INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

MOLECULAR PHYLOGENY OF THE *DROSOPHILA MELANOGASTER* SPECIES GROUP, WITH SPECIAL EMPHASIS ON THE *MONTIUM* SUBGROUP

by

Valerie Ann Schawaroch

A dissertation submitted to the Graduate Faculty in Biology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, the City University of New York

2000

This manuscript has been read and accepted for the Graduate Faculty in Biology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

7/20/2000
Date

Chair of Examining Committee
Dr. Gail Simmons, City College

7/25/2000
Date

Executive Officer
Dr. Richard L. Chappell

Dr. Tadmiri Venkatesh, City College

David Grimaldi, American Museum of National History

Dr. Rob DeSalle, American Museum of Natural History

Dr. Carole Griffiths, Long Island University

Dr. Linda Strausbaugh, University of Connecticut (Reader)

The City University of New York

Abstract

MOLECULAR PHYLOGENY
OF THE *DROSOPHILA MELANOGASTER* SPECIES GROUP,
WITH SPECIAL EMPHASIS ON THE *MONTIUM* SUBGROUP

by

Valerie Ann Schawaroch

Adviser: Dr. Gail M. Simmons

Drosophilids, most notably *Drosophila melanogaster*, are the most well studied and understood eukaryotic organisms in genetics; however, the taxonomic relationships of Drosophilids are poorly understood. *Drosophila melanogaster* is one of 174 species + 3 subspecies in the *melanogaster* species group. Recently (Ashburner *et al.* (1984), Pélandakis *et al.* (1991) and Pélandakis and Solignac (1993) proposed for 8 of the 12 species subgroups within the *melanogaster* species group three major unresolved lineages: the *ananassae* subgroup, the *montium* subgroup, and the *melanogaster* + Oriental subgroups. The Oriental subgroups are the *elegans, eugracilis, ficusphila, suzukii*, and *takahashii* subgroups. This dissertation employs a cladistic analysis of DNA sequence data in an effort to resolve relationships within the *melanogaster* species group and its largest species subgroup – *montium* that contains 87 species.

Initially, 5 different gene regions (alcohol dehydrogenases [*Adh*], cytochrome oxidase II [*co ii*], hunchback [*hb*], 28S ribosomal DNA and 16S ribosomal DNA) were explored for their phylogenetic utility in 24 taxa. Aspects of the data such as relative amount of nucleotide bases, transitions, transversions or different codon positions were tested. A priori evaluation of a gene region's phylogenetic utility seems impossible. Therefore, three gene regions (*Adh*, *co ii*, and *hb*) were chosen because they produced the greatest number of phylogenetically informative characters. A simultaneous analysis produced a single, well-resolved phylogeny for 49 taxa. Of the eight subgroups tested, five contained more than one representative. Monophyly was supported for the *ananassae, melanogaster, montium* and *takahashii* subgroups. The *suzukii* subgroup, whose monophyly has been questioned (Bock and Wheeler, 1972; Bock, 1980; Toda, 1991), was polyphyletic. My dissertation found morphological characters corroborating possible complexes within the *takahashii* subgroup,

polyphyletic affiliations of the *suzukii* subgroup and the sister relationship of the *ananasse* + *montium* subgroups. Many of species relationships within the *montium* subgroup do not agree with previous established complex affiliations. However, Africa appears to be a region of secondary radiation.

ACKNOWLEDGMENTS

"*Give a man a fish,*

You have fed him for a day,

Teach a man to fish,

You have fed him for a lifetime."

Author Unknown

I would like to thank my committee members for training me in a science career. My mentor, Gail Simmons fostered my growth, allowing me to pursue aspects of my project which interested me most and introducing me to researchers who later formed my dissertation committee. Tadmiri "Venky" Venkatesh advised me to maintain two equal foci: one my research, the other my personal life. This advice has held me in good stead. Dave Grimaldi encouraged and challenged me to think critically. I have benefited from Dave overseeing the writing of this dissertation. Dave has generously given his time and expertise to collaborating with me on a NSF funded morphological revision of the melanogaster species group, which will be my postdoctoral project. Rob DeSalle provided me with the opportunity to work in the state of the art molecular laboratory facilities. Rob simplifies the latest technology making molecular lab procedures easy and enjoyable. Rob gave me the time and space to explore questions to my satisfaction. I found to begin writing this dissertation a Herculean task. Carole Griffiths gave me a perspective and a focus. A willing and eager committee participant was Linda Strausbaugh who wrote many letters of support on my behalf. I am very fortunate to have a committee that has been kind, supportive, encouraging and patient.

Other professors and curators have been most helpful. I would most especially like to thank Lee Herman for devoting his time and expertise in matters of taxonomic referencing, and Kumar Krishna who provided me with a microscope and important literature.

