

Identifying a novel and evolutionarily unique bile salt hydrolase from the keystone gut bacterium

Christensenella minuta

Guillaume Déjean¹, Héloïse Tudela^{1,2}, Lisa Bruno¹, Déborah Kissi¹, Georges Rawadi¹ and Sandrine P. Claus^{1,*}

¹ Ysopia Bioscience, 17 Place de la Bourse, 33076 Bordeaux, France; guillaume.dejean@ysopia.bio (G.D.); heloise.tudela@ysopia.bio (H.T.); lisa.bruno@ysopia.bio (L.B.); deborah.kissi@ysopia.bio (D.K.); georges.rawadi@ysopia.bio (G.R.)

² ImmunoConcEpT, 146 rue Léo Saignat, 33076 Bordeaux, France

* Correspondence: Sandrine.claus@ysopia.bio;

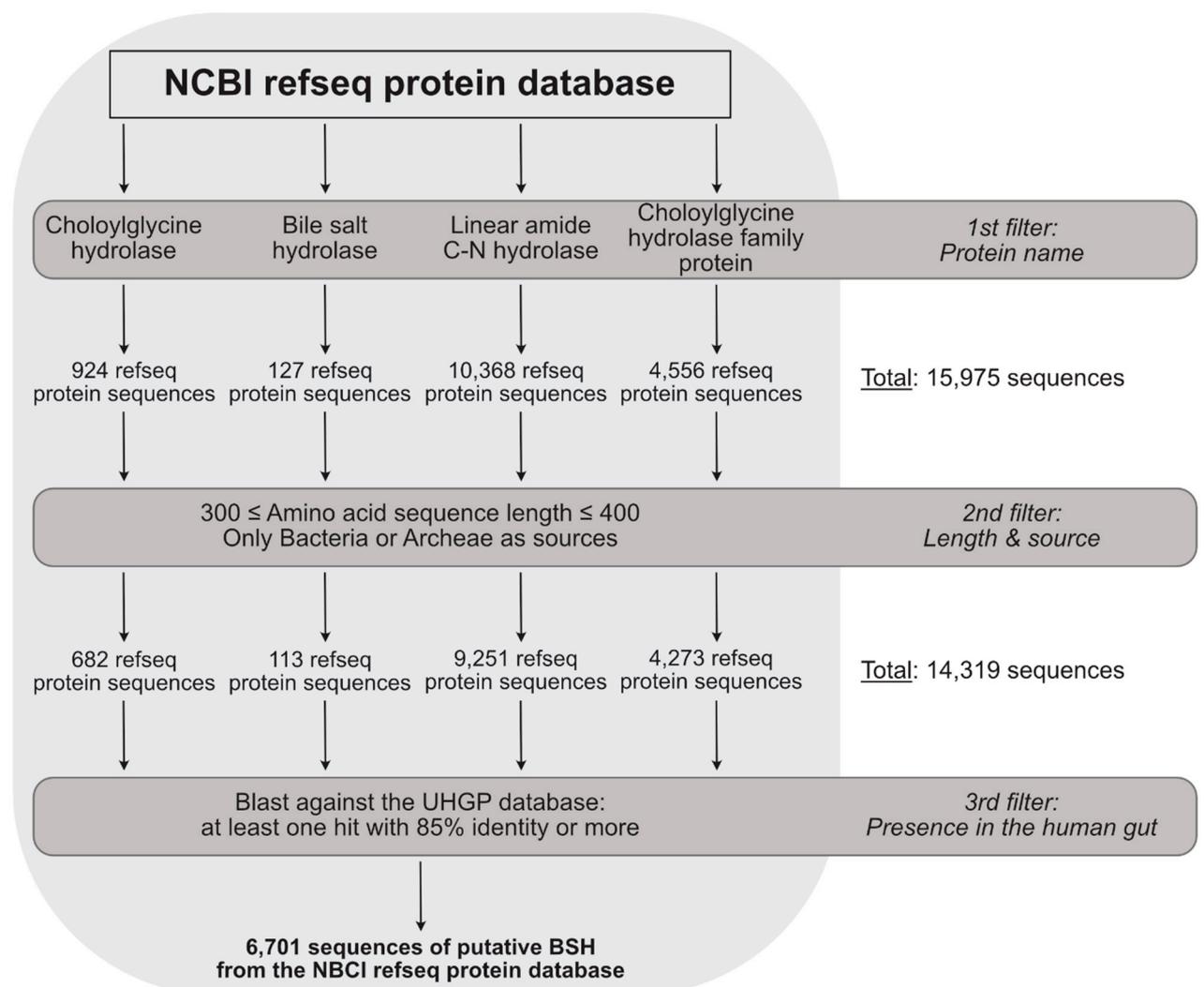
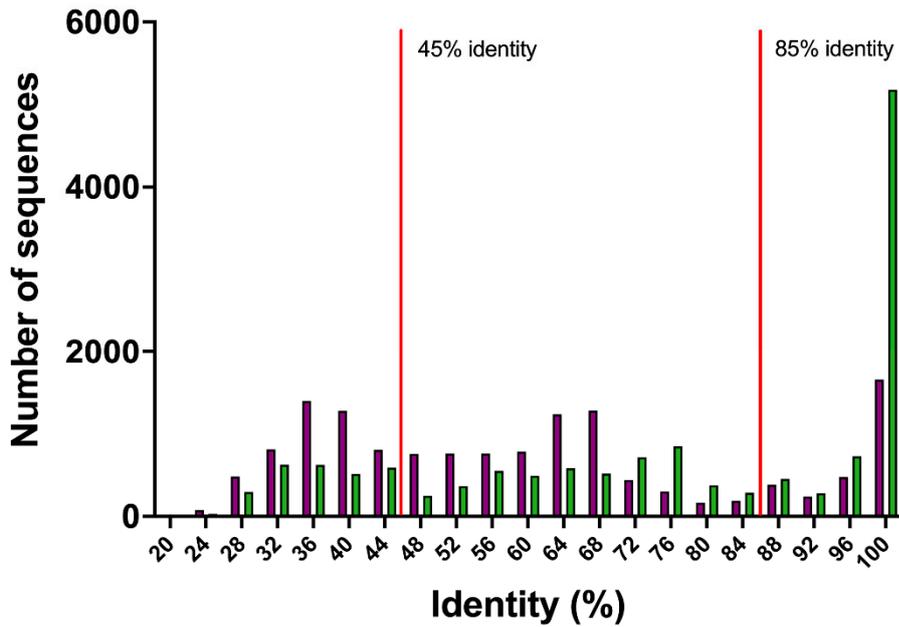


