

# The apparent antiquity and broad environmental diversity of triterpenoid cyclase genes

A. PEARSON<sup>1</sup>, S.R. FLOOD PAGE<sup>1,2</sup> AND S.J. CARTER<sup>1</sup>

<sup>1</sup>Department of Earth, Atmospheric, and Planetary Sciences  
Harvard University, Cambridge MA 02140 USA;  
(pearson@eps.harvard.edu)

<sup>2</sup>Now at: INSTAAR, University of Colorado, Boulder CO  
80309 USA; (sarah.floodpage@colorado.edu)

Hopanoids are among the most abundant biomolecules on Earth and can be found in rocks as old as 2.7Ga. Until now, evidence for the biological sources of hopanoids primarily has come from screening the lipid extracts of pure cultures of prokaryotes, and more rarely, of fungi and higher plants. Other information is available from genomic sequencing projects, in which the gene for squalene-hopene cyclase (SHC) can be identified readily. These approaches require pure or enrichment cultures of the suspected hopanoid producers, and as such they do not capture the fraction of the uncultured microbial community that also may contribute significantly to environmental hopanoids. Here we report sequences for SHC genes obtained from fresh water and marine planktonic communities. Diversity estimators suggest that  $\leq 70$  total SHC phylotypes exist in each of these communities, of which we sequenced successfully only half of the predicted number, despite extensive efforts to obtain greater coverage. Of the clones found, none was more than 86% similar in amino acid sequence to any of the currently available genomic data in public databases. As such, none of them can be classified as belonging to any known species. The data represent the application of six different methods, three of which were designed to favor the amplification of known cyanobacterial sequences and three of which targeted proteobacteria. However, none of the sequences could be assigned definitively to any genus within either of these groups and none at all were affiliated with the cyanobacteria even at the phylum level. The results suggest that the dominant microbial sources of these important lipid biomarkers remain undiscovered. In addition, all hopanoid cyclase genes are related more closely to the steroid cyclases found in bacteria than to those found in eukaryotes. This is consistent with the divergence of the steroid and hopanoid pathways in the last common ancestor (LCA) of eukaryotes. The biosynthetic pathways for polycyclic triterpenoids thus can be inferred to be evolutionarily ancient and inherent to the LCA.