

Meta-analysis of cable bacteria and their associates in existing next-generation sequencing datasets: cables are everywhere

Ian P.G. Marshall¹, Laurine Burdorf^{1,2}, Andreas Schramm¹

1. Center for Electromicrobiology, Department of Bioscience, Aarhus University, Denmark

2. Department of Biology, Antwerp University, Belgium

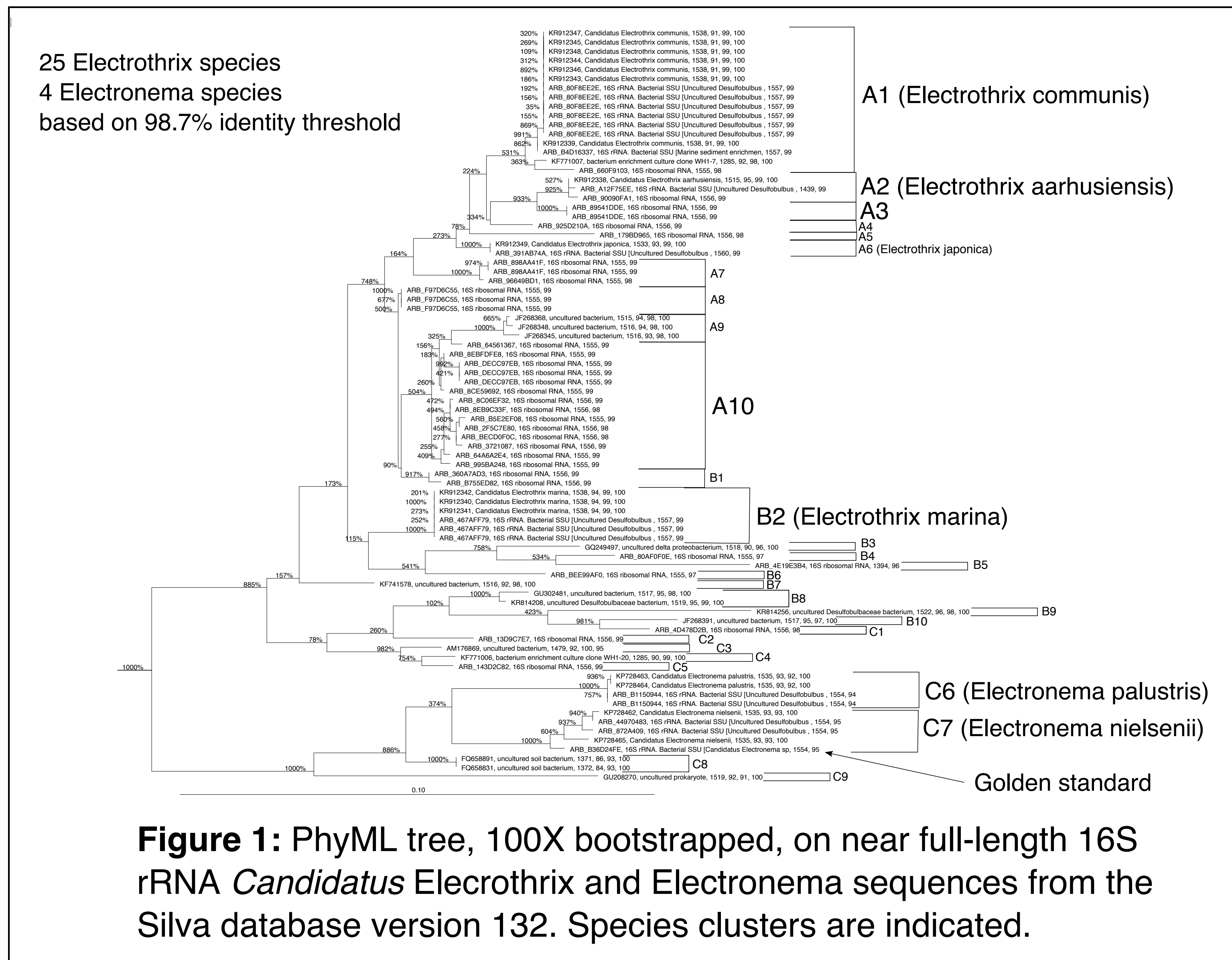


Figure 1: PhyML tree, 100X bootstrapped, on near full-length 16S rRNA *Candidatus* *Electrothrix* and *Electronema* sequences from the Silva database version 132. Species clusters are indicated.

