

---

## A botanical renaissance: state-of-the-art DNA bar coding facilitates an Automated Identification Technology system for plants

---

Steven G. Newmaster\*

Department of Integrative Biology,  
University of Guelph, Ontario, Canada  
Fax: (519) 767-1656  
E-mail: [snewmast@uoguelph.ca](mailto:snewmast@uoguelph.ca)  
\*Corresponding author

Subramanyam Ragupathy

OAC Herbarium, Biodiversity Institute of Ontario,  
University of Guelph, Ontario, Canada  
E-mail: [ragu@uoguelph.ca](mailto:ragu@uoguelph.ca)

John Janovec

Botanical Research Institute of Texas (BRIT),  
Fort Worth, TX 77204-3010, USA  
E-mail: [jjanovec@brit.org](mailto:jjanovec@brit.org)

**Abstract:** Traditional taxonomic practices are insufficient on their own to cope with the growing need for accurate identifications. The recent development of DNA barcoding has been applied to plants. The next step is the development of a high-throughput Automated Identification Technology (AIT) system. Our research indicates that the efficacy of an AIT system equates with savings in time and funding. Given the potential interconnectivity of web-based applications, we suggest an AIT system for plants that uses several existing systems and suggest several applications where AIT could serve as a tool for biologists and for society at large.

**Keywords:** bar coding; bioinformatics; herbarium; plant identification; taxonomy; flora; DNA technology.

**Reference** to this paper should be made as follows: Newmaster, S.G., Ragupathy, S. and Janovec, J. (2009) 'A botanical renaissance: state-of-the-art DNA bar coding facilitates an Automated Identification Technology system for plants', *Int. J. Computer Applications in Technology*, Vol. 35, No. 1, pp.50–60.

**Biographical notes:** S.G. Newmaster received his PhD in Floristic Diversity from University of Alberta in 2000. He is currently an Associate Professor at the Department of Integrative Biology and Director of the OAC Herbarium, Biodiversity Institute of Ontario (BIO), University of Guelph, Ontario, Canada. His current research interests include speciation, plant diversity, DNA bar coding, plant classification and ethnobotany.

S. Ragupathy is a research associate and Curator of floristic databases at the OAC Herbarium, BIO, University of Guelph, Ontario, Canada. He received his PhD in Botany from the University of Madras, India, in 1992. His current research focuses on developing the Flora Ontario Automated Information System (FOIBIS; see <http://www.uoguelph.ca/foibis>), biodiversity database management, molecular taxonomy, Aboriginal classification systems and ethnobotany.

J. Janovec is a senior research scientist at the Botanical Research Institute of Texas (BRIT), Texas, USA. He received his PhD in Botany from Texas A&M University, USA, in 2000. He is actively involved in the development of ATRIUM and conducts research at two scales of tropical plant diversity involving the Myristicaceae (nutmeg) family of trees and the flora of the Andes-Amazon region of southeastern Peru.

---

## 1 Traditional plant identification

Traditional botanists typically devote a portion of their time to identification of plant samples for researchers working in other disciplines in science and to society at large. Usually, this involves a professionally trained taxonomist and the use of a botanical herbarium, which is a museum of preserved reference specimens. The taxonomist uses books from the botanical library, often connected with the herbarium, and reference specimens from the herbarium to identify plants and produce new, updated information products, such as floristic treatments, monographs and revisions. Floristic treatments provide plant nomenclature, descriptions, and keys to document plants of a given region (e.g., local, regional, national and international). Monographs and revisions treat natural groups of plants – usually all the species in a genus or family – and are often the source of documentation of species that are new to science.

The need to know the identity of a particular plant is great and spans research disciplines, government agencies and business sectors. For example, a researcher studying the loss of habitat of a rare animal such as the Woodland Caribou needs to know the identity of all the plants it consumes. Conservation biologists need to know the identity of all the plants in an ecosystem to assess patterns in diversity and to plan and manage conservation areas. Border control officers need to know the identity of plant materials (e.g., seeds) entering our countries. Forest ecosystem managers need to be able to identify the loss of particular species following silvicultural protocols. Farmers need to know the identity of weeds invading their crops. Police require expert plant identifications as forensic evidence in court cases. Hospitals often need the identification of poisonous plants that are consumed by people who harvest wild edibles or children who have consumed an unknown berry or leaf. The food and herbal industry need to provide quality assurance by identifying plants that contaminate their products. These are only a few examples of the need for accurate plant identification, a need which intensifies with the increase of the human population. Unfortunately, existing numbers of taxonomists and herbaria are not sufficient to deal with this growing demand for plant identification.

Traditional taxonomic practices are inadequate on their own to deal with the need for accurate and accessible taxonomic information. The perception that taxonomy is not relevant has led to a significant drop in the funding and training of graduates as professional taxonomists. Many herbaria have closed because of substantial funding cuts for natural science programs in general. This has led to recurring cries of crises in taxonomy and of the need for a rejuvenation of the field (Godfray, 2002a, 2002b; Tautz et al., 2002, 2003). Some researchers (Lipscomb et al., 2003; Wheeler et al., 2004; Will and Rubinoff, 2004) have argued that routine identification should represent a minor part of a taxonomist's work allowing them to focus on alpha taxonomy (the discovery and description of new species). However, it is a fact that taxonomists

complete thousands of routine identifications per year (Newmaster et al., 2006) perpetuating the state of crisis in alpha-taxonomic research (Hebert and Gregory, 2005). As a community we need to do a better job of connecting the alpha-taxonomic research with the broader community of ecologists, biogeographers, agriculture specialists, conservationists and the many governmental and non-governmental organisations.

It is surprising and disturbing that after 250 years of modern systematic biology it is estimated that we have named less than 10% of the species on earth (Hawksworth and Kalin-Arroyo, 1995). The total number of species on earth remains unknown with estimates ranging from 10 million to more than 100 million (May, 1988; Hammond, 1992; Hawksworth and Kalin-Arroyo, 1995). The number of species remaining to be discovered is well beyond the current capacity of our descriptive taxonomists and systematists (Godfray, 2002a; Blaxter, 2004). Current estimates of the losses of biodiversity are greater than our ability to recognise new species. There is an urgent need to develop technology that will expedite our ability to catalogue species (Godfray, 2002b; Blaxter and Floyd, 2003; Godfray and Knapp, 2004). In response, many researchers have called for an automated identification system that will aid taxonomists in routine identifications (Gaston and O'Neill, 2004; Wheeler et al., 2004) and alpha-taxonomic research (Hebert and Gregory, 2005; Newmaster et al., 2006). We believe that species identifications are a rate-limiting step for many ecological and biodiversity investigations, as well as for taxonomic research, and that Automated Identification Technology (AIT) could relieve a burden on taxonomists and fill the current need with important benefits to both taxonomy and biodiversity science. In this paper, we explore identification technology including state-of-the-art DNA bar coding in the development of an automated identification system for plants. We discuss the efficacy of an AIT system used at the Biodiversity Institute of Ontario (BIO) Herbarium and propose applications for society at large.

