Ant Identification in Cyberspace: Tools, Applications, and Challenges

Eli M. Sarnat

The Internet is precipitating the most sweeping revolution in insect identification since the microscope was invented nearly 400 years ago. Advances in digital media, new software tools, and increasingly sophisticated Web sites are actually making insect taxonomy fun. While suffering through 199 couplets of a beetle key may serve as a badge of honor for Introductory Entomology students, few relish the memory. However, jumping into an interactive multimedia key—checking boxes, skipping troublesome questions, clicking on links and pictures, and watching the list of remaining species shrink to one—that's more video gaming than Latin lesson.

New Web-based identification tools are gaining particular popularity in quarantine and pest management communities, where non-specialists are required to make species-level diagnoses for a wide array of insects (Norton and Taylor 2010). Introduced ants, for example, are a group of serious quarantine concern (Lach and Hooper-Bùi 2010), and receive a disproportionate number of identification queries relative to their diversity (Table 1). Invasive ants are causing irreparable harm to natural and agricultural systems across the across the United States, and dozens of exotic ant species are intercepted every year. New species, such as the Caribbean crazy ant (whose taxonomic name remains unknown) continue to grab public attention as they cross national and state borders.

If future invasive ant incursions are to be prevented, non-specialists need better taxonomic tools for species-level diagnosis. This paper discusses how Lucid keys, AntWeb, and the Encyclopedia of Life (EOL) are integrated into an interactive identification guide to over 100 non-native ant species of quarantine concern to North America. The power of these three Web tools is their ability to transform traditional taxonomic information into standardized data that can be imported, exported, and shared across multiple Web sites. Although my decision to use Lucid was based on prior work with the software, and my decision to use AntWeb and EOL was based on recommendations for the ant community to coordinate content delivery to the Web (Kautz and Moreau 2010), many alternatives are available for both interactive keys (Walter and Winterton 2007) and myrmecological Web sites (Klingenberg and Verhaagh 2005; Nash, 2005).

Table 1. Number and percentage of introduced ant species within the top-ranked species pages accessed on AntWeb from October 1, 2010–October 1, 2011. Ranking is based on unique page views reported by Google Analytics from querying "content > content by title" and filtering the resulting pages for those containing "species:" in the title.

Rank	Number introduced	Percent introduced
Top 10	6	60%
Top 20	12	60%
Top 50	18	36%
Top 100	24	24%
Top 200	41	21%
Top 250	47	19%

"Keys are compiled by those who do not need them for those who cannot use them." (Lobanov 2003)

Though somewhat stale, dichotomous keys have been the bread and butter of insect identification for hundreds of years. These traditional keys are cumbersome because each has a single entry point, and only one set of choices produces a correct identification (Walter and Winterton 2007). A new generation of matrix-based, interactive keys solves this problem by allowing the user to start at multiple entry points and then choose from an unfolding set of characters until a diagnosis is complete. In addition to supporting multimedia such as links and digital images, many interactive keys track which characters have been selected, which taxa have been eliminated, and which taxa remain (Fig. 1).

Lucid Keys

The key for the *Introduced Ants of North America* project (the "ant key") is built using Lucid3 software (http://www.lucidcentral.org) and modeled after the *Pacific Invasive Ant Key* (http://www.piakey.com) (Sarnat, 2008). PIAkey was designed to help quarantine personnel identify 44 ant species introduced to the Pacific Islands. Because Lucid3 is database-driven, it was possible to import all relevant taxa, characters, and matrix scores from PIAkey directly into the North American key. A common misconception about Lucid keys is that each taxon must be scored for each character. The ant key, for example, refrains from scoring genera and subfamilies. These higher-taxon groups act as "containers," which are defined exclusively by the scores of the species nested within them. Another practice is to score only taxa for which the given character is informative, and use the "unscoped" feature on taxa for which the character is uninformative (Fig. 2). When properly applied, the unscoped feature causes character choices to appear *only* when all remaining taxa are scored for those characters. Keys designed to unfold in this way protect users from being overwhelmed by an unmanageable number of character choices, of which only a subset are informative for the taxon of interest.

