International Journal of Zoological Research

ISSN 1811-9778

Academic Journals Inc.

www.academicjournals.com

International Journal of Zoological Research 9 (1): 1-16, 2013
ISSN 1811-9778 / DOI: 10.3923/ijzr.2013.1.16

SIrD: A Software Describing a System of Identification for Drosophilidae

[1]Glaiza G. Opema, [3]Rez A. Cabrera, [1]Mark Anthony J. Torres, [2]Jessie G. Gorospe and [1]Cesar G. Demayo

[1]Department of Biological Sciences, Mindanao State University, Iligan Institute of Technology, Iligan City, Philippines

[2]School of Graduate Studies, Mindanao State University-Naawan, Naawan, Misamis Oriental, Philippines

[3]Programming Consultant, Iligan City, Philippines

Corresponding Author: C.G. Demayo, Department of Biological Sciences, Mindanao State University, Iligan Institute of Technology, Iligan City, Philippines

ABSTRACT

The power of computers have enabled the field of biology to make significant leaps in quantitative methods. Fresh ideas from other disciplines in science and technology offer insight on how old techniques can be evolved into new ones for the benefit of biologists worldwide. In this study, a different approach to *Drosophila* species identification is presented by making use of binary coding and computer automation. The developed software, a System of Identification for Drosophilidae (SIrD) uses binary strings to describe species characteristics and use it to calculate the degree of similarity between a sample specimen and the species stored in the database. This method of identification allows for greater tolerance to errors and flexibility in adding new species characters. The linear nature of the binary string enables the software to eliminate the common problems encountered in the couplet system of the dichotomous key, such as, species-specific paths, low tolerance to errors and rigidity in adding new characters. SIrD has a user-friendly interface and is created using open-source web development software, making it Internet-ready, for easy access on the World Wide Web.

Key words: SIrD, Drosophilidae, binary, dichotomous key

INTRODUCTION

Many researches being conducted on the species of Drosophilidae mostly involve species identification and distribution. Field sampling for inventory of flies results in a collection of thousands of specimens. Identifying these samples one at a time using the traditional method, the dichotomous key, can become time-consuming. For this reason, a much more efficient method of identification is necessary.

There are many studies that have used computers to automate species identification. These systems that have been developed are based on the concept of the dichotomous key (Abdulrahaman *et al*., 2010; Kirchoff *et al*., 2008; Rocker *et al*., 2007; Smith-Akin *et al*., 2006; Norton, 2005). The results, however, indicate that automating the dichotomous key may have certain upper limits.

With the dichotomous key, the process of identification must be sequentially followed (Rocker *et al.*, 2007). If the user is unsure of the pair of characters shown, he is forced to choose either one of them. Choosing the wrong option can result to an erroneous species-specific path. It is also pointed out that automation of the dichotomous key works well for small groups of taxa but is unlikely to be very effective with larger groups of taxa (Kirchoff *et al.*, 2008). Existing implementations of automating the dichotomous key do not have a way of quantifying correctness. In this study, a different technique is presented and used for automation. The System of Identification for Drosophilidae (SIrD) is a tool for the ease of identification of the Drosophilidae species, using binary codes instead of the dichotomous key to represent the presence or absence of character states. Encoding character states in binary format, in effect, produces a linear binary string. This binary linearization allows SIrD to calculate, in a straightforward manner, the degree of similarity between the sample specimen and the species in the database and then express it in percentage.

MATERIALS AND METHODS

Several other benefits can be derived from the use of such principle. Binary linearization allows character states to become independent of each other when establishing presence or absence, speeding up the identification of specimens. It also allows greater tolerance to identification errors. If the user made a mistake in establishing the presence or absence of a character, or makes only a partial input, SIrD simply takes this into account and calculates the degree of similarity accordingly. It displays all possible matches in descending order of degree of similarity, thus avoiding the all-or-nothing approach of the dichotomous key. Furthermore, the linear approach of identification makes the character ID easily extensible, allowing new characters to be added at the end of the binary string, especially when new discoveries are found.

Another important aspect of SIrD is that it is created using open-source web development software. The computational algorithms are programmed in PHP, the species data are stored and retrieved using MySQL and the overall user interface is handled by Drupal. These make it possible for SIrD to be compatible with many browsers for the World Wide Web. A prototype has been deployed on the Internet and can be found at http://drosophila.site88.net.

Binary coding and the degree of similarity: The binary code involves only two values, true and false (or yes and no) represented in numeric form as “1” and “0”, respectively. The binary code “1” is assigned to a character state if it is present in the sample being analyzed and binary code “0” if such character state is absent. These binary codes, when lined up, form the character ID that uniquely describes each Drosophilidae species. An identification process based on these binary codes is formulated to determine a possible match between the sample being analyzed and the species in the database.

A crucial factor in establishing a match between two objects is to quantify how similar they are. The technique used in this study is derived from the mathematical model published in 1901 by Jaccard (1901), a professor of botany and plant physiology at the ETH Zurich University. The Jaccard index of similarity, also known as the Jaccard similarity coefficient (Coefficient de communauté), is a statistic used for comparing the similarity and diversity of sample sets:

$$\text{Jaccard index} = \frac{|\text{A intersection B}|}{|\text{A union B}|} = \frac{|\text{A and B}|}{|\text{A or B}|}$$

The degree of similarity, as defined in this study, is the proportion of the actual number of similarities to the possible number of similarities, expressed in percentage, as shown in the mathematical model below:

$$\text{Degree of Similarity} = \frac{\text{Actual No. of similarities}}{\text{Possible No. of similarities}} \times 100$$

With binary coding, the existence of a particular character can simply be established by assigning it a “1” (present) or a “0” (absent). Determining similarity, then, becomes straight forward, as shown in the algorithm below:

- Determine the ordinal position of the present (code “1”) characters in the binary code of each object
- Count the number of ordinal positions for which present (code “1”) characters of each object match. This is the actual number of similarities (numerator)
- Count the number of present (code “1”) characters in the binary code of each object
- Determine which is the larger of the two. This becomes the possible number of similarities (denominator)
- Divide the actual number by the possible number to get the degree of similarity

It is important to keep in mind that, in dealing with proportions, care must be taken in choosing the denominator. Although the resulting quotient may not be mathematically erroneous, it will certainly not be logical (i.e., degree of similarity exceeding 100%).

