The mitochondrial *COI* gene fails as DNA barcoding in the sibling species of *Drosophila buzzatii* cluster.

Santos, M.H.[1], F.F. Franco[2], M.H. Manfrin[1,3] 1.Depto. Genética - FMRP-USP. 2.UFSCar – Campus Sorocaba. 3.Depto. Biologia - FFCLRP-USP.

Abstract

The DNA barcoding using part of the *COI* region from the mitochondrial DNA was proposed to be a faster alternative method to identify species. In this study, we used this methodology to identify the seven cryptic and cactophilic species of the *Drosophila buzzatii* cluster. The identification of these species is made by comparisons of the aedagus morphology, the male reproductive structure, that make impossible to identify immature forms and females. Using three methods of molecular classification (Maximum Likelihood, Maximum Parsimony and Neighbor-Joining), we test the DNA barcoding in identification of 48 samples from these species and *D. martensis* and *D. richardsoni* as outgroups. Our results showed that the use of the DNA barcoding in the fast identification of the *D. buzzatii* species has several limitations and must be used with caution.

Introduction

The identification of species is a ubiquitous problem, considering the different fields and biological groups. The use of DNA-based analysis for species identification, the so called DNA barcoding, was proposed as an alternative path for taxonomic identification and could allow a faster characterization of the species biodiversity than the taxonomic traditional methodologies (Hebert *et al.,* 2003). For animal taxa, fragments of the *cytocrome C oxidase subunit I* (*COI*) mitochondrial gene has gained the designation of "global standard" sequence as the barcoding region for animals (Hebert *et al.,* 2003; Hajibabaei *et al.,* 2007). However, the use of DNA barcoding is controversial, including arguments agreeing (Hebert *et al.,* 2003; Tautz *et al.,* 2003; Hajibabaei *et al.,* 2007) and against (Will and Rubinoff, 2004; Meier *et al.,* 2006; Solé-Cava, 2008) the use of this technique in modern systematics.

The *Drosophila buzzatii* cluster (*D. repleta* group, *D. buzzatii* complex) is composed of seven sibling and cactophilic species: *D. buzzatii*, *D. serido, D. antonietae, D. seriema*, *D. gouveai*, *D. borborema,* and *D. koepferae* (see Manfrin and Sene, 2006). The monophyly and the intra cluster relationship were proposed based on chromosomal inversion and molecular markers (Ruiz *et al.,* 1982; Ehrman and Wasserman, 1987; Tosi and Sene, 1989; Ruiz and Wasserman, 1993; Manfrin *et al., 2001*; Manfrin and Sene, 2006). The cluster is endemic of South America, with the exception of *D. buzzatii* species, which was introduced to other continents along with its host cactus *Opuntia ficus-indica* (Barker *et al.,* 1985).

The species of the *D. buzzatii* cluster are cryptic and their identification is done using comparative analysis of the morphology of the male reproductive apparatus, the aedagus (Vilela and Sene, 1977). This fact has at least three consequences: 1) the immature individual could not be identified; 2) the female identification is possible only by their male progeny with the establishment of female isolines at the laboratory; 3) differences between species are obtained only by morphometrics analyses.

Considering the potential of *COI* gene as DNA barcoding (Hebert *et al.,* 2003; Tautz *et al.,* 2003; Hajibabaei *et al.,* 2007), we test this gene as diagnostic character in *D. buzzatii* cluster and, moreover, it was briefly discussed its use for barcoding in the *Drosophila* species.

Material and Methods

A total 576bp from the 5' end of the *COI* mitochondrial gene of 48 flies from *D. buzzatii* cluster were analyzed: *D. koepferae* (1), *D. antonietae* (7), *D. borborema* (10), *D. buzzatii* (10), *D. gouveae* (6), *D. serido* (8), and *D. seriema* (4). This sample includes 38 *COI* sequences isolated in previous works (Manfrin *et al.,* 2001; Franco 2009) and 10 new sequences. We also included in the analysis *Drosophila martensis* (1) and *D. richardsoni* (1) as outgroups. The template for the mitochondrial DNA was obtained isolated by PCR using the primers TY-J-1460 (Simon *et al.,* 1994; De Brito *et al.,* 2002) and 2191r (De Brito *et al.,* 2002). PCR products were sequenced using the ABI PRISM BigDyeTM Terminator Cycle Sequencing Ready Reaction Kit in an ABI automatic sequencer. The DNA sequences were checked by eye and aligned using the program BioEdit v. 7.0.9.0 (Hall, 1999). The data were analyzed using three algorithms. The first one was made by the program TCS v 1.21 (Clement *et al.,* 2000). This program uses statistical parsimony to construct a haplotype network. The second was a Maximum Parsimony tree (100000 reply) performed by the program PAUP v. 4.0b10 (Swofford, 2001). The third one was a Neighbor-Joining tree (10000 reply) performed by the program Mega v. 4.1 (Tamura *et al.,* 2007).

