# FROM TREES TO DESCRIPTIONS AND IDENTIFICATION TOOLS

**Barry J. Conn**
*National Herbarium of New South Wales*

**Kipiro Q. Damas**
*Papua New Guinea National Herbarium*

There are very few resources available to the timber industry that will enable the accurate identification of major tree species within Papua New Guinea. The unnecessary destruction of rare and otherwise valuable tree species has, in part, resulted from the inability to distinguish these species from the preferred timber species. Furthermore, the mixture of unwanted timber with that from preferred species has frequently resulted in the downgrading of all lumber to wood-chip or round logs.

The *Guide to Trees of Papua New Guinea* project ('PNGtrees') (Conn & Damas, 2006) is a long-term, multi-authored endeavour, with the main collaborators operating from different countries. This project is developing descriptions and interactive identification tools for the common trees of Papua New Guinea so that government and non-government agencies (particularly, foresters, loggers and environmental managers) are able to distinguish readily the important timber species from other non-timber trees. More than 400 tree species have been included in the first edition of PNGtrees.

Field observations and measurements—such as plant habit, bark and leaf features—are digitally captured and managed by a Microsoft Access® database that outputs the data in Descriptive Language for Taxonomy (DELTA) descriptive data format (Dallwitz, 2005). The descriptions are automatically generated and the interactive key is produced from DeltaAccess software (Hagedorn, 2005). New or revised information can also be added from remote sites, via a web interface to the PNGtrees descriptive database. Distribution maps for each species are produced directly from the PNGplants Collection's KE Texpress database (Anonymous, 2006).

## Introduction

One of the major concerns facing the people of Papua New Guinea is the capacity for them to document the rich biodiversity of their country. The documentation of the flora of PNG still relies heavily on the research efforts of scientists who are working outside of the country. However, even with many world experts studying the systematics of this region's flora, the documentation of the vascular plants remains very incomplete. Furthermore, the descriptions of the plants of PNG that have been produced are scattered throughout various publications, many of which are not readily available within the country. The *Handbooks of the Flora of Papua New Guinea* series attempted to document the flora of the region. However, only three volumes of this scholarly work have been published, with most contributions being from non-Papua New Guinean authors (Womersley, 1978; Henty, 1981; Conn, 1995). Although a complete coverage of the region's flora is urgently needed, the *Handbook* series demands a level of research expertise that is unavailable within the country. Therefore, the aim of the *Guide to Trees of Papua New Guinea* ('PNGtrees') project is to develop a simple, structured method for documenting the flora of Papua New Guinea. Since the Papua New Guinea National Herbarium (Lae) has limited resources, including IT capability, the system needs to be simple, being both easy to operate and easy to manage.

This project to document the trees of Papua New Guinea (Conn & Damas, 2006) began in March 2003, with supplementary support from *The Australia & Pacific Biological Foundation* (Filshie, 2006). The initial phase of the project concentrated on preparing an interactive identification tool for the commercial trees of the Morobe Province of Papua New Guinea. This was a three-year project to evaluate the feasibility of preparing a complete account of the trees occurring in the entire country. During 2006, the scope of this project was expanded to include many species that occur in other regions of Papua New Guinea.

The plant identification key being developed will be a tool that—in conjunction with the descriptions, diagrams, botanical illustrations and images—aims to assist users to correctly identify the commercial and common trees of the region. It is hoped that one of the most significant consequences of this interactive identification tool will be a reduction in the unnecessary destruction of forest species caused by incorrect plant identification.

## Software Tools

The DeltaAccess 1.9 software (Hagedorn, 2005) has been used to manage the plant descriptive data because this software is freely available and is based on Microsoft Access 2002® database software. DeltaAccess[1] manages the definition of plant features and other important characteristics. Therefore, it functions as a controlled vocabulary (glossary or data dictionary), which is important for multi-authored and long-term studies. Furthermore, DeltaAccess is able to output these data in DELTA (Descriptive Language for Taxonomy) format (Dallwitz, 2005), which makes these descriptive data widely available for other applications. The DELTA format has been adopted by the International Taxonomic Databases Working Group (TDWG, 2006) as a standard for data exchange. It is a flexible method for encoding taxonomic descriptions for computer processing and management. The DELTA format has been used extensively to produce natural-language descriptions, conventional or interactive electronic identification tools, phenetic or phylogenetic classifications, and information-retrieval systems.

