-strategists) with decreased resource usage efficiency (Stevenson and Schmidt, 2004; Fierer et al., 2007; Roller et al., 2016). Prokaryotes with high resource usage efficiency and simultaneously low maximal growth rates are usually observed in oligotroph environments with the possible exception of obligate intracellular living bacteria: as mentioned above is the evolution of these organism largely affected by genetic drift and they typically live in nutrient rich environments, but tend to feature low growth rates (Couturier and Rocha, 2006; Joseph and Goebel, 2007). We suggest copiotrophic-oligotrophic classifications, at least in free living communities with large effective population sizes that are little impacted by genetic drift, to be aligned with the dimension of resilience and resource usage efficiency. This implies that classifications of free living communities along the copiotrophic-oligotrophic axis are analogously to the resistance axis decoupled from the dimension of resistance related classifications if considering species along the whole range of genome sizes. However, contrasting relationships can occur for instance in aquatic versus soil habitats. In agreement with these theoretical considerations, previous studies actually reported the selection of larger genomes after nutrient addition in aquatic habitats and the opposite in soil habitats (Figure 2). In line with this, a recent study indicated based on habitat dependent covariation patterns between genome size and %GC that genome reduction in soil habitats may not be driven by known mechanisms, such as streamlining due to nutrient limitation (Chuckran et al., 2021).

Importantly, the proposed decoupling of resistance versus resilience and resource usage efficiency dimensions entails that a consistent assignment of traits to the CSR or YAS schemes with validity for all prokaryotes may not be possible: in environments inhabited by prokaryotes with small genomes, such as aquatic habitats a nutrient rich and frequently disturbed habitat should select species with increased CUB and comparably large genomes, while the same scenario in soil environments should instead rather select species with increased CUB, but comparably small genomes. Genome size dependent trait-trait covariation patters might furthermore be the reason for conflicting assignments of genome size either with the C (Fierer et al., 2007) or R category (Krause et al., 2014) and likewise RRN or CUB/growth rate with either the R and S categories (Fierer et al., 2007) or with the C category (Krause et al., 2014) of the CSR schema.

We want to emphasize that the classification of individual species via their genomic traits into life history categories suffers from inaccuracies. This is on the one hand due to the fact that different evolutionary mechanisms and selective forces in combination lead to the selection of genomic trait values, which causes noisy correlations between trait values and the functional characteristics of a species. On the other hand, the assignment of species for instance along the specialist-generalist gradient is highly context-dependent and a resource specialist might be at the same time a salinity generalist (Bell and Bell, 2021). As a consequence, it is not possible to unambiguously characterize prokaryote species along the generalist-specialist continuum or predict their phenotypic response to a specific environmental change based on a simple genomic trait as for instance their genome

size. Still, the probability that a species will be resistant against a specific environmental change increases with its genome size. Similarly, while it is not possible to predict from a high CUB that the corresponding species will actually grow fast in a given environment, a high CUB increases the probability of this species to exhibit high growth rates in this environment. Accordingly, although the outlined genomic traits are imprecise in predicting the phenotypic characteristics of individual species in a given environment, they affect its likelihood to be tolerant or to grow fast in this environment. While probabilities do not allow the prediction of a single event, their predictive power increases with the number of considered events. We therefore suggest that resistance or resilience of individual species in a given situation should preferably be evaluated via the regulation of RNA markers in response to a specific environmental change, as detailed in a recent study (Rain-Franco et al., 2021). We however claim that the predictability of functional consequences from genomic traits increases if simultaneously applied for multiple species in a community and genomic traits values are scaled-up to evaluate their distribution at the community level (BOX 2).

5. Conclusions

Recent publications claimed that trait dimensions that are apparent among heterotroph prokaryotes are due to different physiological constraints and tradeoffs not directly comparable to those of autotroph plants (Malik et al., 2020; Westoby et al., 2021a). Based on our analyses we suggest that physiological constraints and tradeoffs differ even within the microbial cosmos, which precludes a globally consistent assignment of microbial traits in agreement with the CSR or YAS frameworks. In contrast, sorting microbial traits within a resistance/resilience framework increased consistency between trait-trait covariations and earlier reported findings due to variable tradeoffs between resistance and resilience related traits in dependence of the genome size range. It has been argued that varying resistance/resilience relationship have consequences for the stability of communities: In aquatic systems, where our analyses suggest a high likelihood for positive relationship between resistance and resilience levels, disturbances under oligotroph conditions should lead according to ecological theory to a species loss because both resistance and resilience are simultaneously low (Nimmo et al., 2015). Disturbances under high resource availability should instead induce a gain of species if resistance and residence levels are simultaneously high. In contrast, theory suggests a higher degree of stability in response to disturbances in communities featuring a resistance/resilience tradeoff (i.e. soil communities) because these communities are either resistant or resilient (Nimmo et al., 2015).

Even though specific CSR or YAS trait attributions may not generally be applicable for all prokaryotes, they should be applicable with adapted trait associations for communities harboring species within certain ranges of genome sizes, such as aquatic or soil communities. We argue that, beside potential differences between heterotrophs and autotrophs outlined elsewhere (Malik et al., 2020; Westoby et al., 2021a), disturbances and productivity gradients are, analogous to plants, the main drivers for microbial community dynamics. To understand the ecology of microbes and make

predictions about their dynamics it is consequently essential to combine different trait dimensions, concerning their response to disturbances and nutrient availability. We emphasize the need to expose prokaryote communities from different habitats to experimentally crossed disturbance and productivity gradients using full factorial designs and examine genomic trait distributions and diversity patterns. This should ideally be done in combination with functional resistance and resilience measurements to validate the links between genomic traits and community-level functional characteristics of prokaryotes outlined in this study. Such experimental designs will enable to empirically underpin the here presented predictions concerning the assembly of genomic traits and species richness under different scenarios and their relevance for community functioning. The trait table used in this study may hereby serve to extrapolate genomic trait distributions in prokaryotic communities based on taxonomic marker genes as outlined in BOX2.

