



# Simulated rRNA/DNA Ratios Show Potential To Misclassify Active Populations as Dormant

Blaire Steven,<sup>a</sup> Cedar Hesse,<sup>b\*</sup> John Soghigian,<sup>a</sup> La Verne Gallegos-Graves,<sup>b</sup>  
John Dunbar<sup>b</sup>

Department of Environmental Sciences, Connecticut Agricultural Experiment Station, New Haven, Connecticut, USA<sup>a</sup>; Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico, USA<sup>b</sup>

**ABSTRACT** The use of rRNA/DNA ratios derived from surveys of rRNA sequences in RNA and DNA extracts is an appealing but poorly validated approach to infer the activity status of environmental microbes. To improve the interpretation of rRNA/DNA ratios, we performed simulations to investigate the effects of community structure, rRNA amplification, and sampling depth on the accuracy of rRNA/DNA ratios in classifying bacterial populations as “active” or “dormant.” Community structure was an insignificant factor. In contrast, the extent of rRNA amplification that occurs as cells transition from dormant to growing had a significant effect ( $P < 0.0001$ ) on classification accuracy, with misclassification errors ranging from 16 to 28%, depending on the rRNA amplification model. The error rate increased to 47% when communities included a mixture of rRNA amplification models, but most of the inflated error was false negatives (i.e., active populations misclassified as dormant). Sampling depth also affected error rates ( $P < 0.001$ ). Inadequate sampling depth produced various artifacts that are characteristic of rRNA/DNA ratios generated from real communities. These data show important constraints on the use of rRNA/DNA ratios to infer activity status. Whereas classification of populations as active based on rRNA/DNA ratios appears generally valid, classification of populations as dormant is potentially far less accurate.

**IMPORTANCE** The rRNA/DNA ratio approach is appealing because it extracts an extra layer of information from high-throughput DNA sequencing data, offering a means to determine not only the seedbank of taxa present in communities but also the subset of taxa that are metabolically active. This study provides crucial insights into the use of rRNA/DNA ratios to infer the activity status of microbial taxa in complex communities. Our study shows that the approach may not be as robust as previously supposed, particularly in complex communities composed of populations employing different growth strategies, and identifies factors that inflate the erroneous classification of active populations as dormant.

**KEYWORDS** RNA/DNA ratios, activity, dormancy, error, sampling, simulation

Studies of microbial community composition are increasingly concerned with identifying metabolically active populations linked to ecosystem processes. Although amplification and sequencing of rRNA genes effectively document the microorganisms present in a community (1–4), the abundance of an organism in a community DNA survey is an unreliable indicator of its activity. Given the frequent presence of starving cells, spores, or other dormant forms of cells (5–8), nonviable cells with residual DNA (9), or extracellular DNA (10, 11), only a portion of the microbial DNA detected in the environment represents metabolically active cells (12). To identify the metabolically active members of microbial communities, sequencing of rRNA has been employed. Because rRNA is a structural component of ribosomes and ribosomes are expected to

Received 23 March 2017 Accepted 24 March 2017

Accepted manuscript posted online 31 March 2017

**Citation** Steven B, Hesse C, Soghigian J, Gallegos-Graves LV, Dunbar J. 2017. Simulated rRNA/DNA ratios show potential to misclassify active populations as dormant. *Appl Environ Microbiol* 83:e00696-17. <https://doi.org/10.1128/AEM.00696-17>.

**Editor** Frank E. Löffler, University of Tennessee and Oak Ridge National Laboratory

**Copyright** © 2017 Steven et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Blaire Steven, [Blair.Steven@ct.gov](mailto:Blair.Steven@ct.gov), or John Dunbar, [dunbar@lanl.gov](mailto:dunbar@lanl.gov).

\* Present address: Cedar Hesse, Horticultural Crops Research Laboratory, U.S. Department of Agriculture, Corvallis, Oregon, USA.

increase with metabolic activity (13–16), rRNA abundance is a potential marker of growth and metabolism. Additionally, RNA generally degrades faster than DNA after cell death, so detection of rRNA is more likely to indicate active or recently living cells (17).

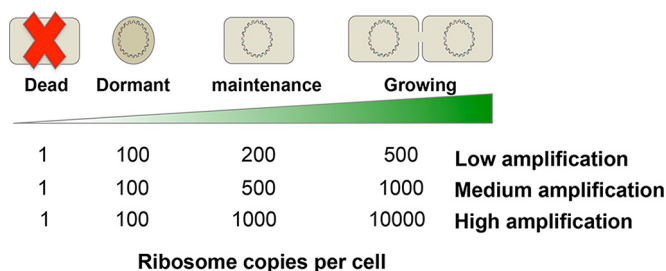
Different approaches have been used to characterize active microbial populations via rRNA sequencing. One approach is to characterize only the rRNA pool and infer that this represents active populations (18, 19). However, the presence of rRNA is not always a reliable indicator of activity because dormant forms of cells harbor ribosomes in order to resume protein synthesis when environmental conditions improve (20). Comparison of rRNA and DNA quantities is a more sophisticated approach. As cellular activity increases, the ratio of rRNA to rRNA genes is expected to increase because ribosome abundance increases much more than the genome copy number in active cells. Consequently, organisms with a higher abundance of rRNA than DNA in community surveys are proposed to be active (21, 22). A critical detail of the rRNA/DNA method used in microbial community analyses is that the true ratios of rRNA to DNA occurring within cells are not measured; instead, the relative abundance of a taxon in a community survey of rRNA is compared to its relative abundance in a community survey of *rRNA* genes. Although a number of studies have used rRNA/DNA ratios to characterize active microbial populations in environmental samples (23–25), there has been little effort to investigate factors that may affect data interpretation.

In this study, we performed simulations to test the effects of community structure, variation in rRNA amplification, and sampling depth on the identification of active populations on the basis of rRNA/DNA ratios. The simulation results were used to guide the interpretation of empirical sequence data generated from forest floor microbial communities. The simulation data revised our original interpretation of the empirical data, demonstrating the potential of these simulations to inform other studies that employ rRNA/DNA ratios.

