COLLECTIONS AS A VALUABLE RESOURCE FOR DATA MINING IN THE DEVELOPMENT OF FUNGAL DIAGNOSTIC TOOLS

By Dr. Sarah Hambleton

TETRAPARTITE 2007

The Biological Collections housed at Agriculture and Agri-Food Canada (AAFC) are of two main types: those dedicated to preserving living organisms or germplasm, Plant Gene Resources of Canada (PGRC), the Glomales *In Vitro* Collection (GINCOcan) and the Canadian Collection of Fungus Cultures (CCFC); and those comprising dried specimens, the Canadian National Collection of Insects, Arachnids, and Nematodes (CNC), the Vascular Plant Herbarium (DAO), and the National Mycological Herbarium (DAOM). Working collections are maintained by individual scientists, and in the broader scientific community collections of all sizes and foci are supported at Canadian universities and industrial research centres. Research scientists also rely on the collections to develop and document expert advice provided to clients through the AAFC Identification Services.

At a minimum, collections provide a repository for voucher material to ensure scientific reproducibility and to document species for biodiversity, public health and quarantine purposes. But they also provide the fundamental raw materials for a broad range of primary research initiatives, including crop improvement for enhanced disease resistance or marketable traits, taxonomy, biodiversity and biogeography studies, screening for secondary metabolites, and the development of improved identification systems and diagnostic tools. An analogy can be made to the mining of mineral resources and we are actually proposing to search for novel enzymes in the main live fungal collection as part of the new AAFC Agricultural Bioproducts Innovation Program. Each accession in a collection embodies a complex array of identification data (name, age, provenance, habitat) and character data (morphological, genetic). The information can be accessed and utilized according to need and technological capability: from host-pest distribution analyses to the development of identification keys to multi-species diagnostic tools based on DNA sequences.

AAFC is at the forefront of diagnostic research with the development of PCR-based DNA hybridization techniques for the detection of plant pathogens, especially in environmental samples (soil, air, plant parts). Macroarray detection systems have already been developed for specific clusters of species such as the bacterial pathogens of potato or the economically important fungal diseases affecting apple production, and for taxonomic groups such as the genus *Pythium*. The latter achievement marked a milestone in plant diagnostics because the single platform assay included diagnostic markers for virtually all the species known in the genus at the time. This would not have been possible without the availability of reference cultures carefully preserved in collections. A sensitive PCR/hybridization assay that can be used with field samples was developed for the potato wart fungus, *Synchytrium endobioticum*, using sequence data generated from samples that included old herbarium specimens, one being a 1903 diseased potato. Under development currently are assays directed at the fungal species infecting cranberries, the soybean root pathogens in *Fusarium*, the *Penicillium* spp. responsible for food and feed spoilage and the obligate fungal pathogens known as rusts. All based on sequences generated from DNA extracted from fungal cultures, or dried plant material infected with a fungal pathogen, preserved in a collection.

These diagnostic tools are based on a hierarchical and internally redundant system of group-specific molecular markers designed to be specific for the target organisms at the required level of taxonomic resolution. To avoid false negative or positive reactions, the marker development must take into account the genetic variability of the target organism(s), by sampling from the known range of morphological and

geographical variants. Therefore the accuracy of this detection method depends on the availability of a large number of reference specimens for study in the development stages. For the *Pythium* array, data for 289 strains in 100 species were analyzed. To set out to assemble such a large and diverse set of samples *de novo* would be a daunting task for any set of organisms. For fungi and other microorganisms, the task is further complicated by the microscopic and in many cases ephemeral stages in their lifecycles.

Aside from the economically important species, collections also preserve rare species and provisionally identified isolates or specimens. Fungi are difficult to identify using traditional methods due to difficulties in interpreting cryptic morphological characters or an absence of taxonomically informative features expressed in cultured isolates or lesions on infected plant parts. As DNA sequencing, in particular of ribosomal genes, becomes routine to aid identifications and connect lifecycle stages, culture collections are moving towards developing their own in-house sequence databases and DNA banks. Studies of soil and root-associated fungi at AAFC discovered new genera of mycorrhizal fungi after molecular data grouped unidentifiable isolates together and connected them to additional isolates recovered in multiple independent studies. In our development of rapid molecular diagnostics for rust fungi causing disease in agricultural crops, the primary source of DNA is dried plant material exhibiting rust infections. Techniques for extracting PCR-amplifiable DNA from older herbarium specimens still need to be improved but our success rate for samples collected during the last century is about 50% overall. To assemble a diverse and representative set of specimens for array development, 7 additional international herbaria have approved specimen loans for DNA sampling.

Collections are the repository of our biological history and as such represent an irreplaceable resource. This is a time when techniques for in-depth data generation are developing at a rapid rate with improved cost-effective molecular techniques and database management. Yet long term logistical and financial support is the current challenge as collections become orphaned due to curator retirements and shifts in priorities at sponsoring institutions. Cooperation already exists among researchers internationally through loans and gifts or exchanges of samples and data. Examples of international collaborative ventures that AAFC mycologists contribute to using collections-based data are GenBank, the publicly available on-line sequence database at the National Centre for Biotechnology Information, the Tree of Life Web Project, providing information about the diversity of organisms on Earth, their evolutionary history (phylogeny) and their characteristics, and the Consortium for the Barcode of Life, devoted to developing DNA barcoding as a global standard in taxonomy. The need for international coordination on standards and management of core collections is urgent, to be able to respond to medical emergencies, quarantine issues and bioterrorism threats that transcend national boundaries.