Results and Discussion

Thirty two (32) haplotypes were generated in network analysis. These haplotypes were allocated in three clades and three isolated haplotypes (Figure 1). The Clade 01 is composed only by samples of *D. buzzatii* species. The Clade 02 comprises all the samples from *D. antonietae* and two *D. gouveae* samples. In the Clade 03 was allocated samples of *D. gouveai*, *D. seriema*, *D. borborema,* and *D. serido* species. This result suggests the absence of diagnostic character for *D. buzzatii* cluster species in this genetic region. *Drosophila koepferae, D. martensis, and D. richadsoni* formed isolated clades.

The MP and NJ trees presented similar topologies, and in both of them we confirm the monophyletic status of the *D. buzzatii* cluster with high bootstrap values (Figures 2 and 3). Two main *COI* lineages were formed within the *D. buzzatii* cluster clade. The first group is formed by *D. buzzatii* and *D. koepferae* (bootstrap value of 79%), and *D. buzzatii* clearly was separated from *D. koepferae* with bootstrap value of 100% (Figure 2). The second subdivision could be divided in two considering the bootsprap values. One division comprises the same species of the Clade 02 generated by the TCS program, with bootstrap values of 91, including all the *D. antonietae* species and three *D. gouveai* samples. The second division is a polytomy including all other species of the group; in this polytomy a division with high bootstrap support (90%) includes the subsample of *D. seriema*, *D. serido,* and *D. gouveai*.

Neighbor-Joining analyses show the same consistency considering the division of the *D. buzzatii* cluster from the external group with a bootstrap value of 100 (Figure 3). Two main groups inside *D. buzzatii* cluster were formed: one containing *D. buzzatii* and *D. koepferae* samples (74% bootstrap) and the other with *D. antonietae*, *D. gouveai*, *D. seriema*, *D. serido,* and *D. borborema* (bootstrap 99). *Drosophila buzzatii* and *D. koepferae* comprised a separate group (bootstrap 100). The samples of *D. antonietae* and three samples of *D. gouveae* were grouped together (bootstrap 92),

as in the TCS and Maximum Parsimony analyses (Figures 1 and 2). The *D. borborema* species encompasses a group with one individual of *D. seriema* (bootstrap value of 63%). The species *D. seriema*, *D. serido,* and *D. borborema* could not be separated.

