SUPPLEMENTARY DATA

Manuscript title: Increased Gut Redox and Low Anaerobic and Methanogenic Prokaryotes in Severe Acute Malnutrition

Author list: Matthieu MILLION, Maryam TIDJANI ALOU, Saber KHELAIFIA, Dipankar BACHAR, Jean-Christophe LAGIER, Niokhor DIONE, Souleymane BRAH, Perrine HUGON, Vincent LOMBARD, Fabrice ARMOUGOM, Julien FROMONOT, Catherine ROBERT, Caroline MICHELLE, Aldiouma DIALLO, Alexandre FABRE, Prof. Régis GUIEU, Cheikh SOKHNA, Prof. Bernard HENRISSAT, Prof. Philippe PAROLA, Prof. Didier RAOULT

Supplementary material and methods

v3v4 small subunit rRNA gene tag sequencing

In order to control for extraction bias, 10 different protocols were compared and only 3 detected the archaea. Protocol 5, the best protocol for the detection of *Euryarcheotae* (pending patent), was retained in association to another protocol (protocol 1). Samples were sequenced for 16S rRNA sequencing on MiSeq technology. Samples were amplified individually for the 16S “V3-V4”regions by PCR, using the Taq Phusion (Thermo Fisher Scientific Inc, Waltham, MA U.S.A) and the surrounding conserved regions V3_V4 primers with overhang adapters (FwOvAd_341F TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG ; RevOvAd_785R GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC). After purification on AMPure beads (Beckman Coulter Inc, Fullerton,CA,USA), concentrations were measuring using High sensitivity Qubit technology (Life technologies, Carlsbad, CA, USA). Using a subsequent limited cycle PCR on 1 ng of each PCR product,

Illumina sequencing adapters and dual-index barcodes were added to each amplicon. After purification on AMPure beads, libraries were then normalized according to the Nextera XT protocol (Illumina Inc, San Diego, CA, USA). The multiplexed samples were pooled into a single library for sequencing on the MiSeq (Illumina Inc, San Diego, CA, USA). Automated cluster generation and paired-end sequencing with dual index reads was performed in a single 39-hour run in a 2x250bp. On the instrument, the global cluster density and the global passed filter per flowcell were generated. The MiSeq Reporter software determined the pourcentage of index for each amplicon. Raw data were configurated in fastaq files for R1 and R2 reads.

The Paired End reads from the raw fastq files of Illumina Miseq were filtered and assembled using pandaseq. The good quality sequences were then extracted from the fasta file, produced by pandaseq, only if the sequence contains both the primers (forward and reverse) being used in PCR amplification. In the next filtering step all the sequences containing N were removed and at the same time sequences shorter than 200 nucleotides were also discarded. Finally, both the primers were removed from the sequences.

In the following steps after filtering the filtered sequences were strictly dereplicated (clustering of duplicate sequences) and were sorted by decreasing number of abundance *(58)*. These sequences in sorted order were then clustered at k = 10 (>= 97% identity) number of differences. In the following step OTUs (representative sequence of each cluster) were extracted, where the OTUs are the unique sequences from each cluster which has the highest abundance during the PCR amplifications. The abundance information of each sequence were calculated during the strict dereplication step as described earlier. OTUs with single representative sequences were discarded.

The release 115 of the SILVA SSU and LSU database were downloaded from SILVA website and from this a local database of predicted amplicon sequences were built. During the construction of our local reference database, we considered only those Silva SSU reference

sequences which contains the both forward and reverse primers (used in our PCR amplifications), by allowing 3 differences between each primer and the Silva reference sequences. Finally, our local reference database contains a total of 479,927 well annotated sequences.