The tree descriptions included in this project have been generated by DeltaAccess, with the output further modified using a perl script to include extra features. These features include a distribution map, specimen images and some aesthetic layout manipulation. Both the distribution map and the specimen images originate from the *Plant Collections from Papua New Guinea* ('PNGplants') database (Conn, Lee & Kiapranis, 2004). The perl script performs queries into this database and inserts the relevant data into the web-based (HTML) description pages.

The character-state values of each feature can be entered either directly into the DeltaAccess PNGtrees database, held at the National Herbarium of New South Wales (henceforth, 'NSW'), or else indirectly, via the web-based data sheets (Conn & Damas, 2006). Since the principal investigators are based in separate countries, this web-access functionality was recognised as an important feature for enabling both researchers to access the single dataset at NSW.

The PNGplants Collections database (Conn, Lee & Kiapranis, 2004) uses KE Texpress software, a single-table object-oriented UNIX database (Anonymous, 2006), held at NSW. The insertion of new records from the Papua New Guinea National Herbarium ('LAE') into this database, plus the ability to edit existing records, is managed through a Web interface. The exchange of data between the databases of participating herbaria—

namely, the Australian National Herbarium ('CANB'); the National Herbarium of Victoria ('MEL'); the National Herbarium of New South Wales ('NSW'); and the Queensland Herbarium ('BRI')—and the PNGplants database follows the protocols of the international data interchange standard HISPID (Conn, 1996; 2000). The interrogation and presentation of these data through the Plants of Papua New Guinea website also uses HISPID protocols. These collection data are also available through the international Global Biodiversity Information Facility (GBIF) (Lane, 2006).

## Controlled Vocabulary

A controlled vocabulary (glossary, data dictionary) is an established list of standardised terminology, which ensures that an object (in this project, a tree) will be described using the same preferred descriptors and, when relevant, appropriate modifiers. Like most classification systems, scientific botanical descriptive terminology is highly developed, with a strongly hierarchical structure so that the relationship between terms is provided. However, botanical terminology has developed over hundreds of years of scientific pursuit, with preferred definitions arising out of frequency of use of particular terms, rather than from the application of an agreed standardised methodology. Frequently, the same term is defined differently, or different terms are used to describe similar features in different plant groups. Since PNGtrees is designed to be a multi-authored project, with the next phase of the project inviting contributions from regional and international specialists, a controlled vocabulary—the data dictionary in PNGtrees (Conn & Damas, 2006)—was developed to ensure that scientific terms are understood and consistently applied. The data dictionary used in PNGtrees is based on the vocabularies used in several interactive identification tools that have been developed. In particular, the features used in PNGtrees were based on the definitions used by FloraBase (2005), Hyland and Whiffin (1993), Jarvie and Ermayanti (2005), Thiele and Adams (1999), and Webb, Jarvie, Schori, Rachman and Mayar (2005). Furthermore, where appropriate many of the internationally accepted data dictionary standards endorsed by TDWG (Berendsohn, 2005), have been used. At this stage, the data dictionary used here is not typical of most controlled vocabularies because it is not designed to be a tagging system that improves the accuracy of free text searching. With the addition of more tree species in future editions of PNGtrees, a controlled vocabulary to improve the accuracy of database searches would be

beneficial. The responsibility for the maintenance and development of the data dictionary rests with the principal authors, and staff at the Papua New Guinea National Herbarium.

To reduce potential confusion, rather than using a precise botanical descriptor each of the conditions or character states of many features was briefly described in simple terms wherever possible. This approach also had the advantage of being more easily understood by users who were not familiar with highly specialised botanical terminology. However, less precise terminology may not distinguish unambiguously between morphologically similar features.