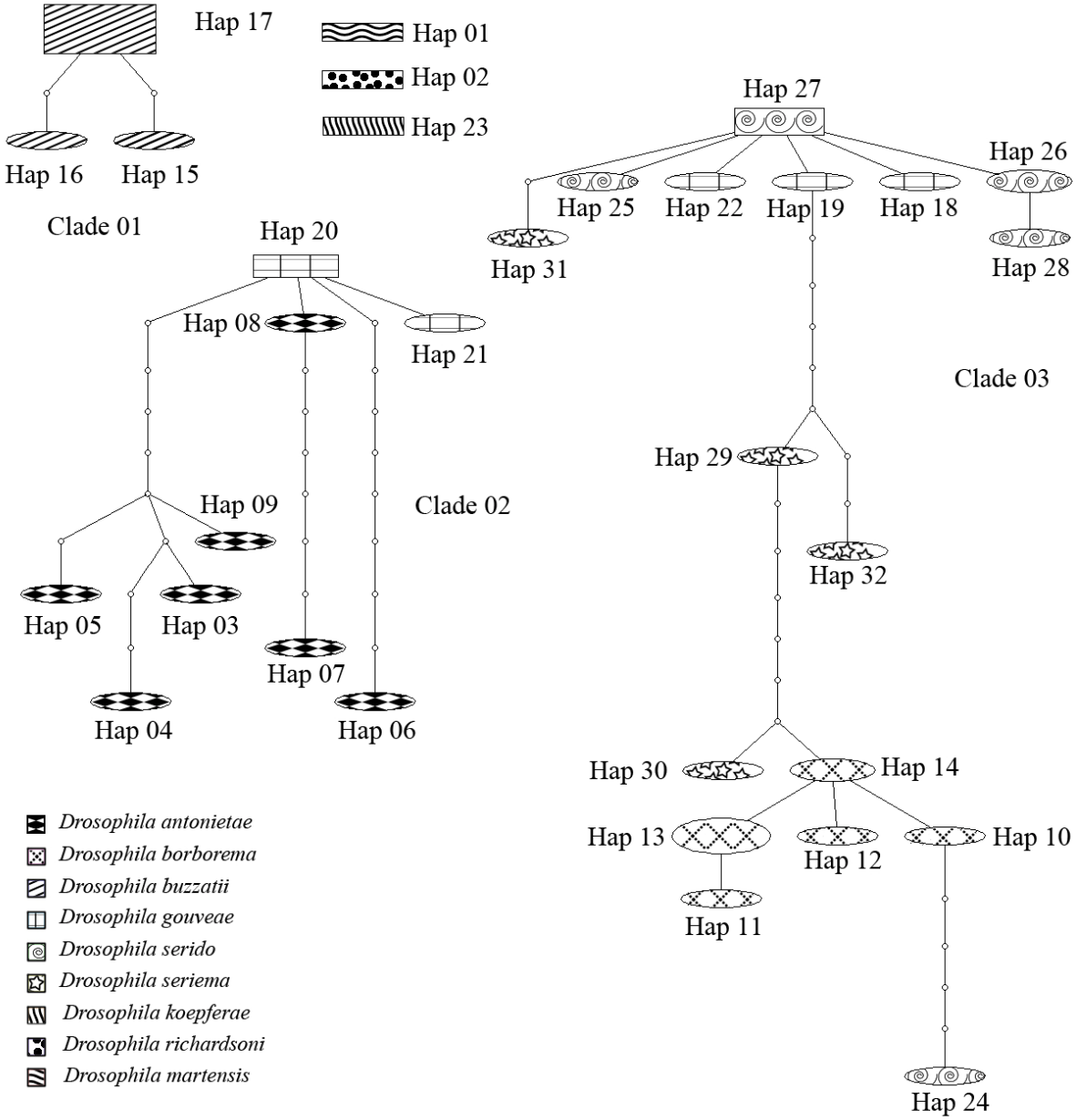


Figure 1. Haplotype network generated by the TCS program. The size of the ellipses represent the frequency of the haplotype. Retangular forms represents ancestral haplotype and the white dots the supposed and not found haplotype.

The results of these analyses are similar. The samples of *D. buzzatii* species, always grouped together as a cohesive entity, even being from different populations, and is more basal on the topology. This is according to other sets of data as morphology and molecular data (Manfrin and

Sene, 2006) that considering this species the more differentiated and most basal composing a more ancient linneage in the cluster (Manfrin and Sene, 2006). In this case, the information in the *COI* mtDNA is a good diagnostic and could be used as a DNA barcoding.

