Method for COGs analysis:

For mapping COG annotations with respective protein IDs, .fasta file of whole proteome *in A. brasiliense* SP7 was acquired using UniProt. The .fasta file of the whole proteome is subsequently submitted to WebMGA (<http://weizhong-lab.ucsd.edu/webMGA/server/cog/>

), an open-source web service for metagenomic analysis1. Specifically, the sequences of each protein were mapped with COG IDs by predicting the relatedness to protein sequences presented in the COG database. Cut-off prediction of 0.05 was used for the mapping algorithm. 5029 proteins in the proteome were mapped to respective COG IDs, and 608 proteins were not mapped. Mapped COG IDs were functionally annotated using the updated 2014 version of the 2003 COG annotation files (<https://www.ncbi.nlm.nih.gov/COG/>)2. General analyses, including the enrichment analyses using Fischer’s exact test, were performed using a customized Python algorithm. The un-mapped proteins were not considered for any COG analyses.

1) S. Wu, Z. Zhu, L. Fu, B. Niu and W. Li, “WebMGA: a Customizable Web Server for Fast Metagenomic Sequence Analysis”, BMC Genomics 2011, 12:444doi:10.1186/1471-2164-12-4.<http://www.biomedcentral.com/1471-2164/12/444>liwz@master:/home/oasis/data/www/home/metagenomic-analysis/wiki/data/pages$

2) Galperin MY, Makarova KS, Wolf YI, Koonin EV. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res*. 2015;43(Database issue):D261-D269. doi:10.1093/nar/gku1223