## RESULTS

**rRNA amplification data.** It is well accepted that the number of ribosomes (and therefore rRNA) in cells increases with the growth rate (26). Yet, the precise number of ribosomes in cells in different metabolic states is very uncertain. Among 18 studies published between 1986 and 2013, the quantity of ribosomes reported in bacterial cells varied 3,600-fold (see Table S1 in the supplemental material). The median quantity in stationary-phase cells was 200 ( $n = 3$  studies; range, 20 to 8,000), and the median quantity in growing cells was about 5,100 ( $n = 18$  studies and 13 species; range, 92 to 72,000). Given the uncertainty and large spread of published estimates, we defined three rRNA amplification models—low, medium, and high—to represent a range of possibilities for the increase in the ribosome content of cells across four metabolic states (Fig. 1). Among the three models, the cellular ribosome content ranged from 1 in dead cells to a maximum of 10,000 in growing cells.

**Effects of community structure and rRNA amplification model on rRNA/DNA ratios.** Simulations of communities with 5,000 populations showed that community structure did not have a significant effect on the accuracy of population activity assessments derived from rRNA/DNA ratios (by analysis of variance [ANOVA], degree of freedom [df] = 2,  $F = 0.002$ , and  $P = 0.99$ ) (Table 1). These results were obtained from simulations (100 runs each) that represented three community structures and the three ribosome amplification models described in Fig. 1. Given a constant rRNA amplification model, the false detection rate (false positives [FP] plus false negatives [FN]) with different community structures was nearly identical (Table 1). For example, the false detection rates ranged from 21.3 to 21.5% for three community structures with the low rRNA amplification model (Table 1; similar results [not shown] were obtained from simulations with 1,000 runs). The data suggest that dramatic differences in community structure would not undermine the use of rRNA/DNA ratios to infer population activity status.



**FIG 1** rRNA amplification and random assignment of cells to metabolic states. Three different rRNA amplification models were used to represent variation in the abundance of ribosomes in cells in four metabolic states. The number of ribosomes per cell in each of the activity metabolic states with the different rRNA amplification models is indicated. Data for the rRNA amplification models were based on data from Table S1. For simulations, each population in a community represented a mixture of cells in different metabolic states. The number of cells in a given population was determined from a community structure model. The cells in a population were randomly assigned to four metabolic states, and the net activity status of each population was calculated.

In contrast, rRNA amplification models had a significant effect on false detection rates (by ANOVA,  $df = 2$ ,  $F = 3,869$ , and  $P < 0.0001$ ). The combined FP and FN rates ranged from 16 to 28%, and the highest rate occurred with the high rRNA amplification model. The distribution of rRNA/DNA ratios in these simulations is illustrated in Fig. 2. The mean of the rRNA/DNA ratios was 1, but the mode shifted from 1 to 1.1 under the high rRNA amplification model. Similarly, the maximum rRNA/DNA ratio increased from ca. 2.4 in the low and medium models to 3.5 in the high rRNA amplification model.

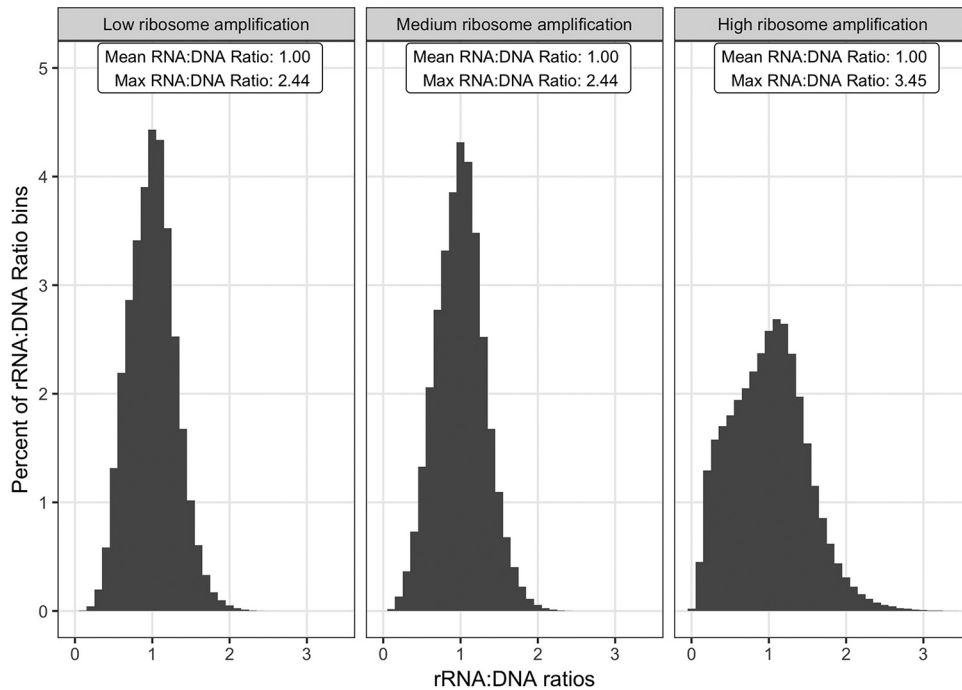
Taken together, these data show that rRNA/DNA ratios are insensitive to variation in community structure, whereas variation in evolved growth strategy (i.e., rRNA amplification) among populations or communities can substantially influence the interpretation of rRNA/DNA ratios.