Binary code score card: A score card (Appendix Table 1) has been established to facilitate the formulation of binary codes necessary to uniquely describe a species. The character states used in the score card are based on the significant identifying characters of the species (Markow and O’Grady, 2006; Wheeler, 1952; Harrison, 1950). A thorough examination of the significant identifying characters is done to ensure that no repetition of character states exists, avoiding that particular pitfall of the dichotomous key. The score card (in printable and spreadsheet formats) and a diagram pointing the locations of the characters are made available for download and reference with the software (http://drosophila.site88.net/?q=downloads).

The primary purpose of the score card is to provide users a guide on the character states needed for the identification of specimens through SIrD. For convenience, the taxonomic characters are classified into the several bodily categories, where each character state is listed accordingly. These are the general body characteristics (13 states), head (59 states), thorax (63 states), legs (74 states), abdomen (66 states), genitalia and anal (91 states) and wings (30 states). There are a total of 396 character states in the score card, all of which can simply be answered by binary codes “1” or “0” (yes or no).

When examining specimens for identification, there are situations where the user is unsure of the character state observed. If the dichotomous key were used, the user would be forced to choose one state and engage in trial-and-error. The dichotomous key may be answered incorrectly, leading to erroneous identification (Osborne, 1963).

With the score card, however, the user can simply place a question mark “?” on that character state to tell SIrD to ignore it during comparison and still allow the software to compute an accurate degree of similarity. This effectively eliminates the problem of dependency. The user’s decision in answering a character state does not affect the decision he makes on any other character state.

Another purpose of the score card is to enable the formulation of binary strings for adding (or editing) established species in the database. In contrast with identification, adding or editing database entries require stricter compliance with binary coding. Only binary codes “1” and “0” are allowed and all other symbols are considered illegal. Also, no character state must be skipped during formulation in order to maintain their ordinal positions relative to each other in the binary sequence.

These strict measures are necessary to ensure the correctness of the resulting character ID for the database. Once the binary coding is completed, the species listed in the software will then have its own unique binary string which becomes the basis for comparison by SIrD when computing the degree of similarity.

RESULTS AND DISCUSSION

SIrD software: SIrD is developed using several open-source programming tools, namely, PHP, MySQL and Drupal. These tools were specifically chosen for their practical merits. PHP is a widely-used open source general-purpose scripting language that is especially suited for web development (http://www.php.net). MySQL is a popular open source database system used by many organizations to save time and money in powering business-critical systems and packaged software (Oracle Corp. 2013). Drupal is an open source content management platform powering millions of websites and applications (http://drupal.org).

A usage guide (Fig. 1) for the software is made available for download and reference (http://drosophila.site88.net/?q=downloads). SIrD is designed with ease-of-use in mind, with a simplistic user interface and avoids clutter to make it pleasant to navigate around. The process of using SIrD is straightforward and can be summed up in only a few steps, as shown:

- Download the binary code score card and the species diagram
  - Use the printable score card and species diagram as a visual aids
  - Use the spreadsheet score card as the answer sheet
- Highlight and copy the binary code portion only of the specific row being answered and paste it on the textbox at the landing page of SIrD
  - There may be white spaces in between the binary codes, but SIrD will simply ignore them upon submission
- Click on the search button to show the results page

SirD automatically displays, in descending order of degree of similarity (expressed in percentage), species that are possible matches to the sample the user is analyzing. Each species listed is presented with its binary identifier grouped into 10 characters for easy counting. The result also highlights the ordinal position of the character state where the sample matches that of a particular species, for review and verification.

SirD also allows users to diagnose specific portions of the specimen (i.e., thorax only, head and abdomen simultaneously, etc). The input binary string for identification is much shorter, but still yields the possible matches. For this to work, the user simply has to replace the binary code with special symbols, at appropriate positions, to effectively ignore some characters, as shown in Table 1 and 2.

