INSTITUT PASTEUR

Research in Microbiology 161 (2010) 192–197

www.elsevier.com/locate/resmic

A systematic assessment of automated ribosomal intergenic spacer analysis (ARISA) as a tool for estimating bacterial richness

Amir Kovacs, Keren Yacoby, Uri Gophna*

Department of Molecular Microbiology and Biotechnology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 69978, Israel

Received 17 November 2009; accepted 7 January 2010
Available online 4 February 2010

Abstract

ARISA (automated ribosomal intergenic spacer analysis) is a commonly used method for microbial community analysis that provides estimates of microbial richness and diversity. Here we investigated the potential biases of ARISA in richness estimation by performing computer simulations using 722 complete genomes. Our simulations based on in silico PCR demonstrated that over 8% of bacterial strains represented by complete genomes will never yield a PCR fragment using ARISA primers, usually because their ribosomal RNA genes are not organized in an operon. Despite the tendency of ARISA to overestimate species richness, a strong linear correlation exists between the observed number of fragments, even after binning, and the actual number of species in the sample. This linearity is fairly robust to the taxon sampling in the database as it is also observed on subsets of the 722 genome database using a jackknife approach. However, this linearity disappears when the species richness is high and binned fragment lengths gradually become saturated. We suggest that for ARISA-based richness estimates, where the number of binned lengths observed ranges between 10 and 116, a correction should be used in order to obtain more accurate ''species richness'' results comparable to 16S rRNA clone-library data.
© 2010 Elsevier Masson SAS. All rights reserved.

Keywords: ARISA; ITS; Richness

1. Introduction

ARISA (automated ribosomal intergenic spacer analysis) is a method of microbial community analysis that provides estimates of microbial richness and diversity and is based on the length heterogeneity of the bacterial rRNA operon 16S–23S intergenic spacer (also known as the internal transcribed spacer, or ITS). Since its introduction in 1999 [5], ARISA has been frequently used for the study of a variety of habitats, including soil, aquatic environments and the human gut [8,2,9]. ARISA has become a very useful tool for comparing community structure across multiple samples based on profile patterns. Nevertheless it can also be used to estimate bacterial richness and diversity. [6] for example, used ARISA to directly quantify species richness in a study examining the latitudinal diversity gradient in planktonic marine bacteria.

The work presented here directly addresses the issue of microbial richness estimation by ARISA. Our goal was to evaluate the errors inherent in ARISA-based richness estimates in relation to actual species richness.

When ARISA was first introduced by Fisher and Triplett, they examined the GenBank database for heterogeneity of the 16S–23S intergenic spacer among cultured organisms, for the purpose of identifying potential biases inherent in the estimates of bacterial diversity that are provided by this method. In their examination, two central conflicting biases were examined. On the one hand, ARISA may underestimate richness because unrelated microorganisms may possess spacer regions of identical length. On the other hand, multiple operons within a single genome may differ in the length of the spacer, thus leading to an overestimation of richness. Fisher and Triplett thus concluded that further investigation will be

* Corresponding author.
E-mail address: urigo@post.tau.ac.il (U. Gophna).

0923-2508/$ - see front matter © 2010 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.resmic.2010.01.006

required in order to gain a clearer understating of ARISA's biases in estimating microbial richness.

In this work, we utilized the hundreds of complete bacterial genomes now available in the NCBI database to perform in silico ARISA and generate a dataset of the bacterial 16S–23S spacers. Using this dataset, we addressed the conflicting biases mentioned above and performed simulations of ARISA for the purpose of quantifying how these biases influence overall accuracy in ARISA's estimations of bacterial species richness.

2. Materials and methods

2.1. Generation of the ITS dataset

722 bacterial complete genomes representing 490 bacterial species were downloaded from the NCBI ftp site (ftp://ftp.ncbi.nih.gov/genomes/Bacteria/) (see Supplementary table S1) and used as templates for in silico PCR using primers 16SF: 5′-GYACACACCGCCCGT-3′ and 23SR: 5′-GGTTBCCCCATTCRG-3′ [4]. A maximum of 2 mismatches was allowed for either primer, and the maximum amplified length was set to 3000 bp. All genomes were organized in an associative array in which each genome points to an array of associated ARISA fragments.