During my tenure at CUNY and AMNH, I have made many friends and colleagues especially in the departments of entomology, herpetology, ichthyology, invertebrates, molecular lab, ornithology and education. In particular I would like to thank: Robin Absher, Rick Baker, Jim Bonacum, Elizabeth Bonwich, Marcelo Carvalho, Tom DiBenedetto, Phaedra Doukakis, Alejandro Espinosa de los Monteros, Paul Goldstein, Cheryl Hayashi, Eileen Keeffe, Varuni Kulasekera, Tam Nguyen, Angel Pimental, Jim Remsen, Carlos Rodriguez, Howard Rosenbaum, Julian Stark, Elisabeth "Lisa" Torres, Mike Whiting, Yale Wyner, and Mike Yudell.

A very special thank you to the following persons. Lori Buckley, my fellow LIRR commuter and CUNY graduate student, who encouraged me to stay at the AMNH. Wenjing Tao my City College lab partner who was most congenial and easy to work with. Susan Klofak instilled in me a greater appreciation for fossils and related many interesting stories. Helpful Mary Egan with whom I enjoyed wonderful lunch breaks observing the bird and plant life in nearby Central Park. I have very much benefited by collaborating and discussing phylogenetic theory with Andy Brower. Mônica Toledo-Piza shared many insightful discussions concerning school, science and life (not necessarily in that order). Mônica introduced me to the enchantment of ballet. Dependable Susan Ratner's "bird by bird" approach helped me write this dissertation. Sue always provided a kind word, a positive out-look and much laughter.

Friends whose annual correspondence provided me with love and support are Jean Stock, John "J.T." Trevor, and Cheri Gratto-Trevor. John and Eileen Carroll sent much appreciated words of encouragement during the difficult hours of writing my dissertation.

I am grateful for having had the opportunity to do my research at the AMNH. The Museum provided an international community that was intellectually stimulating and enriching to my mind and soul.

This dissertation would not have been possible without funds from a NSF Doctoral Dissertation Improvement Grant (NSF DEB 9423508), and two City College Biology

Dissertation Grants. In addition, I received a stipend from the AMNH-CUNY Doctoral Training Program Fellowship. Diane Bynum (Director), Teresa Throckmorton, and Ruben Navarro of the AMNH Office of Grants and Fellowships were always most understanding and helpful.

Important persons in my life who have helped me to come closer to reaching my potential and provide role models are William J. S. Edington, Norton Luger, Lawrence Sheff and Dorit Whiteman.

I am especially grateful for the support of my family. I would like to dedicate this work to my mother, Margaret Schawaroch. She has the ability to see "what is invisible to the eye." She has instilled in me by example a sense of honesty and fairness, a desire for knowledge and pride in a job well done.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Drosophilidae, notably *Drosophila melanogaster* Meigen, are probably the best studied and understood eukaryotic organisms; this does not reflect, however, what is known of their taxonomic relationships. *Drosophila melanogaster* was probably first cultured by entomologist C.D. Woodworth; and in 1906, W.E. Castle was the first to report the use of *Drosophila melanogaster* as a model for genetic experimentation (Sturtevant, 1965; Wheeler 1981; Grimaldi, 1990; Kohler, 1994). From that point on investigations of drosophilids rapidly increased. The Columbia "fly group," which included Thomas Hunt Morgan, Hermann J. Muller, Calvin B. Bridges and Alfred H. Sturtevant, were the founders of modern genetics and their highly collaborative research on *Drosophila melanogaster* established its position as the eukaryotic model organism in this field (Kohler, 1994). Dobzhansky, Mayr, and others used *Drosophila pseudoobscura* and its relatives as a paradigm for the newly formed Neodarwinian synthesis of genetics and evolutionary mechanisms, which also resulted in a de-emphasis of morphological taxonomy (Grimaldi, 1990). Significant taxonomic investigations of *Drosophila* have taken place only within the last two centuries, with the nominal genus, *Drosophila* Fallén, described in 1823. For the remainder of the 19th century there were no or very few taxonomic investigations, and collection numbers were low (Wheeler, 1981). This is probably due to disinterest in small insects like fruit flies, at the time, study of which relies on microscopy for species identification. Sturtevant recognized the need to place research on genetic and evolutionary mechanisms into a phylogenetic context. Besides his numerous genetic investigations, Sturtevant undertook the task of early classification of drosophilids earning him recognition as the "founder of drosophilology" (Grimaldi, 1990).