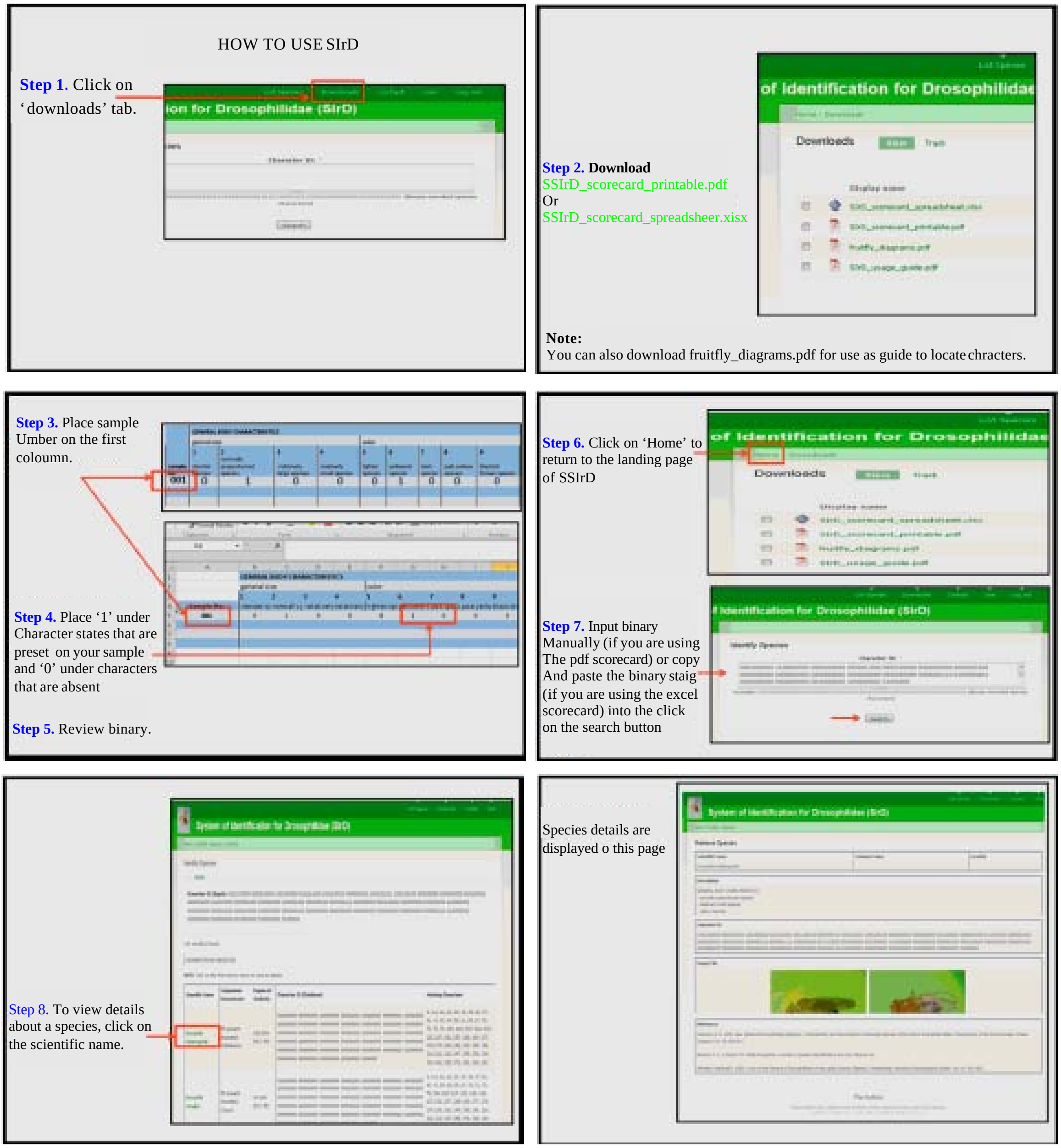


Fig. 1: User guide for system of identification for Drsophilidae (SIrD), a step by step guide in using the software

Table 1 lists the special symbols that will be used when one wishes to ignore one or more character states. Table 2 illustrates sample placements and combinations of these special symbols. The symbols are used to replace the codes for the caracter state(s) that the user wishes to ignore.

Shortcuts to ignore characters in bulk can also be done by category. Each major body part (the category) is assigned a specific alphabetic symbol to act as placeholder if that particular body part needs to be ignored. Basically, the categories being ignored depend on the where the symbols are positioned, as shown in the Table 3 and 4.

Table 1: Special symbols for replacing a certain number of characters

Name	Special symbol	Use
Question mark	?	Ignore one particular character
Number sign	#	Ignore five characters in bulk
Asterisk	*	Ignore ten characters in bulk

Table 2: Positioning of binary string and special symbols

Characters to ignore	Input format of binary string with symbols
Ignore 7 characters before	# ? ? [binary string]
Ignore 48 characters after	[binary string] * * * * # ? ? ?
Ignore 25 characters in between	[binary string] * * # [binary string]
etc	...

Table 3: Alphabetic symbols indicating the particular body part to be ignored

Alphabetic Symbol	Major body part to ignore	No. of characters ignored
A	General body characteristics	13
B	Head	59
C	Thorax	63
D	Legs	74
E	Abdomen	66
F	Genitalia and anal	91
G	Wings	30

Table 4: Positioning of binary string and alphabetic symbols

Major body part being analyzed	Input format of binary string with symbols
General body characteristics	[binary string] B C D E F G
Head	A [binary string] C D E F G
Thorax	A B [binary string] D E F G
Legs	A B C [binary string] E F G
Abdomen	A B C D [binary string] F G
Genitalia and anal	A B C D E [binary string] G
Wings	A B C D E F [binary string]
Head and Abdomen	A [binary string] C D [binary string] F G
Legs and Wings	A B C [binary string] E F [binary string]
etc	...

Table 3 presents the alphabetic symbols representing each body parts. When using these symbols, the user must keep in mind that the symbols should always be placed in order, i.e., the alphabetic symbol A should always be the first symbol in the string as shown in Table 4.

It is important to note that these special and alphabetic symbols are valid only when used during identification. These symbols are not valid when used during adding or updating a character ID in the database, in order to preserve the binary format needed to conveniently and uniquely describe the species.

Aside from being an identification system for the Drosophilidae species, SIrD is also a capable information storage and retrieval system. When the user clicks on a species name, it moves into another page dedicated solely for that particular species.

The species-dedicated page displays the scientific name, its designated binary string, the list of character states present and sample images of the species. It also displays information gathered from the published dichotomous keys and other articles describing the phenotypic characters of the species. These sources of information are listed in the References box in each of the species' dedicated page.

Many of the existing computer-based identification systems utilize the concept of the couplet system of the dichotomous key. Although computerization allows better handling of data, it still fundamentally presents the same problem in which the decision a user makes in a couplet is affected by the decision he made with the previous couplet, because the couplets are dependent on each other. The computer-based applications for identification developed by Rocker *et al.* (2007) and Dallwitz *et al.* (2000), for example, clearly demonstrates such dependence.

