Agro-biodiversity informatics with special reference to spices

Santhosh J. Eapen and Riju A.
Bioinformatics Centre, Indian Institute of Spices Research, Calicut, Kerala – 673 012
Email: sjeapen@spices.res.in

Biodiversity stands for all living things on earth. It refers to the range of variations among a set of entities and is commonly used to describe variety and variability of living organisms in terms of genetic diversity (heritable variations within populations), species diversity (species richness in a habitat) and ecological diversity (biophysical diversity). India, because of its unique biogeographic location embraces three major biological realms, viz. Indo-Malayan, Eurasian and Afro-tropical. It is notable in its species-richness and endemism, and is ranked tenth amongst the biodiversity-rich countries. India is one of the world's 12 Vavilovian centres of origin and diversification of cultivated plants (with 167 species of agrihorticultural crops and 320 species of their relatives known to have originated here). Conservative estimates of species-richness show that around 127,000 species (plants, animals and microorganisms) have been so far reported from India and 400,000 species are yet to be explored (Roy *et al.,* 2002). Biodiversity also occurs at a genetic level. In other words, even within a single species there is variation. This is known as genetic biodiversity. Different populations of the same species can differ in their genetic profiles. It is this genetic biodiversity which facilitates the breeding of different varieties of plants and animals. Diversity in genes provides the basis for continued survival of the species in view of the changing environment.

Why is biodiversity important?

Biodiversity is part of our daily lives and livelihood and constitutes the resources upon which families, communities, nations and future generations depend. The wellbeing and prosperity of earth's ecological balance as well as human society directly depend on the extent and status of biological diversity (CBD, 1992). Plant and animal diversity ensures a constant and varied source of food, medicine and raw material for human populations. Much of traditional medicine is based around plant extracts found in nature. Modern researchers are increasingly looking towards biological resources to find treatments and cures for illnesses. In agriculture, biodiversity provides us with a varied food supply, which is needed for balanced human nutrition. It is important to conserve all the variability in the species gene pool including wild relative, land races and cultivars which are highly vulnerable and prone for extinction. Loss in diversity is a direct threat to the ecological security of the country and the livelihood security of millions of people. As human populations and their demands on the natural world grow, our accumulated knowledge about biodiversity and the environment will become ever more important in the effort to develop a sustainable world. Beside the profound ethical and aesthetic implications, it is clear that the loss of biodiversity has serious economic and social costs.

The biological diversity we see today is the result of billions of years of evolution. While we all welcome our modern thriving economy, we must acknowledge the pressure that increased transportation, construction, inappropriate agricultural practices, poorly managed afforestation and climate change have put on biodiversity. Other factors such as over exploitation and climate change have also had a significant impact on biodiversity. Anthropogenic activities coupled with the burgeoning human population, have led to the grim biodiversity scenario; numerous important plant and animal species are on the verge of extinction, while others are threatened or vulnerable. Biological diversity must be treated more seriously as a global resource, to be indexed, used, and above all, preserved.

Biodiversity informatics

Biodiversity Informatics is the application of informatics to recorded and yet-to-be discovered information specifically about biodiversity, and the linking of this information with genomic, geospatial and other biological and non-biological datasets. The first use of the term can be traced back as far as 1993. According to Soberon and Peterson (2004) Biodiversity Informatics includes the application of information technologies to the management, algorithmic exploration, analysis and interpretation of primary data regarding life, particularly at the species level of organization. It covers the information generated by the fields of systematics (including molecular systematics), evolutionary biology, population biology, behavioural sciences, and synecological fields ranging from pollination biology to parasitism and phytosociology. Presently, the overriding objective of research and other activities in

Biodiversity Informatics is to provide a sound information management infrastructure for biodiversity and global change research.

Biodiversity data refers to scientific information, primarily about biological species and specimens. Biodiversity Informatics has to provide consensus reference systems in structural features (e.g. in database design) and content definitions (e.g. in controlled vocabularies, i.e., lists of applicable terms). Taxon based information systems (or systems using taxon names) must find ways to map individual taxon concepts reliably. The current challenge for biodiversity informatics lies in the creation of descriptive systems (and identification tools based on these), which are useful also for the non-specialist. The electronic representation of voucher specimens (an estimated 2.5 billion specimens world-wide; Duckworth *et al.* 1993) is another priority task for Biodiversity Informatics. Biodiversity information systems also need sophisticated spatial reference systems, which are able to unite several layers of information related to a geographic reference point.