BOX 2: Evaluation of genomic trait distributions from community sequence data

The shape of trait distributions in communities represented by measures such as the community weighted mean (CWM), but also the community weighted variance, skewness or kurtosis belong to the key drivers of community functioning and assembly (Enquist et al., 2015). While there are practical constraints to measure the distribution of physiological traits in microbial communities, genomic trait distributions can be extracted from community sequencing data. For instance, the CWM of the GC contents or genome sizes in microbial communities can be determined directly from the sequenced reads of shot gun metagenomes. In the latter case this can be done by relating the number of reads coding for single copy housekeeping genes to the number of total reads (Nayfach and Pollard, 2015). The same is true for the RRN, as the number of reads coding for 16s rRNA genes can be directly identified from metagenome reads (Kopylova et al., 2012) and related to the number of reads coding for single copy housekeeping genes (Biers et al., 2009). However, while the typically highly conserved housekeeping genes can be identified with high precision from short sequence reads the functional annotation of less conserved genes from short reads lacks accuracy. Longer sequences are therefore necessary for the Hidden Markov Models based annotation of genes encoding transcription factors as well as for the CUB estimation (Vieira-Silva and Rocha, 2010). In order to estimate the gene richness or gene duplication level within genomes as well as the number of prophages, an access to (nearly) full genome sequences is necessary. Longer sequences and even assembled genome sequences from shotgun metagenome data can be obtained via assembly and genome binning approaches. However, both methods (particularly genome binning) are biased towards more abundant sequences and genomes, while some life history traits may prevail in the rare biosphere (Vergin et al., 2013). Accordingly, may genomic traits representing the life histories of species that are rare in a community be underestimated if assembled or binned metagenome data were used for trait detection.

Another option to assess the distribution of genomic traits is to infer genomic traits from taxonomic marker genes of species in a community based on sequenced reference genomes of close relatives (Cébron et al., 2021; Romillac and Santorufo, 2021). This procedure is possible for all traits featuring a sufficiently strong phylogenetic signal (BOX3) and has been applied to determine genome sizes (Barberan et al., 2014) or the CUB (Weissman et al., 2021) of microbial communities based on 16s rRNA gene sequence data. Beside avoiding the possible biases outlined above, this strategy would allow to not only determine CWS, but also the other moments of trait distributions and thereby enable a more thorough evaluation of trait distributions in microbial communities.

The PICRUSt2 software (Douglas et al., 2020) that had been designed to extrapolate the genomic content of uncultured prokaryotes from closely related genomes via taxonomic marker genes can be analogously used to extrapolate the genomic traits outlined here. Our trait table (Table S1) can be applied to predict genomic traits for each species (OTU/ASV) in a sequenced community via the hidden state prediction tool (Louca and Doebeli, 2018) integrated into the PICRUSt2 software and using the default PICRUSt2 species reference database (PICRUSt does not accept trait

601 values outside the range of 0-1000 nor values that have more than two decimal digits, accordingly some of the given
602 trait values need to be re-ranged and rounded). Douglas and colleagues present averaged information of several
603 genomes, if these contained identical 16s rRNA genes (Table S1: picrust.ID = *-cluster), while the trait data presented
604 in Table S1 (with exception of the RRN) refer exclusively to the genomes indicated in column IMG.Genome.ID.
605 However, due to the highly significant phylogenetic signals detected for the presented genomic traits (Table 2), we
606 believe that the variability of these traits among close relatives sharing the same 16s rRNA gene should be limited. In
607 the particular case of RRN and as a consequence of putatively biased RRN values of the JGI/IMG database
608 (Supplementary Figures S1/S2), we propose to use values given in the rrnDB database (Stoddard et al., 2015) in
609 combination with the respective phylogeny to predict RRN values from amplicon sequence data.

610

611 ## BOX 3: Phylogenetic signals of genomic traits

612 All above evaluated genomic traits featured overall significant phylogenetic signals (Table 2). In agreement with the
613 phylogenetic signals, Mantel correlograms illustrated significant positive correlations between phylogenetic and trait
614 value distances at least until a phylogenetic distance of 0.25 (Table 2, Figure 7). The PICRUSt2 software (Douglas et
615 al., 2020) provides phylogenetic distances to the nearest relative in the reference database for each OTU/ASV in a
616 sequenced query community (NSTI values) and a default cutoff level of NSTI<0.2 is used to exclude predictions for
617 OTUs/ASV that are not represented by close relatives in the PICRUSt database. NSTI values can be used in
618 combination with the here reported results from the mantel correlograms to evaluate the reliability for PICRUSt-based
619 trait predictions: for instance the community average NSTI value (weighted mean) in a soil community computed with
620 the PICRUSt1 software was 0.17 (Langille et al., 2013).

621 Due to possible biases of RRN values given in the JGI database (Figure S1,S2) we estimated the phylogenetic signal
622 for RRN based on original entries of the rrnDB in combination with the corresponding rrnDB phylogeny. We assume
623 that the comparably low phylogenetic signals detected for RRN (Table 2, Figure 7) may not be a purely biological
624 signal, but could be due to inaccuracies of this parameter that, although not as pronounced as in the JGI database,
625 seemed to be also inherent in the rrnDB database (Figure S2).

626 One may wonder why genomic traits, such as prophages or % HGT at all exhibit phylogenetic signals, although
627 individual events contributing to these traits are not inherited vertically. However, while the presence of a specific
628 prophage or HGT gene in a genome should indeed not have a phylogenetic signal, it has been argued as outlined above
629 that the characteristic of a genome to host multiple prophages or HGT derived genes is linked to the life-history of the
630 corresponding organism. In agreement with our observation (Table 2), life history traits featured comparably strong
631 phylogenetic signals in an earlier study (Blomberg et al., 2003). Still, in the case of prophages and % HGT it seems
632 biologically meaningful that detected phylogenetic signals were low compared those of other traits (with the exception
633 of RRN).