## 2 Identification technology

Plant identification employs characters for discriminating taxa at several ranks (e.g., family, genus, species, etc). Different types of evidence are utilised in plant classification, such as morphological, chemical and molecular DNA. Each provides characters that are often used in combination when describing a new species or analysing the classification of natural groups of plant species (i.e., genera and families). The study of systematics or the relationships among taxa also utilises many types of characters. Routine plant species identification, however, relies mostly on morphological characters such as the type and position of the gynoecium and androecium. Skilled professional botanists are required to interpret the morphological characters and navigate lengthy and complicated keys to discriminate plant species. This process is expensive and time consuming for professional botanists

who could be using their skills for research in alpha taxonomy. However, there is a certain amount of bias in the present system with respect to the identification of cryptic taxa; species without discrete gross morphological characters are often lumped together. Cryptic taxa may be species, ethnotaxa (Newmaster et al., 2007), varieties, subspecies, ecotypes or cultivars. Furthermore our traditional system relies on a linear approach; recognition of characters of an unknown sample is facilitated by the use of a character key, which leads to a species identity. The keys are often dichotomous proceeding (sometimes biasing) through a structured sequence of paired decision-making steps called couplets. While biologists associate morphological features exhibited by an unknown plant with the features described in the couplets of the key, they are seldom encouraged to consider a variety of characters or the range of variation of morphological features associated with a particular species of plant. The total morphological variation inherent to a particular species is typically not reflected in these keys by the very nature of their construction. Other characters, such as cytology (chromosome counts) or chemicals (produced by only some taxa) require specialised equipment, and for many plants only discriminate to the rank of family or genus.

Several molecular tools have been applied to the detection and identification of biological samples. These are designed to function primarily at the species or population levels of the plant taxonomic hierarchy. Some of the more common markers include Restriction Fragment Length Polymorphism (RFLP), Amplified Fragment Length Polymorphism (AFLP), Microsatellite and Random Amplified Polymorphic DNA (RAPD). These techniques can reveal otherwise hidden genetic variation and can provide markers unique to species or populations and have been used to confirm species boundaries. Genomic fingerprinting or Inter Simple Sequence Repeats (ISSR) is a broadly defined analysis used for testing genomic instability (Leroy et al., 2000; Weising et al., 2005), evaluating genetic diversity (Kantety et al., 1995), cultivar identification (Charters et al., 1996), molecular mapping (Ratnaparkhe et al., 1998), array-based platforms for specimen identification in mammalian species (Hajibabaei et al., 2007a, 2007b) and forensic DNA profiling (Kumar et al., 2001) in plants as well as for sexing in birds (Wink et al., 1998) or identifying hybrids in birds and reptiles (Wink et al., 2001). These molecular tools have been applied to cryptic groups (Wilson, 1995), such as yeast (Kurtzman, 1994), viruses (Allander et al., 2001), bacteria (Hamels et al., 2001), protists (Nanney, 1982; Pace, 1997) and other organisms (Brown et al., 1999; Bucklin et al., 1999; Treweek, 2000; Vincent et al., 2000). Although these methods do yield useful character-based data, they are not conducive to routine plant identifications in conjunction with a high-throughput bioinformatics system. Furthermore, these techniques are less than optimal for the taxonomically wide scope of plant identification because of problems with high within-taxon variability and lack of confident assignment of orthology between markers. DNA sequence

data, however, has the potential to overcome these hurdles (Blaxter, 2004). Some DNA regions are sufficiently conserved to permit the use of 'universal' oligonucleotide primer sets for PCR amplification and yet contain sufficient informative sequence variation to discriminate species. Floyd et al. (2002) recognised that DNA sequence could be used for discriminating species and coined the term Molecular Operational Taxonomic Unit (MOTU), which has also been called 'phylotypes' and 'genospecies'. Hebert et al. (2003) developed the 'DNA Bar code' for high-throughput identification of animal species and has initiated a global movement towards a technologically advanced taxonomic system that will expedite our ability to explore and record the earth's biodiversity (Hebert et al., 2003).

### 3 AIT and DNA bar coding systems

DNA bar coding has proven to be a reliable tool for species identification. Hebert et al. (2003) pioneered the idea of high-throughput DNA bar coding in animals, which is a method of species identification and recognition using DNA sequence data from a standard region. The initial step is the construction of a DNA sequence library using vouchers or type specimens from taxonomic experts. On the basis of the mitochondrial gene, cytochrome c oxidase subunit 1 (COI or *cox1*), DNA bar coding is progressing rapidly in several groups of animals, which can be reviewed online via the Canadian Barcode of Life (<http://www.bolnet.ca>) and the Consortium for the Barcode of Life (CBOL <http://www.barcoding.si.edu>). The development of a bar code system for plants started later and has proceeded cautiously for a number of reasons that have been previously discussed in detail (Kress et al., 2005; Chase et al., 2005; Newmaster et al., 2006; Taberlet et al., 2006; Presting, 2006). The current challenge has been to identify suitable DNA regions for plant bar coding (Chase et al., 2007; Kress and Erickson, 2007; Newmaster et al., 2007; Little and Stevenson, 2007; Pennisi, 2007). Recently, the Canadian Plant Barcoding Group (CPBG, Fazekas et al., 2008) assessed the utility of a number of coding and non-coding genomic regions that have been proposed as plant bar codes for a regional flora. Their results indicate that a multi-region bar code is optimal and that several of the regions could potentially be used in various combinations to effectively bar code plants. The immediate challenge is the transition from experimental research to a high-throughput automated DNA bar coding system.

There are a number of impediments to plant bar code research that must be resolved to match the success of the high-throughput system of animal bar coding. To maximise detectable genetic variation, Kress et al. (2005) proposed the use of a non-coding region, the *trnH-psbA* plastid spacer in a multigene approach. Chase et al. (2005) and others criticised this approach owing to issues of alignment with non-coding regions (discussed by Cowan et al., 2006). These are serious issues for current bioinformatic systems

