Another project brought to you by GenomeCanada

Environmental Barcoding through Massively Parallelized Sequencing

Status	Current
Competition	Development of New Technologies Competition
Sector	Development of New Technologies
Genome Centre	Ontario Genomics Institute
Project Leader	Paul Hebert

Project Description

DNA barcoding represents a novel genomics exercise – the collection of sequence information from a standard gene region across eukaryotic life. A 648 base pair segment of the mitochondrial gene cytochrome *c* oxidase 1 (CO1) has now been selected as the core barcode region for eukaryotes. The horizontal survey of sequence diversity in this gene region is valuable in many contexts; it enables species identification and discovery; it reveals factors influencing rates of molecular evolution and species age; it allows detailed study of evolutionary pathways in the CO1 protein. Motivated by these factors, planning is now underway for a massive international research program to rapidly expand the reference library of CO1 sequences. The International Barcode of Life Project (IBOL), a $150M, 5-year program driven by Canada, will involve researchers from 25 nations. Over this interval, IBOL will deliver barcode records for 500K species and subsequent efforts will produce a barcode reference library for all eukaryotes.

Although completion of a library for all eukaryotes may require 20 years, DNA barcoding is already gaining application as barcode libraries reach closure in varied groups. It is now clear that one particularly important area of application for DNA barcoding will lie in species identifications. It is also apparent that such applications will involve two different technology streams: point-of-contact analysis of single specimens and massive barcode screens. Point-of-contact devices allowing immediate analysis will be critical for species identifications in contexts such as port inspections and pest control. By contrast, the second technology stream, the focus of this application, will enable the analysis of mixed biotic samples, albeit less rapidly. We emphasize that many eukaryotes are too small, too numerous or too conjoined to be analyzed using conventional barcode protocols. Our current application will break this barrier by developing the protocols required for the analysis of any collection of eukaryotes. We term this approach ‘environmental barcoding’ and are sure that its implementation lies in coupling massively parallelized sequencing technologies with new informatics tools. Such analysis will certainly advance biodiversity monitoring. Imagine the newly sophisticated capacity to monitor environmental quality that would result if we could rapidly gain information on the species composition of any environmental sample. However, the implications of our work are broader - environmental barcoding represents the metagenomics tool for eukaryotic life.

This application seeks funding to develop environmental barcoding as a technology stream that can be adopted by sequencing platforms. Our work will make use of new parallelized sequencers, but considerable technological innovation will be required to enable them to support environmental barcoding. We point particularly to the need for new informatics tools to analyze sequence data and to the need for new protocols to enable barcode recovery from large, admixed samples of life. We not only expect to conquer these challenges, but we also plan work to show how environmental barcoding can support a real-world need, the bio-monitoring of Canada's inland waters. We will carry out the latter work in close collaboration with researchers at Environment Canada who not only bring deep expertise in environmental sampling, but will be important end users of this technology. We will also collaborate with colleagues at the Stanford Genome Technology Centre who are world leaders in massively parallelized sequencing technologies. The environmental barcoding technology developed in this project will reinforce Canada's leadership position in DNA barcoding by introducing the first application of this approach to biodiversity monitoring of eukaryote communities.