The dependencies introduced by the dichotomous key concept imply that there is a unique path for each species. The user may need to be an expert in analysing the specimen and be experienced in using the dichotomous key to get the correct identification. This may not be easily accomplished by a novice user. By comparison, the linear binary system of classification implemented in SIrD eliminates the problem of dependency. The user's decision in picking a particular character state, whether it is correct or not, does not affect the decision he makes when choosing other character states.

This problem on dependency has also been addressed in MOSCHweb, an interactive key of the Palearctic tachinidae (Cerretti *et al.*, 2012). MOSCHweb is a more flexible identification software, allowing the user to ignore character states that are difficult to interpret or are inapplicable due to damage. However, the process of identification of MOSCHweb such that when the users selects a character state, it discards all other taxa that do not share that state, eventually narrowing down to a single taxón.

With SIrD, on the other hand, the linear nature of its binary string enables it to calculate the degree of similarity, allowing the software to display any other possible matches. Along with the information on degree of similarity, the ordinal positions of the character states in which they match are displayed to allow for review and verification.

The use of binary codes in this SIrD also allows for updating when new discoveries are found. In Van der Linde and Houle (2008) discovery of new species of Drosophilade fruit flies in the Philippines, she was not able to classify them because of the presence of new characters. With the use of SIrD, new species can be added into the system without having to overhaul or reconstruct the arrangement of characters in the binary strings. New character states can be conveniently added at the end of the binary string without affecting the sequence of binary codes already entered into the database. The binary coding system is extensible enough to easily accommodate new additions.

CONCLUSION

The System of Identification for Drosophilidae offers an efficient way of identification as well as ease of updating for new discoveries. Its binary-based identification system allows straightforward computation of the degree of similarity which currently does not exist in other identification systems and verification of results. The high tolerance to errors by the software allows even novice users to retrieve possible matches. Time and resources consumed in doing research can, therefore, be reduced significantly.

APPENDIX

Appendix Table 1: Score card for the software system of identification for drosophilidae (SSIrD)

1	2	3	4	5
Slender species	Normally proportioned species	Relatively large species	Relatively small species	Lighter species
6	7	8	9	10
Yellowish species	Dark species	Pale yellow species	Blackish brown species	50% of ground color tan
11	12	13	14	15
25% of ground color tan	habitus with stripes	habitus with spots	face brilliant white	face brown
16	17	18	19	20
Face yellowish brown	Fronto-orbital region silver or whitish in color	Arista with only 1 to 2 ventral rays in addition to terminal fork	Arista with 3 branches above, excluding terminal fork	Arista with 4 branches, excluding terminal fork
21	22	23	24	25
Arista with 5 or more branches	3rd antennal segment covered with elongate setulae, roughly 3X the length of ground setulae	Palps light tan, sometimes darker at apex	Palpi yellow to yellowish brown	Palps dark brown to black
26	27	28	29	30
palps more slender and pointed at the tips	Maxillary palps inflated and broadly rounded at the apex	Gena narrow, about 1/4 the width of the eye	Width of gena about 1/3 width of eye	Gena dark brown
31	32	33	34	35
Gena yellowish	Gena yellowish brown to brown	Width of gena about 3/8 width for eye	Gena wide	Ocelli dark
36	37	38	39	40
Ocelli light	Ocellar triangle brown	Eye round in shape	Eyes bright red	Eyes radish
41	42	43	44	45
Eyes dark red or wine in color	Eyes nearly black	Eyes with distinct iridescent highlights	Diameter of eye about 5 times width of cheek	Vittae on head either unicolorous or not present
46	47	48	49	50
Multicolored vittae present on the head	Head with three vitae, one medial and two along the fronto-orbital plate which are bordered by dark brown or orange	Carina broad	Carina narrow	Subcarinal hairs present
51	52	53	54	55
Subcarinal hair absent	Entire frous of males silver to white	Frons brilliant white	Frons dull, dark brown	Frons gloss, with scattered setae that do not form a distinct V-shaped pattern
56	57	58	59	60
Frons distinctly pollinose	Frous with fine, white medial stripe anterior to ocelli	Frons lacking white medial stripe	Middle orbital smaller than the other two	Middle orbital bristle large
61	62	63	64	65
Middle fronto-orbital bristle about half the length of anterior orbital	Ouly one or second oral seta prominent	Second oral bristle less than 1/2 of the length of the first	Second oral bristle 1/2 or more than 1/2 of the length of the first	Second oral bristle equal in length to oral vibrissa
66	67	68	69	70
Anterior reclinate minute or absent	Anterior reclinate present, not minute	Proclinate ronghly eqnal in length to anterior reclinate	Proclinate orbital seta arises posterior to anterior reclinate	Proclinate typically arises anterior to or even with the anterior reclinate orbital