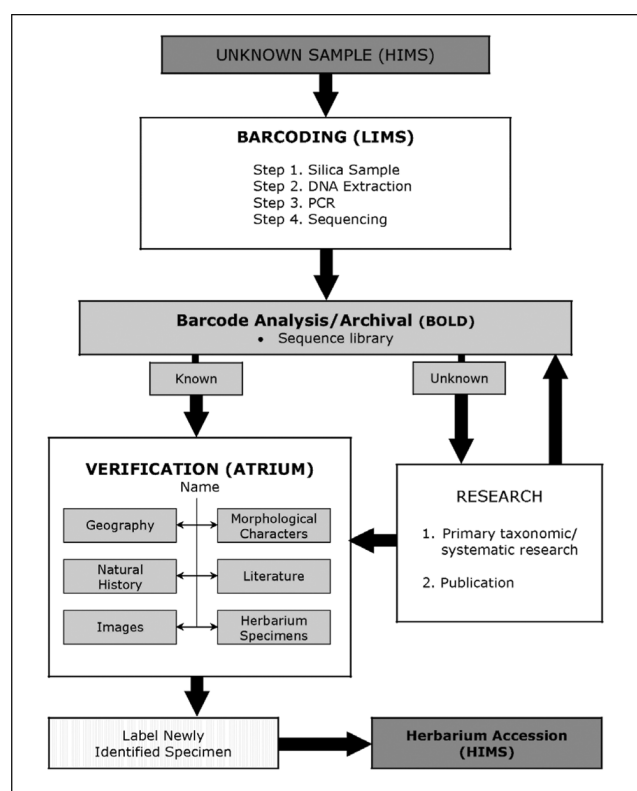
that hinder the standard high-throughput analysis used in Barcode of Life Database (BOLD) (<http://www.boldsystems.org>). Newmaster et al. (2006) introduced the 'tiered concept' or 'nested approach' of analysis as a potential method of facilitating the use of non-coding regions in the bioinformatics processes that the BOLD currently uses. This approach utilises a common, easily aligned coding gene and a more variable non-coding locus in a nested analysis, which could be run in parallel. We currently use a system in which one coding region places an unknown sample within a family, genus or in some cases species. If needed, a more variable region is used to align the unknown sample within a genus or family and then match its sequence with that of a known sequence in our database, thus completing the identification process. Current bioinformatics systems can adapt to this system immediately, but will eventually be replaced by a 'one step' bioinformatics process. We expect this will happen in the near future and encourage our colleagues in computing to collaborate with plant biologists in developing a bioinformatics system that will designate an unknown sample with combined sequence data from coding and non-coding regions.

#### 4 State-of-the-art in AIT bar coding

AIT for plants is anchored in DNA bar coding. Analogous to the UPC bar codes used to identify manufactured goods, DNA bar coding is a tool that facilitates high-throughput identification of organisms. The process of obtaining sequence data (i.e., DNA extraction, PCR, sequencing, databasing) has been automated for animals (Hajibabaei et al., 2005; Richardson et al., 2007) and is being adapted for plants (e.g., Ivanova et al., 2008) at the BIO. Developing an AIT tool begins with the construction of a plant DNA sequence library using accessions from multiple populations for each species, all of which are identified by a taxonomic expert, and with vouchers deposited in a herbarium. Many botanical institutions are in the process of building these databases for entire floras (Newmaster et al., 2006; Kress and Erickson, 2008) with a repository called the Barcode of Life Data System (BOLD; Ratnasingham and Hebert, 2007). BOLD is an informatics workbench aiding the acquisition, storage, analysis and publication of DNA bar code records. By assembling molecular, voucher information, and distributional data, it bridges traditional taxonomy with molecular data in a bioinformatics platform. BOLD has a web-based delivery built on a flexible data security model and is freely available for data deposition to any researcher with interests in DNA bar coding. BOLD also meets the standards needed to gain bar code designation in the global sequence databases: GENbank, National Centre for Biotechnology Information (NCBI), Global Biodiversity Information Facility (GBIF), DNA Data Bank of Japan (DDBJ), European Molecular Biology Laboratory (EMBL). Ratnasingham and Hebert (2007) provide a more detailed discussion on key elements of BOLD. We suggest an automated identification tool for

plants that links via the internet several separate automated processes (see below, Herbarium Information Management System (HIMS), Laboratory Information Management System (LIMS), BOLD, ATRIUM). These separate automated processes start with an unknown plant specimen (from which a small sample is preserved in a package with silica), and end with an identified voucher, which is archived in a herbarium where the data is accessible via the internet (Figure 1). In the future, we will have complete databases for entire floras including voucher records in Herbaria. An unknown plant or part of a plant (e.g., piece of a leaf, root etc.) will become a routine identification and provide a quick and cheap identification service for society at large.

**Figure 1** AIT Flow diagram from an unknown sample to an identified and archived plant voucher

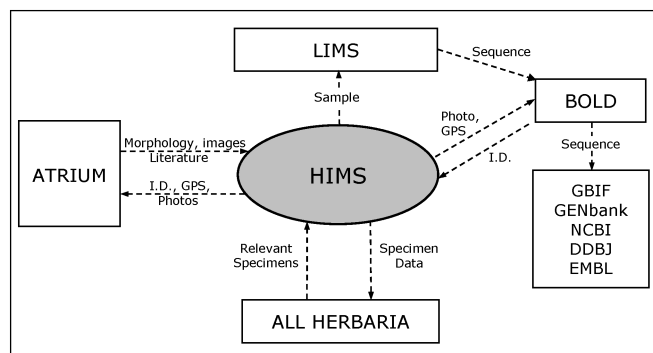


##### 4.1 Herbarium Information Management Systems (HIMS)

HIMS are automated systems developed for creating a database of voucher specimens that can be shared via the internet. HIMS can provide a conduit of specimen information (Figure 2) to the LIMS (Aller, 2005), BOLD, other herbaria and botanical information applications (e.g., Atrium™, see below). HIMS automatically produces labels and provides accession numbers, completing the identification process and archival of a specimen. HIMS can also be linked to other digital herbaria, which share reference data on all voucher specimens. This provides information (photos, locations, etc.) for critical evaluation of the reference specimen in comparison with other specimens of the same species collected and archived

throughout the globe. This system would also provide extensive species information with descriptions, maps, access to herbarium and field fresh images, digital and printed literature, and, if possible, enable connections to ecological and biodiversity data, and conservation programs. Several HIMS programs are available such as BGBASE (2007), BRAHMS (2007) and ETI (2007). Alternatively, individual herbaria may develop their own relational database applications, which meet the specific needs of that institution. We have developed our own web-based relational database application (Newmaster et al., 2008) for introducing novel AIT with direct links to BOLD and the Atrium Biodiversity Information System (see discussion below). This supports our need to process high volumes of plant samples from biodiversity inventories and industrial applications (see below).

**Figure 2** AIT transfer of Information among existing information systems on the web



#### 4.2 Laboratory Information Management System (LIMS)

At the herbarium, a small portion of an unknown sample is placed in a tube and sent to the molecular lab where it enters the process of DNA extraction, PCR and sequencing (Hajibabaei et al., 2005; Richardson et al., 2007; Ivanova et al., 2008). LIMS documents this process and the resultant sequences are automatically uploaded to BOLD (Ratnasingham and Hebert, 2007).