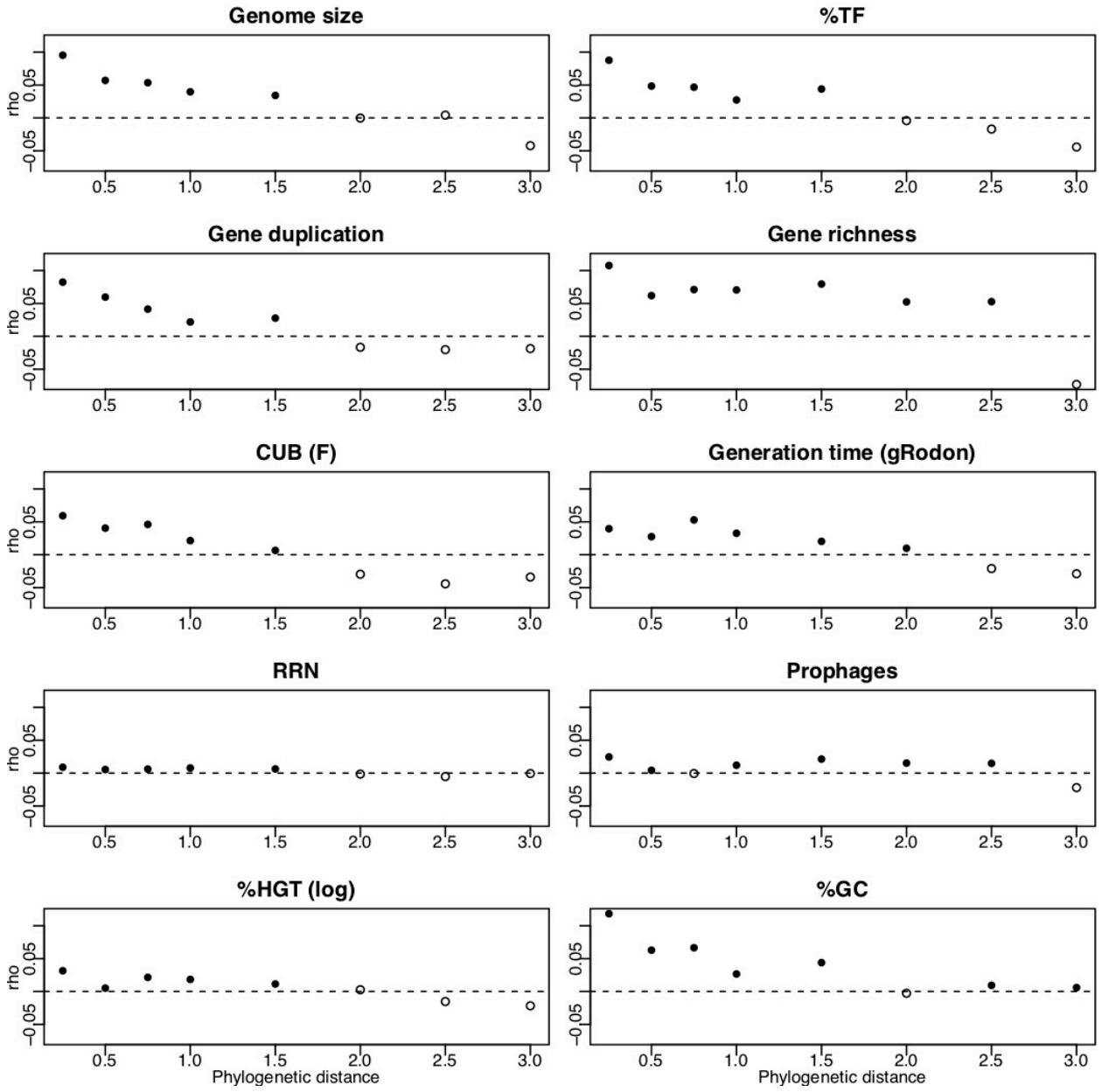

Figure 7: Results from Mantel correlograms. Filled data points indicate positive significant correlations (Bonferroni adjusted p-values Padj < 0.1) of pairwise distances of traits values against pairwise phylogenetic distances among the reference genomes. The Mantel correlograms were computed for 10,000 randomly selected genomes and with 200 permutations. In the specific case of RRN we computed the Mantel correlogram based on trait values of the rrnDB, while for all other traits JGI/IMG trait values as reported in the Supplementary Table S1 were used (%HGT was log(x+0.001) transformed and Generation time was log(x) transformed). We inferred the RRN rrnDB phylogenetic signal from a phylogenetic tree calculated using rrndbDB 16s rRNA gene sequences (FastTree 2, Price et al., 2010) and using the trait values available via the rrnDB (https://rrndb.umms.med.umich.edu/static/download/, rrnDB-5.7). The phylogeny for all JGI/IMG values was inferred using the pro_ref.tre phylogenetic tree, which is used as backbone phylogeny of the PICRUSt2 software. All analyses to compute Mantel correlograms were performed in R (R Core Team, 2021). Scripts with the code for all analyses described in this figure legend are available on GitHub (https://github.com/sarabeier/genomic.traits).

Table 2: Phylogenetic signals of genomic traits

Genomic trait	K[a]	P (K)[a]	Lambda[b]	P (lambda)[b]	Max PD[c]
Genome size (JGI/IMG)[d]	5.89×10^{-4}	<0.001	0.9973	<0.001	1.5
%TF (JGI/IMG)	3.10×10^{-4}	<0.001	0.9905	<0.001	1.5
Gene duplication (JGI/IMG)	4.89×10^{-4}	<0.001	0.9976	<0.001	1.5
Gene richness (JGI/IMG)	5.52×10^{-4}	<0.001	0.9892	<0.001	2.5
CUB (F, JGI/IMG)	1.24×10^{-3}	<0.001	0.9974	<0.001	1.5
Generation time (gRodon, JGI/IMG)	4.66×10^{-4}	<0.001	0.9935	<0.001	2.0
RRN (rrnDB)[d]	1.20×10^{-6}	<0.001	0.9624	<0.001	1.5
Prophages (JGI/IMG)	1.33×10^{-5}	<0.001	0.6403	<0.001	0.5
%HGT (JGI/IMG)	4.22×10^{-5}	<0.001	0.8897	<0.001	1.5
%GC (JGI/IMG)	2.33×10^{-2}	<0.001	0.9999	<0.001	1.5

[a]Blomberg's K statistics.

