

SUPPLEMENTARY MATERIAL

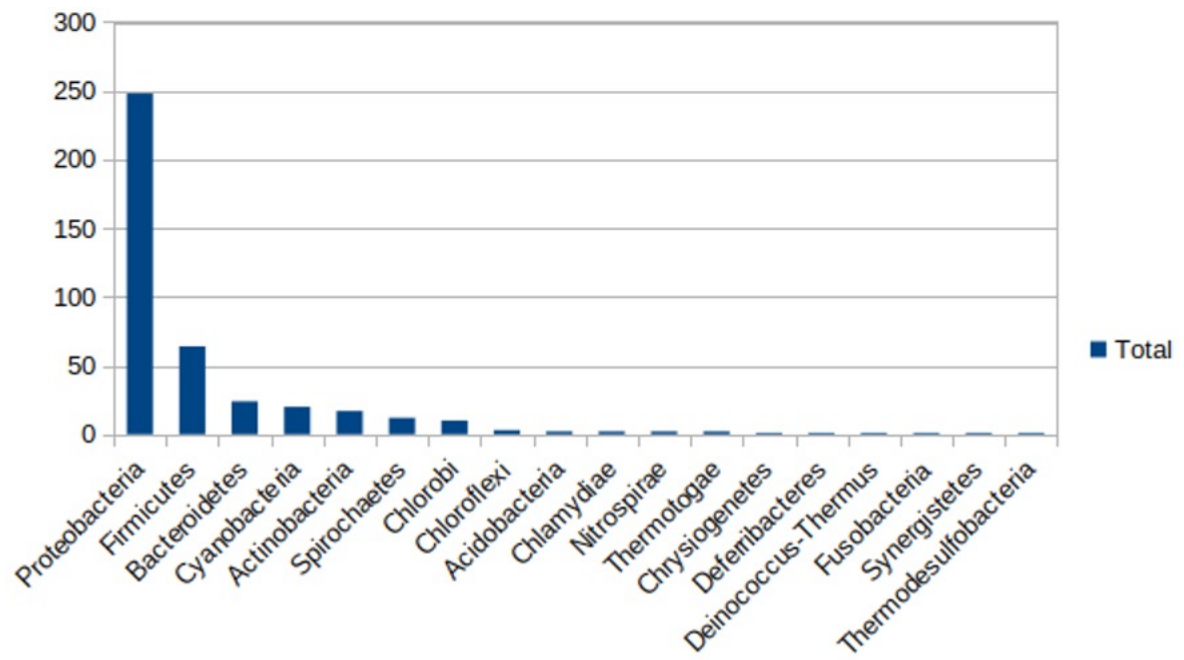
Origin and evolution of nonulosonic acid synthases and their relationship with bacterial pathogenicity revealed by a large-scale phylogenetic analysis

^{a,b}Alexandre Zanatta Vieira, ^bRoberto Tadeu Raittz, ^{a,b}Helisson Faoro*.