#### 4.3 Barcode of Life Data System (BOLD)

The BOLD informatics workbench automatically assembles the DNA sequence with voucher information and distributional data, which could be automatically provided by the HIMS. In BOLD, any query sequence is aligned very quickly to the global alignment through a Hidden Markov Model (HMM) profile of the respective DNA region, followed by a linear search of the reference library (Ratnasingham and Hebert, 2007). On the basis of sequence variation, the identification system on BOLD delivers specimen identification if the query sequence shows a tight match, less than 1% divergence, to a reference sequence. If two or more taxa share sequences with less than 1% divergence, all possible species assignments are displayed. BOLD shares sequence information with several

international bioinformatics and biodiversity data banks (Figure 2).

#### 4.4 Verification

A web-based relational database system could provide a linkage to existing botanical information databases and a means of verifying specimen's identity against a wealth of information about every known taxon. This includes morphometric databases with information of a particular species' geographic range, exact measurements of various morphological characters, chromosome number, images, chemical constituents, taxonomic descriptions, natural history, toxicology, ecology and relevant primary literature. This wealth of information could be used to verify the identity of the unknown plant. Currently, there are several web-based applications such as the Integrated Botanical Information System (IBIS, 2006) in Australia, BGBASE in Britain, the World Taxonomist Database in the UK (ETI, 2007) and Atrium in the USA. We are using Atrium because it is based on standard, modern biodiversity information system technology and it follows an open-source philosophy (see Atrium description here).

As a tool for identification of described species, the AIT system is entirely dependent on existing taxonomy derived from traditional taxonomic methods. Paralleling the traditional methods, sequences and DNA samples are associated with voucher specimens. Where novel sequences or greater than expected variation within species are encountered, AIT acts as a beacon to attract attention to cryptic species that have gone undetected, prompting and directing alpha-taxonomic research. Once this primary research is complete and published, the DNA library is updated in BOLD and information is added to the relational database system (Figure 1).

### 5 The atrium biodiversity information system, version 1.6

The Atrium Biodiversity Information System is one of the most innovative systems available for integrating, managing, sharing and publishing biodiversity and allied data. While there are other proprietary and open-source collection management systems available, Atrium goes far beyond mere collection or specimen management by providing a platform for merging specimen data with literature citations, GIS layers, ecological data, species information, and images. Atrium manages not only plant collections and associated data, but it also manages independent specimen records for each collection. Duplicate specimens deposited in different herbaria are managed independently and each are connected to an associated collection. Checklists can be generated at the level of family, genus and species, as well as by plant habit (life form) and habitat. In Atrium version 1.6 published in November 2007, a new module called Vegetation Survey provides innovative tools for collecting, integrating,

managing, analysing and publishing primary data from quantitative vegetation plots and transects. This new ecological module connects individual trees and their quantitative measurements (i.e., diameter and height) to voucher collection data, images, maps, literature and species pages. Current programming includes the development of a digital flora module that will publish collections, images, descriptions, and maps for specific research sites where floristic treatments are needed (i.e., biological stations, conservation areas and areas of conservation planning). The dynamic, digital floras will connect to the automated plant field guide tools already available in Atrium. This will function as an interactive key to plants.

By providing a web-based interface for a broad spectrum of biodiversity data, Atrium facilitates collaboration, analysis, management and sharing of data on a global scale. The Atrium field database, which is in testing, allows field researchers to search, manage, edit and add to collection data while offline, then synchronise via upload to Atrium servers when online. Atrium allows unprecedented collaboration and rapidity of determinations from experts worldwide, thus vastly decreasing the distance between discovery and dissemination of accurate data.

What the user currently sees at <http://atrium.andesamazon.org> represents (as of November 2007) two years of intensive programming and testing by the BRIT IT team and the scientists, students and others collecting and contributing data. Programming started in June 2005 after nine months of planning the system technology and initial user interface requirements. The current version is Atrium 1.6, which includes a real-time digital herbarium, with data download, label printing, interactive data entry, interactive description interface, online editing, automated family, genus, and species pages, integrated Google Maps and automated colour field guides designed and organised based on users' search strategies.

There are several layers of plant data sources contributing to Atrium, originating through field, laboratory and museum work. The current plant collection and species data that are currently available through the Atrium database include 15,000 collection records and 50,000 associated specimen records in 261 plant families representing close to 5000 species. These specimens are represented by 25,000 images of fresh plant material, habit and habitat, and dried specimens scanned at very high resolution. New plant collections and species data continue to be accumulated through intensive field work being carried out in remote regions of the Andes-Amazon region of Peru, as well as in Papua New Guinea and Melanesia, in general.

### *5.1 Atrium technology standards*

The Atrium server infrastructure follows the industry-standard LAMP model (Linux, Apache, MySQL, PHP) that provides a time-tested operating environment supported by a vast array of supplementary tools and extensions. Many web applications are also based on the LAMP model, which facilitates integration with Atrium.

All core components of Atrium infrastructure are open-source, which provide a stable, peer-reviewed, industry-standard foundation for our application. As there are no licensing costs associated with open-source software, the BRIT team will be able to provide Atrium to other organisations without the significant cost barrier associated with conventional commercial software. The Atrium server and network infrastructure is in place and fully functional with IT support from BRIT Systems Administration.

### *5.2 Atrium GIS data and metadata map server*

Atrium includes a GIS Data and Metadata Server that allows users to browse, search, visualise, download and upload an extensive collection of BRIT's GIS data layers, such as hydrology, collection points, elevation grids and satellite imagery (<http://atrium.andesamazon.org/gis>).

### *5.3 Atrium data standards*

Atrium adheres to data standards established and maintained by TDWG including the data elements defined by Distributed Generic Information Retrieval (DiGIR), Access to Biological Collections Data (ABCD) and the Darwin Core. The system supports bibliographic data imported and exported in ISI and EndNote (XML) formats. Metadata for GIS datasets stored in Atrium complies with ISO 19115 and Federal Geographic Data Committee (FGDC) standards. For interactive display and search of geospatial datasets, Atrium is integrated with University of Minnesota MapServer, which is compliant with the standards set by the Open Geospatial Consortium. Atrium has enabled BRIT to become a data provider for the Global Biodiversity Information Facility (GBIF), a multinational organisation that consolidates taxonomic and collection data into a unified network.

### *5.4 Atrium management and tracking*

The Atrium development team has worked closely with BRIT researchers and other Atrium users to design and implement an application that is uniquely designed to support core scientific goals, with careful planning and prioritisation of features in future versions. Atrium users are from around the world, with heaviest user groups in North, Central and South America.