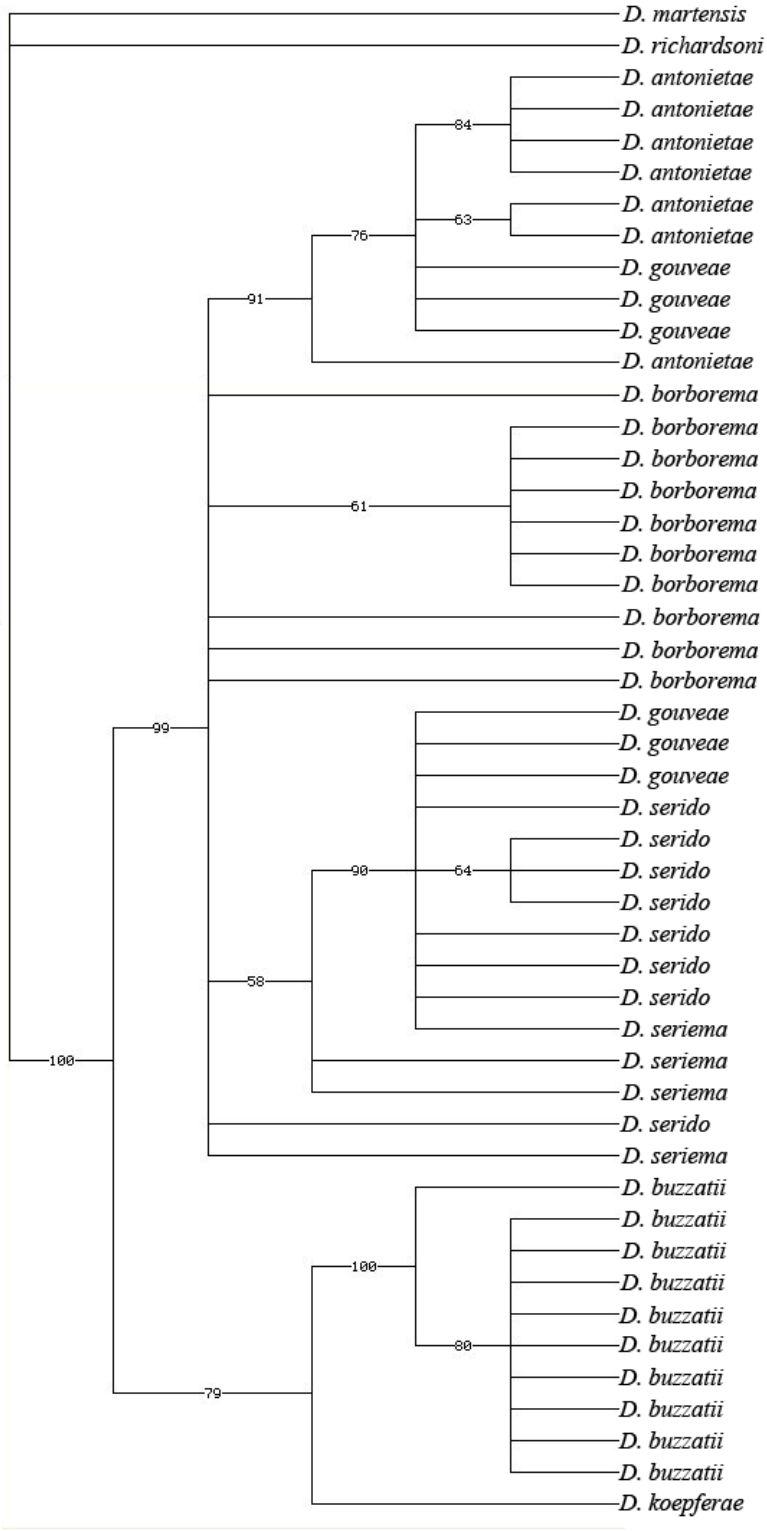
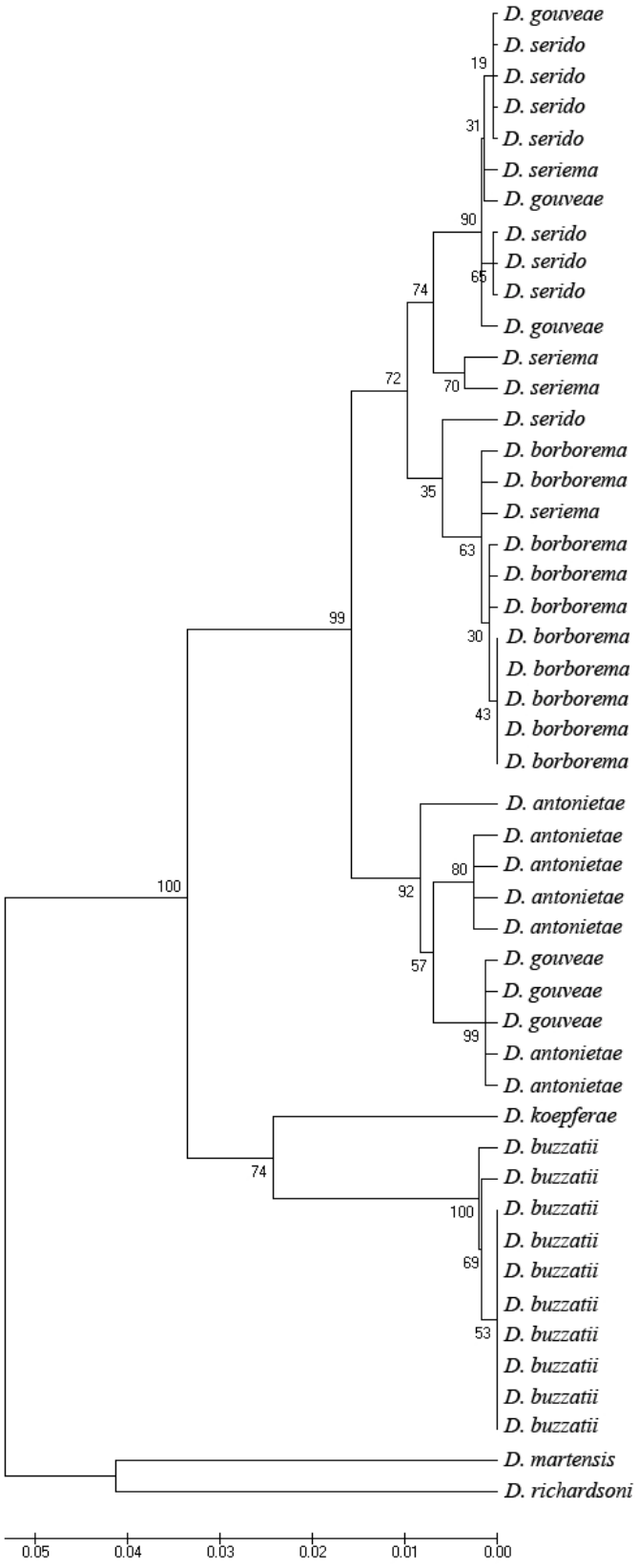