Introduction

16S rRNA gene amplicon sequencing datasets have been produced for years, including from cable bacteria habitats, without being analysed for the presence of the cable bacteria candidate genera *Electrothrix* and *Electronema*. This project is a meta-analysis of existing datasets to identify sites for future cable bacteria fieldwork, assess the extent of cable bacteria diversity and abundance in different types of habitats, and to identify microorganisms living in association with cable bacteria.

Methods

Full-length and near-full-length 16S rRNA gene sequences from *Candidatus* *Electrothrix* and *Electronema* were obtained from the Silva databases version 132, and a phylogenetic reconstruction identified 29 species. One representative of each species was used as query sequence for the integrated microbial next generation sequencing (<http://imngs.org> ~ 300,000 database samples) and MicrobeAtlas (<http://beta.microbeatlas.org/> ~1 million database samples) platforms. Both tools are based on the NCBI Sequence Read Archive.

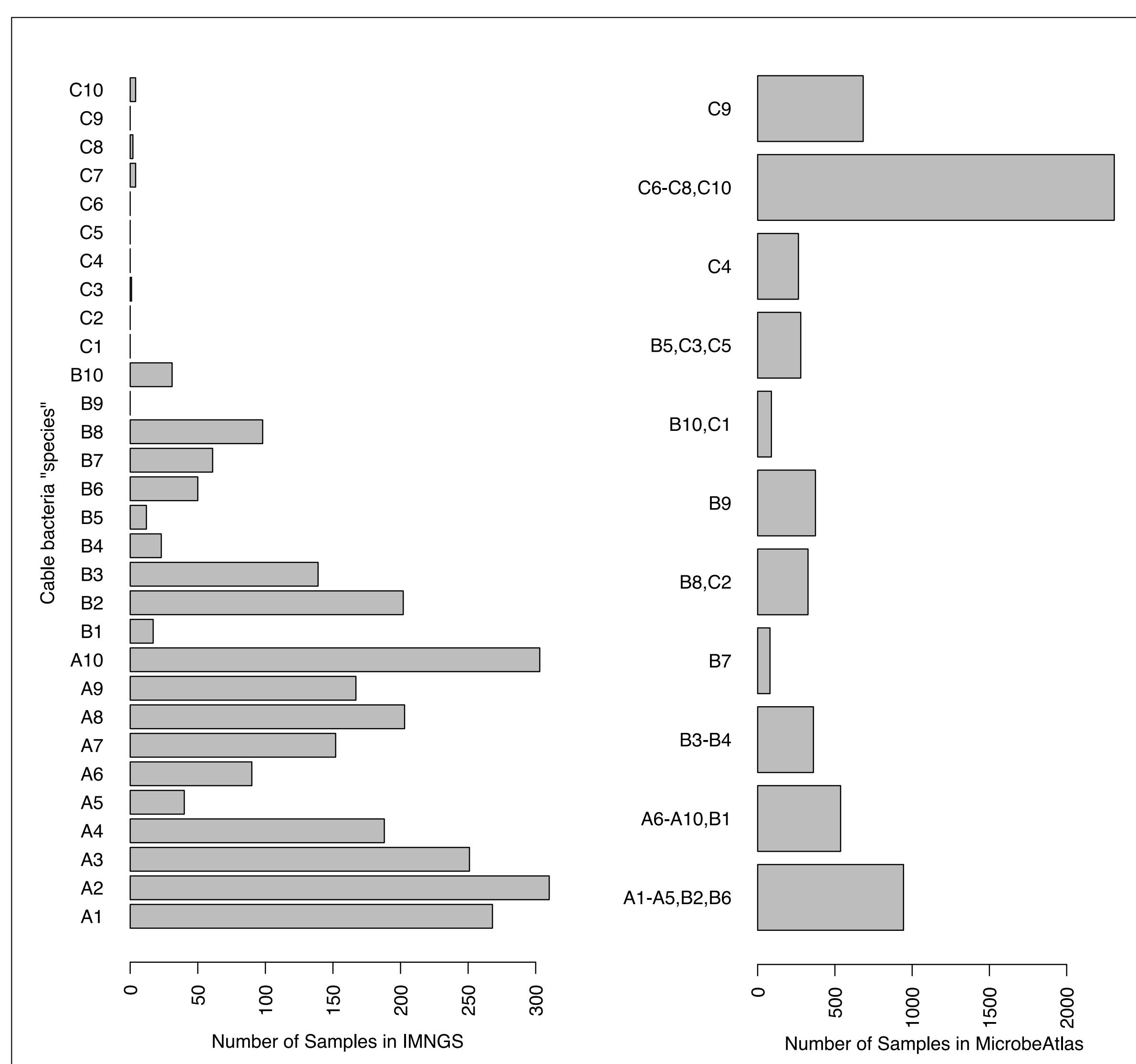


Figure 2: Frequency of identification for each cable bacteria species using imngs (left) and MicrobeAtlas (right).

