

Volume 52, 2007

Keys and the Crisis in Taxonomy: Extinction or Reinvention?*

David Evans Walter¹ and Shaun Winterton²

¹Department of Bioagricultural Sciences & Pest Management, Colorado State University, Fort Collins, Colorado 80523, and Department of Biological Sciences, University of Alberta, Edmonton AB T6G 2E9, Canada; email: dew@ualberta.ca

²Plant Pest Diagnostics Branch, California Department of Food & Agriculture, Sacramento, California 95832-1448; email: SWinterton@cdfa.ca.gov

Annu. Rev. Entomol. 2007. 52:193-208

First published online as a Review in Advance on August 16, 2006

The Annual Review of Entomology is online at ento.annualreviews.org

This article's doi: 10.1146/annurev.ento.51.110104.151054

Copyright © 2007 by Annual Reviews. All rights reserved

0066-4170/07/0107-0193\$20.00

*The U.S. Government has the right to retain a nonexclusive, royalty-free license in and to any copyright covering this paper.

Key Words

computer diagnostics, dichotomous, multiple entry, computer interactive, databases, taxonomic impediment

Abstract

Dichotomous keys that follow a single pathway of character state choices to an end point have been the primary tools for the identification of unknown organisms for more than two centuries. However, a revolution in computer diagnostics is now under way that may result in the replacement of traditional keys by matrix-based computer interactive keys that have many paths to a correct identification and make extensive use of hypertext to link to images, glossaries, and other support material. Progress is also being made on replacing keys entirely by optical matching of specimens to digital databases and DNA sequences. These new tools may go some way toward alleviating the taxonomic impediment to biodiversity studies and other ecological and evolutionary research, especially with better coordination between those who produce keys and those who use them and by integrating interactive keys into larger biological Web sites.

INTRODUCTION

HTML: hypertext markup language

We need well-written taxonomic keys based on morphological characters.... (16)

Keys are compiled by those who do not need them for those who cannot use them. (23)

In recent years, several editorials lamenting the decline in taxonomic expertise and its impact on the rate of production of species descriptions and phylogenies have appeared in prominent general scientific journals (6, 15, 47). Perhaps the least appreciated duty of taxonomists, however, is the construction of tools that allow others to identify the members of a flora or fauna. One reason this important service may be underappreciated is summarized in Lobanov's (23) epigram above: A culture gap often exists between the generator and the end user of a key. Herein we review the history of taxonomic identification, point out advantages and disadvantages of both traditional and more recent approaches to identification, and offer some hypotheses about future needs and trends in the science of arthropod identification.

Science and its funding bodies revere what is new. In the case of systematics, a rough approximation of the current rank order of reverence is new technologies, new phylogenies, and new taxa. This hierarchy is clear from the pages of major scientific journals: New phylogenies highlighting technological innovation are relatively common, but descriptions of new taxa are rare (and almost invariably fossils). For example, we have found only four exceptions for newly described taxa of insects (none for other living arthropods) in the last 20 years of Science and Nature: an unusual physiology (22), a new and major pest (demonstrated by a battery of new technologies) (37), a new order (21), and a correction for a nomen nudum in a report on a snail-eating caterpillar (39). These exceptions prove the rule: The nuts and bolts of taxonomy, i.e., the description of new species, placement in a classification, and production of identification tools (alpha-taxonomy) (28), is not the purview of journals with high Institute for Scientific Information impact factors. Thus,

nomic work are unlikely to thrive in universities or other research institutions where the worth of a publication is increasingly valued by the impact factor of the journal in which it is published. This is a major stumbling block to relieving the taxonomic impediment (i.e., the worldwide shortage of taxonomists and their declining rate of replacement) that inhibits biodiversity research (19).

Alpha-taxonomy may have been displaced to the bottom of the systematic alphabet, but the ability to identify species is a prerequisite for most biological sciences (16), is critical in pest management and quarantine, and in some cases is now required by legislation covering conservation surveys and international transfer of specimens (25). A simple example of the increasing interest in identification tools can be found by using the Internet search engine Google (default settings) on two dates less than a year apart (15 March 2005 and 6 March 2006) and the exact phrases "identification key" (45,000 hits versus 149,00 hits) and "insect key" (1120 hits versus 20,100 hits). Increasingly, the World Wide Web is awash with identification tools, often traditional dichotomous keys with simple hypertext markup language (HTML) links, but also with many examples of new computerbased diagnostics that are revolutionizing biological identification (see Table 1). The Biological Survey of Canada Journal of Arthropod Identification is a new Web-based journal dedicated to the publication of these keys (http://www.biology.ualberta. ca/bsc/ejournal/ejournal.html) (27).

SOLVING THE MYSTERY: WHAT IS THAT BUG?

For many taxonomic specialists, the identification of a specimen may require no more than a brief check of critical characters. Such seemingly quick and easy identifications are what the end users of taxonomic identification tools would like to acquire, but how can this taxonomic knowledge best be shared?

Table 1 Some matrix-based interactive keys available on the Internet (as of March 28, 2006)^a

Interactive key	URLb	
3Iq	http://ctap.inhs.uiuc.edu/dmitriev/3i.asp	
20	http://www.discoverlife.org/	
Lucid ^c	http://www.lucidcentral.org/	
INTkey ^d	http://delta-intkey.com/www/overview.htm	
Navikey ^d	http://www.navikey.net/	
Pollyclave ^d	http://prod.library.utoronto.ca:8090/polyclave/	
X:ID	http://www.ubio.org/services/key.html	
$\mathrm{XID^d}$	http://www.xidservices.com/	
SLIKS	http://www.stingersplace.com/SLIKS/	
ACTkey	http://www.efloras.org/index.aspx	
Visual Key	http://insects.ummz.lsa.umich.edu/beemites/vk_bees/vk_bees.htm	
MEKA ^d	http://ucjeps.berkeley.edu/meacham/meka/	
PANKEY ^d	http://www.exetersoftware.com/cat/pankey/pankey.html	
Linnaeus ∏ ^d	http://www.eti.uva.nl/	
PalmKey	http://www.phylodiversity.net/palmkey/	
Webkey-X	http://www.zin.ru/projects/WebKey-X/	
Pickey	http://www.zin.ru/projects/pickey/	
Lysandra	http://www.bonk.ru/lysandra/english/	
PHPKey	http://www.borealis.nu/exjobb/Index_en.html	
Taxis ^d	http://www.bio-tools.net/index.htm	
Taxy ^d	http://www.collectivesource.com/taxy/taxy.html	
XPER/Xper ²	http://lis.snv.jussieu.fr/apps/xper2/	

^aApologies for any URLs we have missed or that have been posted since we wrote this paper.