The OTUs (representative sequence) extracted in the previous step were searched against our local reference database by using the Needleman–Wunsch global alignment algorithm. The 100 best matches above 80% similarity with each of the representative sequences were extracted from the reference database. These extracted reference sequences were sorted according to the decreasing percentage of similarity. The reference sequences with the highest percentage of similarity (by considering also all the hits within 0.5% similarity of the best hits) with the OTUs were used for taxonomic assignments, and taxonomy to the lowest rank were obtained by applying a majority voting (at least 60% of the hits must agree to the lowest rank). In the case where there are two results with same percentage of similarity, taxonomy to each rank were obtained by the consensus of these taxonomies. For example, a tag with 98% similarity to the class Gammaproteobacteria and Alphaproteobacteria was only assigned to the phylum Proteobacteria. When similarity was 80%, sequences were not assigned. Finally all the tags were clustered to different taxon ranks according to the consensus taxonomy of the unique tags (Representative of each OTUs).

Taxonomic assignment at the species level for studies from the literature

For the study by Ghosh *et al. (27)*, the whole genome shotgun sequence data were downloaded from NCBI Sequence Read Archive (SRA,http://www.ncbi.nlm.nih.gov/Traces/sra) with the following accession numbers SRR1067674, SRR1087919, SRR1068216,SRR1067721, SRR1068217, SRR1068218, SRR1068219,SRR1068596, SRR1068597, SRR1067720, SRR1067719,SRR1087910,

SRR1087911, SRR1067716, SRR1087913, SRR1087914, SRR1087915, SRR1087916, SRR1067717, SRR10879 as described in the paper of Ghosh *(27)*. These data were given as input to MetAmos *(59)* pipeline. In MetAmos Fastqc was used for filtering the low quality reads, velvet *(60)* was used for assembling the good quality sequences into contigs and finally, FragGeneScan *(61)* was used to search and extract the ORFs from the assembled sequence reads.

For taxonomic assignment, we have downloaded the whole NCBI NR *(62)* database and from this we have built our own local database using Pauda *(63)*. ORFs produced by MetAmos were then searched against our local NR database using Pauda *(63)* with –slow option. The search results were then imported to the latest version of Megan *(64)* and taxonomic assignment were performed with Megan *(64)*.

For Smith *et al. (17)*, taxonomic assignment was performed by the authors and available online. For Subramanian *et al. (18)*, individual data were not obtained but as taxonomic assignment and comparison of cases and controls was performed by the authors using a linear mixed model (ruling out the confounding role of age), we choose to directly extract the species enriched or depleted in malnutrition according to supplementary Table 15a of *(18)*.

Measurement of gut pH and redox potential

1 g of stool were diluted in 10 mL of distilled water and centrifugated at 8000 rpm for 10 minutes. A pH and redox meter (PCE-228-R, PCE Instruments, Southampton, United Kingdom) was used to measure the pH and the redox potential according to the manufacturer's instructions.

Bacterial quantification by flow cytometry

For flow cytometry analyses, the BD LSR Fortessa™ Cell Analyzer (Becton Dickinson, Le Pont de Claix, France) was used. We performed bacterial quantification using a suspension of fluorescent microspheres (Cytocount, Dako, Les Ulis, France) as a reference population. The absolute number of bacterial cells (cells/µl) in each sample is calculated using the following equation: (Number of cells counted / Number of Cytocount™ beads counted) × Cytocount™ concentration (1100 beads/µl) × dilution factor *(65)*. The data were analyzed using BD FACSDiva 6.2 Software.