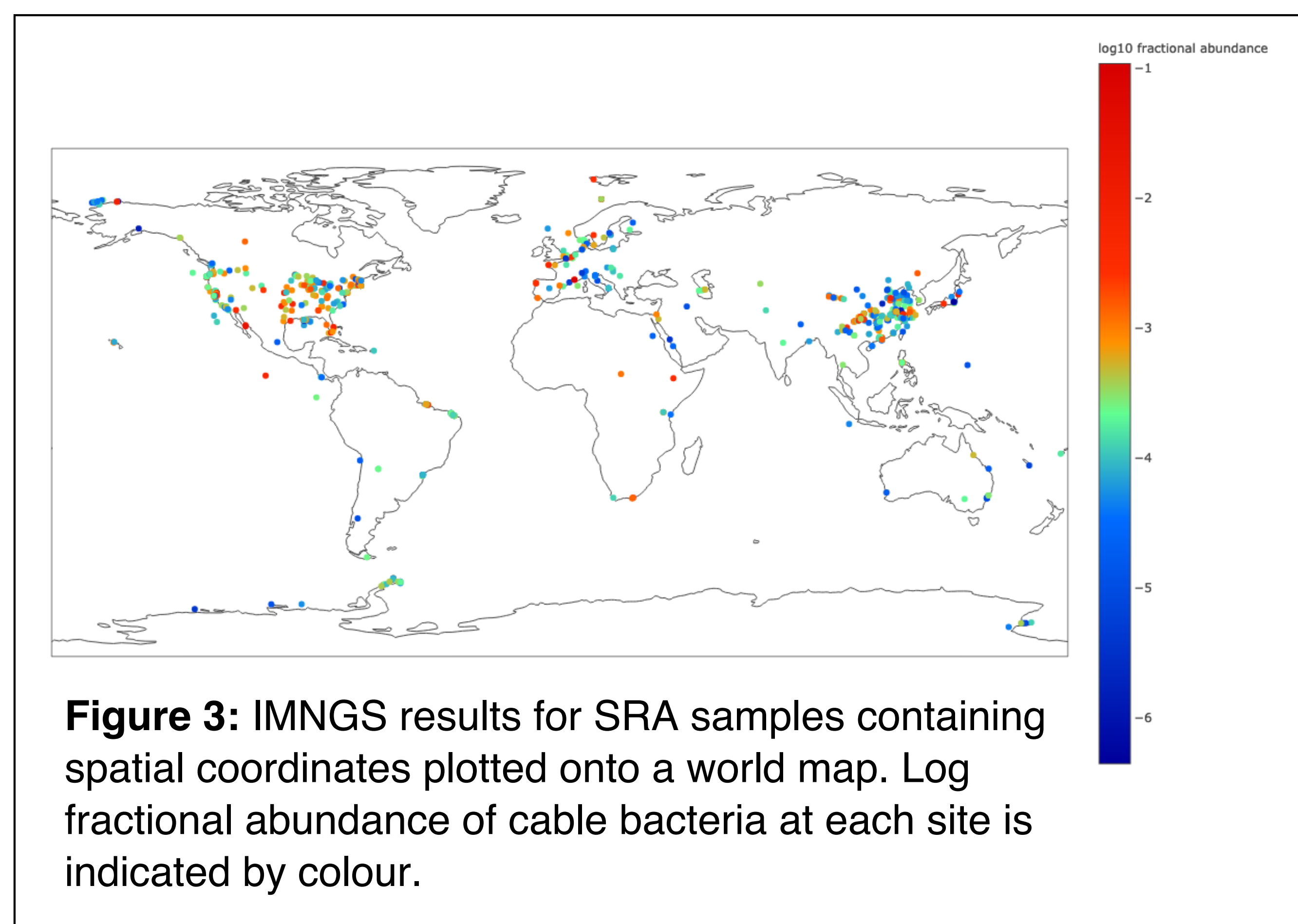
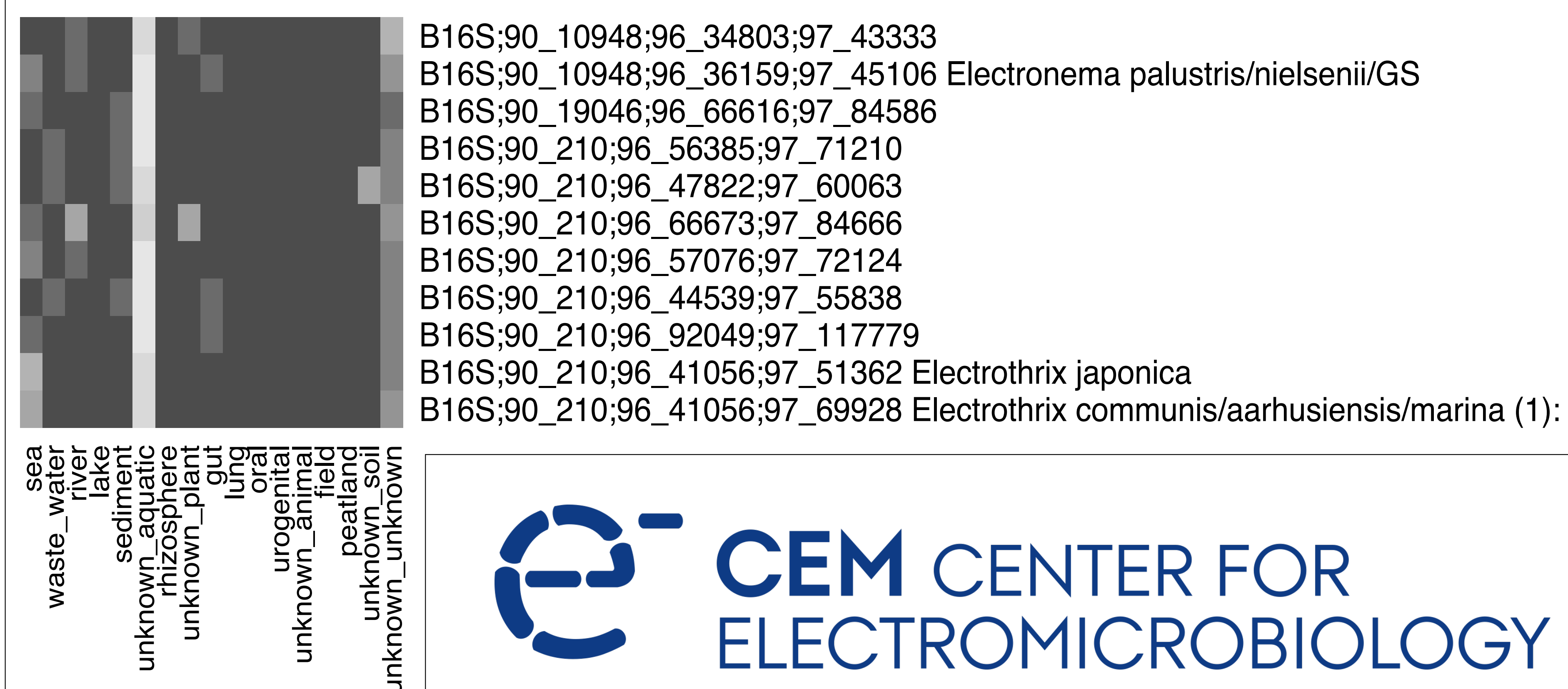


Figure 4: Heatmap of cable bacteria MicrobeAtlas OTUs with environmental type as encoded by MicrobeAtlas. Colouring indicates fractional abundance, with white indicating the highest fraction and black the lowest fraction. Most samples are unhelpfully listed as "unknown aquatic".



Conclusion

Many more samples containing cable bacteria were found using the MicrobeAtlas tool than the imngs tool, most likely due to its larger and more up-to-date database. Especially the *Electronema* species were more readily found using MicrobeAtlas than imngs. However, MicrobeAtlas is still an immature tool without the flexibility of imngs, so for now we will continue to use both tools side-by-side. Metadata in imngs is insufficient to show any niche-specific differences in species abundance based on sample type. The next step will be closer analysis of datasets where cable bacteria were detected using dada2, incorporations of in-house datasets from local sites in Eastern Jutland, and genus-level network analysis to identify likely cable bacteria associates.