



**Additional file 1:** Representation of the location of the genes of Mic-PCC7806 contig328 on the genome of Mic-NIES843. Two proteins are considered to be identical if they share at least 40% similarity on their total length. Large repeats are defined as DNA regions containing more than 1000 bases and a similarity threshold >90%. See the Methods section for the strain identifiers.