Figure S1. Flowchart applied to select BSH sequences. This flowchart was used to select the BSH sequences included in the phylogenetic and clustering analyses.



- Best hit of the putative BSH sequences from the NCBI refseq protein database against the HMP database
- Best hit of the putative BSH sequences from the NCBI refseq protein database against the UHGP database

Figure S4. Comparison of the HMP and UHGP databases. The 14,320 sequences extracted from the NCBI refseq database using the flowchart shown in Figure S1 were blasted against the gastrointestinal tract Human Microbiome Project (HMP) protein and the Unified Human Gastrointestinal Protein (UHGP) databases.

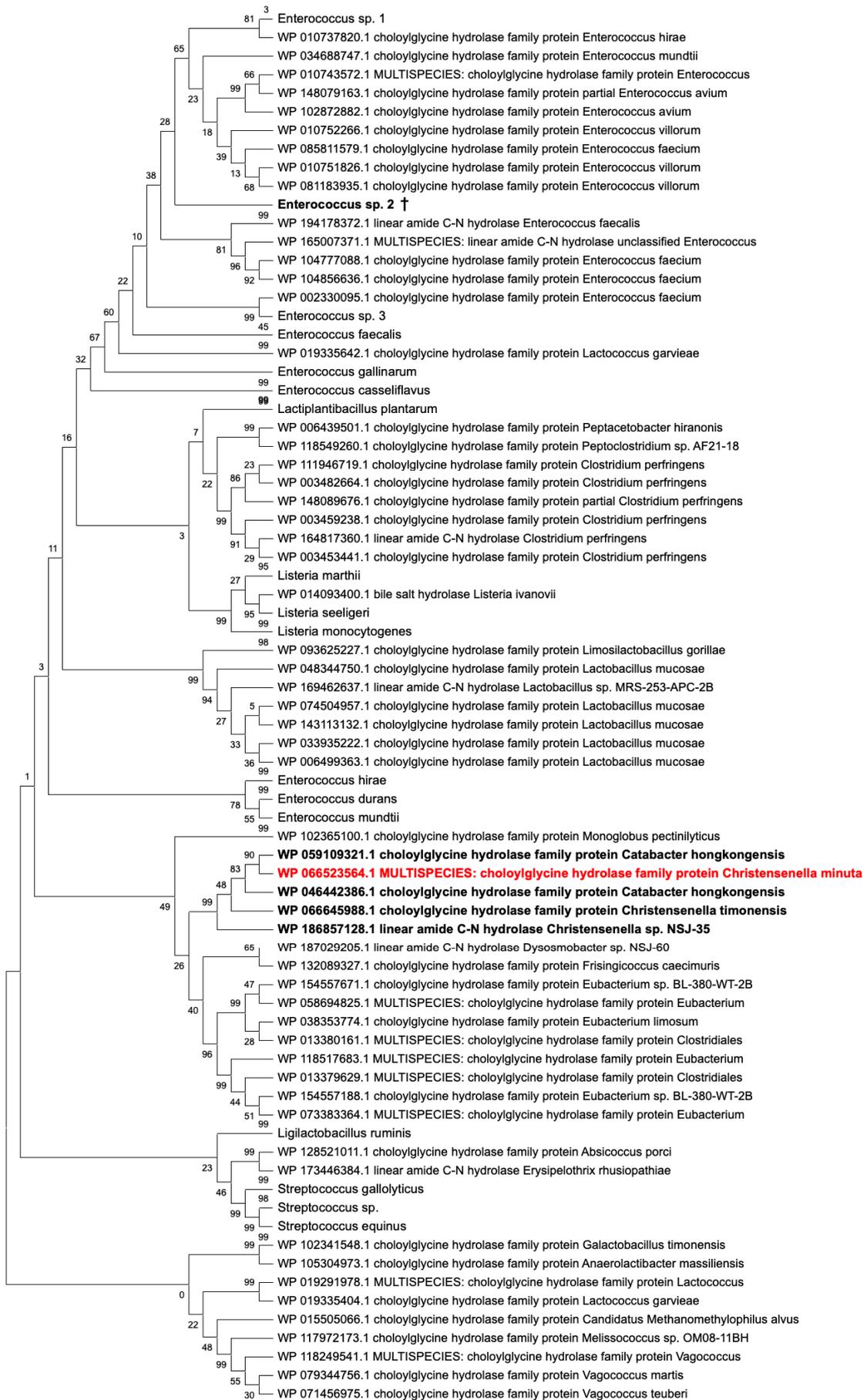


Figure S5. Phylogenetic subtree of the BSH-C1 cluster shown in Figure 6. Bootstrap values are shown on the branches. The composition