**Effect of partial sampling on rRNA/DNA ratios.** Simulations with partial sampling—analogueous to incomplete sampling of natural communities—showed that sampling depth had a significant effect on misclassification rates (Table 2) (by ANOVA,  $df = 4$ ,  $F = 17.65$ , and  $P < 0.0001$ ; similar results [not shown] were obtained from simulations with 1,000 runs). In these simulations, a community with a lognormal population distribution (standard deviation of 1) was used and random samples of increasing size were drawn (Fig. 3; Table 2). The simulations showed two major effects of sampling depth. First, undersampling of the communities produced a spread of rRNA/DNA ratios much greater than 1.0. For example, in simulations with the medium rRNA amplification model, rRNA/DNA ratios as high as 46.8 occurred when the sampling depth was 1 to 100 times the total number of populations (5,000) in the community (Fig. 3). At a sampling depth of 1,000 $\times$ , the maximum rRNA/DNA ratio was only 6.0 (Fig. 3) and the distribution of rRNA/DNA ratios began to resemble the distribution in completely sampled communities (Fig. 2, middle panel) that had a maximum rRNA/DNA ratio of 2.4. Second, undersampling significantly inflated the misclassification rate for activity

**TABLE 1** Effect of community structure and ribosomal amplification on misclassifications

Ribosomal amplification	Community structure	Avg % FP $\pm$ SD <sup>a</sup>	Avg % FN $\pm$ SD <sup>a</sup>
Low	Lognormal $\sigma_1$	10.7 $\pm$ 1.9	10.8 $\pm$ 1.9
Low	Lognormal $\sigma_2$	10.7 $\pm$ 0.6	10.6 $\pm$ 0.6
Low	Even	10.7 $\pm$ 0.5	10.6 $\pm$ 0.4
Medium	Lognormal $\sigma_1$	8.4 $\pm$ 1.8	8.1 $\pm$ 1.8
Medium	Lognormal $\sigma_2$	8.1 $\pm$ 0.7	8.1 $\pm$ 0.6
Medium	Even	8.0 $\pm$ 0.5	8.0 $\pm$ 0.6
High	Lognormal $\sigma_1$	13.9 $\pm$ 1.5	14.1 $\pm$ 1.8
High	Lognormal $\sigma_2$	14.0 $\pm$ 0.6	13.9 $\pm$ 0.6
High	Even	13.9 $\pm$ 0.5	14.0 $\pm$ 0.6

<sup>a</sup>Percentages of the 5,000 input populations in 100 independent runs per simulation are shown.



**FIG 2** Effects of rRNA amplification models on rRNA/DNA ratios. Shown are histograms of rRNA/DNA ratios for each population generated from the simulations. The simulations represent the low, medium, and high rRNA amplification models run under the lognormal  $\sigma_1$  community structure model. Qualitatively similar data were obtained with the  $\sigma_0$  and  $\sigma_2$  community structure models, so the data are not shown. Each histogram represents the cumulative data of 100 independent runs of a simulation.

status (by ANOVA,  $P < 0.001$ ), and the degree of this effect depended significantly on the rRNA amplification model (by ANOVA,  $df = 2$ ,  $F = 10.85$ , and  $P = 0.005$ ). Moreover, partial sampling increased the FN rate compared to the FP rate (by paired-sample two-tailed Student  $t$  test,  $P = 0.006$ ). Increasing the sampling depth from  $1\times$  to  $10,000\times$  the total population richness reduced the combined FN and FP rate from an average of 38% to an average of 22% (Table 2), and the latter level was similar to levels observed in completely sampled communities (Table 1). These findings show that

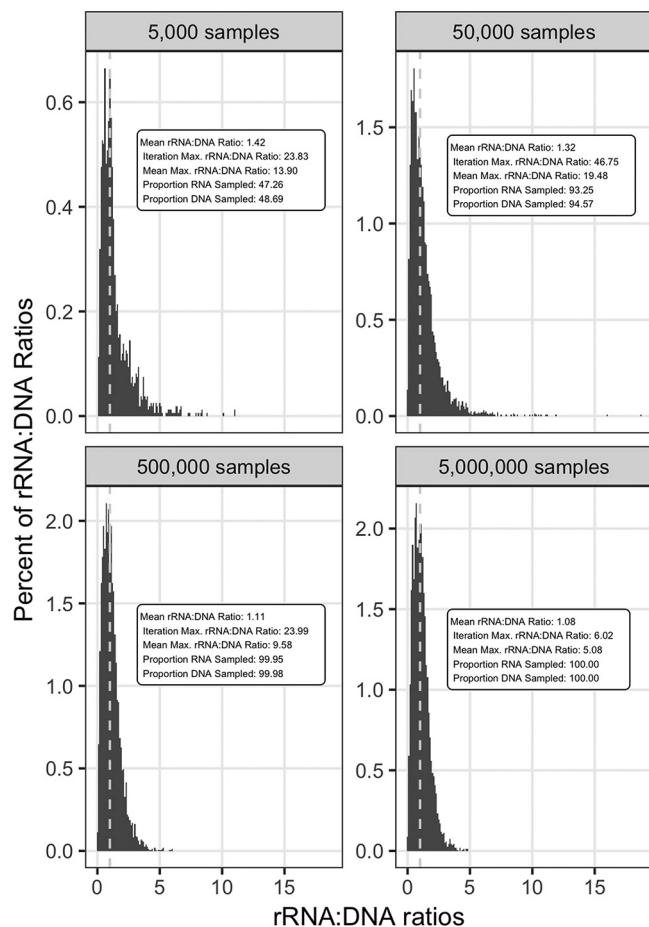
**TABLE 2** Effect of sample size on misclassifications

Sample size	Ribosomal amplification	Avg % FP $\pm$ SD <sup>a</sup>	Avg % FN $\pm$ SD <sup>a</sup>
5,000 ( $1\times$ ) <sup>b</sup>	Low	18.4 $\pm$ 0.6	20.2 $\pm$ 0.7
5,000 ( $1\times$ )	Medium	17.3 $\pm$ 0.7	19.4 $\pm$ 0.7
5,000 ( $1\times$ )	High	17.8 $\pm$ 0.7	20.4 $\pm$ 0.7
50,000 ( $10\times$ )	Low	14.7 $\pm$ 0.6	16.2 $\pm$ 0.6
50,000 ( $10\times$ )	Medium	13.2 $\pm$ 0.6	14.5 $\pm$ 0.6
50,000 ( $10\times$ )	High	15.3 $\pm$ 0.6	16.9 $\pm$ 0.6
500,000 ( $100\times$ )	Low	14.1 $\pm$ 0.6	14.5 $\pm$ 0.7
500,000 ( $100\times$ )	Medium	9.2 $\pm$ 0.6	9.8 $\pm$ 0.5
500,000 ( $100\times$ )	High	14.0 $\pm$ 0.6	14.7 $\pm$ 0.6
5,000,000 ( $1,000\times$ )	Low	10.8 $\pm$ 0.7	10.8 $\pm$ 0.6
5,000,000 ( $1,000\times$ )	Medium	8.3 $\pm$ 0.6	8.4 $\pm$ 0.6
5,000,000 ( $1,000\times$ )	High	14.1 $\pm$ 0.7	14.1 $\pm$ 0.7
50,000,000 ( $10,000\times$ )	Low	10.7 $\pm$ 0.6	10.5 $\pm$ 0.7
50,000,000 ( $10,000\times$ )	Medium	8.1 $\pm$ 0.6	8.1 $\pm$ 0.6
50,000,000 ( $10,000\times$ )	High	14.0 $\pm$ 0.6	14.0 $\pm$ 0.7

<sup>a</sup>The error rates are expressed as percentages of the populations that were detected in both DNA and RNA sampling profiles, and each value is the average of 100 independent iterations  $\pm$  the standard deviation).