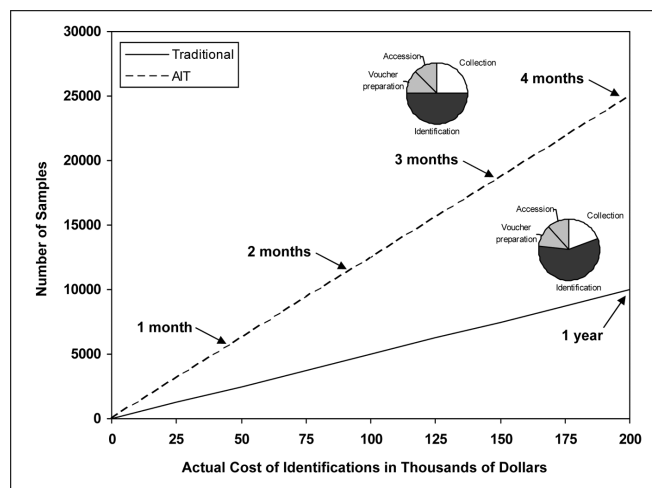
## **6 AIT efficacy**

It is timely for plant bar coding to make the move to an automated high-throughput system. The Canadian Centre for DNA Barcoding (CCDB) at the BIO is at the leading edge of this technology with regular publications (see *Advances* publication – [www.dnabarcoding.ca](http://www.dnabarcoding.ca)) on advances in the research and development of high-throughput DNA bar coding systems. Recently, our research group collaborated with CCDB in the development of a high-throughput DNA extraction protocol for plants at the CCDB (Ivanova et al., 2008). We explored the efficacy

of bar coding against our traditional taxonomic protocols at the BIO herbarium where we annually process over 10,000 samples. This includes collection, identification, voucher preparation (mounting and labelling) and accession (photograph, database and archive) into herbaria. The time required for plant identification using traditional taxonomic methods accounts for about two-thirds of the total cost of processing these samples (i.e., \$200,000 for 10,000 samples, Figure 3). In comparison, it is expected that we can identify all 10,000 samples in less than two months for less than half our current cost of identifications using traditional methods (Figure 3). This efficiency in identification equates with savings in time and funding allowing us to

- increase the number of identifications we can do each year from 10,000 samples to 25,000 samples in a quarter of the time (one season instead of one year, Figure 3)
- generate increased revenue through an increase in routine identifications
- reallocate the time and resources saved from these measures to alpha taxonomy in which bar coding can be used as a tool for the discovery of cryptic taxa (Hebert et al., 2004; Smith et al., 2006) or samples (Hajibabaei et al., 2006; Schneider and Schuettpelz, 2006)
- connect and integrate our alpha-taxonomic research with ecological, biogeographic, and conservation studies.

**Figure 3** Efficacy of AIT and traditional identification (pie diagram represents the proportion of the costs)



## 7 AIT applications

AIT provides a tool that allows us to realistically consider recording all the biodiversity on earth. Janzen (2003) suggested two steps in recording the earth's biodiversity. First, taxonomic organisations with access to the world's biodiversity collections in museums, herbaria and microbe depositories should extract DNA to build bar code libraries.

The second step is the sequencing and characterisation of the yet undiscovered species in the landscape. We propose an additional phase, where the first step is the development of the AIT system that will link together any taxonomist with plant collections, reference material, floristic surveys and societies' need for information of the earth's biodiversity.

Documenting the rest of earth's biota is hindered by the lack of taxonomic expertise to process large collections. Lee (2000) and Mori (1992) discuss how this is becoming more challenging as the number of scientists trained in taxonomy, floristics and systematics drastically declines at a time when this type of research is urgently needed (Tobler et al., 2007). Currently, it is a fact that extensive explorations in remote areas have been limited by the paucity of taxonomic experts who can process large collections in search of new species. Several studies have concluded that the lack of available data limits our ability to conduct a thorough assessment of the diversity and distribution and consequently conservation of plants in the areas such as the Amazonian region (Tobler et al., 2007; Nelson et al., 1990; Williams et al., 1996; Kress et al., 1998). AIT provides a tool for quick identification of known species and flags samples that represent new species or taxa with cryptic diversity. Currently, we are collaborating with the Botanical Research Institute of Texas (BRIT) in a vast survey of the Andes-Amazon region of southeastern Peru, one of the last vast tropical wilderness areas on the planet. This includes collection of specimens by parataxonomists (locals trained as field biologists; Janzen, 2004a, 2004b), identification of samples using AIT and the description of cryptic specimens as flagged by AIT. It also includes the connection of primary biodiversity data with the local, regional and national governments, as well as non-governmental organisations concerned with the conservation of the vast wilderness area in this region, and the rich biodiversity it harbours.

Many industries conducting routine surveys of botanical diversity would benefit from AIT. Environmental consulting firms conduct botanical surveys on potential development properties. Ecosystem management firms conduct plant surveys while auditing for impacts on plant diversity (e.g., forestry impacts) to mitigate the loss of species. Pharmaceutical companies are exploring plant diversity in search of new medicine. Currently, these plant surveys are conducted in the same manner as was done 100 years ago. The bottleneck to completing a plant survey is always identification and documentation of the species because it relies on time-consuming methods and limited resources. AIT can be used as a tool to overcome this bottleneck because it is cost-effective and efficient. The technology is not limited to experts and can be set-up anywhere.

AIT could serve as a useful tool for authenticating products in certain industries. For example, AIT could be used in the herbal industry to identify the presence of a medicinal plant or a poisonous plant within a processing facility. Numerous examples of unintentional contamination

or species substitution in herbal preparations exist (Cole and Fetrow, 2003; Chaffin, 1999; Anonymous/Caregroup, 2006; Dhiman and Singh, 2003). This includes fraudulent substitution in a product with other species that are more easily accessible. AIT could be used as a tool that could guarantee a certain level of purity, a 'gold standard' for herbal products. The cultivation of many medicinal plants requires specific cultural practices and agronomical requirements. There is an urgent need to authenticate and identify the species-specific requirements which depend on soil, water and climatic conditions. Proper identification is needed at several stages of herbal plant processing including, seed selection, propagation, harvesting and step-wise quality control of raw material up to processing and storage stage.

AIT could be used as a tool to control accidental immigration of invasive species. AIT supports the recent call for detection, identification, and monitoring of invasive species (Pheloung et al., 1999; Daehler et al., 2003; Morse et al., 2004; Darling and Blum, 2007). AIT could be used at border control sites to randomly check plant materials entering the country; identifying species in bulk plant materials such as vats of seeds or pulverised plant materials. The management of risks posed by potential invasive species requires, at a minimum, the ability to recognise those species (Mallet and Willmott, 2003; Darling and Blum, 2007). The logistics of early detection and monitoring of invasive species entail that the tools utilised for species recognition be rapidly deployable, cost-effective, technically accessible and accurate; in addition, the best tools will be applicable across a wide range of taxa. Thus, the expensive, time-consuming traditional approach of identification of plants requiring the expertise of multiple taxonomists can now be replaced by AIT, which provides a solution to the limitations of morphological approaches to species identification. Those people who are in close contact with plants on the landscape (e.g., farmers, weed inspectors, foresters, etc.) could send suspect samples to a facility where AIT could provide inexpensive and quick identification leading to early detection and eradication of invasive species.