Identification by Matching: From Type Specimen to Automated Identification

Perhaps the least abstract approach to the identification of a biological specimen is visual comparison to an existing named specimen or to a written description of a particular taxon. This approach is the logical basis of the type concept in taxonomy, i.e., that a species name must be anchored to a designated specimen and differentiated from other species by diagnostic characters. In our experience, visual comparison to a reference collection is often the first choice of a taxonomic specialist faced with an unknown. The matching or comparison approach is feasible when identified specimens (e.g., museum collections) and diagnoses are available and the

number of taxa is manageable, but quickly becomes intractable for any diverse taxonomic group.

Matching can be enhanced by the use of illustrations that highlight diagnostic characters. This approach is used in the popular field guides to birds, other vertebrates, common plants, and butterflies. Similarly, relatively comprehensive field guides to insects have appeared that make extensive use of photobased identification (e.g., the 4000+ color photographs and 28 pictorial keys in Reference 26). Innovations in digital photography, computer graphics, and electronic publishing are leading to the increased use of photographic images in systematic papers in entomological journals as well. For example, although the mean number of taxonomic

^bAll URLs are browser enabled unless otherwise indicated (e.g., Java, XML).

^cStand alone or browser enabled in later versions.

^dStand alone; requires installation of software.

Couplet: a set of contrasting characters (usually a pair) used in a pathway key illustrations per page in the Systematics section of the Annals of the Entomological Society of America (molecular-only papers excluded) in 1995 and 2005 do not differ (2.2 \pm 1.4 versus 2.4 \pm 1.4 SE), there are more than twice as many photographic images in 2005 (1.0 \pm 0.2 per page) than in 1995 (0.4 \pm 0.2; t = -2.195, df = 43, P < 0.034).

Another approach to matching involves passing on the burden of comparison to a computer. Early efforts in the field measured light reflected off wings in flight to estimate wing-beat frequencies and identified targeted mosquito species and distinguished between male and female with a high rate of accuracy (84%) (32, 33). More recent approaches have used digital imaging to train programs to discriminate among a variety of insects, primarily based on wing characteristics, especially Hymenoptera (1, 46, 51) and Lepidoptera (45). One of these systems (digital automated identification system, or DAISY) has been trained and then field-tested against 35 species of Macrolepidoptera out of a digital database for 237 locally light-trapped species. Correct identification of species by DAISY ranged from 35% to 100% (mean = 83%). Although lighting used during imaging, loss of scales, and other damage could account for many of the misidentifications, other problems seemed inexplicable except as inherent differences between human and computer image recognition (45). Advances in optical imaging technology may one day make identification keys obsolete. There are similar hopes for the use of DNA matching (see Molecular Diagnostics, below). At the moment, however, most researchers challenged with identifying a specimen must use a printed taxonomic key—if one is available.

Dichotomous Keys: Origin and Structure

The metaphorical use of the word "key" in English as a solution to something mysterious is long-standing (8). Its special use to mean a written method of identification dates from the mid-1800s, probably from Gray (17) in 1848. However, the logical structure underlying the differentiation of organisms by the use of contrasting statements is apparent in the works of Aristotle (3) and in the bracketed diagrams summarizing classifications that were published by naturalists in the 1600s (44). It was Lamarck in 1778 who realized the power of sets of two contrasting statements about characters for teaching plant identification and who formalized the rules for constructing dichotomous keys (44).

In the pure dichotomous (also dichotomic) form, two sets of contrasting characters called leads (also legs or lugs) form each couplet; the choice of one lead takes a user to an end point (identification) or a pointer to the next couplet. The most common style in the North American entomological literature is called parallel or bracketed because contrasting leads are grouped together (Figure 1). A parallel key is easily converted into a flowchart key by the use of arrows to connect couplets. Flowchart keys often incorporate illustrations of characters within the couplets, a technique that makes a key easier to use and probably more likely to result in a correct identification. In the yoked style, common in the entomological literature of Eastern Europe, the contrasting leads are split so that subsequent dependent leads follow one another in a (usually indented) series.

Another way to view dichotomous keys is as pathways of character state choices (steps) that lead to identifications. This logical structure is both the strongest pillar of traditional keys and one of their greatest flaws. A wellconstructed key limits the number of steps to an identification. Because only one path of logical choices results in a correct identification, each attempt must begin at the same point (i.e., the keys are monoentry) and follow the correct pathway: Any deviation from the one true path results in a failed identification (dead end) or misidentification. Furthermore, if a critical character is obscured or missing at any step along this pathway, the choice required by the couplet is unanswerable (e.g., for

Couplet	Character	to photo	next couplet
1	Antennae absent	c, e	2
	Antennae present	arrow a, d	3
2 (1)	Body distinctly segmented, 3 pairs of legs	е	Protura
	Body not segmented, 4 pairs of legs	С	Acari
3 (1)	Antennae with distal club	а	Coleoptera
	Antennae without club	d	Collembola