## Definition of Characters

Out of a total of 147 characters, 125 are descriptive characters, used to provide descriptions of the tree species included in the PNGtrees project. The interactive identification key uses these characters to distinguish between the species. For a detailed discussion of these descriptive characters refer to the PNGtrees data dictionary section (Conn & Damas, 2006). The remaining 22 characters are non-descriptive and non-diagnostic. These include bibliographical information (characters 1–11), botanical classification (12–16), species distribution (143) and data management (142, 144–147). Some of the descriptive characters and their character states are relatively easy to define and apply (for example, the size of features, based on measurements). However, other features, such as bark aroma and flower (inner perianth) colour, are more difficult to define objectively, even though they are often useful diagnostic features. Users find it more difficult to interpret consistently these types of characters because the various character states are frequently very subjectively perceived.

## Data Gathering

Hand-held personal digital devices (PDAs) have not been used at this stage of the project. Nevertheless, PDAs that are sufficiently robust to withstand the extreme environmental conditions of the tropical forest would potentially offer significant benefits. At this stage, data sheets for current species or blank data sheets for new records are printed from the PNGtrees website, and the characters states are recorded manually (in pencil). Although a low-tech solution, this manual method has the advantage of being water-resistant, unlike many PDAs. Because the within-country IT support available to the project is minimal, the manual

method avoids an over-reliance on a technical solution that might be prone to failure. In any case, insufficient finances are available at this stage to enable trialling of the use of data-loggers to capture electronically the information in the field.

Only relatively low-resolution cameras have been available for the capture of digital images of field characters and other plant features. Most images were captured using a 0.8 Megapixel Sony MVC-FD81 (image size 1024 x 768 pixels), with the images stored to a 1.44MB floppy diskette; each diskette can store 15–20 images (using JPEG compression) and the battery lasts for between 50–60 images. Some images were captured with a 3.1 Megapixel Fujifilm FinePix A310 camera (images 2048 x 1536 pixels), with battery life of about 24 hours, and a 3.2 Megapixel Canon Powershot A400, with a 256 MB memory card that can store between 100–140 colour images (using JPEG compression).. The Canon Powershot camera is powered with a Kodak Rechargeable Digital Camera battery that can last for up to a week and half in the field. The superior images resulting from the recent opportunity of using a 6.1 Megapixel Nikon D70s (images 2240 x 1488 pixels) clearly demonstrated to us the urgent need to upgrade the digital cameras available for PNGtrees.

It is not practical or feasible to develop plant descriptions based on first-hand observations alone. All of the plant collections of a species, as held at various herbaria, contain information that can be used to summarise the morphological features and distributional range of that species. Furthermore, additional information can be extracted from the brief notes provided by the collectors. Two limiting factors in the usefulness of herbarium collections are: (1) impossibility of re-assessing observational errors of features not present in the collected material (such as tree height, which cannot be represented by an herbarium collection); and (2) mis-identification of collections. Mis-identifications frequently arise when the taxonomy of a species is inadequately known, the material collected is inadequate for a full identification, or the collections are inadequately curated because current scientific literature is unavailable. Since a project like PNGtrees must rely heavily on previously collected material and associated information, there is constant pressure to ensure that the collections at LAE are fully and accurately curated, and that this information is made available electronically in the PNGplants database.

## Limits to New Technologies

Although the web-based data sheets within PNGtrees enable remote access to the database, the Papua New Guinea collaborators have not been able to use this facility because their internet access is inadequate. To overcome this problem, the information recorded has been printed in the form of data sheets and then modified manually. This new information is then mailed to NSW and entered into PNGtrees. Not only is this method much slower than the direct electronic method, but there is also an increased chance of additional errors resulting from the re-keying of data.

To make the identification tools as useful as possible, the website was made as image-rich as possible. However, the low-speed internet access at LAE results in these web pages being unacceptably slow to load. In particular, the illustrated data dictionary, critical for ensuring the consistent usage of technical terms, is not referred to as often as might be desired because of slowness of internet access.