Standardization of biological databases has been an important issue and the IUBS Commission for Taxonomic Databases (TDWG 2001) has been a driving force and an important forum for information exchange on this aspect. With respect to taxon names and biological collection data, detailed information models exist which can serve as a base for implementation projects. Some of the hot issues in that realm include concept mapping and data quality and automatic linkage evaluation tools. For collections, attempts are being made to create metadata based systems (e.g. BioCISE and collaborating projects) which are thought as an intermediate step on the way to specimen level information systems (Berendsohn *et al.,* 2000). Biodiversity informatics will also be in the forefront of the development of user interfaces adequately representing highly complex data structures in distributed systems.

Biodiversity database initiatives in India

The development of Biodiversity Informatics resulted in the establishment of a number of biodiversity databases and information resources (Table 1). India has developed a number of excellent biodiversity databases such as the Flora of Karnataka, Traditional Knowledge Digital Library (NISCAIR, 2002) or the National Register of Green Grassroots Innovations and Traditional Knowledge (NIF, 2002). There exist therefore rich possibilities of building upon country's biodiversity resources and associated knowledge; to promote biodiversity-based enterprises in the modern, as well as traditional sectors; to develop biotechnology industries at the cutting edge of new technologies as well as to encourage local level value addition to biodiversity resources.

Table 1. Web resources on Biodiversity Informatics with special reference to plants.

Database Name	**Description about the database (Organism)**	**URL**
Bioversity International	Bioversity is the world's largest international research organization dedicated solely to the conservation and use of agricultural biodiversity.	http://www.bioversityinternational.org/
Botanical Survey of India	Plant database - Three million accessioned specimens of plants	http://envfor.nic.in/bsi/
Plants of India	Plants– Plant family and genus information database	http://www.ecoinfoindia.org/pdb_query.php
FRLHT's Encyclopedia of Indian Medicinal Plants	Indian Medicinal Plants	http://www.frlht.org.in/meta/
Sahyadri: Western Ghats Biodiversity Information System	Western Ghats Biodiversity Information	http://wgbis.ces.iisc.ernet.in/biodiversity/
Botanic Gardens Conservation International (BGCI)	Plant diversity	http://www.bgci.org/worldwide/home/
ETI's World Biodiversity Database	Taxonomic database and information system, which converse a wide variety of organisms. Also has 20 species specific databases.	http://www.eti.uva.nl/tools/wbd.php

Flora of NW Europe	Vascular plants (pteridophytes, gymnosperms and angiosperms)	http://nlbif.eti.uva.nl/bis/flora.php
Ethnoforestry	Indigenous knowledge about forestry	http://www.iifm.ac.in/databank/ef/ethnoforestry.html
Sal Borer Problem in India	Shorea *robusta*- timber species	http://www.iifm.ac.in/databank/problems/salborer.html
Catalogue of the Benthic Marine Algae of the Indian Ocean	All published records of species and infraspecific taxa of benthic marine algae from the Indian Ocean.	http://ucjeps.berkeley.edu/rlmoe/tioc/ioctoc.html
CITES- Species Database	Species database, currently holds 7 million records of trade in wildlife and 50,000 scientific names of taxa.	http://www.cites.org/eng/resources/species.html
Integrated Taxonomic Information System (ITIS)	Taxonomic information on plants, animals, fungi, and microbes of North America and the world.	http://www.itis.gov/
Species 2000	Species database	http://www.species2000.org/
International Plant Name Index (IPNI)	Vascular plants	http://www.ipni.org/
Internet Directory of Botany	Plant names	http://www.botany.net/IDB/
BioCISE	European biological collections	http://www.bgbm.org/BioCise/
Global Biodiversity information facility	Biodiversity information	http://www.gbif.org/
BioCASE	Transnational network of biological collections of all kinds.	http://www.biocase.org/
BioLib (Biological Library)	International encyclopedia of plants, fungi and animals.	http://www.biolib.cz/en/main/
Barley DB	Barley germplasm	http://www.shigen.nig.ac.jp/barley/
Conservation Commons	Open access system to data, information, expertise and knowledge related to the conservation of biodiversity.	http://biodiversity.org/
Biodiversity Information System	Indian Biodiversity Information	http://www.bisindia.org/
Biodiversity Informatics Facility at the American Museum of Natural History's Center for Biodiversity and Conservation	Bioidiversity and Conservation Information	http://biodiversityinformatics.amnh.org/
Chinese Biodiversity Information System (CBIS)	Chinese Biodiversity Information	http://cbis.brim.ac.cn/cbise/index.html
Biodiversity Information System(Plants of Western Ghats)	Information about 5000 species of flowering plants	http://ces.iisc.ernet.in/hpg/cesmg/pew/bis.html
The Biotica Information System (BIOTICA)	The System designed to handle curatorial, nomenclatural, geographical, bibliographical and ecological data.	http://www.conabio.gob.mx/informacion/biotica_ingles/doctos/acerca_biotica.html
Biodiversity, Department of the Environment, Water, Heritage and the Arts.	Biodiversity portal of Australia	http://www.environment.gov.au/biodiversity/index.html
European Natural History Specimen Information Network(ENHSIN)	Specimen databases and resource sharing	http://www.nhm.ac.uk/research-curation/projects/ENHSIN/index.html
FloraBase	Floras Information	http://florabase.calm.wa.gov.au/
GrainGenes	Triticeae and Avena	http://wheat.pw.usda.gov/GG2/index.shtml
International Legume	Fabaceae (Leguminosae)	http://www.ildis.org/