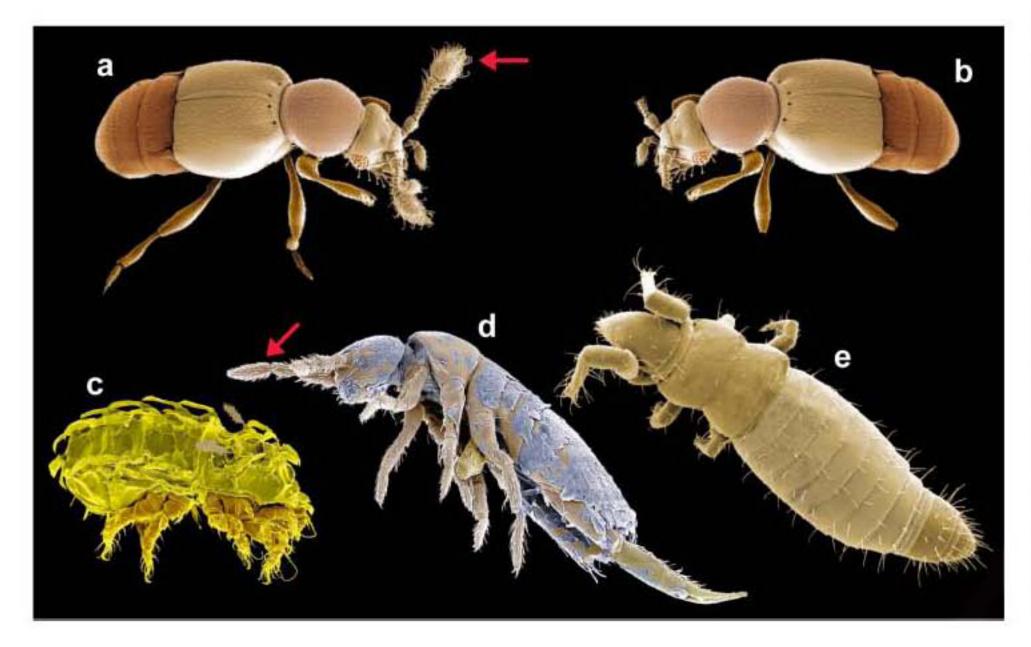


Figure 1

ID or

A dichotomous pathway key in parallel format to four microarthropods from an Australian rainforest. Each leading number defines a couplet of two leads, and each lead ends in identification or a number indicating the next couplet to try. This simple but powerful logical tool allows rapid and accurate identification of four (a, c-e) of the microarthropods in the figure, but results in an unanswerable Couplet 3 for the beetle in panel b. (a) Coleoptera intact, (b) Coleoptera with legs and antennae broken, (c) Acari, (d) Collembola, and (e) Protura (all images by D.E. Walter).

the damaged beetle in Figure 1b, Couplet 3). The more taxa in a key, the longer the average number of steps to an answer, and the more likely the user of the key strays from the path.

Pathway Keys: Problems and Ameliorations

Unanswerable couplets, dead ends, and the momentary distractions that can cause a user to forget his or her position in a key are manageable when using short keys (every couplet can be tried), but students of entomology are often confronted by some truly intimidating keys. For example, in *An Introduction to the Study of Insects* (7) the key to the families of beetles runs for 245 couplets. Although working with such keys undoubtedly builds character, those approaching them outside of a structured learning environment have good reason to look for alternatives.

Match

Pathway keys can be made more user-friendly. For example, in Borror and De Longs's Introduction to the Study of Insects (43), the beetle key has been reduced to 196 couplets that are arrayed in alternating gray and white bands that help a reader retain his or her place in the key. Other ameliorations common in many written keys include providing links to

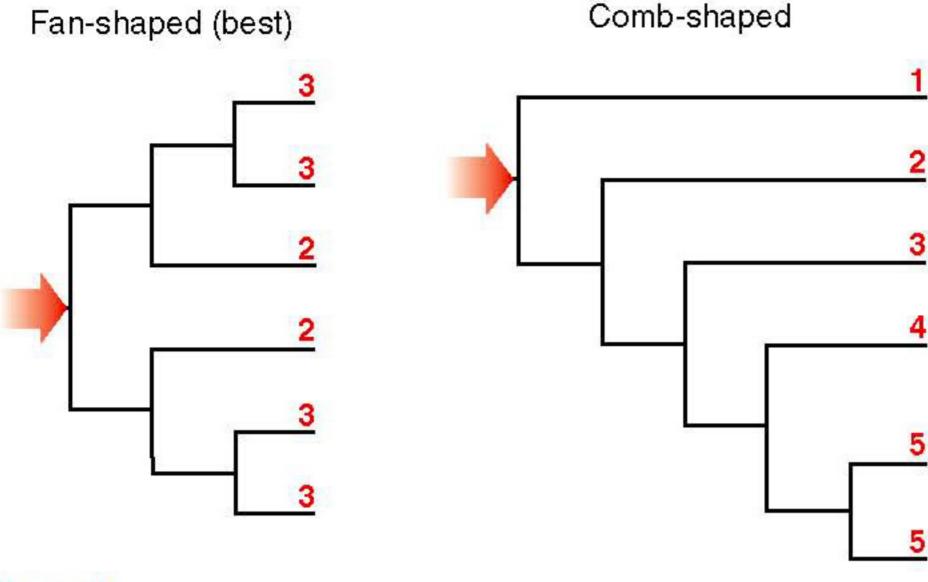


Figure 2

Diagrammatic illustration of the number of steps (red numerals) required to reach the end points (taxon identifications) for a fan-shaped and a comb-shaped dichotomous key to six taxa. Identification of all six taxa requires 20 steps in the comb-shaped key versus 16 in the fan-shaped key; half of the taxa in the comb-shaped key require more steps to an identification than any of the taxa in the fan-shaped key. This difference between fan- and comb-shaped keys rises steeply as the number of end points increases. Similarly, the best choice algorithms used in matrix-based interactive keys typically seek to equally parse remaining taxa. Red arrows indicate entry points for the keys.

return the user to a previous couplet (e.g., parenthetical numbers in Figure 1), using several characters in a couplet, abandoning strict syntopic keys that rigidly reflect a systematic hypothesis (36), and placing images of annotated contrasting character states near the relevant couplets. A particularly good example of the last is Shattuck's key to Australian ants (41). At a structural level, character choices that split off a single taxon at a time result in comb-shaped keys (Figure 2, right) that require significantly more steps to the average identification than those that tend to divide remaining taxa into equal groups and have a fan-shaped structure (e.g., Figure 1 and Figure 2, left). This feature also applies to the best character algorithms in the multiple-entry keys discussed below (11, 12, 14). Our suggestions as to what might be the best practice in the construction of dichotomous keys are presented in Table 2.

Best: an algorithm in an interactive key that attempts to parse remaining taxa equally

Pathway Keys on the World Wide Web

Dichotomous keys have been produced in HTML format for use on computers and for mounting on the World Wide Web for a number of years (35, 50). Using hyperlinks to move between text and images, and from image to image, increases the impact of character state illustrations and the understanding of technical language (via links to illustrative images or glossary definitions). Hyperlinks also can help in backtracking and place-keeping in a pathway key and can somewhat ameliorate the unanswerable couplet problem by allowing easy skipping to subsequent couplets (35). In a fully realized form, HTML keys use hypertext links as pathways to move to the next couplet or as links to additional media sources (e.g., supplementary images, additional text) that are found on other Web pages (Internet flowchart key). Moreover, computer-interactive keys are more easily updated and corrected, and if published on the Web, changes are instantly distributed to potential users.