## Distribution of Information

Since the introduction of printing technology, the book has been an important medium for summarising advances in knowledge and for the distribution of information (UNESCO, 2006a). There is also a close link between the availability of books, scientific literature and other publications with community and economic development (UNESCO, 2006b). However,

> despite the phenomenal increase in world book production, the gains are not evenly shared and the bulk of the demand remains unfulfilled in the developing regions. Over 80 per cent of total book production remains scattered in about thirty-four industrialised countries, which represent only 30 per cent of the world's population. The area of scarcity is spread over all Africa, and Latin America, all Asia excepting Japan, and the Pacific Region excluding Australia and New Zealand (UNESCO, as cited in Evans 2006).

Although there are several initiatives that attempt to overcome the shortage of foreign currency that 'hinders the importation of books, publications and scientific material' (UNESCO, 2006c) into countries like Papua New Guinea (the UNESCO Coupon strategy is one example), the effectiveness of these initiatives is limited by several factors. Two of the most important impediments are the lack of local financial resources and a

lack of awareness of modern literature. However, with respect to descriptive information on trees of Papua New Guinea, the main reason for the unavailability of such publications is that this primary information is very incomplete and fragmentary. Therefore, perhaps the most important aim of the PNGtrees project is not only to publish information on the trees of the country, but also to undertake the research upon which this knowledge can be developed.

> There is fear that cost of books is slipping out of the reach of the ordinary person. Most developing countries lack the necessary purchasing power and the underlying book culture. (Evans, 2006)

Therefore, every effort needs to be made to ensure that the cost of the resultant PNGtrees publication(s) is within the purchasing power of all users who require this information. It had been hoped that a print-on-demand approach, which is currently being considered, together with some level of sponsorship, would reduce production costs sufficiently to make the publications affordable for within-country users. The presentation of this information through development of electronic resources such as the PNGplants website and a planned companion CD-ROM has not been as successful as had been hoped because of the large number of potential users with unreliable and/or inadequate internet access. Furthermore, electronic information is frequently completely inaccessible to the many regional people working in forest camps or with villages and small communities. These users usually do not have internet access and/or do not have computers. Therefore, the information must be provided in hard-copy printed format. Even here, the lack of an established national publication distribution network may make it difficult to get these publications to the users that are not based in the larger cities.

## Project Future

The medium-term scientific objective of PNGtrees is to document all Papua New Guinean tree species (at least 2,000 species) in the next three to five years, with the longer-term objective of using the same methodology to document the total vascular flora (possibly as many as 20,000 species) (Damas, 1998; see also references therein). There are several impediments limiting the potential success of the immediate project and the longer-term plant diversity documentation project. Firstly, to achieve these outcomes, every effort must be made through this and other initiatives to remove the taxonomic impediment (Darwin Declaration, 1998). The lack of taxonomic skills to understand the flora of

this very diversity-rich country is reducing Papua New Guinea's capacity for sound environmental management and conservation of biodiversity. The most cost-effective way of managing a multi-authored, long-term project like PNGtrees is to link all collaborators to the single descriptive database via a web-interface. However, current researchers at LAE do not have adequate internet access for them to be effectively electronically linked to either PNGtrees or to the associated PNGplants database. Therefore, a significant improvement in IT capacity and infrastructure is urgently required at LAE. The more difficult task of ensuring that all future within-country collaborators and users have adequate internet access will be challenging.

Currently, the PNGplants and PNGtrees databases are held and managed at NSW. The medium- to long-term aim is for these databases to be directly managed by LAE, irrespectively of where they are housed. For such a change to be sustainable, a significant investment in IT infrastructure (equipment and personnel) is required by the Papua New Guinea Forest Research Institute (National Forest Authority). The advantage of the current system is that the databases are adequately maintained and managed at NSW, with full digital archiving of these data, while LAE has full internet access to the most current information.

There are many potential synergies between different disciplines and the PNGtrees project. In particular, the incorporation of traditional uses and local nomenclature would enrich the current project significantly, making it more relevant to a broader group of users. Although LAE has a large card-index of traditional uses and local names for many of the trees of Papua New Guinea, it would require considerable editing by anthropologists, linguists and botanists before it would represent reliable information. It should be noted that the incorporation of traditional knowledge into PNGtrees would invoke intellectual, biological and cultural property issues that would need to be resolved prior to the release of this information.