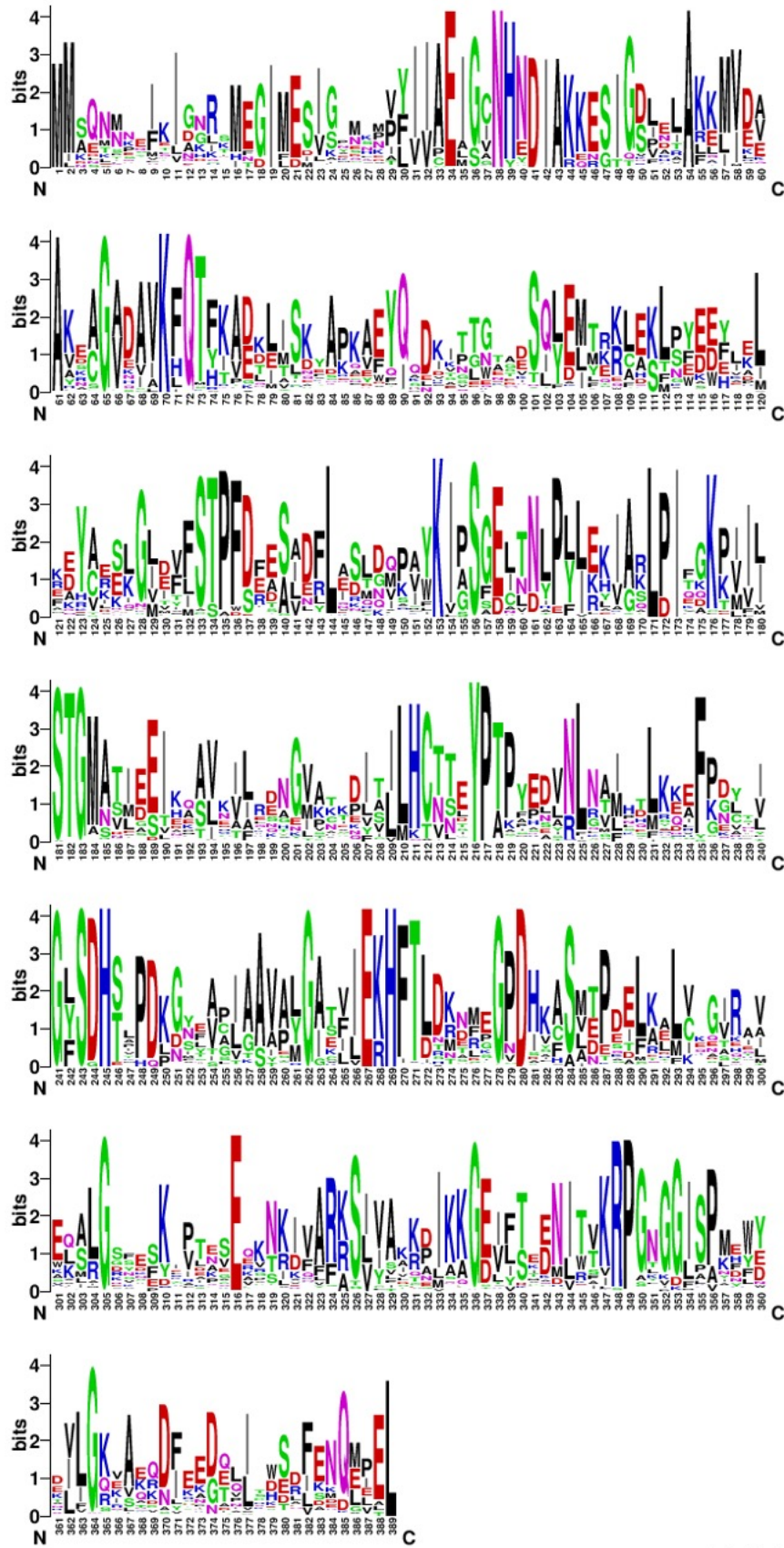
^aLaboratory for Applied Science and Technology in Health, Carlos Chagas Institute, Fiocruz-PR, Algacyr Munhoz Mader street, 3775, Curitiba, Paraná, Brazil, 81350-010.

^bGraduation Program on Bioinformatics – Universidade Federal do Paraná, Alcides Viera Arcoverde street 1225, Curitiba, Paraná, Brazil, 81520-260.

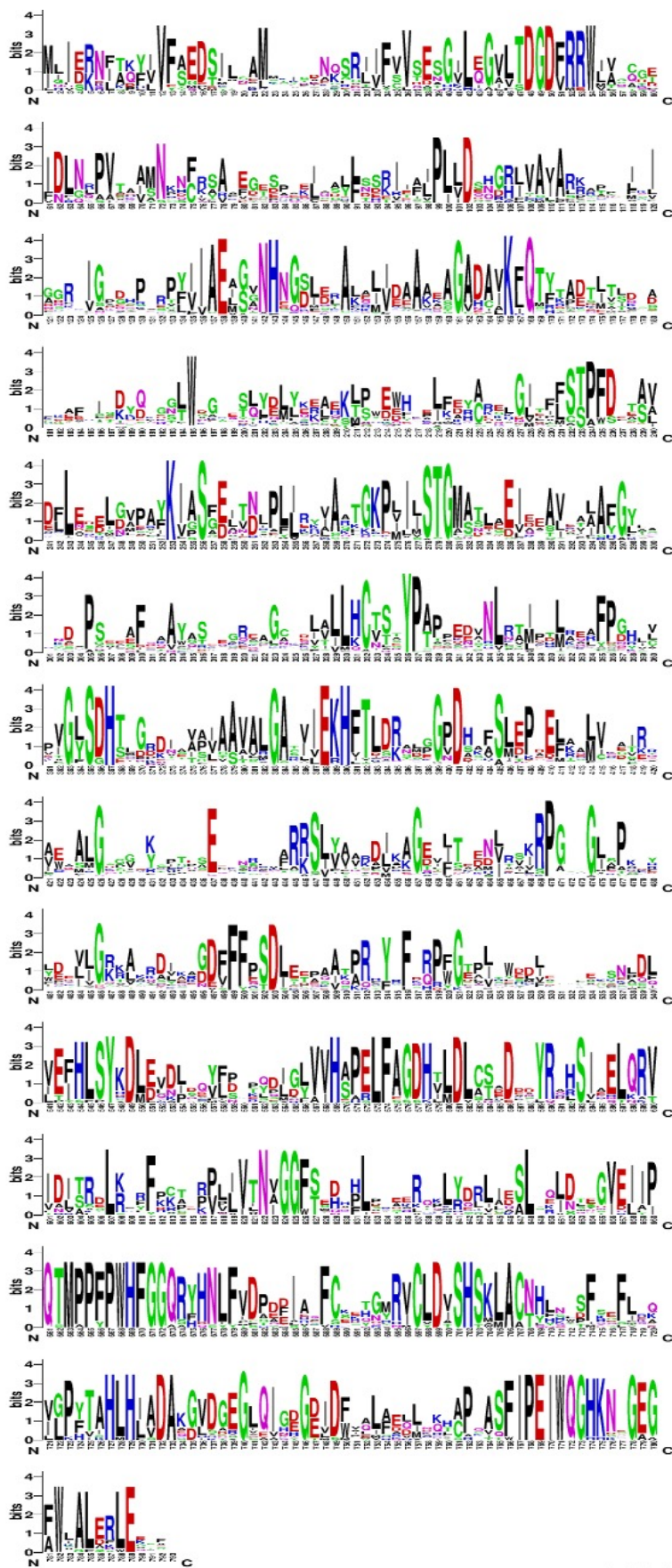
Supplementary Fig. S1. Phyla distribution of the 412 NeuB sequences from KEGG.