Some software programs translate optically scanned written keys and convert them into a computer-interactive format suitable for Web publication or distribution on CDROM. A recent repackaging of the older idea of populating pathway keys with multimedia is Lucid Phoenix (24). Rather than relying on a series of hyperlinked Web pages, Phoenix software provides the same user interface as many matrix-based interactive keys (i.e., a set of windows; see Figure 3) and an evolving list of characters (a pathway) by which sequential selection of character states results in the elimination of names from the taxon window until a single taxon remains. The true utility of Phoenix is found not in its use to build new keys, but in its use to revitalize older, historically important dichotomous keys into an interactive interface by providing additional multimedia not previously available. Older printed keys, i.e., the bulk of available identification tools, are transformed

Table 2 Best practices in the construction of dichotomous keys^a

Practice	Rationale	
1. Avoid the syntopic pitfall: Keys are for identification, not phylogenetic display.	A key allows the end user to quickly and accurately identify a specimen, not to showcase a phylogenetic hypothesis or a taxonomy. Apomorphies are not necessarily the best characters for identification.	
2. Avoid the unanswerable couplet: Monothetic couplets (e.g., Figure 1 , Couplets 1, 3) are weak.	If a couplet is based around a single character, and that character is obscured on a specimen or lost, then the usefulness of the key ends.	
3. Remember your end user: Clear, unambiguous characters and limited jargon result in more correct identifications.	Couplets with overlapping continuous characters, subtle differences, or exceptions are likely to lead to mistakes. Even specialists use special terms in different ways: Clear descriptive language is better.	
4. Construct leads as parallel sets of contrasting characters.	Couplets that use different sets of characters in their leads are likely to seem vague and confusing to a novice.	
5. Illustrate contrasting character states and group contrasting character state images together on the same plate.	If "a picture is worth 10,000 words," then a pair of contrasting character state pictures is worth even more.	
6. Put illustrations near the relevant key couplets (41).	It is easy to lose one's place in a dichotomous key.	
7. Provide a way out of a dead end: Give links to previous couplets or other means of keeping on the path (e.g., the parenthetical numbers in Figure 1).	Backtracking is normal during keying and getting lost in a large key is easy.	
8. Write couplets that evenly divide the remaining taxa. Fan-shaped keys require fewer steps than comb-shaped keys (see Figure 2); the larger the number of taxa in a key, the greater the difference.	The fewer the number of choices an end user must make, the less likely he or she is to make a mistake and the more rapidly he or she is likely to come to a correct identification.	
9. Provide comparative diagnoses for end point taxa in keys so that a user may be confident of their determination.	Anyone can end up at the end point of a pathway key—but the results should be checked against a description and/or image.	
10. Beta-test your key with naïve users before publishing.	We always get the right answer with our own keys, no matter how bad the keys are.	

^aAll 10 points are predicated on the assumption that the purpose of a key is to facilitate the rapid and correct identification of a specimen by someone other than the key's author.

into interactive keys by a process of optical scanning, language interpretation using character-recognition software, and final importation into Phoenix, which also then checks key structure and logic.

MATRICES AND MULTIPLE ENTRY: COMPUTER DIAGNOSTICS

Although pathway keys can be made more user-friendly by running them on computers and serving them on the Web in an interactive format, they are still limited by their logic to a single and often tortuous path to each identification. The alternative is an identification tool

that has many entry points and many paths to an identification.

Tabular Keys: Taxa Versus Character States

A table arranged as a matrix of taxa and characters is a traditional approach to identification that avoids many of the problems of pathway keys. Tabular keys are multiple entry—users can search across the list of characters for one that looks likely or scan up and down columns of character state distributions to look for an instant identification (the bingo approach), or if they think they already know the identity of a specimen, they can easily check its

Bingo: a character unique to a taxon that gives an instant identification; a routine in computer interactive keys that provides a list of unique characters and the taxa they define

Figure 3

Example of a user interface that might be found in an interactive key: character windows (Character list, Character states chosen), taxon windows (Species list, Taxa Discarded), an open character state window (image at top right), and a glossary illustration (image at bottom right). Species in white text on left were discarded from the key when the character state "Sternal shield posterior margin: concave" was chosen. (All images by D.E. Walter.)

Character list

Sternal shield posterior margin

- convex

Gnathotectum

- with single median process
- with median + lateral processes

Fixed digit dentition

- 2 teeth + serrate region
- 3-4 teeth

Posterior marginal-submarginal setation

- 20-30 pairs of R, UR setae
- 10-12 pairs of R, UR setae

Dorsal shield setation

- with 37 pairs of short setae
- with 37 pairs of long setae
- with 37 + 1-2 pairs of px setae

Dorsal shield taper

- gradually curved
- sharply linear

Palp apotele

- with 2 subequal tines
- with 2 tines + small spur

Seta r6

- acicular
- spatulate

Dorsal setae of femora II and IV

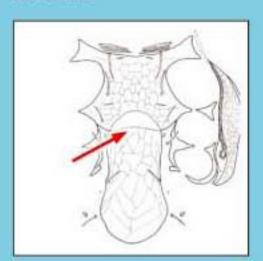
- mostly acicular
- mostly spatulate

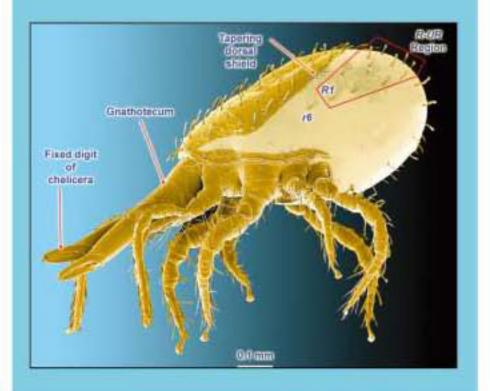
Species list Stratiolaelaps

- S. omatissima Aswegen and Loots
- S. lamington Walter and Campbell
- S. womersleyi Walter and Campbell
- S. Jorna Walter and Campbe
- S manilyn Walter and Campbel
- S. miles (Berlese
- S. scimitus (Womersley)

Character states chosen: 1

Sternal shield posterior margin – concave





Taxa discarded

- S. Iorna Walter and Campbell
- S. marilyn Walter and Campbell
- S. miles (Berlese)
- S. scimitus (Womersley)

characters. Tabular keys can have a box or a circular format (29); both types differ only in the arrangement of cells, either in a familiar table appearance or as a set of concentric rings, respectively. Not surprisingly, the greater the number of taxa in a table, the more difficult it becomes to use—at least until the advent of computers.

Punchcard Keys

Punchcard keys are stacks of thin cardboard cards that were used in early computer systems that recorded information as the presence or absence of holes at predefined locations. Although developed for early data processors, hand-held stacks of punchcards could be taken into the field and the key parsed by inserting a sorting bar (e.g., needle, twig, etc.) through the holes and allowing those cards not held in the stack (i.e., without a hole at the appropriate spot) to fall out, repeating these steps until an identification was obtained. On sunny days, stacks of cards could be held up to the light and sorted visually (D.E. Walter, personal communication).