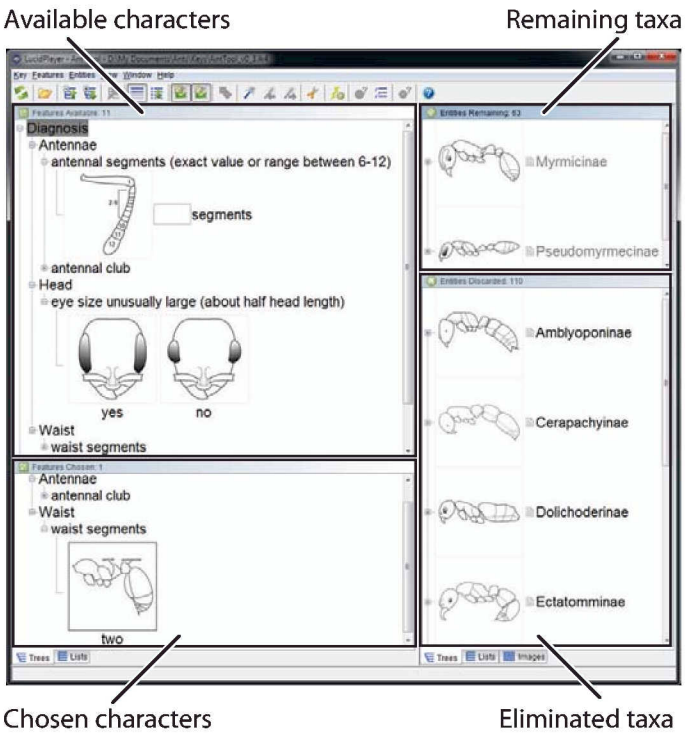

Fig. 1. Screenshot of the *Introduced Ants of North America* Lucid key as it appears in the Lucid Player application. The multiple windows of the graphical interface help the user keep track of characters and taxa during the identification process.

The ant key is also built to allow users access from multiple entry points. For example, treating the container taxa as characters permits users to significantly reduce the number of key steps by selecting the subfamily or genus of their species. Another Lucid strategy worth mentioning is the "best" button. This feature uses a software algorithm to determine which available characters will best divide the remaining taxa into approximately equal groups, thus leading the user down the most parsimonious pathway. The "best" and "next best" buttons can also help matrix builders design both efficiency and redundancy into their keys.

What a Picture is Worth

Perhaps the greatest strength of Lucid keys is their capacity for integrating media. Multiple images and links can be attached to characters and taxa. The ant key uses line drawings to illustrate general character states and specimen photographs to illustrate how those states are expressed by different species (Fig. 3). For example, while most ant species can be coded as spines present/absent, there are some species with minute spines that can arguably be coded either way. Rather than leave the user to guess "present" or "absent," the key builder can decide how to code them and then attach an image of the ambiguous morphology to the corresponding state.

Species Pages

The pillar of Web-based identification is the species page. Species pages are often used in conjunction with identification keys, or even replace keys altogether. Rather than engineer static HTML species pages from scratch, as was done with PIAkey, the *Introduced Ants of North America* project links to dynamic species pages developed by the Encyclopedia of Life (http://eol.org). EOL facilitates the separation of content development from content management so that biologists of all backgrounds can contribute their scientific knowledge to the Web without the responsibility of site design and maintenance. An advantage of separating content from style was illustrated when EOL recently transitioned from the overwhelming "everything on one page" format to a cleaner, more modern tab-based format. Despite major changes made to the site's styling and community features, the content associated with all 753,000 taxa were moved to their