[b]Pagel's Lambda statistics.

Maximal phylogenetic distances (tested for phylogenetic distance classes ≤ 1.5), until which a continuous significant positive correlation against trait distances was detected via Mantel correlograms (significance based on the Bonferroni adjusted p-value Padj < 0.1). See also Figure 6.

[d] In the specific case of RRN we provide phylogenetic signal estimated based on trait values from the rrnDB, while for all other traits JGI/IMG trait values as reported in the Supplementary Table S1 were used (%HGT was log(x+0.001) transformed and Generation time was log(x) transformed). We inferred the RRN rrnDB phylogenetic signal from a phylogenetic tree calculated using rrndbDB 16s rRNA gene sequences (FastTree 2, Price et al., 2010) and using the trait values available via the rrnDB (https://rrndb.umms.med.umich.edu/static/download/, rrnDB-5.7).The phylogeny for all JGI/IMG values was inferred from the pro_ref.tre phylogenetic tree, which is used as backbone phylogeny of the PICRUSt2 software. All analyses to compute phylogenetic signals were performed in R (R Core Team, 2021) and a script is available on GitHub (https://github.com/sarabeier/genomic.traits.

6. Data availability and availability of detailed method descriptions

Sequence data for all genomes included in this study are available via the JGI/IMG platform (https://img.jgi.doe.gov/). The scripts used for bioinformatic data processing are available via GitHub (https://github.com/sarabeier/genomic.traits).

7. Acknowledgements

We thank Christiane Hassenrück for providing code to extract raw read data for genomes stored in the IMG/JGI database. We also Sara Vieira-Silva for her advices to setup scripts for the estimation of codon usage biases and the associated generation times from genome data. The study was supported by a grant from the German Science Foundation (DFG) awarded to SB (BE 5937/2-1). Bioinformatic analyses were supported by the BMBF-funded de.NBI Cloud within the German Network for Bioinformatics Infrastructure (de.NBI) (031A537B, 031A533A, 031A538A, 031A533B, 031A535A, 031A537C, 031A534A, 031A532B).

8. References

Alneberg, J., Bennke, C., Beier, S., Bunse, C., Quince, C., Ininbergs, K., et al. (2020). Ecosystem-wide metagenomic binning enables prediction of ecological niches from genomes. *Commun. Biol.* 3, 1–10. doi: 10.1038/s42003-020-0856-x.

Barberan, A., Ramirez, K. S., Leff, J. W., Bradford, M. A., Wall, D. H., and Fierer, N. (2014). Why are some microbes more ubiquitous than others? Predicting the habitat breadth of soil bacteria. *Ecol. Lett.* 17, 794–802. doi: 10.1111/ele.12282.

Bardgett, R. D., and Caruso, T. (2020). Soil microbial community responses to climate extremes: resistance, resilience and transitions to alternative states. *Philos. Trans. R. Soc. B-Biol. Sci.* 375, 20190112. doi: 10.1098/rstb.2019.0112.

Bell, T. H., and Bell, T. (2021). Many roads to bacterial generalism. *FEMS Microbiol. Ecol.* 97. doi: 10.1093/femsec/fiaa240.

Bentkowski, P., Van Oosterhout, C., and Mock, T. (2015). A Model of Genome Size Evolution for Prokaryotes in Stable and Fluctuating Environments. *Genome Biol. Evol.* 7, 2344–2351. doi: 10.1093/gbe/evv148.

Biers, E. J., Sun, S. L., and Howard, E. C. (2009). Prokaryotic genomes and diversity in surface ocean waters:interrogating the Global Ocean Sampling metagenome. *Appl. Environ. Microbiol.* 75, 2221–2229.

Botzman, M., and Margalit, H. (2011). Variation in global codon usage bias among prokaryotic organisms is associated with their lifestyles. *Genome Biol.* 12, R109. doi: 10.1186/gb-2011-12-10-r109.

Cébron, A., Zeghal, E., Usseglio-Polatera, P., Meyer, A., Bauda, P., Lemmel, F., et al. (2021). BactoTraits – A functional trait database to evaluate how natural and man-induced changes influence the assembly of bacterial communities. *Ecol. Indic.* 130, 108047. doi: 10.1016/j.ecolind.2021.108047.

Charoensawan, V., Wilson, D., and Teichmann, S. A. (2010). Genomic repertoires of DNA-binding transcription factors across the tree of life. *Nucleic Acids Res.* 38, 7364–7377. doi: 10.1093/nar/gkq617.

Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., and Parks, D. H. (2020). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36, 1925–1927. doi: 10.1093/bioinformatics/btz848.

Chen, I.-M. A., Chu, K., Palaniappan, K., Ratner, A., Huang, J., Huntemann, M., et al. (2021). The IMG/M data management and analysis system v.6.0: new tools and advanced capabilities. *Nucleic Acids Res.* 49, D751–D763. doi: 10.1093/nar/gkaa939.

Chen, Y., Neilson, J. W., Kushwaha, P., Maier, R. M., and Barberán, A. (2020). Life-history strategies of soil microbial communities in an arid ecosystem. *ISME J.*, 1–9. doi: 10.1038/s41396-020-00803-y.

Christie-Oleza, J. A., Fernandez, B., Nogales, B., Bosch, R., and Armengaud, J. (2012). Proteomic insights into the lifestyle of an environmentally relevant marine bacterium. *Isme J.* 6, 124–135. doi: 10.1038/ismej.2011.86.