Appendix Table 1: Continue

71	72	73	74	75
Postocellar setae well developed	Postocellar setae small and inconspicuous, less than one-third	Achrostical setulae eularged at sutures	Acrostichal setulae in 2-4 rows	Acrostichal setulae in 6 or more rows
76	77	78	79	80
Dark brown stripes present inside each dorsocentral row	1 unpaired dark brown to black median streak or spot in front of anterior dorsocentral	1-2 pairs of dorsocentral setae	3 sets of dorsocentrals present	Mesonotum without markings
81	82	83	84	85
Mesonotum tan	50% of ground color tan	25% of ground color tan	ground color distinctly bluish gray	Ground color light yellowish to light brown
86	87	88	89	90
Mesonotum grayish	Mesonotum grayish brown	Mesonotum polished	Mesonotum unicolorous brown to black	Mesonotum pollinose
91	92	93	94	95
Mesonotum with 2 prominent stripes	Stripes on mesonotum indistinct	Mesonotum with 3 longitudinal stripes	Mesonotum with brown stripe or stripes	Mesonotum with indistinct grayish stripes
96	97	98	99	100
Mesonotum between the 2 middle stripes black	Spots in mesonotum sometimes fused in stripes or other patterns	Mesonotum indistinctly patterned	Mesonotum with dark brown spots	Pattern of pigment on mesonotum light
101	102	103	104	105
Mesonotum with dark gray spots	Seta on mesonotum arising from a single dark spot	Pleura with longitudinal stripes	Pleura without lateral stripes	Pleura yellowish
106	107	108	109	110
Pleura significantly darker than mesonotum	Pleura entirely dark brown to black	Pleura with distinct brownish stripe	Thorax light yellowish brown or dull yellow to brown in color	Thorax reddish brown dorsally
111	112	113	114	115
Thorax light grey to black	Thorax ground color has a distinct metallic tinge	Thorax with pattern	Pattern of pigment on thorax dark	Thorax with distinct pattern of tan, gray and black
116	117	118	119	120
Notum light orange	Vittae on notum either unicolorons or not present	Notum with 1 or more orange vittae present	Scutellum with white apical spot	Scutellum with very thick U-shaped dark marking
121	122	123	124	125
Scutellum with an X-shaped dark marking	Scutellum same colour as notum	Middle katepisternal seta either absent or mnch smaller than anterior and posterior katepisternals	3 subequal katepisternal setae present	Prescutellar setulae present
126	127	128	129	130
Sterno index 0.3 to 0.5	Sterno index 0.6 - 0.7	Sterno index 0.8-0.9	5X index 1.0-1.1	5X index about 1.3
131	132	133	134	135
Anterior scutellar setae divergent	Anterior scutellar setae convergent	Proclinate inserted posterior to anterior reclinate	Anterior reclinate inserted posterior to proclinate orbital	Spots on posterolateral corner of lateral areas irregularly triangular in shape

Appendix Table 1: Continue

136	137	138	139	140
Legs entirely yellowish	Legs yellowish tan to brown or radish brown	Legs pale grayish yellow	Legs brownish to black	Legs tinged with brown on apices of tarsi
141	142	143	144	145
Hooked setae present on mid-legs of males	Legs banded	Mid and hind femora and tarsi paler	Femura dark	Femur banded
146	147	148	149	150
Inner margin of femur with a row of stout, peg-like setae	Forefemur with short stout knob or tubercule located near middle of the posteroventral surface	Row of peg-like setae on Forefemur poorly developed	Tibia dark	Hind tibia darker near base
151	152	153	154	155
Mesothracic tibia with apical and preapical bristles	Tibia banded	Apical setae on 1st and 2nd tibia, preapicals on all 3	2 heavy setae on 1st tibia, 1 on 2nd, preapicals on all 3	1st tarsomere equal in length to 2nd tarsal segment
156	157	158	159	160
1st tarsal segment with single seta	1st tarsal segment with 2 rows of sex comb	1st tarsal segment with 3 or 4 rows	1st tarsal segment with 5 rows of setae	2-3 rows of sex comb on 2nd tarsal segment
161	162	163	164	165
Sex comb on 2nd tarsomere with 4-17 teeth	Sex comb on 2nd tarsomere with more than 17 teeth	3 or 4 rows of sex comb on 2nd tarsal segment	Single seta present on 3rd tarsomere	2 rows of setae on 3rd tarsal segment
166	167	168	169	170
2 teeth on 3rd tarsal segment	Foretarsi with some degree of modification	Long recurved hairs on medial side of male foretarsi	Sex comb on foretarsus composed ouly of 1 or 2 setae	Distal tarsal joints black
171	172	173	174	175
Terminal tarsal segments slightly darker than the remainder of the leg	Basal branch of posterior paramere long	Proximal sex combs with 10 or less teeth	Proximal sex combs with greater than 10 teeth	Basitarsus distinctly longer than 2nd tarsal segment
176	177	178	179	180
Basitarsi dark	Basitarsus with single row of sex comb	Basitarsus with 2-3 transverse rows of sex combs	Sex comb covering ouly tip of the the basitarsus	Sex comb covering 2/3 of basitarsus
181	182	183	184	185
Apex of basitarsus with claw-like setae	Distal section of basitarsus with stout setae	Basitarsal sex comb small	Basitarsasl sex comb large	2 long basal teeth and 5 long distal teeth on basitarsal segments
186	187	188	189	190
3 teeth on basitarsal sex comb	3 teeth on each row of basitarsal sex comb	4-6 teeth on each basitarsal sex comb	7-10 teeth on basitarsal sex comb	Basitarsal sex comb with 11-20 teeth
191	192	193	194	195
Basitarsal sex comb with 21-29 teeth	Basitarsal sex comb with 30 or more teeth	Sex comb on basitarsus ouly	Sex comb on basitarsus and 2nd tarsomere	Sex comb on 1st and 2nd tarsal segments
196	197	198	199	200
Sex comb present on 1st and 3rd tarsomeres	Distal sex comb less than 10 teeth	Distal sex comb with 9-13 teeth	Sex combs long	Sex combs short