Figure 1. Parsimony tree generated by the program PAUP, using the Maximum Parsimony approach. The bootstrap values were obtained using 10000 replications.

The sample for *D. koepferae* was small and it is a limit to discuss the use the *COI* sequence as a DNA barcoding for this species. However, considering the entire previous data that proposed that this species is older in the cluster (Manfrin *et al.*, 2006), we believe that the data are a good indicator for the presence of diagnostic characters in sequence and it could be used as a DNA barcoding.

The results for *D. antonietae* samples from different geographic areas showed that they encompass a consistent group indicating that the sequence of *COI* has diagnostic character for the species being a DNA barcoding. The hypotheses to explain the clustering of the sequences of three individuals of *D. gouveai* together with *D. antonietae* is asymmetrical introgressive hybridization resulting from secondary contact between populations of this species (Manfrin *et al.*, 2001). This information confirmed that mtDNA sequences can have diagnostic character for one species, but we have to use it with caution.

The species *D. borborema, D. serido, D. seriema,* and *D. gouveai* were grouped together but with unresolved relationships



among them. These results show that the sequences of *COI* could not be used as a DNA barcoding for these species. One possible explanation for this fact is incomplete lineage sorting as the cladogenetics events are recent for this species and do not have time to acquire reciprocal monophyly (Manfrin *et al.,* 2001; Franco, 2009), this way, to have a DNA barcoding is a time depending event. As these species are sympatric in several locations we also could assume that the sharing of mtDNA sequences could be the result of introgression. In this sense, other molecular markers are required to evaluate introgressive events and give support or not to the sequences of *COI* to be used as DNA barcoding. Besides this possibility, assuming that the divergent times of cladogenetics events among these species are recent (3-6 Myr according to Manfrin *et al.,* 2001), the sharing of DNA sequences could be an ancestral polymorphism shared by those species which obscures any historical patterns maintained in the nuclear genome.

Figure 2. Similarity tree generated by the program Mega, using the Neighbor-Joining approach and bootstrap values using 10000 replications.