With low sampling depths, the number of detected populations was less than the 5,000 input populations.

<sup>b</sup>The parenthetical factors indicate the sampling depth as a multiple of the number of input populations (5,000) in the simulated community.



**FIG 3** Effects of partial sampling on rRNA/DNA ratios. Each panel represents a different sampling depth, ranging from random samples of 1 to 1,000 times the input population richness. The simulation data presented are for a lognormal  $\sigma 1$  community structure with the medium rRNA amplification model and 5,000 populations. Data from other model permutations were qualitatively similar (data not shown). Each histogram represents illustrative data from a single run of a simulation. Summary statistics of combined data from 100 independent runs are included in the insets. The reported proportions of RNA and DNA sampled refer to the average proportions of the 5,000 input taxa that were detected in the samples.

sampling depth is an important concern when using rRNA/DNA ratios to infer the activity of microbial populations.

**Effect of mixed rRNA amplification on rRNA/DNA ratios.** Given that species have evolved different growth strategies, rRNA amplification is expected to vary among populations in natural communities. For example, a species with an inherently low growth rate is expected to have a lower number of ribosomes per cell (i.e., low ribosomal amplification) than a species with an inherently high growth rate when both species are in the active growth phase (15). To evaluate this effect, we constructed communities with a mixture of three rRNA amplification models. In these simulations, populations were randomly assigned to one of the three rRNA amplification models, yielding a community in which populations can have identical activity levels (i.e., they are equally active) but different ribosome counts. These simulations showed a dramatic increase in the total misclassification rate (47% versus 16 to 28%) of population activity status from rRNA/DNA ratios in comparison to simulations with a single rRNA amplification model (Table 3 versus Table 1). The increased error rate arose mainly from an increase in the FN rate to 34%, which was more than twice the 13% FP rate from a single rRNA amplification model (Table 3). When the populations that comprised the FP and FN detections were identified, 100% of the FP populations had the high rRNA amplification model, whereas only 1.4% of the FN populations harbored the high

**TABLE 3** Effect of a mixed community on misclassifications

Error	% of total misclassification <sup>a</sup>	% of error from ribosome amplification model <sup>b</sup>		
		Low	Medium	High
FP	12.8	0	0	100
FN	33.8	49.2	49.4	1.4

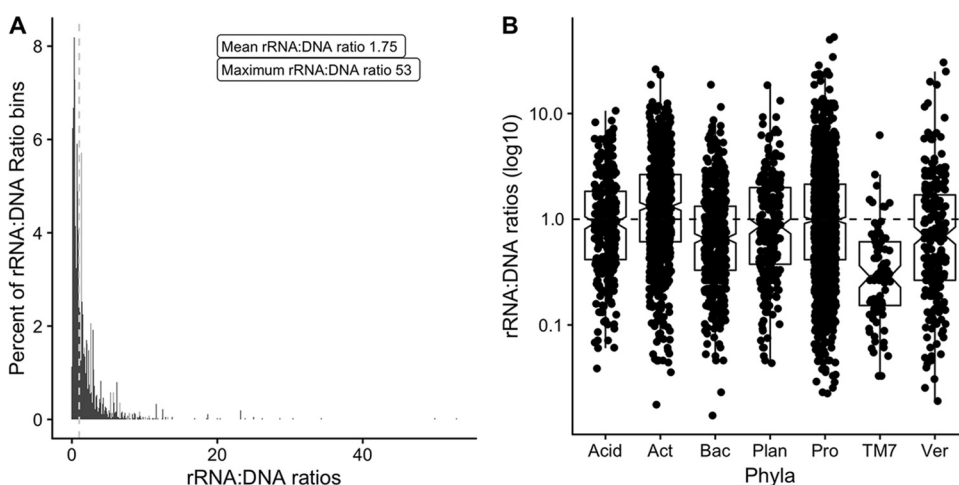
<sup>a</sup>Percentage of misclassifications in the entire community (5,000 populations, 100 runs per simulation).

<sup>b</sup>Percent contributions of the different amplification models to the FP and FN error rates.

amplification model (Table 3). Conversely, ca. 99% of the FN populations were from the low and medium rRNA amplification models. These results suggest that a high FN rate is a likely characteristic of rRNA/DNA ratios from natural communities where growth strategies vary among populations, and therefore, “dormant” status may often be an incorrect inference. In contrast, the majority of populations classified as “active” from rRNA/DNA ratios are likely to be valid, but “active” status may be biased toward species with high rRNA amplification growth strategies.

**Empirical sequencing data and rRNA/DNA ratios.** rRNA/DNA ratios were calculated for operational taxonomic units (OTUs) found in each of 12 forest floor soil communities that were sampled from four sites across a 500-km transect in Michigan, USA. The sequence surveys represented only a portion of the total diversity of the samples, as indicated by Good’s coverage (27), which ranged from ca. 61 to 70% across the sequence libraries. The rRNA/DNA ratios calculated from the surveys ranged from 0.003 to 53 (Fig. 4A). Qualitatively, the distribution of rRNA/DNA ratios appeared similar to the distributions from undersampled communities *in silico* (Fig. 3).