Punchcard keys are divided into two basic types on the basis of the arrangement of

punch holes and whether the card represents a taxon (i.e., edge punched) or a character (i.e., center or body punched) (36, 38). In centerpunched cards, each card typically represents a single character, with all of the taxa represented by fields in the card. All taxa that exhibit a character have their field punched out. In edge-punched cards, each card typically represents a taxon, with characters indicated as fields around the edge of the card. Such keys have many limitations, including the expense of reproduction for mass distribution, character limitations (e.g., inapplicable and variable characters cannot be used), and the impossibility of backtracking through the key if a misinterpretation is suspected (36).

Computer-Based Taxonomic Programs

The structured nature of taxonomic description is ideal for its incorporation into computerized programs for subsequent output in a variety of formats such as interactive keys, natural language taxonomic descriptions, or phylogenetics. Despite this structure, there is great variation in the organization and content of descriptive data that prevents the seamless exchange of data between electronic formats. There have been attempts to standardize taxonomic descriptive data. The descriptive language for taxonomy (DELTA) format (9, 10, 12) was developed both as a standard format for taxonomic descriptions for computer programs and as a database for matrix-based interactive keys. DELTA format essentially consists of scored data and directives. Directives define parameters for the file and are used by programs that generate interactive keys (10, 14) and natural language taxonomic descriptions (13, 14). DELTA format has numerous types of directive syntax but basically comprises a character list, taxon list, character types, implicit values, and character dependencies (14). The first two directives are self-explanatory, but the others require some explanation. Character type describes whether the particular character is or-

dered or unordered, numeric integers or fractions, or simply text. Implicit values are used to improve and refine the output of natural language taxonomic descriptions. Character dependencies are specific sets of characters applicable only when certain character states are selected (14). Although DELTA format was developed in 1971 and has a long history of usage, it is rather inflexible in format and is closely tied to particular software applications, limiting its usefulness outside of these implementations. Designed as a replacement for DELTA format is the Structure of Descriptive Data (SDD), developed by the Taxonomic Databases Working Group (TDWG) (42). This extended markup language (XML)based format is promoted as an international standard for descriptive data similar to that used in DELTA format, but without the implementation dependency of the latter.

Matrix-Based Key Builders and Players

Matrix-based interactive keys are composed of three basic parts: (a) a database in the form of a matrix of character states scored for a set of taxa, (b) a program that queries the database, and (c) a user interface that allows for information input and/or user choices (11, 12). Matrix-based interactive keys allow a user to select a single character state at a time or multiple character states in one pass (using checkboxes) before rejected taxa are deleted (see programs in Table 1). Sorting programs search the database for the selected character state(s) and then prune those taxa not exhibiting that state(s). The user interface of most interactive keys essentially comprises a character list, taxon list, and often lists of taxa and characters discarded by user choices. Many interactive keys take the form of subdivided windows displaying the above elements (Figure 3).

With advances in computing power, matrix-based interactive keys have become easier to construct and use in diagnostics, particularly as Web-based tools. Two basic types DELTA: descriptive language for taxonomy

XML: extended markup language

GenBank: a publicly available database of >100 gigabases of gene sequences

of Web-based interactive keys are (a) standalone programs that require the installation of specific software that is often tied to a particular operating system, and (b) browser programs written in Java language or XML. Stand-alone programs have their own user interface, while the browser programs run in an Internet browser. Both types of programs can access scored data served online during the identification process. The main advantages of browser over stand-alone programs are that browser programs require little or no prior download of software and run on most operating systems. The primary disadvantage of Web-based programs is that access to the data and program requires connection to the Internet, restricting use of the program in areas without an Internet connection. Key operation is also slowed by the speed of the Internet connection, especially for larger, image-rich keys. In these cases downloading the source data and media is advantageous.

MOLECULAR DIAGNOSTICS

In the field of diagnostics the distinction between traditional and interactive keys and analytical diagnostic tools is becoming blurred by the use of molecular data. DNA sequences are characters, and as long as positional homology of nucleotides or amino acids is not violated, these sequences can be used in a comparative way to deduce identifications of unknown specimens from a reference library of sequences (18). This comparison or matching type of diagnostic protocol is often employed using a phenogram based on genetic distance (e.g., a neighbor-joining tree). The Barcoding of Life Data System (18) uses this type of matching protocol by building a neighbor-joining tree based on a set of sequences of a small portion of the mitochondrial gene, cytochrome oxidase I (COI). The genetic distance between an unknown specimen and taxa in the tree is calculated and the closest match represents an identification. The result is an interactive key by which the input is sequence data and the output is an identification based on

genetic distance. Obvious limitations of this protocol include incomplete sequence reference libraries, ambiguous alignments, and gene sequences that are not limited to one taxon (4, 5, 40). BLAST (basic local alignment search tool) (2) is an algorithm for comparing sequence data in large sequence databases (e.g., GenBank) in a similar way. Molecular characters in the form of random fragment length polymorphism banding patterns have also been used in an interactive key format by Moritz et al. (34).

Recently, DNA-barcoding has been promoted as a one-fix cure for the taxonomic impediment, especially in biodiversity inventories (20). However, as Marshall (25) has pointed out, "[S]imilar claims have been made by champions of emerging technologies (quantitative phenetics, cuticular hydrocarbons, gel electrophoresis, DNA-DNA hybridisation) in the past, in each case generating a flurry of interest (and funding) but never making a significant impact on how organisms are identified." Molecular techniques do hold significant promise, especially for uncovering cryptic species and for associating immature stages and sexes within a species, but the logical primacy of a particular COI haplotype over other species diagnostic characters is difficult to support (30). Moreover, recent research indicates that, at least in poorly studied or rapidly speciating lineages, DNA-barcoding is not a reliable identification tool (30, 31, 40).

CONCLUSIONS

Although automated matching to digital databases of images or gene sequences is currently feasible for some groups, the implementation of these systems is not likely to happen soon for most taxonomic identifications. Both approaches require extensive expenditures of time, money, and taxonomic expertise, and the last is in increasingly short supply. Most existing taxonomic identification tools are stored in printed publications that may be difficult to access and to use, but

increasing numbers of digital interactive keys are appearing. Do these newer interactive tools offer a short-term solution to the taxonomic impediment?

