A single betaproteobacterium dominates the microbial community of the crambescidine-containing sponge *Crambe crambe*

Julie Croué [1,2], Nyree J. West [1,3], Marie-Line Escande [1,4], Laurent Intertaglia [1,3], Philippe Lebaron [1,2] and, Marcelino T. Suzuki * [1, 2]

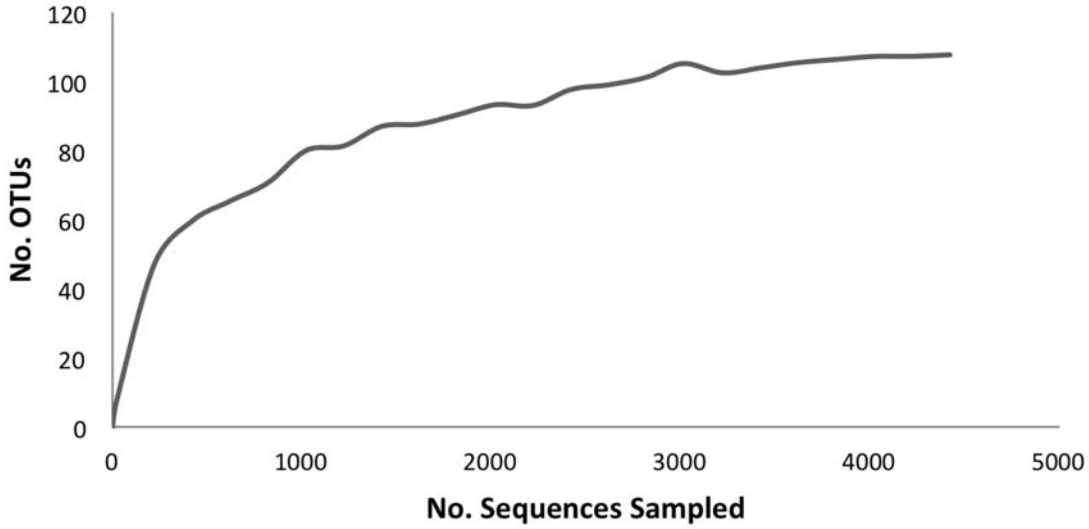
[1] UPMC Univ. Paris 06, UMR 7621, LOMIC, UMR 7232, BIOM, UMS 2348 (Plate-forme Bio2Mar), Observatoire Océanologique, F-66650 Banyuls-sur-Mer, France;

[2] CNRS, UMR 7621, LOMIC, F-66650 Banyuls-sur-Mer, France;

[3] CNRS, UMS 2348 (Plate-forme Bio2Mar), Observatoire Océanologique, F-66650, Banyuls/Mer, France;

[4] CNRS, UMR 7232, BIOM, Observatoire Océanologique, F-66650, Banyuls/Mer, France.

Supplementary Figure S1. Rarefaction curve of bacterial OTUs associated with *Crambe crambe*. OTUs were binned at 97% sequence identity.



Supplementary Methods

Chimera checking pipeline for 16S rRNA gene clone library sequences and for the pyrosequencing data

After assembly of the partial 16S rRNA gene sequences with the gap4 program[1], the edited contigs were exported in fasta format including the number of reads per contig as weights, and checked for chimeras using a stringent chimera checking pipeline consisted of 1) uchime analysis[2], options -minh 0.8, -noskipgaps –noskipgaps2; thereafter chimeric sequences were removed and the analysis was repeated with the sequences used option –abskew=1 (i.e. chimeras and parents can have equivalent weights) 2) Bellerophon[3] and 3) As 1 and 2 above yielded singletons as putative chimeras and parents, we performed a semi-automated analysis where all contigs were queried by blastn (options -W7 -r2 -G5 -E 2 -F F -b10 -v10 –e1) against a curated version of the Silva database SSURef_111_SILVA_ NR_98_26_07[4] using the parameters: silva quality>99 and silva pintail value=100 and including the Silva taxonomy. Blast results were parsed to output the number of mismatches and gaps per blastn alignment line, using shell scripts and the resulting file visually inspected to define anomalous regions (i.e. all mismatches situated in 5', center or 3'). Individual regions of putative chimeric sequences were queried separately by blastn as above, and sequences with two regions assigned to different Silva taxonomic clades were considered as chimeric.

454-pyrosequencing data analysis pipeline

Multiplex raw sff files were analyzed using a hybrid analysis pipeline. Flowgram files were generated using sffinfo (Roche). Recovery of sequences, and denoising analysis was performed using binaries and scripts in AmpliconNoise V1.25[5] implemented via the ampliconnoise.py and workflow.py scripts of Qiime V1.5[6]. All parameters were the default for Titanium chemistry, except that chimera analysis by Perseus was not done. Instead the weighted fasta output of SeqDist was parsed with awk to create an input file for *de novo* chimera detection and removal by the uchime module[2] of Usearch V5.0[7] using default parameters and a minh value of 0.80. Sequences deemed non-chimeric by uchime were manually analyzed as described for the clone library and also removed from

the analysis. Sequences were unweighted using unweight.py, and grouped into OTUs using uclust with 97% sequence identity. OTU representatives were identified using the pic_rep_set.py tool in Qiime selecting the most abundant sequences and metrics of alpha diversity (chao1; shannon; equitability) calculated using alpha_diversity.py. OTUs of planktonic origin were identified by blastn searches as above.

References

1. Bonfield, J.K., Smith, K.F. and Staden, R. A new DNA sequence assembly program. *Nucleic Acids Res* **23**, 4992-4999 (1995).
2. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194-2200 (2011).
3. Huber, T., Faulkner, G. & Hugenholtz, P. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* **20**, 2317-2319 (2004).
4. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590–D596.
5. Quince, C., Lanzen, A., Davenport, R. J. & Turnbaugh, P. J. Removing noise from pyrosequenced amplicons. *BMC bioinformatics* **12**, 38 (2011).
6. Caporaso, J. G. *et al*. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**, 335-336 (2010).
7. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-2461 (2010).