NOVEL PHYLOGENETIC MARKERS FOR COST-EFFICIENT AND ACCURATE TYPING OF *PROTOTHECA* SPP.

Marcin Świstak[1], Zofia Bakuła[1], Paweł Siedlecki[2], Jan Gawor[3], Robert Gromadka[3], Tomasz Jagielski[1]

[1]Department of Applied Microbiology, Institute of Microbiology, Faculty of Biology, University of Warsaw, Poland

[2]Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland; Department of Systems Biology, Institute of Experimental Plant Biology and Biotechnology, University of Warsaw, Poland

[3]DNA Sequencing and Oligonucleotides Synthesis Unit at the Institute of Biochemistry and Biophysics, Polish Academy of Sciences, Warsaw, Poland

BACKGROUND

The genus *Prototheca* comprises unicellular, achlorophylous, yeast-like algae widely distributed in the environment. Four out of seven currently postulated species are pathogenic to humans and animals, being the causative agents of protothecosis.

State-of-the art methods of *Prototheca* identification and discrimination include microscopic assesment of cell morphology, evaluation of carbohydrate and alcohol assimilation patterns, and three PCR-based methods, which focus on analysis of rDNA operon. These methods have been designed for identification of *P. blaschkeae*, *P. zopfii* gen 1 and *P. zopfii* gen 2 and are not suited for differentiation of other postulated species.

Moreover, our studies have shown (not published) that each *Prototheca* species has multiple divergent copies of rDNA operon in its genome. This poses technical problems for analysis of PCR products and sequencing.

OBJECTIVES

The study aims at proposing an accurate, easy to use, quick and cheap method for differentiation of all *Prototheca* species as an alternative approach to current typing methods.

METHODS AND RESULTS

Based on Whole Genome Sequencing of 8 reference strains (not published) eight draft genomes have been assembled for which genes have been predicted *in silico*.

METHODS AND RESULTS

Cross-examination of the drafts identified 187 genes shared among all eight strains. Out of the set of 187 common genes, two single-copy genes have been selected based on their discriminative power in phylogenetic analyses.

For these genes primer pairs have been designed, PCR products amplified and sequenced using Sanger technology. Additionally, in order to compare new method with previously available 18S rDNA and internal transcribed spacers (ITS) of 8 reference strains have also been sequenced.

Sequences have been assembled using SeqMan Pro, aligned using MAFFT software with default parameters. Phylogenetic trees have been infared using MrBayes with GTR substitution model with gamma-distributed rate variation across sites and a proportion of invariable sites. For each set of sequences 2,000,000 generations of MCMC algorithm have been run in 3 runs with 4 chains each. Sample frequency have been set to 500.

Two representative phylogenetic trees have been presented on Figure 1. It can be seen that cytochrome B tree captures differences between different *Prototheca* species as well as reference tree constructed upon alignment of 18S rDNA. Moreover, sequencing cytB requires only 2 Sanger reads in contrast with up to 20 in case of ITS or 18S rDNA sequences.

This makes cytB the phylogenetic marker which is easier and faster to use than conventional markers and has the same discriminative power.

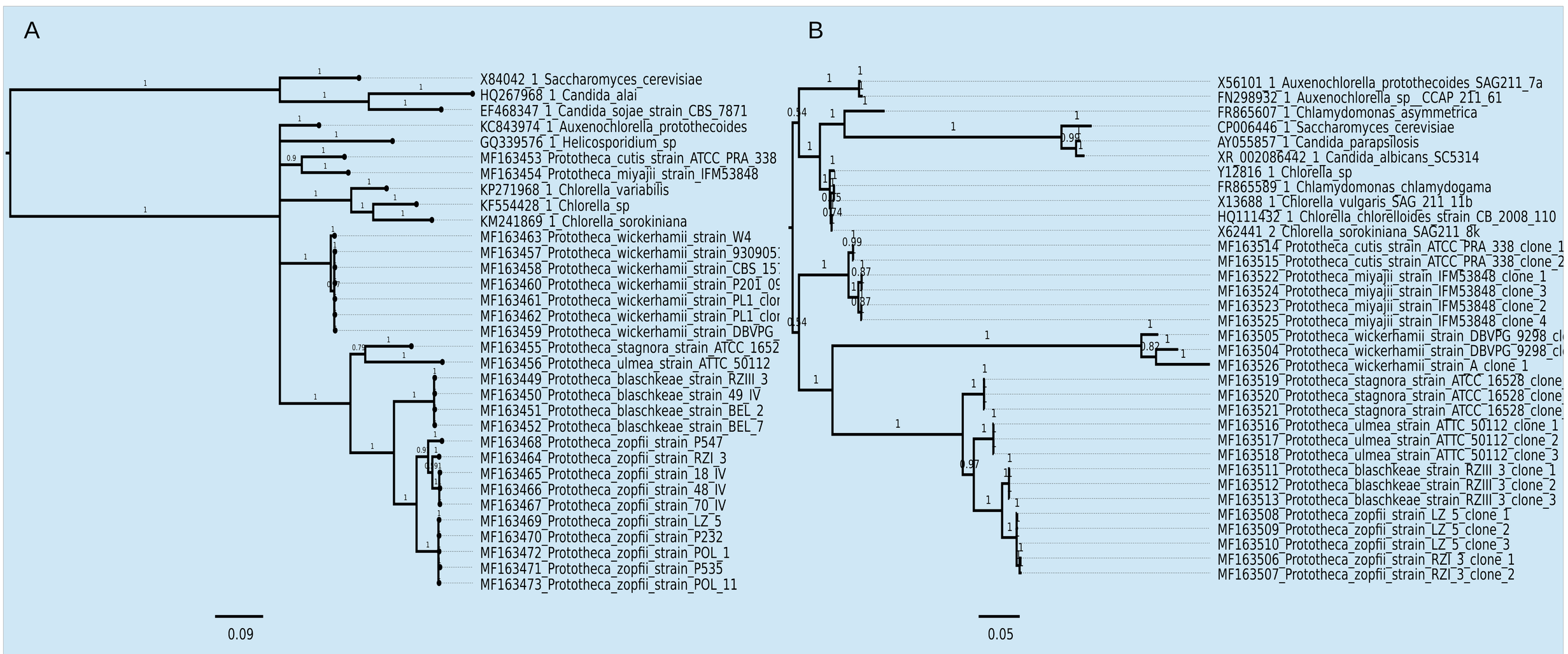

Figure 1. Comparison of phylogenetic trees obtained for cytochrome B sequence alignment (A) and 18S rDNA sequence alignment (B). Above branches are shown posterior probabilities of subtrees. For each tree a scale bar is given below.

CONCLUSIONS

A new PCR-based method based on single-copy gene: cytB is proposed. This phylogenetic marker enables distinction of all seven known *Prototheca* species and also genotypes of *Prototheca zopfii*. It's usage is much faster and cheaper than sate of the art typing methods.