Database & Information Service(ILDIS)		
International Organization for Plant Information (IOPI)	Plant taxonomic information	http://plantnet.rbgsyd.nsw.gov.au/iopi/iopihome.htm
Centre for Plant Biodiversity Research and Australian National Herbarium	Botanical Names and Biodiversity information	http://www.anbg.gov.au/cpbr/index.html
KOMUGI (Integrated wheat Science Database)	Wheat Science	http://www.shigen.nig.ac.jp/wheat/komugi/top/top.jsp
National Plant Germplasam system	Genetic diversity in plants	www.ars-grin.gov/npgs
NatureServe	Biodiversity and conservation information	http://www.natureserve.org/
OryzaBase	Rice Science	http://www.shigen.nig.ac.jp/rice/oryzabase/top/top.jsp
IABIN	The Inter-American Biodiversity Information Network	http://www.iabin-us.org/
The System-wide Information Network for Genetic Resources (SINGER)	Germplasm information exchange network of the Consultative Group on International Agricultural Research (CGIAR) and its partners.	http://www.singer.cgiar.org/
The *Arabidopsis* Information Resource (TAIR)	Genetic and molecular biology data for the model higher plant *Arabidopsis thaliana*.	http://www.arabidopsis.org/index.jsp
The Tree of Life (TOL)	Diversity of organisms on Earth, their evolutionary history (phylogeny), and characteristics.	http://www.tolweb.org/tree/

The Indian Institute of Spices Research, Calicut has played a phenomenal role in collecting and conserving the genetic resources of spices, which include cultivated, wild, hybrids and several endangered species. The national repository of spice germplasm maintained in *ex situ* and *in situ* conservatories are enriched regularly by undertaking collection surveys in primary and secondary centres of origin. Some of the valuable collections in the germplasm are endangered species like *Piper barberi* and *P. arboretum. P. silentvalleyensis, P. sugandhi* and *P. nigrum var. hirtellosum* are three new taxa identified and reported. *Vanilla anadamanica*; *P. colubrinum*, a source of resistance against *Phytophthora, pollu* beetle and *Radopholus similis*; multibranch types and natural *katte* resistant lines of cardamom; king cloves; putative wild types of ginger and high curcumin types of turmeric were also collected and conserved. The institute also conserves diverse strains of microorganisms associated with spices. Several information resources were developed pertaining to these germplasm collections and micro organisms associated with spices (Table 2).