In conclusion, our data suggest that the mt*COI* has a limited use as a DNA barcoding and should be used with caution. In the *D. buzzatii* cluster these sequences have diagnostic information allowing its use as a DNA barcoding in species that split for long evolutionary period of times, but not when they are recent. Moreover, the so common event of introgression (Manfrin *et al.,* 2001) should be considered when DNA barcoding testing is done, which could generate many erroneous results and could not be performed for these purposes.

Acknowledgments: We like to thanks CAPES, CNPq and FAPESP for financial support, Profa. Dra. Nilce Maria Martinez Rossi (FMRP/USP) for making available their laboratory to perform the sequencing reactions and Mendelson Mazucato (FMRP/USP) for your essential assistance in the sequence process.

References: Barker, J.S.F., F.M. Sene, P.D. East, and M.A.Q.R. Pereira 1985, Genética 67: 161-170; Clement, M., D. Posada, and K.A. Crandall 2000, Molecular Ecology 9(10): 1657-1659; De Brito, R.A., M.H. Manfrin, and F.M. Sene 2002, Genetics and Molecular Biology 25(2): 161-171; Ehrman, L., and M. Wasserman 1987, Evolutionary Biology 21: 1-20; Franco, F.F., 2009, História evolutiva do "cluster" *Drosophila buzzatii* (grupo *D. repleta*): eventos históricos e diversificação de espécies no Brasil. Ribeirão Preto, Universidade de São Paulo Faculdade de Medicina de Ribeirão Preto: 197; Hajibabaei, M., G.A.C. Singer, P,D.N. Hebert, and D.A. Hickey 2007, Trends in Genetics 23(4): 167-172; Hall, T., 1999, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT; Hebert, P.D.N., A. Cywinska, S.L. Ball, and J.R. DeWaard 2003, Proceedings of the Royal Society of London Series B-Biological Sciences 270(1512): 313-321; Manfrin, M.H., R.A. De Brito, and F.M. Sene 2001, Annals of the Entomological Society of America 94(3): 333-346; Manfrin, M.H., and F.M. Sene 2006, Genetica 126(1-2): 57-75; Meier, R., K. Shiyang, G. Vaidya, and P.K. Ng 2006, Syst. Biol. 55(5): 715-28; Ruiz, A., A. Fontdevila, and M. Wasserman 1982, Genetics 101: 503-518; Ruiz, A., and M. Wasserman 1993, Heredity 70: 582-596; Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu, and P. Flook 1994, Annals of the Entomological Society of America 87(6): 651-701; Solé-Cava, A.M., 2008, Código de Barras do DNA: o rabo que abana o cachorro. Ciência Hoje. Brasil, Ciência Hoje. 41: 65-67; Swofford, D., 2001, PAUP* 4.0 beta 5: Phylogenetic Analysis Using Parsimony (and other methods), Sinauer Associates Sunderland, MA; Tamura, K., J. Dudley, M. Nei, and S. Kumar 2007. Molecular biology and evolution 24(8): 1596; Tautz, D., P. Arctander, A. Minelli, R.H. Thomas, and A.P. Vogler 2003, Trends in Ecology & Evolution 18(2): 70-74; Tosi, D., and F.M. Sene 1989, Revista Brasileira de Genética 12: 729-745; Vilela, C.R., and F.M. Sene 1977, Papéis Avulsos de Zoologia 30: 295-299; Will, K.W., and D. Rubinoff 2004, Cladistics-the International Journal of the Willi Hennig Society 20(1): 47-55.

Studies on Drosophilids (Diptera: Drosophilidae) of Gujarat State in India.

Kumar, Rabindra, and Ajai Kumar. Genetics and Molecular Biology Laboratory, Department of Zoology, Feroze Gandhi College, Raebareli-229001; E-mail: rabindra_kumar10@yahoo.com; ajaikumar_rbl2002@yahoo.com

Introduction

During recent years, considerable data have been accumulated regarding faunal composition of drosophilid species as a result of extensive field collections in different ecological habitat by Ayala (1970). The drosophilid flies thus obtained have been utilized for various studies *viz*., taxonomic, ecological, genital, behavioral, and its distribution record. As a result more than 290 drosophilid species have been reported so far from different ecogeographical areas in India (Gupta, 2005), and most of them are new to the world of science. These data prompted us to undertake such studies from previously unexplored forest region of Gujarat state in India.