## 8 Future directions

AIT could become a state-of-the-art system that will revolutionise biology and have considerable impacts on society. Advances in both molecular and bioinformatics technology will advance AIT so that it is available to society. Scientists (Pennisi, 1994; Tautz et al., 2002) have discussed the application of computer identifications and recent innovations in animal and plant bar coding (Pennisi, 2007; Fazekas et al., 2008). Research has now shifted from an exploratory phase to a high-throughput phase based on these recent innovations. The 'tricorder' remains science fiction for now, but a hand-held species identification device could be available in the near future.

Hand-held computers link us to the internet and have considerable data storage; PCR and sequencing technology (microfluidic devices) are currently at the size (nanolitre scale) that could fit in a very small portable field unit. We expect that field biologists will some day be using a hand-held AIT system in which an unknown plant leaf is inserted for identification yielding via wireless connection to the internet a plethora of information such as scientific name, images, range maps, ecology, chemistry, food value, etc. Ecologists could use this as a tool for quick surveys as could environmental consultants and conservationists. Border control stations could search for invasive species within materials such as wood products entering the country. Industry could implement protocols to identify contaminants in food or health products. Farmers and gardeners could quickly identify a weed and learn how to control it. Naturalists could explore wetlands and children could explore their back yard. All of the biodiversity data from these applications could be fed into a data repository that would help us to understand, appreciate and conserve our natural world.

## Acknowledgement

This research was supported by Genome Canada through the Ontario genomics institute and the Canadian Foundation for Innovation. Special thanks to Paul Hebert, Sujeevan Ratnasingham and the Barcode of Life Data System team for supporting plant bar coding on BOLD. We thank Aron Fazekas, Candice Newmaster, Tim Zitnak anonymous reviewers and the editors of this Journal for their editorial review.

## References

- Allander, T., Emerson, S.U., Engle, R.E., Purcell, R.H. and Bukh, J. (2001) 'A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species', *Proceedings National Academy of Science USA*, Vol. 98, pp.11609–11614.
- Aller, R.D. (2005) 'CAP TODAY's annual laboratory information systems line-up', *System Review Series*, Vol. 32, pp.28–56.
- Anonymous/Caregroup (2006) *Adulteration and Contaminants in Chinese Patent Medicines*, [http://home.caregroup.org/clinical/altmed/interactions/Herb\\_Groups/Adulteration\\_an.htm](http://home.caregroup.org/clinical/altmed/interactions/Herb_Groups/Adulteration_an.htm)
- ATRIUM 1.6 (2007) <http://atrium.andesamazon.org/> (the database) and <http://www.atrium-biodiversity.org> (the central website about Atrium).
- BG-BASE Inc. and BG-BASE Ltd. (1997–2007) *BG-BASE Collection Management Software. Royal Botanical Garden, USA and UK*, <http://rbg-web2.rbge.org.uk/bg-base/>
- Blaxter, M. and Floyd, R. (2003) 'Molecular taxonomics for biodiversity surveys: already a reality', *Trends in Ecology and Evolution*, Vol. 18, pp.268–269.
- Blaxter, M.L. (2004) 'The promise of a DNA taxonomy', *Philosophical Transactions of the Royal Society of London Series B Biological Science*, Vol. 359, pp.669–679.