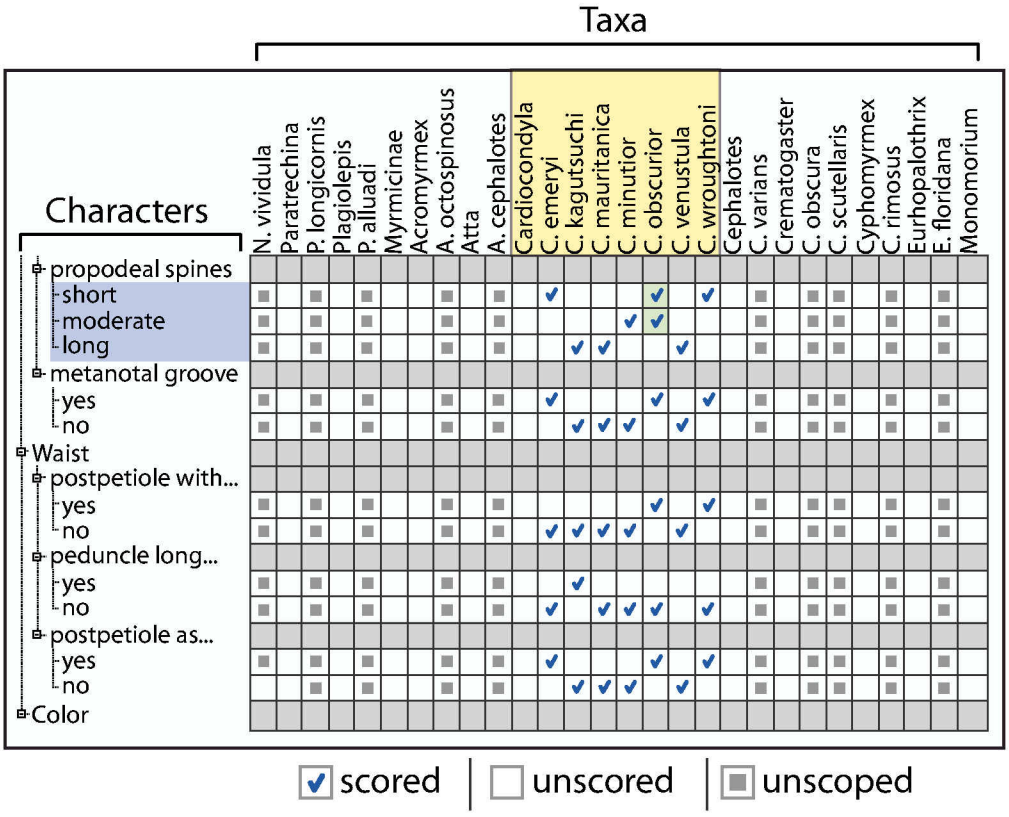

Fig. 2. Lucid3 score matrix illustrating strategies for well-structured keys (adapted from a portion of the Lucid key to *Introduced Ants of North America*). All of the characters shown here are only informative for diagnosing the seven *Cardiocondyla* species (highlighted in yellow). Species for which these characters are uninformative are coded as "unscoped." Genera and subfamilies are left unscored, allowing them to act as "container" taxa. Characters can have more than two states, as exemplified by the three "propodeal spines" states (highlighted in blue). Taxa are allowed to be scored for more than one character state, as exemplified by *C. obscurior*, which is scored for propodeal spines "short" and "long" (highlighted in green).