Although the high cost of undertaking field work in the more inaccessible areas of Papua New Guinea make the completion of this project a challenge, a greater threat comes from outbreaks of civil unrest. Field work has already been cancelled or postponed in several regions because of civil unrest (for example, because of the declared state of emergency in the Southern Highlands province) even though tree species that are unknown in other parts of the country occur there.

## Conclusion

The implementation of the PNGtrees descriptive database assists staff at LAE—particularly relatively inexperienced, junior staff—to document the trees of Papua New Guinea according to a set of protocols and within the framework of an accepted controlled vocabulary (data dictionary). This database provides a structured environment in which staff can rapidly gain the highly advanced skills required to describe plant species succinctly and accurately. Approximately 145 species have been described each year of the current project. An increase in the number of collaborators would dramatically reduce the amount of time required to document the common trees of Papua New Guinea. At the present rate, it will take 10 to 11 years to document the common trees of PNG. It must be acknowledged that it will become increasingly difficult to complete the documentation process once the most common species are completed.

The distribution of the completed documentation to all potential within-country users remains a challenge irrespective of whether the knowledge is published as a printed book or provided electronically via the internet and/or CD-ROM. The advantages of the electronic medium include the low cost of publication and the ease with which the information can be maintained to reflect current taxonomic opinion.

## Acknowledgements

We wish to acknowledge Linn Linn Lee (formerly NSW) who assisted with the design and development of the websites *Plant Collections from Papua New Guinea* (Conn, Lee & Kiapranis, 2004) and *Guide to Trees of Papua New Guinea* (Conn & Damas, 2006). She also developed the PNGplants Collections database. Her excellent programming has enabled LAE to start data processing their important collections and to make these records available via the GBIF website (Lane, 2006). Gary Chapple (NSW) has consistently provided prompt database management advice. We thank Kaigube Fazang, Tory Kuri and Oliver Paul (all LAE) for providing considerable field assistance for the PNGtrees project, as well as assistance with the identification of plant material. Jessie Waibauru (Papua New Guinea Forest Research Institute) provided computer support for the PNGplants database. Balpina Tiki, Brenda Paul and Dubi Damas (all from LAE) have specifically data processed many herbarium records from Papua New Guinea so that these records are available for the PNGtrees project. We thank Billy Bau (LAE) for co-ordinating this data processing initiative. The PNGplants Collections database has been extensively used throughout this project to check the plant collector's notes for details of morphological features and to verify the distribution of the species within Papua New Guinea. Robert Kiapranis (formerly LAE), Roy Banka (LAE) and Nalish Sam (PNG Forest Research Institute) have consistently provided full managerial support to the above projects. We are extremely

grateful to the Australian herbaria (namely, CANB, BRI, MEL and NSW) that generously made their herbarium data available to the PNGplants project. The PNGtrees project was generously supported by The Australia & Pacific Biological Foundation (Filshie, 2006). Financial support for the repatriation of plant collection data from Australian herbaria to the Papua New Guinea National Herbarium was provided by the Global Biodiversity Information Facility (Lane, 2006). We thank the staff from the Forest Research Institute, Lae, and the Wildlife Conservation Society, Goroka, for providing an independent review of the PNTtrees online data dictionary.

## Endnotes

[1] The next version of DeltaAccess will be named DiversityDescriptions (Hagedorn, 2005).

## References

Anonymous. (2006). *KE Texpress*. Retrieved September 2006 from http://www.kesoftware.com/texpress/index.html

Berendsohn, W. (Ed.) (2005). *TDWG Subgroup on Biological Collection Data. Standards, Information Models, and Data Dictionaries for Biological Collections*. Retrieved October 2006 from http://www.bgbm.fu-berlin.de/TDWG/acc/Referenc.htm

Conn B.J. (Ed.). (1995). *Handbooks of the Flora of Papua New Guinea, vol. 3*. Carlton, VIC: Melbourne University Press.