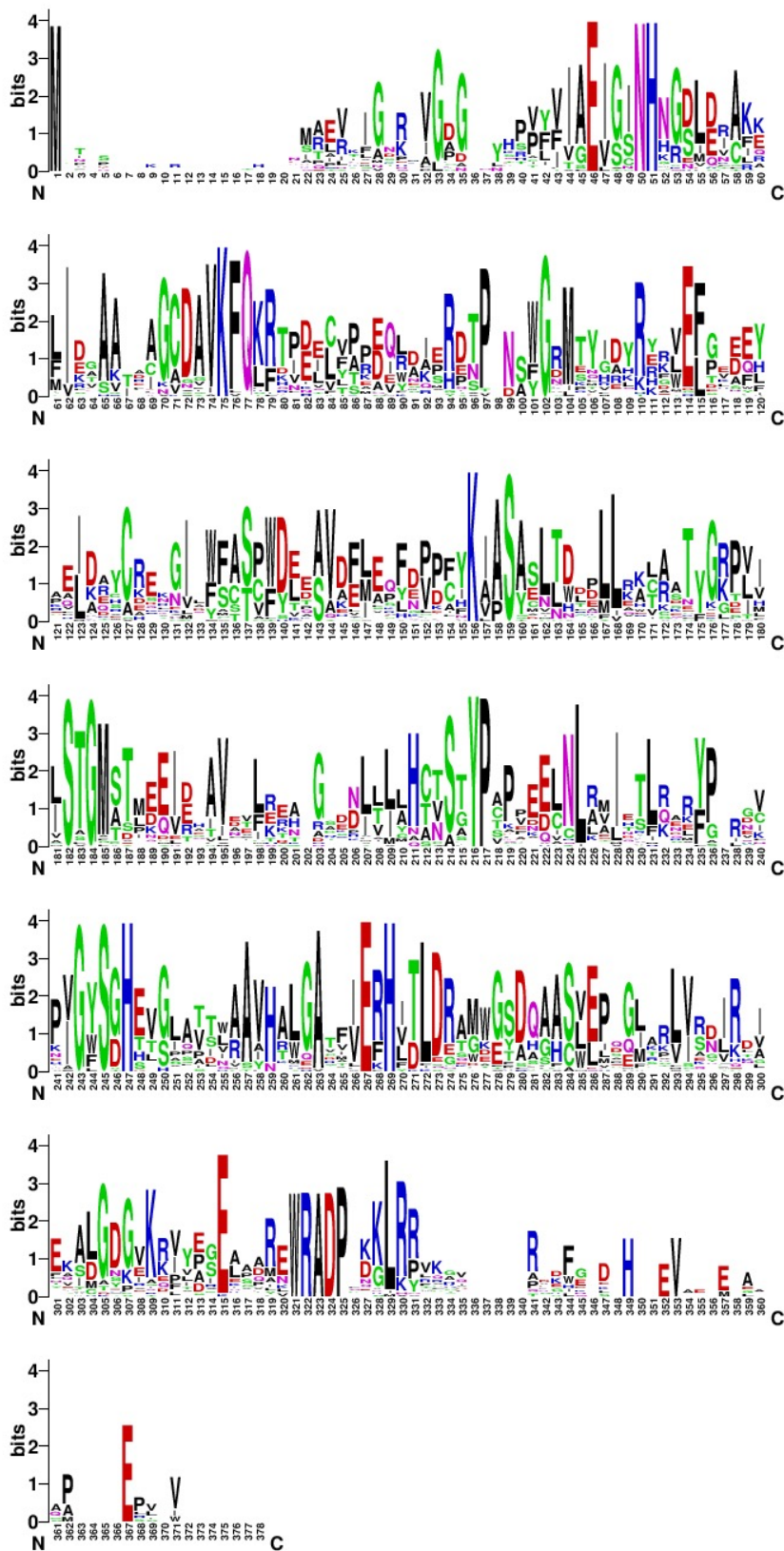
Supplementary Fig. S2. Sequence WebLogos of pathogenic NeuB homologues profile.