2.2. Data analysis

2.2.1. Internal genomic spacer variability

This was assessed by counting the number of discrete ARISA fragments associated with each genome.

2.2.2. Length-specific species representation

This refers to the number of species that are associated with a certain spacer amplicon length. This analysis was done at the species level for the purpose of comparing ARISA-based OTU determination to a species-level OTU count (for a parallel genome-level analysis see Supplementary Fig. S6).

2.2.3. Binning

ARISA fragments were assigned to bins of 3 bp (±1 bp) for fragments up to 700 bp length, bins of 5 base pairs for fragments between 700 and 1000 bp length and bins of 10 base pairs for fragments above 1000 bp [1].

2.3. Simulations

The dataset of 722 genomes and their in-silico-amplified PCR products (if any, see above) was used for simulations of ARISA. For the purpose of exploring a range of theoretical environments, varying in their microbial richness, 71 different simulations were performed differing in the number of genomes sampled in each one, reaching saturation of the dataset (5, 10, 20, 30…690, 700). Each such simulation involved 1000 iterations. In each of the 1000 iterations, the corresponding number of genomes was randomly sampled from the 722 genomes. The number of distinct species sampled, as well as the number of discrete ARISA fragments associated with these species, were recorded for each of the 1000 iterations, and mean values of these two parameters were calculated for each specific simulation. The error in species-level richness estimation was calculated by subtracting the mean value of sampled species from the mean value of ARISA fragments.

In order to assess the robustness of our findings in relation to the present sampling of genomes in the NCBI database, a jackknife simulation was carried out. This simulation involved the creation of 20 sets of 300 randomly chosen genomes (out of the 722). Each of these sets was treated as a new database, and 71 corresponding simulations were carried out for each of the sets.

All the simulations were performed using PERL scripts, which are available upon request.

2.4. ARISA

PCR reactions were performed in duplicate and each 25 μl reaction contained 1.25 U of Taq DNA polymerase (BIOTAQ™ by Bioline, UK), 3 mM of $MgCl_2$, 2.5 μl 10× PCR buffer, 0.1 mM of each dNTP, ultra-pure water (Biological Industries, Israel) and 1 pmol of both primers (16S–1392F and a 5′ TET labeled 23S–125R, sequences detailed above). The reaction was performed as follows: preliminary denaturation, 3 min at 94 °C; 32 cycles of 1 min denaturation at 94 °C, 1 min annealing at 52 °C, 1.5 min elongation at 72 °C; and 6 min final elongation at 72 °C.

2.4.1. Fragment analysis

PCR products were analyzed using the ABI PRISM 3100 Genetic Analyzer. The labeled fragments were separated on the capillary sequencer along with a custom-made ROX-labeled 250–1150 bp standard (Bioventures, USA).

2.4.2. Richness estimation

Raw data generated by the ABI PRISM 3100 Genetic Analyzer was initially analyzed using GeneMarker™ (SoftGenetics, USA). After performing accurate size calling using the program, all data were exported to Microsoft Excel for further analysis. In Excel, first, all OTUs with fluorescence intensity of 10 RFUs (relative fluorescence units) or lower were excluded. Second, all OTUs were binned (as detailed in the previous section under 'binning') and intensities were summed up for each bin. Next, relative intensities for each binned OTU in a certain sample were calculated, and binned OTUs which contributed less than 0.5% to the total intensity of the sample were excluded. Duplicates were compared to each other, and OTUs that appeared in only one of the duplicates were excluded, in order to minimize the effect of false peaks resulting from instrument noise, thus preventing overestimation of richness. Finally, richness was determined by counting the number of remaining OTUs in each sample.

2.5. Experimental testing of the improved 23SR primer

The aquatic samples were taken from the Yarqon River, an urban stream crossing Tel Aviv, and the gut samples were

Download English Version:

https://daneshyari.com/en/article/4359079

Download Persian Version:

https://daneshyari.com/article/4359079

Daneshyari.com