- BRAHMS and BRAHMS online (1985–2007) *University of Oxford*, <http://herbaria.plants.ox.ac.uk/BOL/software/default.aspx>
- Brown, B., Emberson, R.M. and Paterson, A.M. (1999) 'Mitochondrial COI and II provide useful markers for *Weiseana* (Lepidoptera, Hepialidae) species identification', *Bulletin of Entomological Research*, Vol. 89, pp.287–294.
- Bucklin, A., Guarnieri, M., Hill, R.S., Bentley, A.M. and Kaartvedt, S. (1999) 'Taxonomic and systematic assessment of planktonic copepods using mitochondrial COI sequence variation and competitive, species-specific PCR', *Hydrobiology*, Vol. 401, pp.239–254.
- Chaffin, J.A. (1999) 'Contamination in herbal preparations', *Minnesota Medicine*, Vol. 82, pp.1–7.
- Charters, Y.M., Robertson, A., Wilkinson, M.J. and Ramsay, G. (1996) 'PCR analysis of oilseed rape cultivars (*Brassica napus* L. ssp. *oleifera*) using 5¢-anchored simple sequence repeat (SSR) primers', *Theoretical Applied Genetics*, Vol. 92, pp.442–447.
- Chase, M.W., Cowan, R.S., Hollingsworth, P.M., van den Berg, C., Madriñán, S., Petersen, G., Seberg, O., Jørgensen, T., Cameron, K.M., Carine, M., Pedersen, N., Hedderson, T.A.J., Conrad, F., Salazar, G.A., Richardson, J.E., Hollingsworth, M.L., Barraclough, T.G., Kelly, L. and Wilkinson, M. (2007) 'A proposal for a standardised protocol to barcode all land plants', *Taxon*, Vol. 56, No. 2, pp.295–299.
- Chase, M.W., Salamin, N., Wilkinson, M., Dunwell, J.M., Prasad Kesanakurthi, R., Haidar, N. and Savolainen, V. (2005) 'Land plants and DNA barcodes: short-term and long-term goals', *Philosophical Transactions of the Royal Society of London Series B Biological Sciences*, Vol. 360, pp.1889–1895.
- Cole, M.R. and Fetrow, C.W. (2003) 'Adulteration of dietary supplements', *American Journal of Health-System Pharmacy*, Vol. 60, pp.1576–1580.
- Cowan, R.S., Chase, M.W., Kress, W.J. and Savolainen, V. (2006) '300,000 species to identify: problems, progress, and prospects in DNA barcoding of land plants', *Taxon*, Vol. 55, pp.611–616.
- Daehler, C.C., Denslow, J.S., Ansari, S. and Kuo, H.C. (2003) 'A risk assessment system for screening out invasive pest plants from Hawaii and other pacific islands', *Conservation Biology*, Vol. 18, pp.360–368.
- Darling, J.A. and Blum, M.J. (2007) 'DNA-based methods for monitoring invasive species: a review and prospectus', *Biological Invasions*, Vol. 9, pp.751–765.
- Dhiman, B. and Singh, M. (2003) 'Molecular detection of cashew husk (*Anacardium occidentale*) adulteration in market samples of dry tea (*Camellia sinensis*)', *Planta Medica*, Vol. 69, pp.882–884.
- ETI (2007) (*World Biodiversity Database, World Taxonomist Database*) and the *Linnaeus II Software Modules*, ETI Information Services Ltd., UK, <http://www.eti.uva.nl/tools/wbd.php>
- Fazekas, A.J., Burgess, K.S., Kesanakurti, P.R. Percy, D.M., Hajibabaei, M., Graham, S.W., Newmaster, S.G., Husband, B.C. and Barret, S.C.H. (2008) 'Assessing the utility of coding and non-coding genomic regions for plant DNA barcoding', *PLOS I*, Vol. 3, pp.1–12.
- Floyd, R., Eyualem, A., Papert, A. and Blaxter, M.L. (2002) 'Molecular barcodes for soil nematode identification', *Molecular Ecology*, Vol. 11, pp.839–850.
- Gaston, K.J. and O'Neill, M.A. (2004) 'Automated species identification: Why not?', *Philosophical Transactions of the Royal Society of London Series B*, Vol. 359, pp.340–387.
- Godfray, H.C.J. (2002a) 'Challenges for taxonomy', *Nature*, Vol. 417, pp.17–19.
- Godfray, H.C.J. (2002b) 'Towards taxonomy's 'glorious revolution'', *Nature*, Vol. 420, p.461.
- Godfray, H.C.J. and Knapp, S. (2004) 'Introduction. in taxonomy for the twenty-first century—special issue', *Philosophical Transactions of the Royal Society of London Series B*, Vol. 359, pp.559–569.
- Hajibabaei, M., deWaard, J.R., Ivanova, N.V., Ratnasingham, S., Dooh, R.T., Kirk, S.L., Mackie, P.M. and Hebert, P.D.N. (2005) 'Critical factors for assembling a high volume of DNA barcodes', *Philosophical Transactions of the Royal Society Series B*, Vol. 360, pp.1959–1967.
- Hajibabaei, M., Singer, G.A.C., Clare, E.L. and Hebert, P.D.N. (2007a) 'Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring', *BMC Biology*, Vol. 5, p.24.
- Hajibabaei, M., Singer, G.A.C., Hebert, P.D.N. and Hickey, D.A. (2007b) 'DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics', *Trends in Genetics*, Vol. 23, pp.167–172.
- Hajibabaei, M., Smith, M.A., Janzen, D.H., Rodriguez, J.J., Whitfield, J.B. and Hebert, P.D.N. (2006) 'A minimalist barcode can identify a specimen whose DNA is degraded', *Molecular Ecology Notes*, Vol. 6, pp.959–964.
- Hamels, J., Gala, L., Dufour, S., Vannuffel, P., Zammateo, N. and Remacle, J. (2001) 'Consensus PCR and microarray for diagnosis of the genus *Staphylococcus*, species, and methicillin resistance', *BioTechniques*, Vol. 31, pp.1364–1372.
- Hammond, P. (1992) 'Species inventory', in Groombridge, B. (Ed.): *Global Biodiversity: Status of the Earth's Living Resources*, Chapman and Hall, London, pp.17–39.
- Hawksworth, D.L. and Kalin-Arroyo, M.T. (1995) 'Magnitude and distribution of biodiversity', in Heywood, V.H. (Ed.): *In Global biodiversity Assessment*, Cambridge University Press, UK, pp.107–191.
- Hebert, P.D.N. and Gregory, T.R. (2005) 'The Promise of DNA barcoding for taxonomy', *Systematic Biology*, Vol. 54, No. 5, pp.852–859.
- Hebert, P.D.N., Cywinska, A., Ball, S.L. and deWaard, J.R. (2003) 'Biological identifications through DNA barcodes', *Proceedings of the Royal Society of London Series B*, Vol. 270, pp.313–321.
- Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H. and Hallwachs, W. (2004) 'Ten species in one: DNABarcoding reveals cryptic species in the neotropical skipper butterfly *Astrapes fulgerator*', *Proceedings of the National Academy of Science USA*, Vol. 101, pp.14812–14817.
- IBIS (1993–2006) *Australian National Botanic Gardens-Australian National Herbarium*, Australia, <http://www.eti.uva.nl/tools/wbd.php>
- Ivanova, N.V., Fazekas, A.J. and Hebert, P.D.N. (2008) 'Semi-automated, membrane based protocol for DNA isolation from plants', *Plant Molecular Biology Reporter*, Vol. 26, pp.186–198.
- Janzen, D.H. (2004a) 'Setting up tropical biodiversity for conservation through non-damaging use: participation by parataxonomists', *Journal of Applied Ecology*, Vol. 41, pp.181–187.

