

Supplementary Information

for

Scaffold-level validation is needed for phage-host inference from metagenome-assembled genomes

Ling Yuan (袁凌)¹, Karthik Anantharaman², Bin Ma (马斌)^{3,4}, Antonio Perdo Camargo^{5,*}, LinXing Chen (陈林兴)^{1,*}

¹ State Key Laboratory of Advanced Environmental Technology, the Department of Environmental Science and Engineering, University of Science and Technology of China, Hefei, China

² Department of Bacteriology, University of Wisconsin-Madison, Madison, WI, USA

³ State Key Laboratory of Soil Pollution Control and Safety, Zhejiang University, Hangzhou, 310058, China

⁴ Zhejiang Provincial Key Laboratory of Agricultural, Resources and Environment, College of Environmental and Resource Sciences, Zhejiang University, Hangzhou, 310058, China

⁵ Department of Biochemistry, Institute of Chemistry, University of São Paulo, São Paulo, SP 05508-060, Brazil

*Corresponding author:

Antonio Pedro Camargo, antonio@iq.usp.br

LinXing Chen, linxingchen@ustc.edu.cn

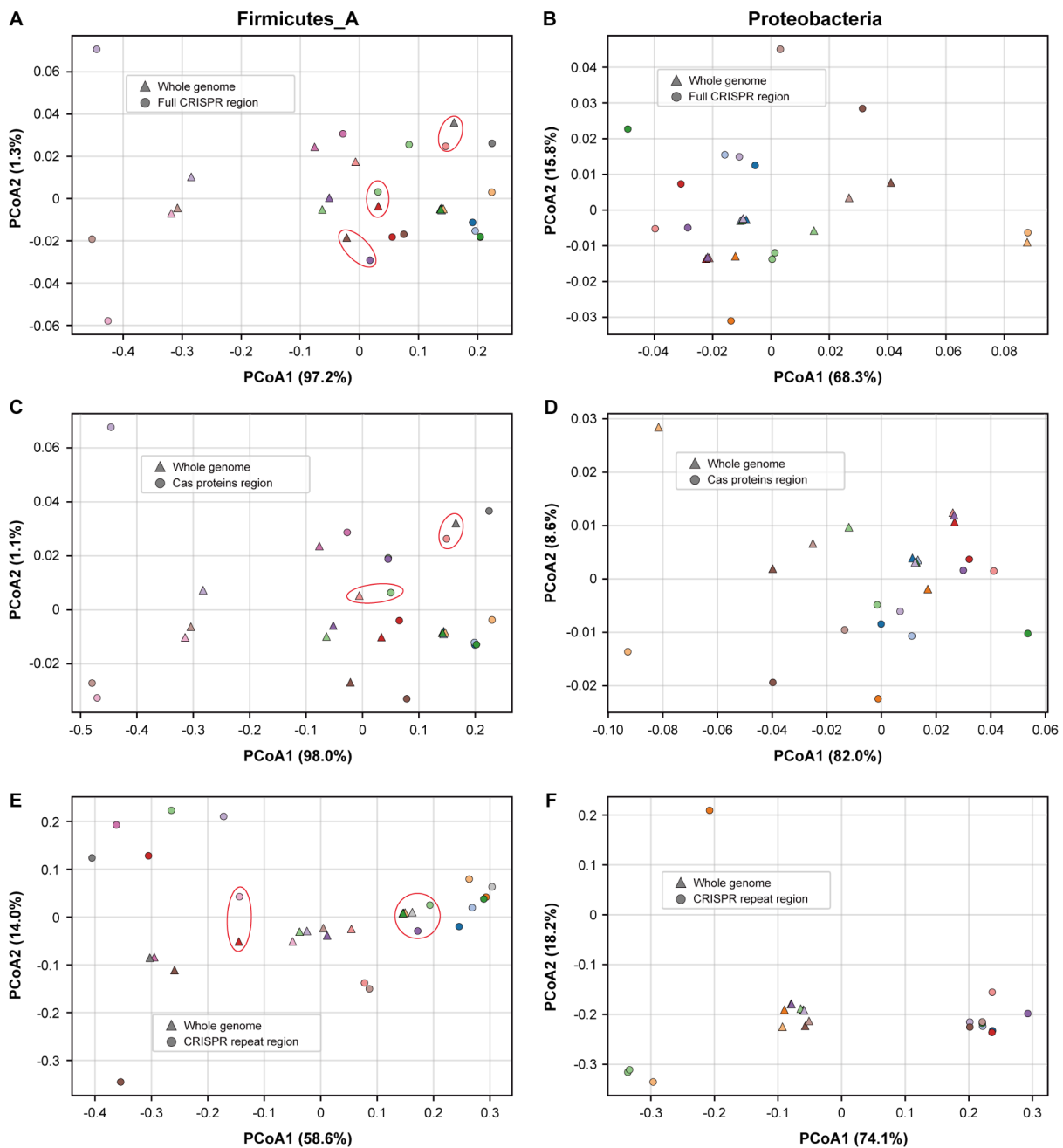


Figure S1 | TNF-based PCoA comparison of CRISPR-related sequences against the whole genomes at the genome level. Comparison of the whole genomes against (A) and (B) the full CRISPR-cas system regions, (C) and (D) the cas proteins regions, (E) and (F) the CRISPR repeat regions, for those from the phylum of (A) (C) (E) Firmicutes_A and (B) (D) (F) Pseudomonadota. Red circles indicate sequences of whole genomes and CRISPR-related regions from different genome, which could be binned together if with comparative sequencing depth and presented in the same community.