To investigate possible relationships between taxonomy and rRNA/DNA ratios, every OTU in the empirical sequencing data was assigned to a phylum level taxonomic bin. The rRNA/DNA ratios within each of the seven numerically dominant phyla (defined as those consistently present at >5% abundance in the sequence libraries) are presented in Fig. 4B. In this analysis, OTUs within the phyla *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, and *Proteobacteria* mostly had mean rRNA/DNA ratios close to or equal to 1.0,



**FIG 4** rRNA/DNA ratios in empirical sequence data. (A) Histogram of the ratios of OTUs from forest floor sequence libraries. Ratios were calculated individually for every OTU from 12 paired replicate sequence libraries. (B) Relationship between rRNA/DNA ratios and taxonomy. Each point represents an individual OTU. Only phyla that were present across all of the sequence libraries and accounted for >5% of the sequences are displayed. Given the large range of rRNA/DNA ratios, the y axis is displayed as a log scale. The dashed line represents an rRNA/DNA ratio of 1. OTUs above the line would be classified as active, and those below the line would be classified as dormant. The boxes denote the interquartile range of the rRNA/DNA ratios, with the notch denoting the mean. Phylum abbreviations: Acid, *Acidobacteria*; Act, *Actinobacteria*; Bac, *Bacteroidetes*; Plan, *Planctomycetes*; Pro, *Proteobacteria*; Ver, *Verrucomicrobia*.



whereas the rRNA/DNA ratios of OTUs within the phylum TM7 (*Candidatus Saccharibacteria* [28]) were almost exclusively  $<1.0$  (Fig. 4B).

## DISCUSSION

Our simulations verified that rRNA/DNA ratios can identify active populations but false detections can occur. False detection rates increased as the sampling depth decreased (Table 2). More importantly, the false detection rate was dramatically higher (48% versus 16 to 28%) when communities had populations that differed in rRNA amplification (Table 3). We did not test the combined effect of undersampling and mixed rRNA amplification, but the combined factors are likely to inflate false detection rates even more. Given the high diversity of most terrestrial microbial communities and the range of physiologies and growth rates among populations, high false detection rates are anticipated when rRNA/DNA ratios are used to infer activity status in complex environmental samples. The strong bias toward inflation of FN over FP (Tables 2 and 3) shows that classification of populations as “dormant” is less reliable than classification as “active.”

We expected the misclassification error to be higher in communities with a strong dominance pattern—that is, a less even species abundance distribution—but simulations did not support this hypothesis. The logic of our hypothesis was that increased dominance would exacerbate measurement errors in the RNA or DNA abundance of rare types, giving rise to greatly distorted rRNA/DNA ratios of rare types. Why? Because a statistical characteristic of community survey data is that the relative measurement error of taxon abundance is greater for rare taxa than for abundant taxa in community surveys. We can only speculate that a community structure effect was not observed because the misclassification of activity status depends foremost on the distribution of activity levels among species, not species abundance. Our simulations used a normal distribution of activity levels, owing to the representation of each taxon as a random mixture of four metabolic states. This is an important caveat for future studies: a different distribution of activity levels may produce different misclassification error rates.

Given the propensity for FN to occur with rRNA/DNA ratios from complex communities (Tables 2 and 3), inferring the dormancy of a specific taxon in natural communities should be done with caution. In our forest floor samples, OTUs within the phylum TM7 almost exclusively showed rRNA/DNA ratios of  $<1$  (Fig. 4B), possibly indicating consistent dormancy along the 500-km transect we sampled. Our simulations, however, suggest the alternative explanation that TM7 bacteria may simply be more prone to misclassification as FN. This phylum may be misclassified as dormant if it has evolved a low rRNA amplification strategy, whereas co-occurring active taxa have medium to high rRNA amplification (Table 3). This simulation-informed insight demonstrates the value of using modeling to guide the interpretation of empirical data, especially when a foundation of empirical ground truth is absent.

The findings from simulations improve the interpretation of rRNA/DNA ratios of natural communities. A characteristic feature of rRNA/DNA ratios derived from our natural communities and found in other empirical studies is a wide spread of ratios much greater than 1. The highest ratio we observed was 53 (Fig. 4A), and ratios of up to 122 were reported in aerosol samples (29). Although it is tempting to move beyond the classification of taxa as active or dormant and use rRNA/DNA ratios to distinguish degrees of activity (also known as “specific activity” [29]), several considerations suggest that the practice is generally unsound. First, insufficient sampling depth can dramatically distort rRNA/DNA ratios, as our simulations showed (Fig. 3). Second, for some species, rRNA/DNA ratios are misleading indicators of their general activity status (active versus dormant), much less the degree of activity, because of eccentric biology. For example, the cyanobacterium *Aphanizomenon ovalisporum* forms dormant cells called akinetes that contain a greater abundance of ribosomes than vegetative cells do (30), and among members of the genus *Bacillus*, high ribosome content may be a prerequisite to enter dormancy (20). Finally, the species-specific interplay among

genomic factors (discussed below), ribosome content, and growth strategy makes fine-scale interpretation of rRNA/DNA ratios difficult. The scaling of rRNA/DNA ratios with specific activity levels is likely to vary considerably among species, such that two species with the same rRNA/DNA ratio can have different activity levels or vice versa. Recently, Blazewicz et al. (31) reviewed various biological factors that affect the use of rRNA as a general indicator of activity. Our simulations showed that both methodological and biological factors can confound the use of rRNA/DNA ratios to infer specific activity.

Our simulations captured the effects of several biological factors that could contribute to rRNA and rRNA gene counts in natural communities. The use of different community structures captured variation in rRNA gene abundance among populations that may arise not only from differences in cellular abundance but also from several genomic factors, namely, variation in the copy number of *rrn* genes within genomes, genome ploidy, and growth rate. In bacteria, rRNA gene copy numbers vary from 1 to as many as 15 per genome (32, 33) and some species of bacteria are polyploids that harbor multiple copies of their genome (34, 35). "*Epulopiscium*" sp. is an extreme polyploid case with thousands of copies of the genome per cell during normal growth (36). The copy number of rRNA genes per cell can also increase transiently as the growth rate increases owing to the presence of multiple genome replication forks per cell (37, 38). All of these factors contribute to the observed abundance distribution of rRNA gene sequences—the community structure—in targeted metagenomic surveys of natural communities.