Supplementary Table S1: Ten matched pairs used to control the age bias

Study	Center	Pair number	Name of the children with SAM	Age (months)	Sex	Name of the control	Age (months)	Sex
Million, 2015	Dakar	1	KW04_S	4	F	S4_S	4	F
Million, 2015	Dakar	2	KW03_S	6	F	S2_S	6	F
Million, 2015	Dakar	3	KW01_S	7	F	S3_S	7	F
Million, 2015	Dakar	4	KW07_S	15	M	S9_S	19	M
Million, 2015	Dakar	5	KW02_S	49	F	S7_S	38	F
Million, 2015	Dakar	6[a]	KW09_S[a]	12.5	F	MR03_S[a]	11	F
Million, 2015			**Mean age ± SD**	**15.6 ± 16.9**	**5F/6**		**14.2 ± 12.8**	**5F/6**
Ghosh, 2014	Birbhum district	7	976AB	35	M	AA6	33	M
Ghosh, 2014	Birbhum district	8	A1	48	M	683PD	48	M
Ghosh, 2014			**Mean age**	**41.5**	**2M/2**		**40.5**	**2M/2**
Smith, 2013	Mbiza	9	K56B.2KwPre	18.1	F	H56A.1[b]	17.2	F
Smith, 2013	Chamba	10	K268B.4KwPre	8.5	M	H268A.4Pre	8.5	M
Smith, 2013			**Mean age**	**13.3**	**1F/2**		**12.8**	**1F/2**

[a]Gut total bacterial concentration could not be assessed on this pair because sample MR03_S was exhausted. [b]This sample (H56A.1) was selected instead of that (H56A.2Pre) collected at the time of the admission of the severely undernourished co-twin (K56B.2KwPre). Indeed, at the time of H56A2.Pre collection, the 'healthy co-twin' was actually suffering from moderate stunting (HAZ = -2.01).

Supplementary Table S2: Detection of *Methanobrevibacter smithii*

	qPCR		v3v4 pyrosequencing[a]	
	SAM	**Control**	**SAM**	**Controls**
Dakar, Senegal	0/10 (0%)[b]	7/11 (63%)	0/10 (0%)	1/11 (9.0%)
Niamey, Niger	0/10 (0%)[b]	8/9 (88.8%)	0/10 (0%)[b]	7/9 (78%)
All children	0/20 (0%)[b]	15/20 (75%)	0/20 (0%)[b]	8/20 (40%)

[a]Detection was considered without any normalization (1 read or more were considered as positive). [b]Significantly different from controls (Two-sided Barnard bilateral test).

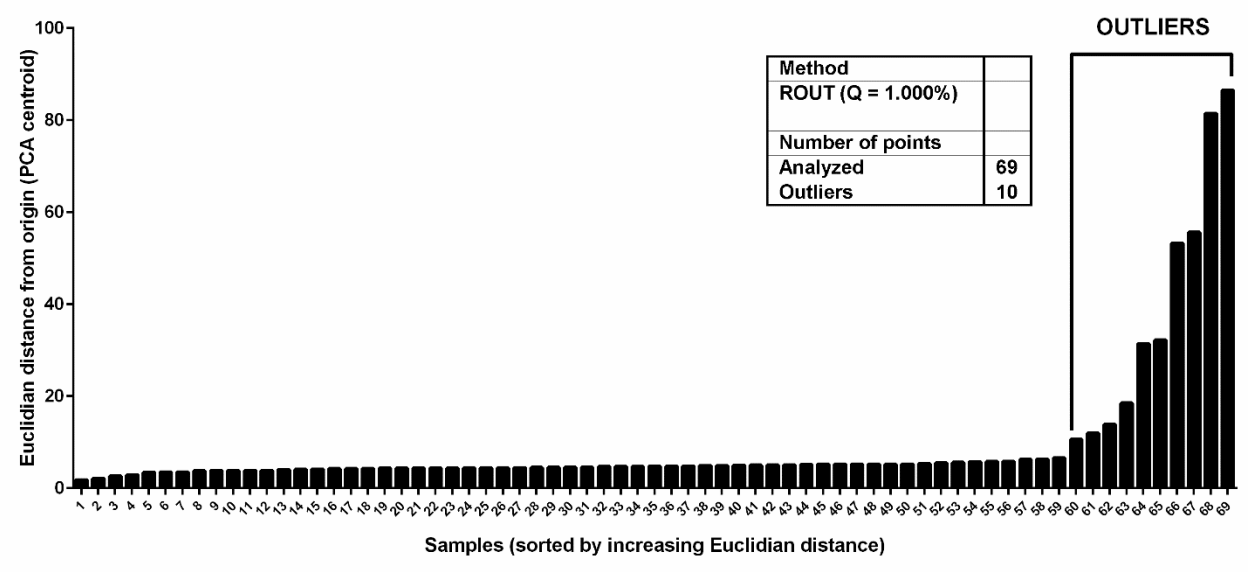
Supplementary Table S3: PICOS criteria for inclusion and exclusion of studies and individuals