Appendix Table 1: Continue

201	202	203	204	205
Sex comb transverse	Sex comb longitudinal	Small teeth on sex comb	Large teeth on sex comb	Sex comb with only 2 strong setae
206	207	208	209	210
Sex comb composed of 2-3 stout, claw-like setae on apex of the basitarsus	Sex comb composed of a single row of 8-9 stout setae inserted on distal section of basitarsus	Sex combs in 2 rows, each with about 3 setae	Row of peg-like setae well developed	Abdomen dark yellow to rofous
211	212	213	214	215
Abdomen yellow with indistinct brown basal bands	Abdomen of male entirely pale brown or yellowish brown	Abdomen polished, unicolorous dark brown, gray to black	Abdomen black posteriorly	Ground coloration of abdomen has a distinct metallic tinge
216	217	218	219	220
Abdominal tergites with dark posterior bands in both sexes	Abdomen lighter on basal segment in midline	Opaque areas present on tergites of females	Basal 2 tergites yellowish	Abdominal tergites broken or interrupted at the mid-dorsal line, this characteristic is often difficnlt to see in dark species
221	222	223	224	225
2nd to 5th abdominal tergites with posterior dark bands not broken in mid-dorsal line	Abdominal tergites Brown with dark posterior bands in both sexes	6th tergite of female with opaque area in each side, near the posterior margin	Ventral margin of 9th tergite with prominent posteroventral projections	ventral margins of the 9th tergite nearly vertical
226	227	228	229	230
Spiracles 6 and 7 always found at ventral or posteroventral margin of tergite six	Spermatheca with invagination	Spermatheca with thin invagination	Spermatheca with long invagination	Spermatheca with collar at base
231	232	233	234	235
Sternites lack pigment or are light yellow brown	Sternites with heavy black pigment	Sternite 4 square and intact, not split	Sternite 5 square and intact, not split	Sternite 5 rectangular and split medially
236	237	238	239	240
6th sternite with thin, deeply incised V-shaped penis	Median region of 6th sternite broad	6th sternite of females somewhat pointed at apices	6th sternite of females rounded at apices	7th sternite of females pointed at apices
241	242	243	244	245
7th sternite with sharply cut V-shape median	7th sternite with a more U-shaped median	Lateral areas of tergites with light grayish areas near lateral margin, especially on 1st 3 tergites	Posterior margins or tergites with sparce rows of large setae	Sparce rows of large setae absent from posterior margins of abdominal tergites
246	247	248	249	250
6th tergite of female without a median opaque area	Abdominal bands distinct	Abdominal bands less distinct	Abdominal banding light and diffuse	Abdominal banding blackish in a complex pattern created by a dark

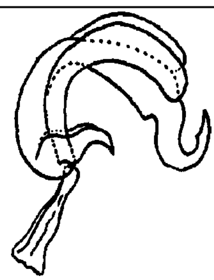
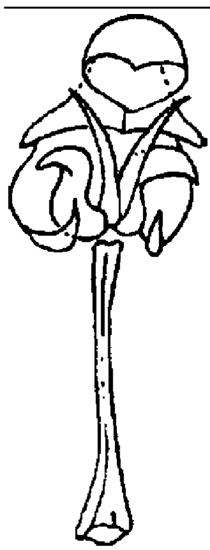
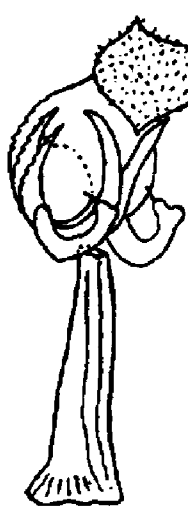
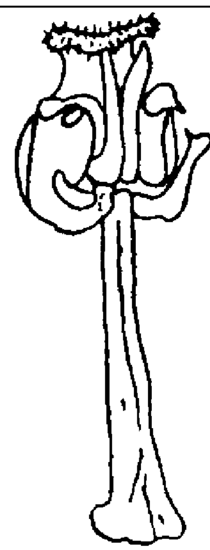
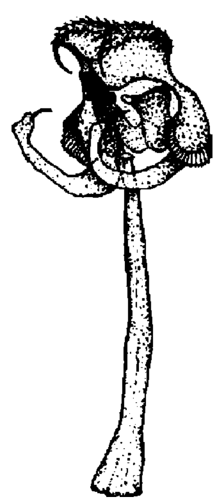
Appendix Table 1: Continue

				marginal bar that is thickest in the posterior corner of tergite 2,3 and 4 and usually connects with posterior band
251	252	253	254	255
Abdominal tergites with a very faint diffuse apical band	Dark banding appears at 2nd abdominal segment	Some abdominal bands show lightened areas on mid-dorsal line	Dark abdominal bands reach lateral margius	Dark abdominal banding thins out toward lateral margins
256	257	258	259	260
Abdomen in males and females elongate	Abdominal tergites with a very faint diffuse apical band	Apical bands on abdomen dark and distinct	Apical band on tergites greatly widened to cover nearly all of the lateral area	Apical bands of abdominal segments 3/4 or greater of the width of the tergite on either side of the mid-dorsal line
261	262	263	264	265
Apical bands of abdominal segments narrow, half or less of the width of the tergite on either side of the mid-dorsal line	All lateral areas of abdominal segments almost completely covered by expansions of apical bands, solid color without internptions or light areas	Some or all lateral areas of abdominal segments from the more medial portion of the apical band near the angle of the tergite, lateral areas not completely covered with dark markings	Lateral areas of abdominal tergites much darker than medial portious of apical bands	Color pattern in lateral areas of abdominal Segments faint, faded and diffuse
266	267	268	269	270
Color pattern in lateral areas of abdominal segments brown or black, distinct	Lateral areas of abdominal tergite more or less solid but with distinctly lighter areas	Lateral areas of abdominal tergite broken into distinct spots	Abdomen with distinct pattern	Abdomen with some pattern
271	272	273	274	275
Dark pigment on dorsal surface of abdomen forms distinct spots or dots	Dark markings are contiguous from irregularly triangle-shaped region in posterolateral corner of abdominal tergites to the margin of tergite	Dark makings are only narrowly contiguous from triangle shaped region in posterolateral corner of abdominal tergites to margin of tergite	Darkk brown spot present only on tergite 6	Triangular areas in posterolateral corners about the saine shade or lighter than apical bands on median sections of tergite, these regious are joined by a narrow band
276	277	278	279	280
Primary claspers present	Secondary claspers present	Secondary claspers seta large	Secondary claspers either partially or completely fused to anal plate	Secondary claspers connected to anal plate via a thin, membranous connection
281	282	283	284	285
Secondary claspers partially connected, but without membranous connection	Secondary claspers elongate and ovoid with 1 large curved medial seta	Secondary claspers nearly circnlar, with a cluster of 7-15 setae	Secondary claspers nearly oval in shape	Secondary claspers with a single large, black seta

