

Metagenomics insights reveals functional patterns among soil microbial communities of global biomes

Melline Fontes Noronha¹, Gileno Vieira Lacerda Junior¹, Jack A. Gilbert ^{2,3} and

Valéria Maia de Oliveira¹

1 Microbial Resources Division, Research Center for Chemistry, Biology and Agriculture (CPQBA), University of Campinas, 2 The Microbiome Center, Department of Surgery, University of Chicago, Chicago, IL, USA and 3 The Microbiome Center, Bioscience Division, Argonne National Laboratory, Lemont, IL, USA

A biome is a geographical unit characterized according to its vegetation type, macroclimate, soil, and specific elevation. In contrast, a microbiome is a mix of microorganisms that coexist in a defined space. Although soil microbial communities have shown to vary across many spatial scales, soils between ecosystems showed to be leading by some biogeographical trends. In order to investigate functional convergence within soils from the same biome type, thirty publically-available metagenomes from 11 globally distributed biomes were selected and clustered by biome groups (i.e. forest, grasslands, tundra, semiarid and desert) based on vegetation features. Functional analyses revealed a close pattern among biome groups, in which DNA repair, central carbohydrate metabolism, and antibiotic resistance were the most statistically different metabolism annotated by SEED subsystems among biome groups. In order to provide a better analytical resolution of those metabolisms, metagenomic reads were annotated using the Carbohydrate-Active enZymes database (Cazy), Antibiotic Resistance gene DataBase (ARDB) and, additionally, the Heat Shock Protein Information Resource (HSPiR). Carbohydrate-active enzyme analyses showed that biomass degradation, sucrose and starch metabolism, cell wall biosynthesis and alginate degradation were overrepresented in forest and grasslands soils. As expected, desiccation and other stress resistance genes were more abundant in deserts and semiarid soils. Antibiotic Resistance Genes (ARGs) were prevalent in forest and grassland soils, where multidrug efflux pumps were the most abundant ARG class, with the majority of the reads assigned to Proteobacteria. Heat Shock Proteins (HSPs) were more abundant in tundra, semiarid and desert soils. Although HSP70 and HSP100 were uniformly distributed across biomes, while HSP60 and HSP20, which are predominantly from the Archaea, were more abundant in the Saline Desert soils. Our results suggest that local environmental conditions select for the enrichment of specific functions important for survival in those ecosystems.