Supplementary Fig. S3. Sequence WebLogos of environmental NeuB homologues with SAF domain profile.

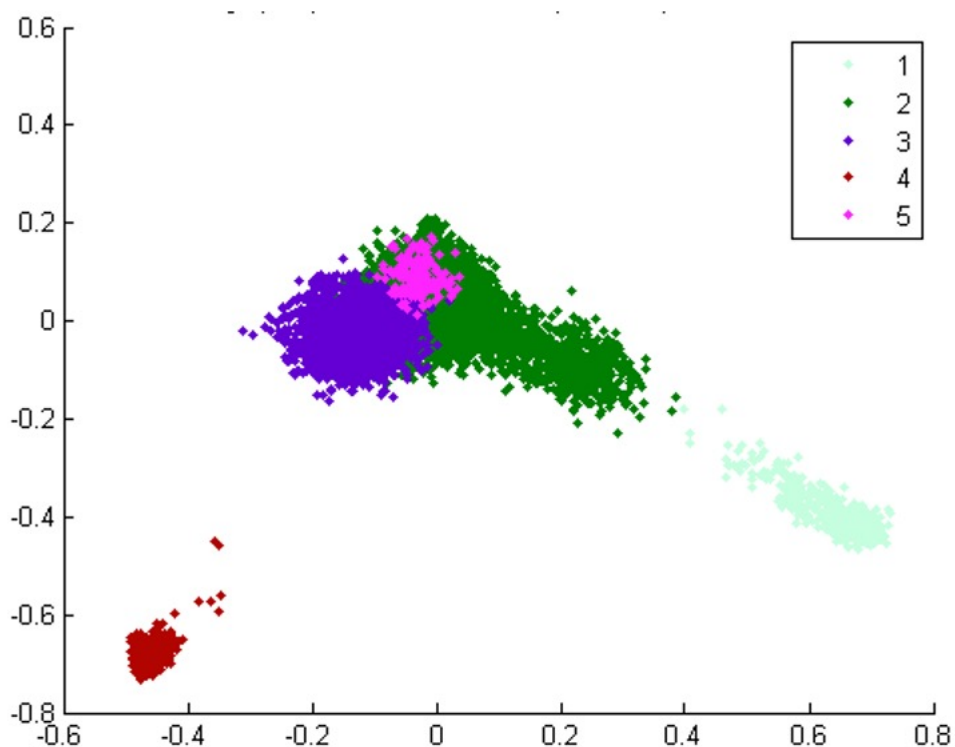


Supplementary Fig. S4. Sequence WebLogos of environmental NeuB homologues without SAF domain profile.