Advantages of Interactive Matrix-Based Keys

We present our suggestions for best practice in the construction of matrix-based interactive keys in Table 3, but these keys have clear advantages over traditional pathway keys. Matrix-based keys provide unrestricted character use in both order of selection and availability (14). This feature is one of the greatest benefits of matrix-based keys, enabling the user to select characters based on any criterion, such as character availability (thus avoiding unanswerable couplets) or familiarity (thus avoiding guesses when characters are unfamiliar or confusing). Such flexibility may confuse a novice, especially if there is a diversity of characters to choose among, but algorithms that parse the remaining taxa efficiently, best character options, reduce the number of required character state selections dramatically, even for large keys. Error tolerance to account both for errors made by the end user and errors introduced by the key builder also can be built into keys either through character state weighting schema or by compensatory scoring.

In addition, new types of characters are readily incorporated into interactive keys. For example, in paper-based keys numerical characters typically are useful only when presented as ranges. However, discrete numerical characters are easily accommodated by interactive keys and provide an important means of identification, as individual taxa can be scored directly. Similarly, gene sequences or representations of agarose gel banding patterns (of PCR products, isozymes, etc.) are easily incorporated into interactive keys as discrete or numerical characters defined by distance traveled on a calibrated gel.

Finally, dichotomous keys published in journals and books are fixed in time and place.

One important advantage of all types of interactive keys (not just matrix-based) is that they are updatable as taxonomy changes. As new taxa are described they can be incorporated into the key with minimal effort and may be instantaneously distributed via the Internet.

Flies in the Ointment: Problems with Adoption and Use

One problem faced by present developers of all types of interactive keys is that the interactive format is new both in terms of technology and in basic approaches to identification. For example, one advantage of traditional pathway keys to builders is that they can use their knowledge of a group to deploy a limited number of character state choices in a welltried logical format. In contrast, a diversity of programs, each with its own internal logic, exists for developing interactive keys. Moreover, the production of a fully realized matrix of characters and taxa requires knowing the distribution of character states in taxa outside of their traditional area of use. The builder of a matrix key would have to hunt down the obscure character states in the literature or determine them by examining specimens. Similar problems occur for the end user: He or she must set aside a lifetime of experience with traditional keys and learn to use often complicated new software—a daunting task if training programs are not available. Although naïve users may prefer interactive keys, they do not necessarily perform better with them (50).

Another problem specific to Web-based keys is the perceived ephemeral nature of Internet resources. While Web page addresses do change over time, significant taxonomic resources such as interactive keys are unlikely to disappear completely. Although many taxonomists, particular those of the younger generation, are embracing interactive keys as important taxonomic tools, it is our experience that often the development of these keys is not valued as much as traditional publications by the wider taxonomic community. We feel that

Table 3 Suggested best practices for constructing matrix-based keys^a

Practice	Rationale	
1. Use key-building software programs that provide 'Best' sorting of characters.	'Best' preferentially selects characters that provide the most equal division of the remaining taxa (see Figures 2 and 3).	
2. Partition characters into discrete sets and use software that can code characters for dependencies, i.e., so that character choices by the user narrow or expand the characters available for identification.	A profusion of characters may be daunting to end users. Dividing characters into logical subsets (e.g., wing venation, mouthparts, tarsal) facilitates entry into the key and improves the chance that damaged specimens will be correctly identified. Negative character dependencies actively simplify matrix keys as they are used. Positive dependencies reduce the chance of a misidentification by bringing up additional characters for checking.	
3. Reach a balance between the number of characters versus the number of character states per character.	Binary characters choices are useful, but multiple states are more efficient—within reason. Characters with too many states may be better divided into multiple characters, each with fewer states.	
4. Make ample use of supplementary information for characters (e.g., definitions, illustrations) and taxa (e.g., fact sheets, image galleries).	Page space is not a constraint in a Web-based key and supplementary information and hypertext links to Internet search engines are effective diagnostic aids.	
5. Provide an illustrated glossary of terminology.	Jargon and complicated terminology should be avoided, but necessary terminology should be fully explained and illustrated in a glossary.	
6. Clearly label and annotate figures with relevant structure(s) identified.	Do not rely on the end user to know the critical differences on your figures: identify characters (e.g., arrows, borders, etc.) and label your figures.	
7. Use software with a scoring system.	Scoring systems rank remaining taxa by similarity. Thus, if a single taxon is not recovered from the selected character states, a list of high-scoring taxa is available for comparison.	
8. Use software that will build in error tolerance and allow error tolerance to be switched off.	Misinterpretation of some characters is inevitable in naïve users. Programs that allow scoring character states for misinterpretations are more likely to result in correct identifications.	
9. Use software that is minimally platform dependent: Keys readily available for use regardless of the user's current software are those that are most likely to be used.	Keys should be ready to use instantly without prerequisite installation of software. Keys should be freely available on the Internet and cross-platform so that they will run in an Internet browser on any computer.	
10. Never overestimate the attention span of the end user: Keep it simple.	Sophisticated software and complex key structure may look impressive, but it does slow down performance of a key (especially over the Internet) and can be confusing to naïve users.	

^aAll 10 points are predicated on the assumption that the purpose of a key is to facilitate the rapid and correct identification of a specimen by someone other than the key's author.

as the number of interactive keys produced by taxonomists increases a critical mass will be reached and their production will become an essential part of the taxonomic process.

One method for making interactive keys more attractive to taxonomists and the wider scientific community is to embed them within larger topic-based Web sites. Until we have an all-encompassing interactive key to life, interactive key development is going to be piecemeal and individual keys must be placed in context for an end user to understand their relevance and utility. This can be achieved by embedding interactive keys (without prerequisite installation of software) within larger Web sites that present a variety of information (e.g., bibliographies, ecology, and taxonomic databases) (48, 49) as well as an interactive key.

Communication of research is a fundamental responsibility of scientists. For taxonomists, the identification of specimens and the provision of identification or diagnostic tools (e.g., published keys) are critical to this responsibility. Presently, there is a plethora of interactive key-building programs available, each vying for wider adoption by the taxonomic community and each varying in sophistication, ease of use, cross-platform compatibility, and overall utility (e.g., **Table 1**). Many of these programs share similar routines and others have unique applications;

however, we are unaware of any systematic studies of how useful these routines may be to end users. We suggest that it is time to adopt a set of guidelines for best practice based on experimental studies using naïve end users. This would help to ensure that scarce resources are not wasted through a natural selection of key format adoption, with only the strongest (i.e., most user-friendly or bestmarketed) programs gaining wider use. Moreover, because of this myriad of interactive key formats and the absence of links among them, focus should be placed on integrating identification keys with the immense taxonomic resources available online (chiefly databases) using relational databases, portals, and search engines. The final product, we hope, would culminate in a single user interface for an Identification Key to All of Life.