Genome factors and other regulatory phenomena also contribute to differences among species in growth strategy—a multifaceted evolutionary trait that includes how species manage ribosome content (rRNA abundance). Our use of three different rRNA amplification models reflected the uncertainty in the current state of knowledge concerning the ribosome content of cells in different physiological states, as well as the variation in growth strategy among species in natural communities. Detailed information about rRNA amplification among species with different growth strategies is a large knowledge gap that merits further attention and careful physiological studies.

Given our observations and the paucity of data on the growth strategies of most bacterial species, the safest application of rRNA/DNA ratios may be time series studies, in which the activity of a population is assessed relative to an initial baseline instead of relative to an assumption of a universal (across microbial phyla) scaling of rRNA/DNA ratios with population activity status. A substantial change in the rRNA/DNA ratio of a taxon can indicate a shift in activity level, reducing dependence on a universal threshold value (e.g., 1.0) for classification of populations as active or dormant. Time series analysis also mitigates the problem of legacy effects, wherein an rRNA/DNA point measurement reflects a prior activity status and fails to reveal populations that may be transitioning to a new activity status (12).

**Conclusion.** With a modest simulation effort, we gained important insights into the interpretation of the rRNA/DNA ratios of real communities. Our study showed that highly asymmetric FP and FN error rates are possible when rRNA/DNA ratios are applied to complex communities to infer population activity status. Inference of active status appears generally sound, whereas inference of dormant status is undermined by a potentially high misclassification error rate. Insufficient sampling depth and evolved physiological variation among species significantly inflate the FN rate (i.e., active populations misclassified as dormant). Validation of rRNA/DNA ratios with orthogonal techniques to measure population activity, such as stable isotope probing (39) or nanoscale secondary ion mass spectrometry (40), would improve the interpretation of metabolic activity. Finally, additional physiological studies that accurately characterize the relationship between ribosome content and specific growth rates, or activity status, among different species would also improve the modeling and interpretation of rRNA/DNA ratios.



## MATERIALS AND METHODS

**Simulations.** Simulations and statistical analyses were performed with the R software package (41). Every simulation represented a community made up of cells belonging to 5,000 populations. Each simulation involved three parameters, described below.

**Community structure (i.e., abundance distribution of populations).** Three community structures were tested. The community structures differed in the extent of population dominance, with dominance ranging from high to none. The community structures were as follows: a lognormal distribution with a standard deviation of 2, a lognormal distribution with a standard deviation of 1, and a uniform distribution with a standard deviation of 0 (i.e., equal abundance of all populations). These population abundance distributions were used to represent the abundance of cells (and *rrn* gene abundance) in each population.

**Metabolic states.** Each population was represented as a collection of cells in different metabolic states. This approach was used because microbial populations in natural environments are seldom expected to be metabolically homogeneous, owing to factors such as fluctuating and patchy resource distributions and complex ecological gradients created by nonuniform distributions of other interacting populations. For modeling, each cell in a population was assigned to one of four metabolic states—dead, dormant, maintenance, or growing (Fig. 1). The key characteristics of the metabolic states are as follows: (i) dead, cells contain DNA but very limited rRNA; (ii) dormant, cells have low ribosome numbers but are capable of resuscitation (e.g., spores); (iii) maintenance, cells are alive and metabolically active but maintain a low metabolic rate with little, if any, reproduction; (iv) growing, cells have high activity and are multiplying. The four metabolic states were based on those proposed by Blazewicz et al. (31). For each population, the fraction of cells in each metabolic state was assigned randomly, with the constraint that the sum of the fractions equals 1 (Fig. 1). Random assignment of the metabolic states within each population created an approximately normal distribution of activity levels among the 5,000 populations in each community.

**rRNA amplification.** We created three rRNA amplification models describing how ribosome abundance changes among the four cellular metabolic states (Fig. 1A). We derived the models on the basis of data from studies examining ribosome abundance in stationary-phase or growing cells of 13 bacterial species (see Table S1). Given a specific rRNA amplification model and the fractions of a population in the four metabolic states, the rRNA count of a population could be explicitly calculated as the sum of ribosomes in cells in the four metabolic states. The resulting relative abundance of rRNA of each population in a community could then be compared with the relative *rrn* gene abundance from the community structure model to calculate the rRNA/DNA ratio. Populations with an rRNA/DNA ratio of  $>1$  were classified as active, and those with an rRNA/DNA ratio of  $<1$  were classified as dormant, as in a previous study (21).

The simulations consisted of the following seven sequential steps. (i) Assign a community structure. (ii) Randomly assign cells within each population to four metabolic states. (iii) Select an rRNA amplification model. (iv) Calculate the resulting rRNA/DNA ratios explicitly, or create random subsamples of the rRNA and *rrn* gene pools, and use these surveys to calculate rRNA/DNA ratios. (v) Use the ratios to classify populations as active versus dormant. (vi) Calculate misclassification (FN and FP) rates. (vii) Repeat the simulation over 100 iterations.

To assess the accuracy of rRNA/DNA ratios for identifying population activity status (active versus dormant), misclassification rates were calculated. FP were populations that were known to be inactive (i.e.,  $>50\%$  of the cells were in the dead or dormant state) but had observed rRNA/DNA ratios of  $>1$ . Conversely, FN were populations that were known to be active (i.e.,  $>50\%$  of the cells were in the maintenance or growing state) but had rRNA/DNA ratios of  $<1$ . With these metrics, it was possible to quantify the effects of community structure, rRNA amplification model, and sampling depth on the accuracy of activity assessments derived from rRNA/DNA ratios.

The R code to perform these simulations is included in the supplemental material.

**Empirical sequence data.** Forest floor samples were collected in October 2011 at four sites (three replicates per site) over an approximately 500-km latitudinal gradient. Extraction of DNA and RNA and nucleic acid sequencing were described previously (22, 42). The primers employed to amplify bacterial 16S rRNA genes from DNA and cDNA were targeted to the V5 and V6 region of the gene as described by Claesson et al. (43). Sequencing was performed via 454 FLX Titanium at the Los Alamos National Laboratory. The resulting sequences were quality checked in the mothur software package v.1.27.0 (44) by using the PyroNoise algorithm (45). Potentially chimeric sequences were identified by using the mothur implementation of Perseus (46), and all prospective chimeras were removed from further analysis. The bacterial 16S rRNA sequences were aligned in mothur against the SILVA bacterial reference alignment (47), and OTUs were determined by clustering sequences by average neighbor clustering in mothur with a 97% sequence identity threshold.