of collapsed branches is described in the Supplementary Materials. The branch of the cluster containing the characterized BSH from *Enterococcus faecalis* is displayed in bold († : [WP_002355428.1](https://pubmed.ncbi.nlm.nih.gov/31111111/) with PDB structure [4WL3](https://pubmed.ncbi.nlm.nih.gov/31111111/)). The BSH from *C. minuta* is shown in red; the BSHs from other *Christensenellaceae* and *Catabacteriaceae* are shown in bold.

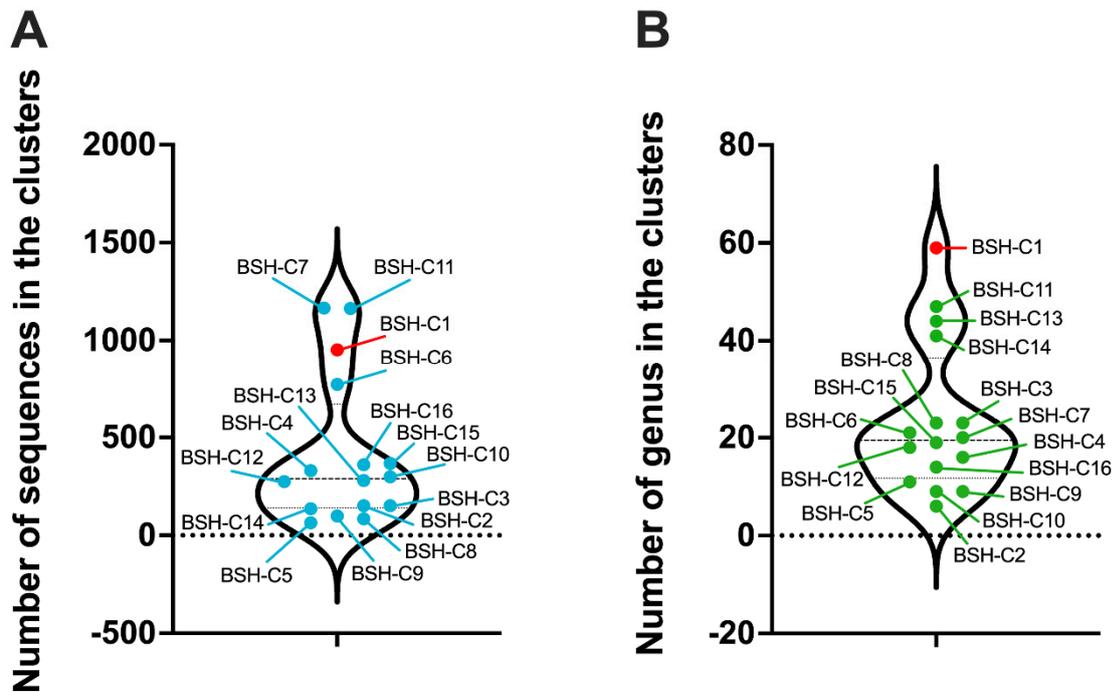


Figure S6. Diversity of genus and number of sequences in the 16 major BSH clusters identified in Figure 4. Each dot corresponds to the number of BSH sequences (panel **A**) or the number of genera (panel **B**) contained in each major cluster from Figure 4. The BSH-C1 cluster is shown in red in both panels. The violin plots were generated using GraphPad Prism v9.1.0.