SUMMARY POINTS

- Pathway keys, typically in a dichotomous format, have been the basis of biological identification for more than two centuries; however, they have an inherent structural problem—each key has a single entry point and only one set of choices produces a correct identification.
- 2. Matrix-based, computer-interactive keys ameliorate or solve many of the problems inherent in pathway keys; however, they have their own set of problems, especially the learning curves associated with new technology, resistance to adoption by both builders and users, and a lack of accepted standards. Best practices in matrix-based keys are currently the source of considerable contention among proponents of different systems and would benefit from some experimental testing of hypotheses using naïve end users.

DISCLOSURE STATEMENT

We have published keys in DELTA, Lucid, and traditional dichotomous formats.

ACKNOWLEDGMENTS

We would like to thank the numerous students, colleagues, and anonymous referees who over the years have forced us to pay more attention to the structure of our keys and to remember that we are writing those keys for them.

LITERATURE CITED

1. ABIS. Automatic identification of bees. http://www.informatik.uni-bonn.de/projects/ABIS/ABIS_Publications.html

- 2. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-10
- 3. Aristotle. History of Animal, Vol. IV. Transl. D'AW Thompson, 1910. http://www.netlibrary.com.login.ezproxy.library.ualberta.ca/Reader/
- Ballard JWO, Whitlock MC. 2004. The incomplete natural history of mitochondria. Mol. Ecol. 13:729–44
- Bensasson D, Zhang X, Hartl DL, Hewitt GM. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol. Evol. 16:314–21
- 6. Blackmore S. 2002. Biodiversity update: progress in taxonomy. Science 298:365
- 7. Borror DJ, Triplehorn CA, Johnson NF. 1989. An Introduction to the Study of Insects. Fort Worth, TX: Harcourt Brace Jovanovich. 875 pp. 6th ed.
- 8. Brown L, ed. 1993. The New Shorter Oxford English Dictionary on Historical Principles, Vol. 1 A-M. Oxford, UK: Clarendon. 1876 pp.
- 9. Dallwitz MJ. 1980. A general system for coding taxonomic descriptions. Taxon 29:41-46
- 10. Dallwitz MJ. 1993. DELTA and INTKEY. In Advances in Computer Methods for Systematic Biology: Artificial Intelligence, Databases, Computer Vision, ed. R Fortuner, pp. 287–96. Baltimore: Johns Hopkins Univ. Press
- 11. Dallwitz MJ. 2005. A comparison of interactive identification programs. http://delta-intkey.com/www/comparison.htm
- 12. Dallwitz MJ. 2006. Programs for interactive identification and information retrieval. http://delta-intkey.com/www/idprogs.htm.
- 13. Dallwitz MJ, Paine TA, Zurcher EJ. 1998. Interactive keys. In *Information Technology*, *Plant Pathology and Biodiversity*, ed. P Bridge, P Jeffries, DR Morse, PR Scott, pp. 201–12. Wallingford, UK: CAB Int. 496 pp.
- 14. Dallwitz MJ, Paine TA, Zurcher EJ. 2000. Principles of interactive keys. http://delta-intkey.com
- 15. Godfray HCJ. 2002. Challengers for taxonomy. Nature 417:17–19
- 16. Gotelli NJ. 2004. A taxonomic wish-list for community ecology. *Philos. Trans. R. Soc. London* B 359:585-97
- 17. Gray A. 1848. A Manual of the Botany of the Northern United States: From New England to Wisconsin and South to Ohio and Pennsylvania Inclusive. Boston: James Monroe. 710 pp.
- 18. Hebert PDN, Cywinska A, Ball SL, de Waard JR. 2003. Biological identifications through DNA barcodes. *Proc. R. Soc. London B* 270:313–21
- 19. Hoagland K. 2000. The taxonomic impediment and the convention on biodiversity. http://www.sasb.org.au/TaxImp.html
- Janzen DH, Hajibabaei M, Burns JM, Hallwachs W, Remigio E, Hebert PDN. 2005. Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding.
 Philos. Trans. R. Soc. London B 360:1835–45
- 21. Klass KD, Zompro O, Kristensen NP, Adis J. 2002. Mantophasmatodea: a new insect order with extant members in the Afrotropics. *Science* 296:1456–59
- 22. Kohsima S. 1984. A novel cold-tolerant insect found in a Himalayan glacier. *Nature* 310:225-27
- 23. Lobanov AL. 2003. Keys to beetles and biological diagnostics. http://www.zin.ru/Animalia/Coleoptera/eng/syst8.htm
- 24. Lucid Phoenix. http://www.lucidcentral.org/phoenix/
- Marshall SA. 2003. The real cost of insect identification. Newsl. Biol. Surv. Can. (Terr. Arthropods). Vol. 22 No. 1, Spring 2003