The rRNA/DNA ratios of the empirical sequencing data were calculated after randomly subsampling paired DNA and rRNA libraries to the same number of reads to remove potential biases due to differences in sampling depth. The proportional abundance of each OTU was then determined (dividing the number of reads for each OTU by the number of reads in the sample). Only OTUs that were detected in both the *rrn* gene survey and the rRNA survey were employed to calculate rRNA/DNA ratios. The rRNA/DNA ratio of each OTU was determined by dividing the relative abundance of the OTU in the rRNA survey by its relative abundance in the *rrn* gene survey.

The taxonomic identities of OTUs in the empirical sequencing data were determined with the mothur-implemented Bayesian classifier (48) to compare OTU representative sequences against the

SILVA reference database (47), release 123. Only taxonomic assignments with a confidence score of  $\geq 70\%$  are reported.

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/AEM.00696-17>.

**SUPPLEMENTAL FILE 1**, PDF file, 0.2 MB.

## REFERENCES

- Weisburg WG, Barns SM, Pelletier DA, Lane DJ. 1991. 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* 173:697–703. <https://doi.org/10.1128/jb.173.2.697-703.1991>.
- Case RJ, Boucher Y, Dahllöf I, Holmstrom C, Doolittle WF, Kjelleberg S. 2007. Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. *Appl Environ Microbiol* 73:278–288. <https://doi.org/10.1128/AEM.01177-06>.
- Lane DJ, Pace B, Olsen GJ, Stahl DA, Sogin ML, Pace NR. 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci U S A* 82:6955–6959. <https://doi.org/10.1073/pnas.82.20.6955>.
- Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hemsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF. 2016. A new view of the tree of life. *Nat Microbiol* 1:16048. <https://doi.org/10.1038/nmicrobiol.2016.48>.
- Oliver JD. 2005. The viable but nonculturable state in bacteria. *J Microbiol* 43:93–100.
- Kaprelyants AS, Gottschal JC, Kell DB. 1993. Dormancy in non-sporulating bacteria. *FEMS Microbiol Rev* 10:271–285. <https://doi.org/10.1111/j.1574-6968.1993.tb05871.x>.
- Lennon JT, Jones SE. 2011. Microbial seed banks: the ecological and evolutionary implications of dormancy. *Nat Rev Microbiol* 9:119–130. <https://doi.org/10.1038/nrmicro2504>.
- Stevenson LH. 1977. A case for bacterial dormancy in aquatic systems. *Microb Ecol* 4:127–133. <https://doi.org/10.1007/BF02014283>.
- Grégori G, Citterio S, Ghiani A, Labra M, Sgorbati S, Brown S, Denis M. 2001. Resolution of viable and membrane-compromised bacteria in freshwater and marine waters based on analytical flow cytometry and nucleic acid double staining. *Appl Environ Microbiol* 67:4662–4670. <https://doi.org/10.1128/AEM.67.10.4662-4670.2001>.
- Dlott G, Maul JE, Buyer J, Yarwood S. 2015. Microbial rRNA:rDNA gene ratios may be unexpectedly low due to extracellular DNA preservation in soils. *J Microbiol Methods* 115:112–120. <https://doi.org/10.1016/j.mimet.2015.05.027>.
- Steinberger RE, Holden PA. 2005. Extracellular DNA in single- and multiple-species unsaturated biofilms. *Appl Environ Microbiol* 71:5404–5410. <https://doi.org/10.1128/AEM.71.9.5404-5410.2005>.
- Blagodatskaya E, Kuzyakov Y. 2013. Active microorganisms in soil: critical review of estimation criteria and approaches. *Soil Biol Biochem* 67:192–211. <https://doi.org/10.1016/j.soilbio.2013.08.024>.
- Koch AL. 1993. Biomass growth rate during the prokaryote cell cycle. *Crit Rev Microbiol* 19:17–42. <https://doi.org/10.3109/10408419309113521>.
- McCarthy BJ. 1962. The effects of magnesium starvation on the ribosome content of *Escherichia coli*. *Biochim Biophys Acta* 55:880–889. [https://doi.org/10.1016/0006-3002\(62\)90901-0](https://doi.org/10.1016/0006-3002(62)90901-0).
- Ramos C, Mølbak L, Molin S. 2000. Bacterial activity in the rhizosphere analyzed at the single-cell level by monitoring ribosome contents and synthesis rates. *Appl Environ Microbiol* 66:801–809. <https://doi.org/10.1128/AEM.66.2.801-809.2000>.
- Sturani E, Costantini MG, Zippel R, Alberghina FAM. 1976. Regulation of RNA synthesis in *Neurospora crassa*. *Exp Cell Res* 99:245–252. [https://doi.org/10.1016/0014-4827\(76\)90580-2](https://doi.org/10.1016/0014-4827(76)90580-2).
- Maiväli Ü, Paier A, Tenson T. 2013. When stable RNA becomes unstable: the degradation of ribosomes in bacteria and beyond. *Biol Chem* 394:845–855. <https://doi.org/10.1515/hsz-2013-0133>.
- Anderson IC, Parkin PL. 2007. Detection of active soil fungi by RT-PCR amplification of precursor rRNA molecules. *J Microbiol Methods* 68:248–253. <https://doi.org/10.1016/j.mimet.2006.08.005>.
- Felske A, Wolterink A, Van Lis R, Akkermans AD. 1998. Phylogeny of the main bacterial 16S rRNA sequences in Drentse A grassland soils (The Netherlands). *Appl Environ Microbiol* 64:871–879.
- Nanamiya H, Sato M, Masuda K, Sato M, Wada T, Suzuki S, Natori Y, Katano M, Akanuma G, Kawamura F. 2010. *Bacillus subtilis* mutants harbouring a single copy of the rRNA operon exhibit severe defects in growth and sporulation. *Microbiology* 156:2944–2952. <https://doi.org/10.1099/mic.0.035295-0>.
- Jones SE, Lennon JT. 2010. Dormancy contributes to the maintenance of microbial diversity. *Proc Natl Acad Sci U S A* 107:5881–5886. <https://doi.org/10.1073/pnas.0912765107>.
- Mueller RC, Gallegos-Graves L, Zak DR, Kuske CR. 2016. Assembly of active bacterial and fungal communities along a natural environmental gradient. *Microb Ecol* 71:57–67. <https://doi.org/10.1007/s00248-015-0655-y>.
- Schostag M, Stibal M, Jacobsen CS, Bælum J, Taş N, Elberling B, Jansson JK, Semenchuk P, Priemé A. 2015. Distinct summer and winter bacterial communities in the active layer of Svalbard permafrost revealed by DNA- and RNA-based analyses. *Front Microbiol* 6:399. <https://doi.org/10.3389/fmicb.2015.00399>.
- Campbell BJ, Yu L, Heidelberg JF, Kirchman DL. 2011. Activity of abundant and rare bacteria in a coastal ocean. *Proc Natl Acad Sci U S A* 108:12776–12781. <https://doi.org/10.1073/pnas.1101405108>.
- Wilhelm RC, Niederberger TD, Greer C, Whyte LG. 2011. Microbial diversity of active layer and permafrost in an acidic wetland from the Canadian High Arctic. *Can J Microbiol* 57:303–315. <https://doi.org/10.1139/w11-004>.
- Elser JJ, Acharya K, Kyle M, Cotner J, Makino W, Markow T, Watts T, Hobbie S, Fagan W, Schade J, Hood J, Sterner RW. 2003. Growth rate-stoichiometry couplings in diverse biota. *Ecol Lett* 6:936–943. <https://doi.org/10.1046/j.1461-0248.2003.00518.x>.
- Good IJ. 1953. The population frequencies of species and the estimation of population parameters. *Biometrika* 40:237–264.
- Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol* 31:533–538. <https://doi.org/10.1038/nbt.2579>.
- Klein AM, Bohannan BJM, Jaffe DA, Levin DA, Green JL. 2016. Molecular evidence for metabolically active bacteria in the atmosphere. *Front Microbiol* 7:772. <https://doi.org/10.3389/fmicb.2016.00772>.
- Sukenik A, Kaplan-Levy RN, Welch JM, Post AF. 2012. Massive multiplication of genome and ribosomes in dormant cells (akinetes) of *Aphanizomenon ovalisporum* (Cyanobacteria). *ISME J* 6:670–679. <https://doi.org/10.1038/ismej.2011.128>.
- Blazewicz SJ, Barnard RL, Daly RA, Firestone MK. 2013. Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. *ISME J* 7:2061–2068. <https://doi.org/10.1038/ismej.2013.102>.
- Klappenbach JA. 2001. rrndb: the ribosomal RNA operon copy number database. *Nucleic Acids Res* 29:181–184. <https://doi.org/10.1093/nar/29.1.181>.
- Lee ZM-P, Bussema C, Schmidt TM. 2009. rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. *Nucleic Acids Res* 37:D489–D493. <https://doi.org/10.1093/nar/gkn689>.
- Griese M, Lange C, Soppa J. 2011. Ploidy in cyanobacteria. *FEMS Microbiol Lett* 323:124–131. <https://doi.org/10.1111/j.1574-6968.2011.02368.x>.
- Gross L. 2006. The more the merrier: multiple genomes in a parasitic prokaryote. *PLoS Biol* 4:e212. <https://doi.org/10.1371/journal.pbio.0040212>.
- Mendell JE, Clements KD, Choat JH, Angert ER. 2008. Extreme polyploidy in a large bacterium. *Proc Natl Acad Sci U S A* 105:6730–6734. <https://doi.org/10.1073/pnas.0707522105>.