Parameter	Inclusion criteria	Exclusion criteria
Patients	Cases: Children ≤ 60 months with severe acute malnutrition	Animal studies
	Controls: Apparently healthy children ≤ 60 months without moderate acute malnutrition, severe wasting or severe stunting	Children > 60 months and adults
	Cases and controls from the same geographical location (city or region)	Cases and controls from different geographical locations
		Children with moderate acute malnutrition
		Recent or ongoing antibiotic treatment[a]
		Controls with severe wasting or severe stunting[b], infection, diarrhea[a]
Intervention	No intervention	Children under ready-to-use therapeutic food
		Children under antibiotics
Comparator	Prokaryotic composition of the gut microbiota analysed by metagenomics analysis	Absence of metagenomics analysis
		Studies focusing on eukaryotes
		Studies using different metagenomics chemistry analysing cases and controls
Outcomes	Studies with metagenomics data allowing assignment at the species level or comparison at the species level between cases and controls	Studies without metagenomics data available up to the species level
	Observational studies	Reviews

Study design	Cohort studies	Case reports
	Longitudinal studies	Editorial & opinion pieces

PICOS: patients, intervention, comparator, outcomes, study design. [a]According to the definition of authors of each study. [b]Severe wasting and stunting was not a reason for exclusion of cases. [a]Cases with diarrhea were included as diarrhea is a frequent complication of severe acute malnutrition.

Supplementary Figure 1. Euclidian distance to the framework origin of each gut metagenome (Pearson primary component analysis)



Euclidian distance ($\sqrt{F1^2 + F2^2 + F3^2}$) from origin (PCA centroid) was calculated for each sample based on the first three component of a Pearson primary component analysis performed on the relative abundance of 2245 species detected in at least one of the 69 analysed sample with at least 100,000 reads. The ROUT method (Q =1%) *(24)* identified 10 outliers (6 healthy child from Dakar, Senegal (S01_S, S04_S, S05_S, S06_S, S09_S, S10_S), 2 children with SAM from Dakar, Senegal (MR06_S, KW08_S), and 2 children with SAM from Niamey, Niger (KW13_N, KW14_N). Analysis performed with Graphpad Prism 6.07 for Windows.

References

58. T. Stoeck, *et al.* Massively parallel tag sequencing reveals the complexity of anaerobic marine protistan communities. *BMC Biol*. **7**, 72- (2009).
59. T.J. Treangen, *et al.* MetAMOS: a modular and open source metagenomic assembly and analysis pipeline. *Genome Biol*. **14**, R2 (2013).
60. D.R. Zerbino and E. Birney, Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*. **18**, 821-829 (2008).
61. M. Rho, H. Tang, and Y. Ye, FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res*. **38**, e191 (2010).
62. E.W. Sayers, *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. **37**, D5-15 (2009).
63. D.H. Huson and C. Xie, A poor man's BLASTX - high-throughput metagenomic protein database search using PAUDA. *Bioinformatics*. **30**, 38-39 (2014).
64. D.H. Huson, S. Mitra, H.J. Ruscheweyh, N. Weber, and S.C. Schuster, Integrative analysis of environmental sequences using MEGAN4. *Genome Res*. **21**, 1552-1560 (2011).
65. M.M. Khan, B.H. Pyle, and A.K. Camper, Specific and rapid enumeration of viable but nonculturable and viable-culturable gram-negative bacteria by using flow cytometry. *Appl Environ Microbiol*. **76**, 5088-5096 (2010).