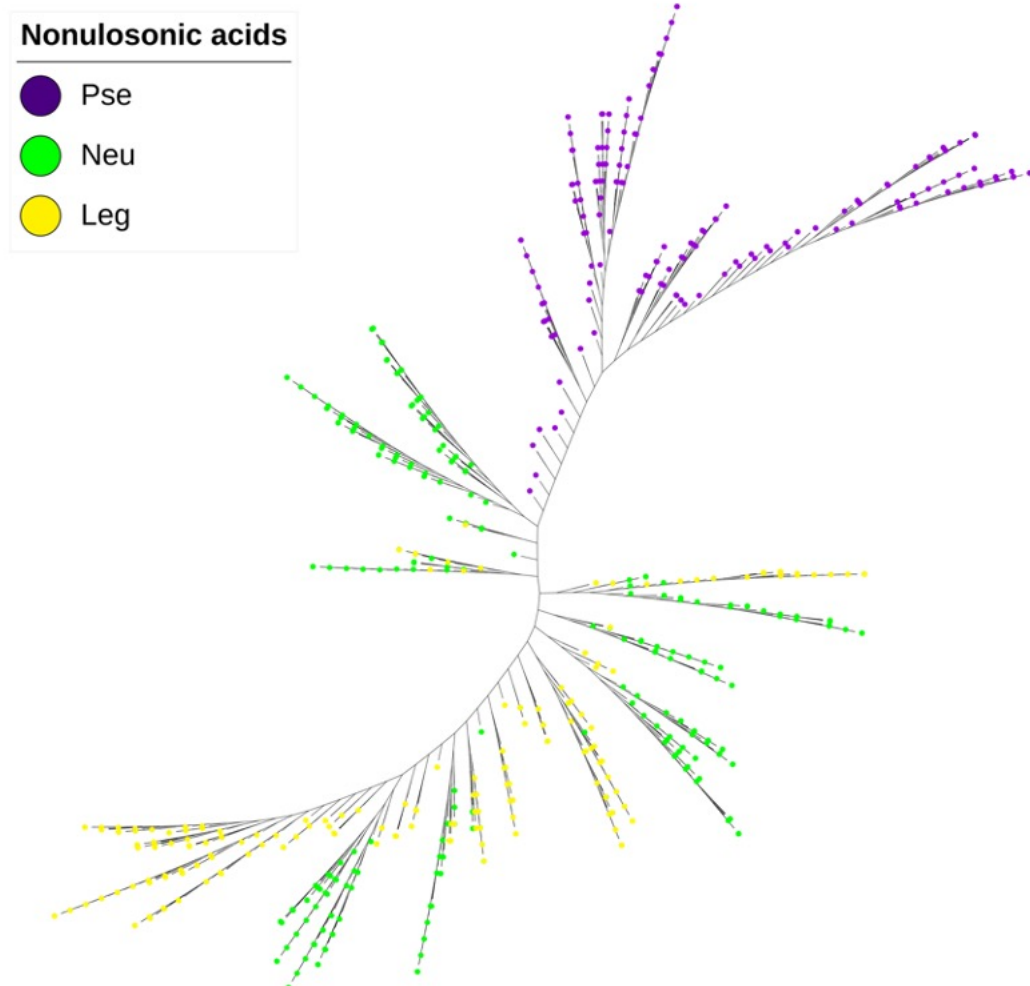
Supplementary Fig. S5. K-means clustering of the 13,941 NeuB sequences and its homologues using SweeP alignment-free clustering tool. All sequences were trimmed leaving only the NeuB domain. The clustering has three distinct main clusters. The largest of the clusters still has three divisions. The three main clusters indicate a clear separation between the 3 bacterial sialic acids. 1-NeuB, 2-Legl, 3-Psel, 4-outros, 5-NC e 6-NanS

1- LegI Majority (2.5.1.101); 2- Mixture of NeuB, NanS and LegI (2.5.1.56, 2.5.1.57 and 2.5.1.101); 3- Mixture of NeuB, NanS and Psel (2.5.1.56, 2.5.1.57 and 2.5.1.97); 4- Majority Psel (2.5.1.97); 5- Majority NeuB and NanS (2.5.1.56, 2.5.1.57)



Supplementary Fig. S6. Phylogenetic analysis of bacterial NeuB and NeuB homologues extracted from KEGG. Phylogenetic tree created with sequence of NeuB (K01654), PseI (K15898) and LegI (K18430) extracted from KEGG database. the presence of the NeuB domain was confirmed by comparison with the CDD database. The sequences were aligned using MAFFT and a maximum likelihood tree was inferred using the RaXML tool with 1,000 fast bootstrap replicates.

Tree scale: 1 



Supplementary Fig. S7. Phylogenetic analysis of bacterial NeuB and NeuB homologues. The NeuB HMM profile, constructed with the sequences from the KEGG database, was used to mine NeuB sequences and their homologues in the UniprotKB database. A total of 13,941 sequences were recovered and the presence of the NeuB domain was confirmed by comparison with the CDD database. The NeuB domains (~ 250 aa) were separated from the rest of the sequences and used in phylogenetic analysis. The Neighbor-joining tree was calculated using SWeeP, an alignment-free clustering tool. The orange dots represent bacterial sequences marked with the E.C Number 2.5.1.57, NanS. The colored clades represent the pathogenic organisms identified through GOLD and KEGG databases.

Tree scale: 1 

Nonulosonic Acids

 Nan

Pathogen

 Pathogen