Appendix Table 1: Continue

286	287	288	289	290
Secondary claspers has 2 large setae, roughly equal in size	Secondary claspers with 2-3 large setae, dorsally larger	Secondary clasper lacks large lateral setae, medial setae only present	Secondary clasper posseses prominent lateral setae, each roughly equal in size to the typical medial setae on this structure	Secondary clasper large with ventral projection
291	292	293	294	295
Secondary clasper with large medial seta inserted dorsally	Secondary calsper setae 2; 1 large, sub-equal in size and 3rd seta smaller	Secondary clasper with 4 or less medial setae	Secondary clasper with 2 lateral setae	Secondary clasper with 3 lateral setae
296	297	298	299	300
Secondary clasper on anal plate with 2 heavy,elougate setae at tip	Secondary clasper with 4 medial and 2 lateral setae	Secondary clasper with 5 medial setae	Surstylus C-shaped	Surstylns large and cu-sized, laterally composed and containing a very short and square looking comb with 6-8 prensisetae
301	302	303	304	305
Surstylns flatter	Surstylns with a row of 4 teeth	surstylus with a row of 5-6 teeth	Surstylus with 7 prensisetae	Surtylus with only 2 regions of prensisetae
306	307	308	309	310
Surtylus with 3 distinct regions of prensisetae	Primary surstylus with ventrolateral comb of long black setae, only a few teeth dorsolaterally	Primary surstylns large, with several sets of distinctly different teeth	Prensisetae on surstylus arranged in a concave row	Prensisetae on surstylus arranged in a straight row
311	312	313	314	315
Prensisetae 6 or less in surstylus	Lower portion of genital arch densely setose	Lower portion of genital arch with few setae	Anal plates knee-shaped in lateral view, flat, ventral process	Anal plate with cilia or setae
316	317	318	319	320
Anal plate with a series of thickened spines	Anal plate with dense setae especially at the apex	Anal plate with large black seta	Anal plates with strong teeth	Anal plate with 1 very large ventro-medial tooth
321	322	323	324	325
Anal plate with 2 prominent setae	Anal plate with 1, upwardly curved spine	Anal plate partially fused to genital arch	Anal plates with ventral process or stroug teeth	Hypandrium with large hook-shaped process on inner margin
326	327	328	329	330
Inner margin of hypandrium lackig large, hook-shaped process, with only a small tubercule	Hypandrium with 7 or less setae	Hypandrium with 8 or more setae	Testes yellow to blnnt orange in color	Testes red
331	332	333	334	335
Testes large	Testes short and only slightly coiled	Ventral margin of cercus with patch of short dense setulae	Ventral margin of cercus with long setulae	External process of epandrium ronnded and bulging at base and drawn out into a thin projection

Appendix Table 1: Continue

336	337	338	339	340
Epandrial expansion trapezoid in shape	Epandrium with posterior expansions	Epandrial expansion large, semicircular, clearly viewed laterally	Epandrial expanion longer than wide	Epandrial expansion wider than long
341	342	343	344	345
Epandrial expansion narrow, finger like	Epandrial expansion elaborate, sinous and slightly eularged at apex and pointed ventrally	Epandrium with a small, triangular expansion inserted anterior to clasper	Epandrial lobes symmetrical, short and narrowed to thin points	Epandrial lobes assymetrical with left lobe longer than the right
346	347	348	349	350
Epandrial lobe large and elaborate	Epandrial lobes short and rounded with 6 or 7 prensisetae	Aedeagus with lateral expansions	Aedeagus laterally flattened, at least apically and are approximately bilobed at least partially	Tip of eadeagus sharply cut
351	352	353	354	355
Tip of eadeagus slender	 aedeagus as in Fig. A	Paraphyses slender	Paraphyses broadened	Penis cylindrical with lateral flaps
356	357	358	359	360
 Penis as in Fig. B	 Penis as in Fig. C	 Male genitalia as in Fig. D	 Male genitalia as in Fig. E	Phallus pointed

Appendix Table 1: Continue

361	362	363	364	365
Phallus truncated, boomerang-shaped with small spines	Phallus slender, curved dorsally	Phallus large, straight, gutter-like with saw-shaped fringes	Basal branches of posterior paramere very short	Basal branch of posterior paramere long and serrated
366	367	368	369	370
Basal branch of posterior paramere pointed apically	Wings hyaline	Wings hyaline along coastal margin	Apex of the costal segment jet black	Costal index greater than 3.0
371	372	373	374	375
Costal index 3.0 or less	Wings with distinct infuscations	Wings of male infuscated	Heavy infuscatious at the tips of each long vein at the wing apex	Crossveins infuscated
376	377	378	379	380
Apex of wing infuscated	Eularged setae present at regular intervals on the costa	Setulae on costal margin darker and thicker than typical	3rd costal section with heavy setulae on basal 1/3	Basal 2/3 of 3rd costal section with heavy seta
381	382	383	384	385
3rd costal section with heavy setae on basal half or more	3rd costal section with heavy setae on basal 3/4	3rd costal section with heavy setulae on basal 1/5, 2/5 or more	Heavy setulae on basal 3/5 or more of 3rd costal section	Apices of 3rd and 4th longitudinal veins distinctly convergent
386	387	388	389	390
Apices of 3rd and 4th longitudinal veins distinctly divergent	Discal and 2nd basal cells separated by a rudimentary vein	Discal and 2nd basal cells united	Costal lappet present at subcostal break	Crossveins hyaline