Chuckran, P. F., Hungate, B. A., Schwartz, E., and Dijkstra, P. (2021). Variation in genomic traits of microbial communities among ecosystems. *FEMS Microbes* 2, xtab020. doi: 10.1093/femsmc/xtab020.

Cobo-Simon, M., and Tamames, J. (2017). Relating genomic characteristics to environmental preferences and ubiquity in different microbial taxa. *Bmc Genomics* 18, 499. doi: 10.1186/s12864-017-3888-y.

Cordero, O. X., and Hogeweg, P. (2009a). Regulome size in Prokaryotes: universality and lineage-specific variations. *Trends Genet.* 25, 285–286.

Cordero, O. X., and Hogeweg, P. (2009b). The impact of long-distance horizontal gene transfer on prokaryotic genome size. *Proc. Natl. Acad. Sci. U. S. A.* 106, 21748–21753. doi: 10.1073/pnas.0907584106.

Couturier, E., and Rocha, E. P. C. (2006). Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. *Mol. Microbiol.* 59, 1506–1518. doi: https://doi.org/10.1111/j.1365-2958.2006.05046.x.

Dall, S. R. X., and Cuthill, I. C. (1997). The information costs of generalism. *Oikos* 80, 197–202. doi: 10.2307/3546535.

de Vries, F. T., and Shade, A. (2013). Controls on soil microbial community stability under climate change. *Front. Microbiol.* 4, 265. doi: 10.3389/fmicb.2013.00265.

DeLong, J. P., Okie, J. G., Moses, M. E., Sibly, R. M., and Brown, J. H. (2010). Shifts in metabolic scaling, production, and efficiency across major evolutionary transitions of life. *Proc. Natl. Acad. Sci.* 107, 12941–12945. doi: 10.1073/pnas.1007783107.

Douglas, G. M., Maffei, V. J., Zaneveld, J. R., Yurgel, S. N., Brown, J. R., Taylor, C. M., et al. (2020). PICRUSt2 for prediction of metagenome functions. *Nat. Biotechnol.* 38, 685–688. doi: 10.1038/s41587-020-0548-6.

Duffy, J. E., Godwin, C. M., and Cardinale, B. J. (2017). Biodiversity effects in the wild are common and as strong as key drivers of productivity. *Nature* 549, 261-+. doi: 10.1038/nature23886.

Enquist, B. J., Norberg, J., Bonser, S. P., Violle, C., Webb, C. T., Henderson, A., et al. (2015). "Chapter Nine - Scaling from Traits to Ecosystems: Developing a General Trait Driver Theory via Integrating Trait-Based and Metabolic Scaling Theories," in *Advances in Ecological Research* Trait-Based Ecology - From Structure to Function., eds. S. Pawar, G. Woodward, and A. I. Dell (Academic Press), 249–318. doi: 10.1016/bs.aecr.2015.02.001.

Ferenci, T. (2016). Trade-off Mechanisms Shaping the Diversity of Bacteria. *Trends Microbiol.* 24, 209–223. doi: 10.1016/j.tim.2015.11.009.

Fierer, N. (2017). Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat. Rev. Microbiol.* 15, 579–590. doi: 10.1038/nrmicro.2017.87.

Fierer, N., Bradford, M. A., and Jackson, R. B. (2007). Toward an ecological classification of soil bacteria. *Ecology* 88, 1354–1364. doi: 10.1890/05-1839.

Freilich, S., Kreimer, A., Borenstein, E., Yosef, N., Sharan, R., Gophna, U., et al. (2009). Metabolic-network-driven analysis of bacterial ecological strategies. *Genome Biol.* 10, R61. doi: 10.1186/gb-2009-10-6-r61.

Gadgil, M., and Solbrig, O. T. (1972). The Concept of r- and K-Selection: Evidence from Wild Flowers and Some Theoretical Considerations. *Am. Nat.* 106, 14–31.

Garcia, M. O., Templer, P. H., Sorensen, P. O., Sanders-DeMott, R., Groffman, P. M., and Bhatnagar, J. M. (2020). Soil Microbes Trade-Off Biogeochemical Cycling for Stress Tolerance Traits in Response to Year-Round Climate Change. *Front. Microbiol.* 11. doi: 10.3389/fmicb.2020.00616.

Giovannoni, S. J., Thrash, J. C., and Temperton, B. (2014). Implications of streamlining theory for microbial ecology. *Isme J.* 8, 1553–1565. doi: 10.1038/ismej.2014.60.

Giovannoni, S. J., Tripp, H. J., Givan, S., Podar, M., Vergin, K. L., Baptista, D., et al. (2005). Genome streamlining in a cosmopolitan oceanic bacterium. *Science* 309, 1242–1245. doi: 10.1126/science.1114057.

Gophna, U., Kristensen, D. M., Wolf, Y. I., Popa, O., Drevet, C., and Koonin, E. V. (2015). No evidence of inhibition of horizontal gene transfer by CRISPR–Cas on evolutionary timescales. *ISME J.* 9, 2021–2027. doi: 10.1038/ismej.2015.20.

Grilli, J., Bassetti, B., Maslov, S., and Lagomarsino, M. C. (2012). Joint scaling laws in functional and evolutionary categories in prokaryotic genomes. *Nucleic Acids Res.* 40, 530–540. doi: 10.1093/nar/gkr711.

Grime, J. (1977). Evidence for Existence of Three Primary Strategies in Plants and Its Relevance to Ecological and Evolutionary Theory. *Am. Nat.* 111, 1169–1194. doi: 10.1086/283244.

Grzymski, J. J., and Dussaq, A. M. (2012). The significance of nitrogen cost minimization in proteomes of marine microorganisms. *ISME J.* 6, 71–80. doi: 10.1038/ismej.2011.72.

Hastings, A. (1980). Disturbance, coexistence, history, and competition for space. *Theor. Popul. Biol.* 18, 363–373. doi: 10.1016/0040-5809(80)90059-3.