12. An excellent source of information on various aspects of interactive key principles and development.

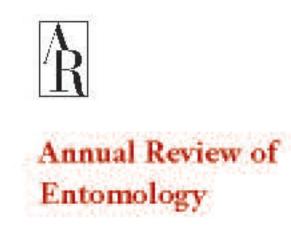
- 26. Marshall SA. 2006. Insects Their Natural History and Diversity: With a Photographic Guide to Insects of Eastern North America. Richmond Hill, Can.: Firefly Books. 704 pp.
- 27. Marshall SA, Danks H. 2005. Project update: the *Biological Survey of Canada Journal of Arthropod Identification*: a modular identification e-journal of the *Biological Survey of Canada*. Newsl. Biol. Surv. Can. (Terr. Arthropods). Vol. 24 No. 2, Fall 2005
- 28. Mayr E. 1969. Principles of Systematic Zoology. New York: McGraw-Hill. 428 pp.
- 29. Mayr E, Linsley EG, Usinger RL. 1953. Methods and Principles of Systematic Zoology. New York: McGraw-Hill. 328 pp.
- Meyer CP, Paulay G. 2005. DNA barcoding: error rates based on comprehensive sampling. PLOS Biol. 3(12):e422
- 31. Monagham MT, Balke M, Pons J, Vogler AP. 2006. Beyond barcodes: complex DNA taxonomy of a South Pacific Island radiation. *Proc. R. Soc. London B* 273:887–93
- 32. Moore A. 1991. Automated identification of insects in flight. *Micronesica Suppl.* 3:129–33
- 33. Moore A, Miller JR, Tabashnik BE, Gage SH. 1986. Automated identification of flying insects by analysis of wingbeat frequencies. J. Econ. Entomol. 79:1703-6
- 34. Moritz M, Mound L, Morris D, Goldarazena R. 2001. Pest Thrips of the World: An Identification and Information System Using Molecular and Morphological Methods. Melbourne: CSIRO Publ. CD-ROM
- 35. Norton G. 2002. Multi-media/internet keys and the taxonomic crisis. Antenna 26:245-48
- 36. Pankhurst RJ. 1991. *Practical Taxonomic Computing*. Cambridge, UK: Cambridge Univ. Press. 214 pp.
- 37. Perring TM, Cooper AD, Rodriguez RJ, Farrar CA, Bellows TSJ. 1993. Identification of a whitefly species by genomic and behavioral studies. *Science* 259:74–77
- 38. Quicke DLJ. 1993. Principles and Techniques of Contemporary Taxonomy. London: Blackie Acad. Prof. 311 pp.
- 39. Rubinoff D, Haines WP. 2006. Hyposmocoma molluscivora description. Science 311:1377
- 40. Rubinoff D, Holland BS. 2005. Between two extremes: Mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. Syst. Biol. 54:952–61
- 41. Shattuck SO. 1999. Australian Ants: Their Biology and Identification. Collingwood, Aust.: CSIRO Publ. 226 pp.
- 42. Taxonomic Databases Working Group. 2004. http://www.diversitycampus.net/ Projects/TDWG-SDD/
- 43. Triplehorn CA, Johnson NF. 2005. Borror and De Longs's Introduction to the Study of Insects. New York: Thomson. 864 pp. 7th ed.
- 44. Voss EG. 1952. The history of keys and phylogenetic trees in systematic biology. J. Sci. Lab. Denison Univ. 43:1-25
- 45. Watson AT, O'Neill MA, Kitching IJ. 2003. Automated identification of live moths (Macrolepidoptera) using Digital Automated Identification System. Syst. Biodiv. 1:287–300
- 46. Weeks PJD, O'Neill MA, Gaston KJ, Gauld ID. 1999. Automated insect identification: exploring the limitations of a prototype. J. Appl. Entomol. 123:1–8
- 47. Wheeler QD, Raven PH, Wilson EO. 2004. Taxonomy: impediment or expedient? Science 303:285
- 48. Winterton SL. 2004. Aquarium and pond plants of the world. Lucid key. APHIS, USDA, Version 1. http://www.lucidcentral.org/keys/appw/

36. An important publication that has served as a foundation for interactive key development.

44. A delightful and erudite review of the evolution of keys and trees.

45. A detailed field study of the DAISY system, the successes in identification, and the discordance between some computergenerated misidentifications and how people see the same moths.

- 49. Winterton SL, Skevington JH, Lambkin CL. 2005. Stiletto flies of Australasia (Diptera: Therevidae). Lucid3 key. Calif. Dep. Food, Agric.; Agric. Can.: CSIRO. Version 1. http://www.cdfa.ca.gov/phpps/ppd/Entomology/Lucid/Therevidae/Austherevid/key/Austherevid/Media/Html/opening_page.html
- 50. Wright JF, Morse DR, Tardivel GM. 1995. An investigation into the use of hypertext as a user-interface to taxonomic keys. Comp. Appl. Biosci. 11:19–27
- 51. Yu DS, Koko EG, Barron JR, Schaale GB, Gowen BE. 1992. Identification of ichneumonid wasps using image analysis of wings. Syst. Entomol. 17:389–95



Volume 52, 2007

Contents

Frontispiece Charles D. Michener	. xiv
The Professional Development of an Entomologist Charles D. Michener	1
Insect/Mammal Associations: Effects of Cuterebrid Bot Fly Parasites on Their Hosts Frank Slansky	. 17
Phenology of Forest Caterpillars and Their Host Trees: The Importance of Synchrony Margriet van Asch and Marcel E. Visser	. 37
Arthropod Pest Management in Organic Crops Geoff Zehnder, Geoff M. Gurr, Stefan Kühne, Mark R. Wade, Steve D. Wratten, and Eric Wyss	. 57
The Sublethal Effects of Pesticides on Beneficial Arthropods Nicolas Desneux, Axel Decourtye, and Jean-Marie Delpuech	81
Impact of Extreme Temperatures on Parasitoids in a Climate Change Perspective Thierry Hance, Joan van Baaren, Philippe Vernon, and Guy Boivin	107
Changing Paradigms in Insect Social Evolution: Insights from Halictine and Allodapine Bees Michael P. Schwarz, Miriam H. Richards, and Bryan N. Danforth	127
Evolutionary Biology of Centipedes (Myriapoda: Chilopoda) Gregory D. Edgecombe and Gonzalo Giribet	151
Gene Regulation by Chromatin Structure: Paradigms Established in Drosophila melanogaster Sandra R. Schulze and Lori L. Wallrath	171
Keys and the Crisis in Taxonomy: Extinction or Reinvention? David Evans Walter and Shaun Winterton	193
Yellow Fever: A Disease that Has Yet to be Conquered Alan D.T. Barrett and Stephen Higgs	209

Molecular Mechanisms of Metabolic Resistance to Synthetic and Natural Xenobiotics Xianchun Li, Mary A. Schuler, and May R. Berenbaum
Group Decision Making in Nest-Site Selection Among Social Insects P. Kirk Visscher
The Role of Allatostatins in Juvenile Hormone Synthesis in Insects and Crustaceans Barbara Stay and Stephen S. Tobe
Nectar and Pollen Feeding by Insect Herbivores and Implications for Multitrophic Interactions Felix L. Wäckers, Jörg Romeis, and Paul van Rijn
Biology and Evolution of Adelgidae Nathan P. Havill and Robert G. Foottit
Biology of the Bed Bugs (Cimicidae) Klaus Reinhardt and Michael T. Siva-Jothy
The Use of Push-Pull Strategies in Integrated Pest Management Samantha M. Cook, Zeyaur R. Khan, and John A. Pickett
Current Status of the Myriapod Class Diplopoda (Millipedes): Taxonomic Diversity and Phylogeny Petra Sierwald and Jason E. Bond
Biodiversity Informatics Norman F. Johnson
Cockroach Allergen Biology and Mitigation in the Indoor Environment *J. Chad Gore and Coby Schal**
Insect Conservation: A Synthetic Management Approach Michael J. Samways
Interactions Between Mosquito Larvae and Species that Share the Same Trophic Level Leon Blaustein and Jonathan M. Chase
Indexes
Cumulative Index of Contributing Authors, Volumes 43–52
Cumulative Index of Chapter Titles, Volumes 43–52
Errata
An online log of corrections to <i>Annual Review of Entomology</i> chapters (if any, 1997 to the present) may be found at http://ento.annualreviews.org/errata.shtml

viii Contents