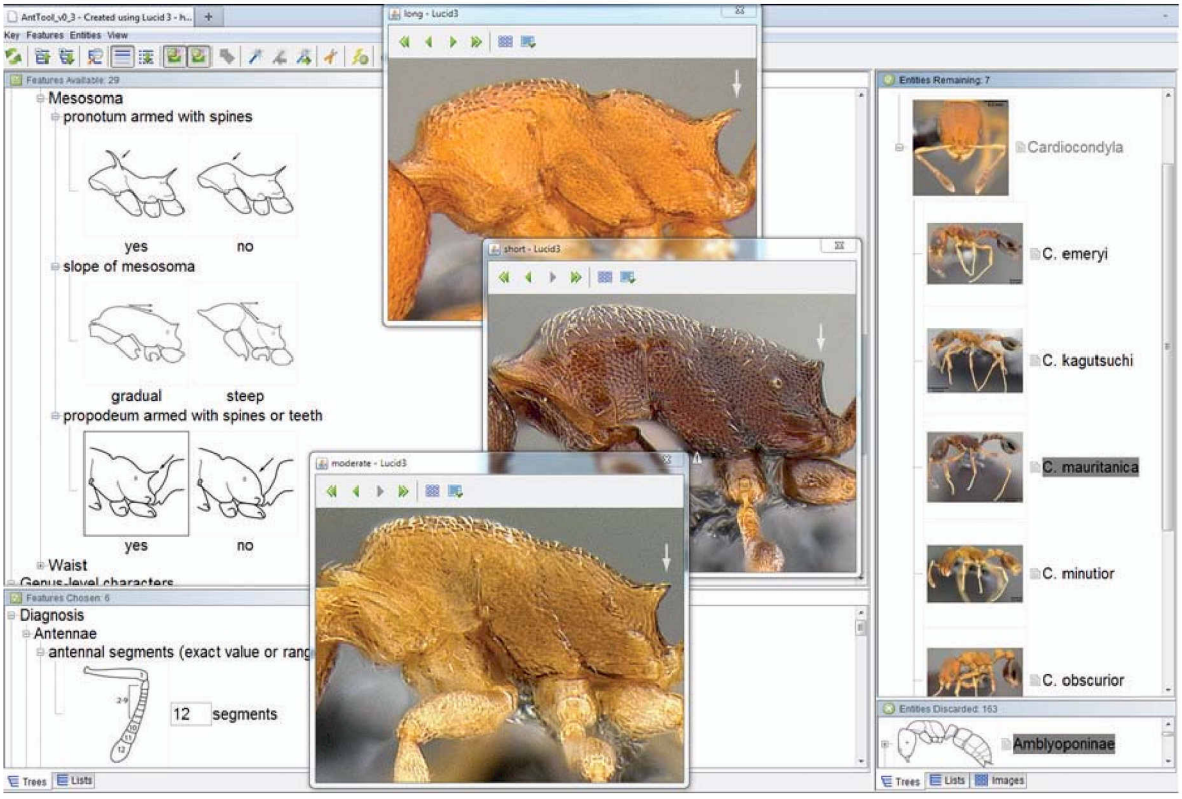

Fig. 3. Screenshot of the Introduced Ants of North America Lucid key demonstrating the use of both line drawings and specimen photographs for character states. The three specimen photographs are illustrating examples of the three character states available for *Cardiocondyla* propodeal spines.

respective tabs by executing an elegant bit of coding programmed by the EOL engineers.

Analogous rearrangement of PIAkey's content (which was not designed on a database foundation) must be done manually, so it is unlikely that the site will adapt as gracefully to the Internet's evolving environment. EOL also has a dynamic process for handling changes to taxonomic names based on digital classifications. For example, when dozens of species in the ant genus *Paratrechina* were recently moved to *Nylanderia*, all of the content associated with the former names was associated with the new names, and queries in EOL for either name now return the same species page.

Data Sharing

Although EOL shows great promise as an online resource for the study and appreciation of all life on earth, many taxonomic communities have already invested in online resources tailored to their members' needs. For example, AntWeb received ~77,000 unique page views during September 2011, of which 16,834 were species pages. By contrast, EOL received only 2,709 unique views of ant species pages during the same period. Rather than compete with AntWeb and other taxon-specific Web resources, EOL is implementing content partnerships whereby satellite Web sites allow their data to be harvested by EOL and dynamically imported to the appropriate EOL page. Serving 14,762 taxa, 4,151 articles and 40,017 images, AntWeb is among EOL's most productive content partners. The partnership allows species pages for *Introduced Ants of North America* to be developed directly on AntWeb, which receives the majority of myrmecological traffic. The AntWeb pages are then served to EOL and aggregated with content harvested from other trusted sites or provided directly by EOL curators. As an EOL curator, I can add additional media (such as video clips), organize content, and add species to EOL "collections" and "communities" where the species pages can be browsed and discussed by quarantine personnel or any other interested person with an Internet connection.

Specimens vs. Names

The partnership between EOL and AntWeb serves as a great model for entomologists developing Web-based identification tools. Whereas EOL is analogous to a traditional encyclopedia, AntWeb is more similar to a traditional specimen-based insect collection. For example, images and collection data of ant specimens intercepted at ports of entry across the United States are uploaded to AntWeb as part of the *Introduced Ants* project. Quarantine staff from anywhere in the world can view and compare these online specimens as if they were browsing through a physical museum-quality reference collection. Another critical feature of specimen-based Web resources is their capacity for making type specimens—the ultimate arbitrator for any identification—available online. Unlike EOL, which uses unique taxon identifiers as the primary key for its database, AntWeb uses unique specimen identifiers. Thus, when the white-footed ant (*Technomyrmex albipes*) was recently split into three separate species, it was possible for AntWeb to associate the proper images with the proper names to the extent that the databased specimens were identified correctly. Had the images been associated with the name, as is true for EOL, there would be no *a priori* way of knowing which images to associate with which species.