Hellweger, F. L., Huang, Y., and Luo, H. (2018). Carbon limitation drives GC content evolution of a marine bacterium in an individual-based genome-scale model. *ISME J.* 12, 1180–1187. doi: 10.1038/s41396-017-0023-7.

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9, 5114. doi: 10.1038/s41467-018-07641-9.

Joseph, B., and Goebel, W. (2007). Life of Listeria monocytogenes in the host cells' cytosol. *Microbes Infect.* 9, 1188–1195. doi: 10.1016/j.micinf.2007.05.006.

Kassen, R. (2002). The experimental evolution of specialists, generalists, and the maintenance of diversity. *J. Evol. Biol.* 15, 173–190. doi: 10.1046/j.1420-9101.2002.00377.x.

Klappenbach, J. A., Dunbar, J. M., and Schmidt, T. M. (2000). RRNA operon copy number reflects ecological strategies of bacteria. *Appl. Environ. Microbiol.* 66, 1328–1333. doi: 10.1128/AEM.66.4.1328-1333.2000.

Koch, A. L. (2001). Oligotrophs versus copiotrophs. *Bioessays* 23, 657–661. doi: 10.1002/bies.1091.

Kondrashov, F. A. (2012). Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc. R. Soc. B-Biol. Sci.* 279, 5048–5057. doi: 10.1098/rspb.2012.1108.

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

Konstantinidis, K. T., and Tiedje, J. M. (2004). Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc. Natl. Acad. Sci. U. S. A.* 101, 3160–3165. doi: 10.1073/pnas.0308653100.

Konstantinidis, K. T., and Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 102, 2567–2572. doi: 10.1073/pnas.0409727102.

Kopylova, E., Noe, L., and Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* 28, 3211–3217. doi: 10.1093/bioinformatics/bts611.

Kostadinov, I., Kottmann, R., Ramette, A., Waldmann, J., Buttigieg, P. L., and Glöckner, F. O. (2011). Quantifying the effect of environment stability on the transcription factor repertoire of marine microbes. *Microb. Inform. Exp.* 1, 9. doi: 10.1186/2042-5783-1-9.

Krause, S., Le Roux, X., Niklaus, P. A., Bodegom, P. V., Lennon, J. T., Bertilsson, S., et al. (2014). Trait-based approaches for understanding microbial biodiversity and ecosystem functioning. *Aquat. Microbiol.* 5, 251. doi: 10.3389/fmicb.2014.00251.

Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* 31, 814-+. doi: 10.1038/nbt.2676.

Lauro, F. M., McDougald, D., Thomas, T., Williams, T. J., Egan, S., Rice, S., et al. (2009). The genomic basis of trophic strategy in marine bacteria. *Proc. Natl. Acad. Sci. U. S. A.* 106, 15527–15533. doi: 10.1073/pnas.0903507106.

Leff, J. W., Jones, S. E., Prober, S. M., Barberan, A., Borer, E. T., Firn, J. L., et al. (2015). Consistent responses of soil microbial communities to elevated nutrient inputs in grasslands across the globe. *Proc. Natl. Acad. Sci. U. S. A.* 112, 10967–10972. doi: 10.1073/pnas.1508382112.

Long, A. M., Hou, S., Ignacio-Espinoza, J. C., and Fuhrman, J. A. (2021). Benchmarking microbial growth rate predictions from metagenomes. *Isme J.* 15, 183–195. doi: 10.1038/s41396-020-00773-1.

Louca, S., and Doebeli, M. (2018). Efficient comparative phylogenetics on large trees. *Bioinformatics* 34, 1053–1055. doi: 10.1093/bioinformatics/btx701.

Malik, A. A., Martiny, J. B. H., Brodie, E. L., Martiny, A. C., Treseder, K. K., and Allison, S. D. (2020). Defining trait-based microbial strategies with consequences for soil carbon cycling under climate change. *ISME J.* 14, 1–9. doi: 10.1038/s41396-019-0510-0.

Markowitz, V. M., Chen, I.-M. A., Palaniappan, K., Chu, K., Szeto, E., Grechkin, Y., et al. (2010). The integrated microbial genomes system: an expanding comparative analysis resource. *Nucleic Acids Res.* 38, D382–D390. doi: 10.1093/nar/gkp887.

Marshall, K. T., and Morris, R. M. (2013). Isolation of an aerobic sulfur oxidizer from the SUP05/Arctic96BD-19 clade. *ISME J.* 7, 452–455. doi: 10.1038/ismej.2012.78.

Miyashita, K., Fujii, T., and Sawada, Y. (1991). Molecular cloning and characterization of chitinase genes from Streptomyces lividans 66. *J. Gen. Microbiol.* 137, 2065–2072.

Molina, N., and van Nimwegen, E. (2009). Scaling laws in functional genome content across prokaryotic clades and lifestyles. *Trends Genet.* 25, 243–247. doi: 10.1016/j.tig.2009.04.004.

Mouquet, N., Hoopes, M. F., and Amarasekare, P. (2005). "The world is patchy and heterogeneous!," in *Metacommunities, Spatial Dynamics and Ecological Communities*, eds. M. Holyoak, M. A. Leibold, and R. D. Holt (The University of Chicago Press), 237–262.

Mukherjee, S., Stamatis, D., Bertsch, J., Ovchinnikova, G., Sundaramurthi, J. C., Lee, J., et al. (2021). Genomes OnLine Database (GOLD) v.8: overview and updates. *Nucleic Acids Res.* 49, D723–D733. doi: 10.1093/nar/gkaa983.

Nayfach, S., and Pollard, K. S. (2015). Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. *Genome Biol.* 16, 51. doi: 10.1186/s13059-015-0611-7.

Nemergut, D. R., Knelman, J. E., Ferrenberg, S., Bilinski, T., Melbourne, B., Jiang, L., et al. (2016). Decreases in average bacterial community rRNA operon copy number during succession. *Isme J.* 10, 1147–1156. doi: 10.1038/ismej.2015.191.