391	392	393	394	395	396
Ouly posterior crossveius clouded	Crossveins clouded	Posterior crossvein slightly curvate	Posterior crossvein strongly curvate	Crossveins ouly narrowly clouded	Anterior crossvein black

General body characteristics, 1-4: General size, 5-11: Color, 12-13: Pigmentation, Head characters: 14-17: Face and fronto orbital region, 18-21: Arista, 22: Antennae, 23-27: Palps, 28-34: Gena, 35-37: Ocelli, 38-44: eyes, 45-47: Vittae, 48-61: Carina, 62-65: Oral bristle, 66-67: Anterior reclinate, 68-70: Proclinate, 71-72: Postocellar setae, Thorax: 73-79: Achrostical setulae, 80-102: Mesonotum, 103-108: pleura, 109-115: Thorax general color and patteru, 116-118: Notum, 119-122: Scutellum, 123-124: Katepisterual setae, 125: Prescutellar setulae, 126-130: Sterno index, 131-132: Scutellar setae, 133-134: Position of reclinate, 135: Distinct spots, Legs: 136-142: General color and characteristics, 143-148: Femur, 149-154: Tibia, 155-166: Tarsals, 167-169: Foretarsus, 170-171: Distal tarsals, 172: Posterior paramere, 173-174: Proximal sex comb, 175-196: Basitarsus, 197-198: Distal sex comb, 199-209: Sex comb and teeth, Abdominal characters: 210-215: General characteristics of the abdomen, 216-223: Abdominal tergites, 224-225: Ventral margin, 226: Spiracles, 227-230: Spiracles, 231-232: Sternites, 233: 4th sternite, 234-235: 5th sternite, 236-239: 6th sternite, 240-242: 7th sternite, 243-246: Tergites, 247-255: Abdominal banding, 256: Abdominal shape, 257-261: Apical banding, 262-268: Lateral areas of the abdomen, 269-275: Markings on the abdomen, 276: Primary claspers, 277-298: Secondary claspers, 299-308: Surstylus, 309-311: Preusisetae, 312-313: Genital arch, 314-324: Anal plate, 325-328: Hypandrium, 329-332: Testes, 333-334: cercus, 335-347: Epandrium, 348-352: Aedeagus, 353-354: Paraphyses, 355-359: Penis, 360-363: Phallus, 364-366: Paramere, Wings: 367-371: General characteristics of the wings, 372-376: Infuscatious, 377-384: Costal cell setulae, 385-386: Longitudinal veins, 387-388: Basal cells, 389: Costal lappet, 390-396

REFERENCES

Abdulrahaman, A.A., I.B. Asaju, M.O. Arigbede and F.A. Oladele, 2010. Computerized system for identification of some savanna tree species in Nigeria. J. Horticul. For., 2: 112-116.

Cerretti, P., H.P. Tschorsnig, M. Lopresti and F. di Giovanni, 2012. MOSCHweb-A matrix-based interactive key to the genera of the palaearctic tachinidae (Insecta, Diptera). ZooKeys, 205: 5-18.

Dallwitz, M.J., T.A. Paine and E.J. Zurcher, 2000. Principles of interactive keys. http://delta-intkey.com/www/interactivekeys.January 1, 2013. htm

Harrison, R.A., 1950. New Zealand drosophilidae (Diptera): I-introduction and descriptions of domestic species of the genus *Drosophila* fallen. Trans. Royal Soc. New Zealand, 79: 505-517.

Jaccard, P., 1901. Distribution de la flore alpine dans le bassin des Dransesetdansquelquesregionsvoisines. Bull. de la Societe Vaudoise des Sci. Naturelles, 37: 241-272.

Kirchoff, B.K., D.L. Remington, L. Fu and F. Sadri, 2008. A new type of image-based key. Proceedings of the International Conference on BioMedical Engineering and Informatics, Volume 1, May 28-30, 2008, Sanya, Hainan, China.

Markow, T.A. and P.M. O'Grady, 2006. Drosophila: A Guide to Species Identification and Use. Elsevier Inc., New Yok, Pages: 247.

Norton, G., 2005. Invasive species: The role of lucid identification keys. Centre of Biological Information Technology, The University of Queensland. http://www.agnet.org/htmlarea_file/library/20110712090617/eb561b.pdf

Osborne, D.V., 1963. Some aspects of the theory of dichotomous keys. New Phytologist, 62: 144-160.

Rocker, J., C.M. Yauch, S. Yenduri, L.A. Perkins and F. Zand, 2007. Paper-based dichotomous key to computer-based application for biological identification. J. Comput. Sci. Coll., 22: 30-38.

Smith-Akin, K.A., S. McLane, T.M. Craig and T.R. Johnson, 2006. Application of cognitive engineering principles to the redesign of a dichotomous identification key for parasitology. Proceedings of the American Medical Informatics Association Annual Symposium, November 11-15, 2006, Washington, DC., pp: 739-743.

Van der Linde, K. and D. Houle, 2008. A supertree analysis and literature review of the genus *Drosophila* and closely related genera (Diptera, Drosophilidae). Insect Syst. Evol., 39: 241-266.

Wheeler, M.R., 1952. A key to the genera of Drosophilidae of the pacific Islands (Diptera). Proc. Hawaiian Entomol. Soc., 14: 421-423.