37. Condon C, Philips J, Fu Z-Y, Squires C, Squires CL. 1992. Comparison of the expression of the seven ribosomal RNA operons in *Escherichia coli*. *EMBO J* 11:4175.
38. Cooper S, Helmstetter CE. 1968. Chromosome replication and the division cycle of *Escherichia coli* B/r. *J Mol Biol* 31:519–540. [https://doi.org/10.1016/0022-2836\(68\)90425-7](https://doi.org/10.1016/0022-2836(68)90425-7).
39. Dumont MG, Murrell JC. 2005. Innovation: stable isotope probing, linking microbial identity to function. *Nat Rev Microbiol* 3:499–504. <https://doi.org/10.1038/nrmicro1162>.
40. Herrmann AM, Ritz K, Nunan N, Clode PL, Pett-Ridge J, Kilburn MR, Murphy DV, O'Donnell AG, Stockdale EA. 2007. Nano-scale secondary ion mass spectrometry, a new analytical tool in biogeochemistry and soil ecology: a review article. *Soil Biol Biochem* 39:1835–1850. <https://doi.org/10.1016/j.soilbio.2007.03.011>.
41. R Core Team. 2013. R: a language and environment for statistical computing. R Foundation, Vienna, Austria.
42. Kuske CR, Hesse CN, Challacombe JF, Cullen D, Herr JR, Mueller RC, Tsang A, Vilgalys R. 2015. Prospects and challenges for fungal metatranscriptomics of complex communities. *Fungal Ecol* 14:133–137. <https://doi.org/10.1016/j.funeco.2014.12.005>.
43. Claesson MJ, Wang QO, O'Sullivan O, Greene-Diniz R, Cole JR, Ross RP, O'Toole PW. 2010. Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res* 38:e200. <https://doi.org/10.1093/nar/gkq873>.
44. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Horn DJV, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 75:7537–7541. <https://doi.org/10.1128/AEM.01541-09>.
45. Quince C, Lanzén A, Curtis TP, Davenport RJ, Hall N, Head IM, Read LF, Sloan WT. 2009. Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat Methods* 6:639–641. <https://doi.org/10.1038/nmeth.1361>.
46. Quince C, Lanzén A, Davenport RJ, Turnbaugh PJ. 2011. Removing noise from pyrosequenced amplicons. *BMC Bioinformatics* 12:38. <https://doi.org/10.1186/1471-2105-12-1>.
47. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 41(Database issue):D590–D596. <https://doi.org/10.1093/nar/gks1219>.
48. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 73:5261–5267. <https://doi.org/10.1128/AEM.00062-07>.