Neuenschwander, S. M., Ghai, R., Pernthaler, J., and Salcher, M. M. (2018). Microdiversification in genome-streamlined ubiquitous freshwater Actinobacteria. *ISME J.* 12, 185–198. doi: 10.1038/ismej.2017.156.

Nimmo, D. G., Mac Nally, R., Cunningham, S. C., Haslem, A., and Bennett, A. F. (2015). Vive la résistance: reviving resistance for 21st century conservation. *Trends Ecol. Evol.* 30, 516–523. doi: 10.1016/j.tree.2015.07.008.

Okie, J. G., Poret-Peterson, A. T., Lee, Z. M., Richter, A., Alcaraz, L. D., Eguiarte, L. E., et al. (2020). Genomic adaptations in information processing underpin trophic strategy in a whole-ecosystem nutrient enrichment experiment. *eLife* 9, e49816. doi: 10.7554/eLife.49816.

Parter, M., Kashtan, N., and Alon, U. (2007). Environmental variability and modularity of bacterial metabolic networks. *Bmc Evol. Biol.* 7, 169. doi: 10.1186/1471-2148-7-169.

Piton, G., Foulquier, A., Martinez-García, L. B., Legay, N., Arnoldi, C., Brussaard, L., et al. (2021). Resistance–recovery trade-off of soil microbial communities under altered rain regimes: An experimental test across European agroecosystems. *J. Appl. Ecol.* 58, 406–418. doi: https://doi.org/10.1111/1365-2664.13774.

Poindexter, J. (1981). Oligotrophy - Fast and Famine Existence. *Adv. Microb. Ecol.* 5, 63–89.

Poisot, T., Mouquet, N., and Gravel, D. (2013). Trophic complementarity drives the biodiversity–ecosystem functioning relationship in food webs. *Ecol. Lett.* 16, 853–861. doi: 10.1111/ele.12118.

Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2-Approximately Maximum-Likelihood Trees for Large Alignments. *Plos One* 5. doi: 10.1371/journal.pone.0009490.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing Available at: https://www.R-project.org/.

Rain-Franco, A., Mouquet, N., Gougat-Barbera, C., Bouvier, T., and Beier, S. (2021). Niche breadth affects bacterial transcription patterns along a salinity gradient. *Mol. Ecol.* n/a. doi: 10.1111/mec.16316.

Roller, B. R. K., Stoddard, S. F., and Schmidt, T. M. (2016). Exploiting rRNA operon copy number to investigate bacterial reproductive strategies. *Nat. Microbiol.* 1, 1–7. doi: 10.1038/nmicrobiol.2016.160.

Romillac, N., and Santorufo, L. (2021). Transferring concepts from plant to microbial ecology: A framework proposal to identify relevant bacterial functional traits. *Soil Biol. Biochem.* 162, 108415. doi: 10.1016/j.soilbio.2021.108415.

Sexton, J. P., Montiel, J., Shay, J. E., Stephens, M. R., and Slatyer, R. A. (2017). "Evolution of Ecological Niche Breadth," in *Annual Review of Ecology, Evolution, and Systematics, Vol 48*, ed. D. J. Futuyma (Palo Alto: Annual Reviews), 183–206.

Shade, A., Peter, H., Allison, S. D., Baho, D. L., Berga, M., Buergmann, H., et al. (2012). Fundamentals of microbial community resistance and resilience. *Front. Microbiol.* 3, 417. doi: 10.3389/fmicb.2012.00417.

Shrestha, P. M., Noll, M., and Liesack, W. (2007). Phylogenetic identity, growth-response time and rRNA operon copy number of soil bacteria indicate different stages of community succession. *Environ. Microbiol.* 9, 2464–2474. doi: 10.1111/j.1462-2920.2007.01364.x.

Sriswasdi, S., Yang, C., and Iwasaki, W. (2017). Generalist species drive microbial dispersion and evolution. *Nat. Commun.* 8, 1162. doi: 10.1038/s41467-017-01265-1.

Steen, A. D., Crits-Christoph, A., Carini, P., DeAngelis, K. M., Fierer, N., Lloyd, K. G., et al. (2019). High proportions of bacteria and archaea across most biomes remain uncultured. *ISME J.* 13, 3126–3130. doi: 10.1038/s41396-019-0484-y.

Stevenson, B. S., and Schmidt, T. M. (2004). Life history implications of rRNA gene copy number in Escherichia coli. *Appl. Environ. Microbiol.* 70, 6670–6677. doi: 10.1128/AEM.70.11.6670-6677.2004.

Stoddard, S. F., Smith, B. J., Hein, R., Roller, B. R. K., and Schmidt, T. M. (2015). rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. *Nucleic Acids Res.* 43, D593–D598. doi: 10.1093/nar/gku1201.

Takemoto, K. (2013). Does Habitat Variability Really Promote Metabolic Network Modularity? *Plos One* 8, e61348. doi: 10.1371/journal.pone.0061348.

Tilman, D. (1994). Competition and Biodiversity in Spatially Structured Habitats. *Ecology* 75, 2–16. doi: 10.2307/1939377.

Touchon, M., Bernheim, A., and Rocha, E. P. C. (2016). Genetic and life-history traits associated with the distribution of prophages in bacteria. *Isme J.* 10, 2744–2754. doi: 10.1038/ismej.2016.47.

van Nimwegen, E. (2003). Scaling laws in the functional content of genomes. *Trends Genet.* 19, 479–484. doi: 10.1016/S0168-9525(03)00203-8.

Vergin, K. L., Done, B., Carlson, C. A., and Giovannoni, S. J. (2013). Spatiotemporal distributions of rare bacterioplankton populations indicate adaptive strategies in the oligotrophic ocean. *Aquat. Microb. Ecol.* 71, 1-U129. doi: 10.3354/ame01661.