Everyone Wants a Web Site

What about the entomologists who don't belong to a larger online community or don't have programming skills, but still want to create dynamic Web sites for their research taxa? Scratchpads (http://scratchpads.eu) is emerging as a powerful, user-friendly application

that enables researchers to manage, share, and even publish taxonomic data online. Scratchpad Web sites are built around a taxonomy either supplied by the users or imported from EOL. Custom templates and modules are used to add content (e.g., specimen data, literature, images, etc.) that is coded for easy export to EOL. Other modules can be used to import content to the Web site from databases such as Genbank, Morphbank, GBIF, Biodiversity Heritage Library, Google Scholar, and Wikipedia. Because Scratchpad is based on Drupal (an open-source content management system), an active developer community is continuing to update existing modules and to add new ones, such as the matrix key module and the phylogeny module.

New Tools, New Challenges

For all of the opportunities promised by Web-based insect taxonomy, the burgeoning field is not without its challenges. As cybertaxonomy evolves, a primary goal is to move towards information centralization without losing the innovation of independent developers. EOL's content partner model is one solution to this challenge. This content partner model will succeed to the extent that independent projects shift towards database-powered Web sites compliant with emerging biodiversity data standards. However, few incentives exist to draw researchers into taxonomy's new frontier. In a profession that deals in the hard currency of high-impact taxonomy, innovation trades at a miserably low exchange rate. New Web-based journals like *Zookeys*, however, are investing in taxonomic pioneers and advancing the taxonomic revolution by building a 21st-century publishing platform capable of delivering interactive keys and interfacing with new applications like Scratchpads (Penev et al. 2009; Smith et al. 2009).

Exciting advances in Web-based insect identification make it easy to forget that increasing taxonomic literacy requires making identification guides that people will actually use. As bright a future as interactive matrix keys might have, the printed dichotomous key is an elegant tool that has survived hundreds of years, requires no computer or Internet connection, and continues to be used by the vast majority of biologists. Even as we march towards the brave new world, let's not forget the simple functionality of the old one.

Literature Cited

Kautz, S. and C.S. Moreau. 2010. Creating Encyclopedia of Life's species pages for ants (Hymenoptera: Formicidae): What we have done and what remains to be done. Myrmecological News 14: 69-72.

Klingenberg, C. and M. Verhaagh. 2005. Myrmecology in the Internet: Possibilities of information gathering (Hymenoptera, Formicidae). Beiträge zur Entomologie 55: 485-498.

Lach, L. and L.M. Hooper-Bùi. 2010. Consequences of ant invasions, pp. 261-286. In: Lach, L., Parr, C. & Abbott, K. (Eds.) Ant Ecology. Oxford University Press, Oxford, UK.

Lobanov, A. 2003. Keys to beetles and biological diagnostics. http://tinyurl.com/6k93r2c.

Nash, D.R. 2005. Web resources for myrmecologists. Myrmecologische Nachrichten 7: 95-102.

Norton, G. and M. Taylor. 2010. What pest is that? Recent developments in digital pest diagnostics. Outlooks on Pest Management 21: 236-238.

Penev, L., M. Sharkey, T. Erwin, S. van Noort, M. Buffington, K. Seltmann, N. Johnson, M. Taylor, F.C. Thompson, and M.J. Dallwitz. 2009. Data publication and dissemination of interactive keys under the open access model ZooKeys working example. Zookeys 1-17.

Sarnat, E.M. 2008. PIAkey: Identification guide to ants of the Pacific Islands, Edition 2.0, Lucid v. 3.4. USDA/APHIS/PPQ Center for Plant Health Science and Technology and University of California–Davis. 8 October 2011. http://tinyurl.com/67cnfzu.

Smith, V.S., S.D. Rycroft, K.T. Harman, B. Scott, and D. Roberts. 2009. Scratchpads: a data-publishing framework to build, share and manage information on the diversity of life. BMC Bioinformatics 10: Suppl. 14:S6.

Walter, D.E. and S. Winterton. 2007. Keys and the crisis in taxonomy: extinction or reinvention? Annual Review of Entomology 52: 193-208.

Dr. Eli Sarnat is a postdoctoral research associate at University of Illinois, Urbana-Champaign. In addition to developing taxonomic tools for the identification of introduced ants, his research interests include the systematics, biogeography and conservation of the Pacific Island ant fauna.