Table 2. Biodiversity databases developed at IISR, Calicut

Biodiversity Database / Tools Name	**Crop / Organism**	**Availability**
Spice Genes I	Black pepper	Online and CD-ROM
Spice Genes II	Curcuma Species	CD-ROM
Spice Genes III	Nutmeg	CD-ROM
Piper base	*Piper* species	CD-ROM
PhyDish	*Phytophthora* spp.	Online

PIR databases	*Phytophthora* spp.	Online
PLASBID	Plant associated bacteria	Online

Black pepper

Black pepper (*Piper nigrum* L.) is one of the most important spices and India (Western Ghats) is the centre of origin and diversity for *Piper nigrum*. India leads in area and/ or production of this commodity. The Indian Institute of Spices Research, Kozhikode is the National Repository for black pepper germplasm and it has a collection of 1075 wild accessions and 1272 cultivar accessions in the gene bank. This comprises of about nineteen indigenous species and six exotic species besides more than 80 local cultivars. A portion of this germplasm is also maintained in the *in vitro* repository in the Institute. Wide variability occurs in the genus *Piper* with regard to morphological and reproductive characters like leaf size, shape, pubescence, nature of spike (size, colour, erect or pendent), fruit size etc. The germplasm at IISR, Kozhikode is characterized for about 50 characters based on IPGRI descriptor. The data on these characters are recorded and a database has been prepared using MS-Access –2000 programme. This database "Spice Genes I" is used for entering and retrieving the data. This is useful for researchers and all those who are involved or interested in *Piper*. A unique spike proliferating black pepper accession collected from a farmers nursery is being maintained in the pepper repository. The database is available on CD-ROM as well as online.

Another database, PiperBase on *Piper* species in India is also available on CD-ROM from the Institute. It includes botany, taxonomy, agronomy, biochemistry and medicinal properties of various *Piper* species present in India.

***Curcuma* species**

The genus *Curcuma* belonging to the family *Zingiberaceae* has a wide spread occurrence in the tropics of Asia to Africa and Australia. Apart from *Curcuma longa* or C. *domestica*, the common culinary turmeric, there are several other species of *Curcuma* which are mainly used as coloring agents, for production of arrowroot and for medicinal purposes. Species such as *C. aromatica* (Kasturi manjal), *C. caesia* (black turmeric) *C. amada*, *C. zedoaria*, *C. purpurescens*, *C. mangga*, *C. heyneana*, *C. xanthorrhiza*, *C.aeruginosa*, *C. phaeocaulis* and *C. petiolata* are also cultivated in different places and regions. The genus *Curcuma* consists of about 117 species whereas from India around 40 species are reported (Velayudhan *et al*., 1999). *Curcuma* species differ in floral characters, aerial morphology and under ground rhizome features (Valton, 1918; Velayudhan *et al.* 1999) besides the chemical traits. Spice Genes II, A database on *Curcuma* species has been prepared using Visual Basic and MS Access. The database is available on CD-ROM.

Nutmeg

Nutmeg belong to the family Myristicaceae with about 18 genera and 300 species of which genus *Myristica* is the most primitive (Sinclair, 1958). At present Myristicaceae is considered as a member of Magnolides or its taxonomic equivalents (Cronquist, 1981, Dajlgren, 1983). *Myristica fragrans* Houtt is the cultivated nutmeg tree and gives two spices – nutmeg and mace. The major nutmeg growing areas are Indonesia and Grenada and to a limited extent in Sri Lanka, India, China, Malaysia, Western Sumatra, Zanzibar, Mauritius and the Solomon Islands. Though *M. fragrans* is indigenous to the Banda islands in the Moluccas, many related species of *Myristica* are distributed from India and South East Asia to North America and the Pacific Islands. Sinclair (1958) listed a total of 72 species distributed in these areas. Nutmeg was introduced to Southern India in 18th century where it was naturalised. Nutmeg is dioecious with male and female flowers occurring on different trees. About 12 species of *Myristica* occurs in Indo Malayan region (Krishnamoorthy and Rema 1994). A collection of 465 accessions of *Myristica* including cultivated types and related species are conserved at IISR gene bank. The present day nutmeg population in India has evolved from a few sources trees introduced originally. The biodiversity information on *Myristica* germplasm accessions and related species has been organized as searchable database called Spice Genes III using MS Access and Visual Basic.