Vieira-Silva, S., and Rocha, E. P. C. (2010). The Systemic Imprint of Growth and Its Uses in Ecological (Meta) Genomics. *Plos Genet.* 6, e1000808. doi: 10.1371/journal.pgen.1000808.

Violle, C., Navas, M.-L., Vile, D., Kazakou, E., Fortunel, C., Hummel, I., et al. (2007). Let the concept of trait be functional! *Oikos* 116, 882–892. doi: 10.1111/j.2007.0030-1299.15559.x.

Ward, C. S., Yung, C.-M., Davis, K. M., Blinebry, S. K., Williams, T. C., Johnson, Z. I., et al. (2017). Annual community patterns are driven by seasonal switching between closely related marine bacteria. *Isme J.* 11, 1412–1422. doi: 10.1038/ismej.2017.4.

Weissman, J. L., Hou, S., and Fuhrman, J. A. (2021). Estimating maximal microbial growth rates from cultures, metagenomes, and single cells via codon usage patterns. *Proc. Natl. Acad. Sci.* 118. doi: 10.1073/pnas.2016810118.

Westoby, M., Gillings, M. R., Madin, J. S., Nielsen, D. A., Paulsen, I. T., and Tetu, S. G. (2021a). Trait dimensions in bacteria and archaea compared to vascular plants. *Ecol. Lett.* n/a. doi: https://doi.org/10.1111/ele.13742.

Westoby, M., Nielsen, D. A., Gillings, M. R., Litchman, E., Madin, J. S., Paulsen, I. T., et al. (2021b). Cell size, genome size, and maximum growth rate are near-independent dimensions of ecological variation across bacteria and archaea. *Ecol. Evol.* doi: 10.1002/ece3.7290.

Whitman, W. B., Coleman, D. C., and Wiebe, W. J. (1998). Prokaryotes: The unseen majority. *P Natl Acad Sci USA* 95, 6578–6583.

Supplementary Material

Figure S1: A) Correlation of RRN data from the JGI/IMG database (RRN_IMG) against RRN values that were extrapolated from genome level RRN values available via the rrnDB database indicate that a bias in the JGI/IMG values is mainly due to RRN underestimation for entries with high rrnDB RRN values. B) The standard deviation (sd) of within genus (NCBI taxonomy) RRN values for 666 genera of the rrnDB database with more than one entry per genus indicate values of sd < 1 in the large majority of genera.

Figure S2: The JGI database provides information about NCBI accession, bioproject or biosample accession numbers, but no direct link to the raw read data. While for the majority of JGI entries listed in Table S1 it was not possible to identify runids by command-line based matches, we could extract raw read data for some of the entries. We applied the MicrobeCensus software to estimate the genome size as well as the number of sequenced genome equivalents based on the number of reads coding for single copy housekeeping genes relative to the number of total reads (Nayfach and Pollard, 2015). We excluded genomes with <2,000,000 sequenced reads and < 50 sequenced genome equivalents. We further applied the SortMeRNA software (Kopylova et al., 2012) to identify reads coding for the 16s rRNA and used the output in combination with the indicated number of sequenced genome equivalents to estimate the number of RRN per genome. A) Pearson correlation between the read based RRN estimate and the JGI RRN estimate. B) Pearson correlation between the read based RRN estimate and the rrnDB RRN estimate as given in Table S1. C) Pearson correlation between the read based genome size estimate and the genome size given in the JGI database.

While both the read-based and assembly-based estimation of RRN or genome size may suffer from biases, these biases are different and independent from each other. We therefore interpreted the strength of correlation between assembly-based and read-based estimates for RRN and genome size as measure for the quality of values provided in the rrnDB and JGI/IMG databases. The strongly improved correlation of read based RRN estimates against the rrnDB derived RRN values compared to a correlation against the JGI values, indicates that the presented rrnDB derived RRNs are exposed to reduced biases. However, the even better correlation between the genome size estimates in combination with low phylogenetic signals of RRN (Table 2, Figure 7) suggests that also the rrnDB derived RRN are likely still less accurate that are assembly-based genome estimates from the JGI/IMG database.

Figure S3: Percent contribution to variability of principal components PC1 to PC9.

Figure S4: The random removal of 1,2,3 or 4 variable from the principal component analyses presented in Figure 3 demonstrates that the spatial patterns of resistance versus resilience related genomic traits are robust against the removal of individual variables from the dataset.

Table S1: Overview table for 17,856 genomes available via the JGI/IMG platform (https://img.jgi.doe.gov/, downloaded in August 2021), which are integrated in the reference database of the PICRUSt2 software (v2.1.2-b, Douglas et al., 2020). Some genomes present in the original PICRUSt2 reference database had meanwhile been replaced or removed. In the case of replaced genomes we indicate the new IMG genome ID in the column IMG.Genome.ID, while in the column PICRUSt.ID the former IMG genome ID is given. Douglas and colleagues furthermore present averaged information of several genomes, if these contained identical 16s rRNA genes (picrust.ID = *-cluster). However, the trait data presented in Table S1 refer exclusively to the genomes indicated in column IMG.Genome.ID.. All traits values were either directly extracted from the JGI/IMG database or computed with scripts that are available on GitHub (https://github.com/sarabeier/genomic.traits). RRN_IMG and RRN_rrnDB represent RRN values given by

941 the JGI/IMG database and values extrapolated at the genus level (NCBI taxonomy) from the rrnDB,
942 respectively. It was not possible to extract the information for all traits from all genomes and missing trait
943 values are indicated by NA. Strains that could not be unambiguously binned at the species level via the
944 GTDB taxonomy were classified via the FastANI software (Jain et al., 2018) into species level bins using
945 an average nucleotide identity >94% (column FastANI_species). Information of all remaining columns is
946 available via the JGI/IMG database.

947 Table S2: Read based estimation for RRN and genome size of JGI entries with >2,000,000 sequenced reads
948 and >50 sequenced genome equivalents.

949