Conn B.J. (Ed.). (1996). *HISPID3 - Herbarium Information Standards and Protocols for Interchange of Data, Version 3 (Royal Botanic Gardens: Sydney)*. Retrieved September 2006 from http://plantnet.rbgsyd.nsw.gov.au/HISCOM/HISPID/HISPID3/hispidright.html

Conn B.J. (Ed.). (2000). *HISPID4 - Herbarium Information Standards and Protocols for Interchange of Data, Version 4 (Royal Botanic Gardens: Sydney)*. Retrieved September 2006 from http://plantnet.rbgsyd.nsw.gov.au/Hispid4

Conn B.J. & Damas, K. (2006). *Guide to Trees of Papua New Guinea*. Retrieved September 2006 from http://www.pngplants.org/PNGplants/PNGtrees

Conn B.J., Lee L.L. & Kiapranis R. (2004). *Plant Collections from Papua New Guinea*. Retrieved September 2006 from http://www.pngplants.org/PNGplants/PNGdatabase

Dallwitz M.J. (2005, 13 September). *Overview of the DELTA System*. Retrieved August 2006 from http://delta-intkey.com/www/overview.htm

Damas, K.Q. (1998). The present status of plant conservation in Papua New Guinea. In C.-I Peng & P.P. Lowry II (Eds). *Rare, Threatened and Endangered Floras of Asia and the Pacific Rim* (pp. 171-179). Academia Sinica Monograph Series 16.

Darwin Declaration (1998). *Global Taxonomy Initiative – The Darwin Declaration*. Retrieved September 2006 from http://www.biodiv.org/programmes/cross-cutting/taxonomy/darwin-declaration.asp

Evans, J. (2006). *PNG Buai – Publications in Papua New Guinea*. Retrieved September 2006 from http://www.pngbuai.com/000general/publishers/papuanewguinea/pngpubs00.html

Filshie B.K. (2006). *The Australia & Pacific Science Foundation*. Retrieved October 2006 from http://apscience.org.au/background.html

FloraBase (2005). *The Western Australian Flora*. Retrieved September 2005 from http://florabase.calm.wa.gov.au

Hagedorn G. (2005, 30 March). *DiversityDescriptions (DeltaAccess) (version 1.9): Documentation of the information model*. Retrieved September 2006 from http://www.diversityworkbench.net/OldModels/Descriptions/Model/2005-03-30/DModelDD19.html

Henty, E.E. (Ed.). (1981). *Handbooks of the Flora of Papua New Guinea, vol. 2*. Carlton, VIC: Melbourne University Press.

Hyland, B.P.M. & Whiffin, T. (1993). *Australian Tropical Rain Forest Trees*. East Melbourne: CSIRO Publications.

Jarvie, J.K. & Ermayanti (2005). *The Tree and Shrub Genera of Borneo*. Retrieved September 2006 from http://www.phylodiversity.net/borneo/delta

Lane, M. (2006). *GBIF – Global Biodiversity Information Facility*. Retrieved September 2006 from http://www.gbif.org

TDWG (2006). *Taxonomic Database Working Group*. Retrieved October 2006 from http://www.nhm.ac.uk/hosted_sites/tdwg/index.html

Thiele, K.R. & Adams, L.G. (1999) *The Families of Flowering Plants of Australia, Version 1.0*. Collingwood, VIC: CSIRO Publishing

UNESCO (2006a). *UNESCO e-Book project*. Retrieved September 2006 from http://www.vidyanidhi.org.in/ebook

UNESCO (2006b). *Cultural Sector, UNESCO*. Retrieved September 2006 from http://portal.unesco.org/culture/en/...&URL_DO=DO_TOPIC&URL_SECTION=201.html

UNESCO (2006c). *UNESCO Coupons*. Retrieved September 2006 from http://www.unesco.org/general/eng/about/coupon/coupons.shtml

Webb, C. Jarvie, J.K. Schori, M. Rachman, I & Mayar, U. (2005). *Trees of Borneo*. Retrieved September 2006 from http://www.phylodiversity.net/borneo

Womersley, J.S. (Ed.). (1978). *Handbooks of the Flora of Papua New Guinea, vol. 1*. Carlton, VIC: Melbourne University Press.