***Phytophthora* spp.**

Apart from these biodiversity databases, the institute is having germplasm of *Phytophthora*, an important pathogen of horticultural crops. *Phytophthora* root rot is a serious, widespread and difficult to control fungal disease affecting a wide range of plants across the world. *Phytophthora* species are mostly pathogens of dicotyledons, and are relatively host-specific parasites. Many species of *Phytophthora* are plant pathogens of considerable economic

importance. *Phytophthora infestans* was the infective agent of the potato blight that caused the Great Irish Famine (1845-1849). The foot rot of black pepper is caused by *P. capsici*. The Institute has developed two databases, which deals with *Phytophthora* information namely PIR and PhydisH. PIR database contains information on 64 species and an expert identification tool based on morphological characteristics of each species. Extensive information about literature collection and molecular information details are also included. The other database PhyDisH (*Phytophthora* Diseases of Horticulture Crops) deals with the major diseases caused by *Phytophthora* spp. on horticultural crops in India. One can find different species of *Phytophthora* affecting different crops and also their management methods. Information on 447 *Phytophthora* isolates belonging to nine species obtained from 30 different hosts that are maintained in the National Repository of *Phytophthora* (NARPH) is also made available in PhyDisH. It also integrates a literature database which contains abstracts of publications on *Phytophthora* from Pubmed database. This database is developed primarily from the information generated through a National Network Project on *Phytophthora* Diseases of Horticultural Crops (PHYTONET) funded by the Indian Council of Agricultural Research (ICAR), New Delhi. Both these databases are available online.

PLASBID

Bacteria associated with plants have been observed frequently to form assemblages referred to as aggregates, microcolonies, symplasmata, or biofilms on leaves and on root surfaces and within intercellular spaces of plant tissues. Bacteria associated with plants are diverse in their ability to affect plant health, their genotypic and phenotypic characteristics and their phylogeny. A database named PLASBID (Plant Associated Bacteria Identification system) has been developed by consolidating information available on bacteria reported to be associated with plants. The core information made available is nucleotide sequences belonging to 16S rRNA, primer information, G + C content, NCBI link and their taxonomy. It has a unique sequence similarity search facility using NCBI BLAST to find the similar sequence in the database with respect to query sequence. The sequence provided by the user is searched against PLASBID 16S rRNA nucleotide database and identifies the species based on their sequence similarity. Available primer information also has been compiled. Another feature of the database is Restriction Analyzer, which is designed to provide the information regarding restriction enzymes and their cutting sites based on the sequence entered by the user. The database is available online.

References

1. Berendsohn, W. G., Costello, M. J., Emblow, C., Güntsch, A., Hahn, A., Koenemann, J., Thomas, C., Thomson, N. & White, R. 2000: Concepts for a European Portal to Biological Collections. Pp. 59-70
2. Bisby, F.A., 1998: Putting names to things and keeping track: the Species 2000 programme for a coordinated catalogue of life. - Pp. 59-68 in: Bridge, P., Jeffries, P., Morse, D. R. & Scott, P. R. (ed.), Information technology, plant pathology & biodiversity. - Oxon, New York.
3. Cronquist A., 1981 An Integrated System of Classification of Flowering Plants. Columbia, University Press, New York.
4. Dajlgren R., 1983. General aspects of angiosperm evolution and macro sytematics. *Nordic Journal Botany* 3: 119-149.
5. Duckworth, W. D., Genoways, H. H. & Rose, C. L. 1993: Preserving natural science collections: Chronicle of our environmental heritage. - Washington.
6. Krishnamoorthy B and Rema J 1994, Genetic resources of Tree Spices. In. K.L Chadha and P Rethinam(eds) Advance in
7. Horticulture, Vol. 9. Plantation Crops and Spices. Malhotra Publishing House, New Delhi. Pp 169-192.
8. Roy, P. S. Sarnam Singh and Toxopeus, A. G. (eds). Indian Institute of Remote Sensing, Kalidas Road, Dehra Dun. 219 pp.
9. Sinclair J., 1958. A revision of the Malaysian Myristicaceae. *Singapore Gard. Bull*, 16:205-472.
10. Soberón, J., and A. T. Peterson. 2004. Biodiversity informatics: Managing and applying primary biodiversity data.
11. *Philosophical Transactions of the Royal Society of London B*, 359:689-698.
12. Velayudhan K.C., Muralidharan, V.K., Amalraj, V.A.,Gautham, P.L., Mandal, S and Dineshkumar 1999. *Curcuma* Genetic
13. Resources. Scientific Monograph, No. 4, National Bureau of Plant Genetic Resources, New Delhi, India, P-149.