- Janzen, D.H. (2004b) 'Now is the time', *Philosophical Transactions of the Royal Society of London Series B*, Vol. 359, pp.731–732.
- Janzen, E. (2003) *Taxonomy, DNA, and the Barcode of Life web Publication*, [http://phe.rockefeller.edu/BarcodeConference/archives/2003\\_09.html](http://phe.rockefeller.edu/BarcodeConference/archives/2003_09.html)
- Kantety, R.V., Zeng, X., Bennetzen, J.L. and Zehr, B.E. (1995) 'Assessment of genetic diversity in dent and popcorn (*Zea mays* L.) inbred lines using inter-simple repeat (ISSR) amplification', *Molecular Breeding*, Vol. 1, pp.365–373.
- Kress, J. and Erickson, D.L. (2008) 'DNA barcodes: genes, genomics, and bioinformatics', *Proceedings National Academy Science USA*, Vol. 105, pp.2761–2762.
- Kress, W.J. and Erickson, D.L. (2007) 'A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnHpsbA* spacer region', *PLoS ONE*, Vol. 6, pp.1–10.
- Kress, W.J., Heyer, W.R., Acevedo, P., Coddington, J., Cole, D., Erwin, T.L., Meggers, B.J., Pogue, M., Thorington, R.W., Vari, R.P., Weitzman, M.J. and Weitzman, S.H. (1998) 'Amazonian biodiversity: assessing conservation priorities with taxonomic data', *Biodiversity and Conservation*, Vol. 7, pp.1577–1587.
- Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A. and Janzen, D.H. (2005) 'Use of DNA barcodes to identify flowering plants', *Proceedings of the National Academy of Science USA*, Vol. 102, pp.8369–8374.
- Kumar, L.D., Kathirvel, M., Rao, G.V. and Nagaraju, J. (2001) 'DNA profiling of disputed chilli samples (*Capsicum annum*) using ISSR-PCR and FISSR-PCR marker assays', *Forest Science*, Vol. 116, pp.63–68.
- Kurtzman, C.P. (1994) 'Molecular taxonomy of the yeasts', *Yeast*, Vol. 10, pp.1727–1740.
- Lee, M.S.Y. (2000) 'A worrying systematic decline', *Trends in Ecology and Evolution*, Vol. 15, pp.346–346.
- Leroy, X.J., Leon, K. and Branchard, M. (2000) 'Plant genomic instability detected by microsatellite-primers', *Electronic Journal of Biotechnology*, Vol. 3, pp.140–148.
- Lipscomb, S., Platnick, N. and Wheeler, Q. (2003) 'The intellectual content of taxonomy: a comment on DNA taxonomy', *Trends in Ecology and Evolution*, Vol. 18, pp.65–66.
- Little, D.P. and Stevenson, W.M.D. (2007) 'A comparison of algorithms for identification of specimens using DNA barcodes: examples from gymnosperms', *Cladistics*, Vol. 23, pp.1–21.
- Mallet, J. and Willmott, K. (2003) 'Taxonomy: renaissance or tower of babel?', *Trends in Ecology and Evolution*, Vol. 18, pp.57–59.
- May, R.M. (1988) 'How many species are there on Earth?', *Science*, Vol. 241, pp.1441–1449.
- Mori, S.A. (1992) 'Neotropical floristics and inventory: Who will do the work?', *Brittonia*, Vol. 44, pp.372–375.
- Morse, L.E., Randall, J.M., Benton, N., Hiebert, R. and Lu, S. (2004) *An Invasive Species Assessment Protocol: Evaluating Non-native Plants for their Impact on Biodiversity*, Version 1, Nature Serve, Arlington, Virginia.
- Nanney, D.L. (1982) 'Genes and phenes in Tetrahymena', *Bioscience*, Vol. 32, pp.783–788.
- Nelson, B.W., Ferreira, C.A.C., Dasilva, M.F. and Kawasaki, M.L. (1990) 'Endemism centers, refugia and botanical collection density in Brazilian Amazonia', *Nature*, Vol. 345, pp.714–716.
- Newmaster, S.G., Fazekas, A., Steeves, R. and Janovec, J. (2008) 'Testing candidate plant barcode regions in the myristicaceae', *Molecular Ecology Notes*, Vol. 8, pp.480–490.
- Newmaster, S.G., Fazekas, A.J. and Ragupathy, S. (2006) 'DNA barcoding in the land plants: evaluation of *rbcL* in a multigene tiered approach', *Canadian Journal of Botany*, Vol. 84, pp.335–341.
- Pace, N.R. (1997) 'A molecular view of microbial diversity and the biosphere', *Science*, Vol. 276, pp.734–740.
- Pennisi, E. (1994) 'Name that fly – computers help make species identification child's play', *Science News*, Vol. 145, pp.108–111.
- Pennisi, E. (2007) 'Wanted: a barcode for plants', *Science*, Vol. 318, pp.190–191.
- Pheloung, P.C., Williams, P.A. and Halloy, S.R. (1999) 'A weed risk assessment model for use as a bio-security tool evaluating plant introductions', *Journal of Environmental Management*, Vol. 57, pp.239–251.
- Presting, G.G. (2006) 'Identification of conserved regions in the plastid genome: implications for DNA barcoding and biological function', *Canadian Journal of Botany*, Vol. 84, pp.1434–1443.
- Ratnaparkhe, M.B., Tekeoglu, M. and Muehlbauer, F.J. (1998) 'Inter-Simple-Sequence-Repeat (ISSR) polymorphisms are useful for finding markers associated with disease resistance gene clusters', *Theoretical Applied Genetics*, Vol. 97, pp.515–519.
- Ratnasingham, S. and Hebert, P.D.N. (2007) 'BOLD: the barcode of life data system ([www.barcodinglife.org](http://www.barcodinglife.org))', *Molecular Ecology Notes*, Vol. 7, No. 3, pp.355–364.
- Richardson, D.E., Vanwye, J.D., Exum, A.M., Cowen, R.K. and Crawford, D.L. (2007) 'High-throughput species identification: from DNA isolation to bioinformatics', *Molecular Ecology Notes*, Vol. 7, pp.199–207.
- Schneider, H. and Schuettelpelz, E. (2006) 'Identifying fern gametophytes using DNA sequences', *Molecular Ecology Notes*, Vol. 6, pp.989–991.
- Smith, M.A., Woodley, N.E., Janzen, D.H., Hallwachs, W. and Hebert, P.D.N. (2006) 'DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae)', *Proceedings of the National Academy of Sciences USA*, Vol. 103, pp.3657–3662.
- Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., Vermet, T., Corthier, G., Brochmann, C. and Willerslev, E. (2006) 'Power and limitations of the chloroplast *trnL* (UAA) intron for plant DNA barcoding', *Nucleic Acids Research*, Vol. 35, No. e14, pp.1–8 (doi:10.1093/nar/gkl938).
- Tautz, D., Arctander, P., Minelli, A., Thomas, R.H. and Vogler, A.P. (2002) 'DNA points the way ahead in taxonomy', *Nature*, Vol. 418, p.479.

- Tautz, D., Arctander, P., Minelli, A., Thomas, R.H. and Vogler, A.P. (2003) 'A plea for DNA taxonomy', *Trends in Ecology and Evolution*, Vol. 18, pp.70–74.
- Tobler, M., Honorio, E., Janovec, J. and Reynel, C. (2007) 'Implications of collection patterns of botanical specimens on their usefulness for conservation planning: an example of two neotropical plant families (Moraceae and Myristicaceae) in Peru', *Biodiversity and Conservation*, Vol. 16, pp.659–677.
- Trewick, S.A. (2000) 'Mitochondrial DNA sequences support allozyme evidence for cryptic radiation of New Zealand Peripatoides (Onychophora)', *Molecular Ecology*, Vol. 9, pp.269–282.
- Vincent, S., Vian, J.M. and Carlotti, M.P. (2000) 'Partial sequencing of the cytochrome oxidase-b subunit gene. I. A tool for the identification of European species of blow flies for post mortem interval estimation', *Journal of Forensic Science*, Vol. 45, pp.820–823.
- Weising, K., Nybom, H., Wolff, K. and Kahl, G. (2005) *DNA Fingerprinting in Plants*, 2nd ed., CRC Press, Boca Raton, FL.
- Wheeler, Q.D., Raven, P.H. and Wilson, E.O. (2004) 'Taxonomy: Impediment or Expedient?', *Science*, Vol. 303, No. 5656, pp.285, 286.
- Will, K.W. and Rubinoff, D. (2004) 'Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification', *Cladistics*, Vol. 20, pp.47–55.
- Williams, P.H., Prance, G.T., Humphries, C.J. and Edwards, K.S. (1996) 'Promise and problems in applying quantitative complementary areas for representing the diversity of some neotropical plants (families Dichapetalaceae, Lecythidaceae, Caryocaraceae, Chrysobalanaceae and Proteaceae)', *Biological Journal of the Linnean Society*, Vol. 58, pp.125–157.
- Wilson, K.H. (1995) 'Molecular biology as a tool for taxonomy', *Clinical Infectious Diseases*, Supplement 20, pp.192–208.
- Wink, M., Guicking, D. and Fritz, U. (2001) 'Molecular evidence for hybrid origin of *Mauremys iversoni*', in Pritchard et McCord, 1991, and *Mauremys pritchardi* McCord, 1997 (Reptilia: Testudines: Bataguridae)', *Zoologische Abhandlungen/Staatliches Museum für Tierkunde in Dresden*, Vol. 51, pp.41–50.
- Wink, M., Sauer-Gürth, H., Martinez, F., Doval, G., Blanco, G. and Hatzofe, O. (1998) 'The use of (GACA)<sub>4</sub> – PCR to sex Old World vultures (Aves: Accipitridae)', *Molecular Ecology*, Vol. 7, pp.779–782.