As of 2000, the family Drosophilidae contains approximately 3,300 species (Wheeler, 1981; Wheeler, 1986; Grimaldi, pers. comm.). There have been four large studies on intrafamilial relationships or classifications (Duda, 1924, Throckmorton, 1962, 1966,and 1975; Okada, 1989 and Grimaldi, 1990). The traditional classification of the family contained 65 genera (Wheeler, 1981, 1986). Throckmorton, in his influential 1975

paper, focused on the subgenera and species groups of the genus *Drosophila*, and concluded that paraphyly is widespread among drosophilids (Grimaldi, 1990). In a cladistic, generic revision for the family Drosophilidae, Grimaldi (1990) revised the genus *Drosophila* by elevating some subgenera to generic level, to produce a monophyletic *Drosophila* and a taxonomy that reflects phylogenetic relationships (Grimaldi, 1990). At this time, the family Drosophilidae is divided into 57 genera, one being the genus *Drosophila* (Grimaldi, 1990).

The genus *Drosophila* (*sensu* Grimaldi, 1990) contains approximately 1,000 species (Wheeler, 1981; Wheeler, 1986; Grimaldi, pers. comm). This genus was traditionally divided into 15 subgenera (see Grimaldi, 1990 for review). As a result of Grimaldi's (1990) revision there are 12 subgenera in the genus, since *Hiritodrosophila*, *Lordiphosa* and *Scaptodrosophila* were elevated to generic rank, and the Hawaiian drosophilids formerly placed in the subgenus *Drosophila* were replaced to the genus *Idiomyia* Perkins as well as the genera *Ateledrosophila* and *Nudidrosophila* (Grimaldi, 1990). Since then, several DNA sequence studies have indicated that Hawaiian species (in *Drosophila* and *Scaptomyza*) are closely related to the genus *Scaptomyza* (DeSalle, 1992; Thomas and Hunt, 1993). Although this differs from Grimaldi's (1990) results that Hawaiian "Drosophila" are related to *Hirtodrosophila*, it is a strong indication that Hawaiian drosophilids are not *Drosophila*. Within each subgenus different geographic regions contain major radiations of species (e.g., Bock and Wheeler, 1972). Some subgenera are further subdivided into the informal taxonomic categories of species group, species subgroup, and complexes within species subgroups (Patterson and Stone, 1952; Vilela and Bächi, 1991). In evolutionary biology, a great deal of research has been dedicated to detailed comparisons of species within species groups, such as the *obscura* group (Dobzhansky and Epling, 1944; Dobzhansky and Powell, 1975), *repleta* group (Wasserman, 1982; Heed and Mangan, 1986), and others.

Sophophora is one of the 12 currently recognized subgenera within the genus *Drosophila*. *Sophophora* is subdivided into eight species groups - four were established by Sturtevant in 1942, a fifth (*fima*) added by Burla in 1954, the sixth (*dentissima*) elevated from a species subgroup by Tsacas (1979, 1980), the seventh is either *dispar* (Lemeunier *et al.*, 1986; Scouras, 1995) or *populi* (Ashburner, 1989) depending on the source, and the remaining species are non-classified (Scouras, 1995) (Figure 1). One of these species groups is the *Drosophila melanogaster* group. The *D. melanogaster* group represents one of the most explosive radiations within the genus *Drosophila*, with the possible exception of the Hawaiian species (Bock, 1980). There have been 5 major studies on the *melanogaster* group (Hsu, 1949; Okada, 1954; Bock and Wheeler, 1972; Bock 1980; and partial revision by Toda 1991) (Figure 1).

In 1949, Hsu initially divided the *melanogaster* group into five subgroups - *ananassae*, *melanogaster*, *suzukii*, *takahashii* and *montium*. Okada (1954) added two subgroups - *nipponica* and *ficusphila*. Bock and Wheeler (1972) added the four subgroups *denticulata*, *elegans*, *eugracilis* and *dentissima*, plus five species *incertae sedis*. Bock (1980) added the twelfth subgroup - *flavohirta*. In his revision of the subgenus *Sophophora*, Tsacas (1979, 1980) elevated the *dentissima* subgroup of Bock and Wheeler (1972) to a species group based on the functional morphology of the male genitalia. The *nipponica* subgroup was removed when Okada (1984) synonymized the *nipponica* subgroup with the *L. miki* species group of what is now the genus *Lordiphosa* (then a subgenus of *Drosophila* [Grimaldi, 1990]). Recently, Toda (1991) added two more subgroups - *rhopaloa* and *longissima*, thus making a total of twelve subgroups within the *melanogaster* group as currently recognized. All of these revisions relied on morphological characters and employed a traditional evolutionary taxonomic approach, (e.g., Wiley, 1981) with the possible exception of Okada (1954), who presented a phenetic classification for these subgroups. The number and sources of characters have dramatically increased from the initial to the most recent study. Sources of the morphological characters are male

external and internal genitalia, the dimorphic male sex comb, internal female genitalia, and setae on the palp, oral region and head. Morphological studies of Hsu (1949), Okada (1954), Bock and Wheeler (1972), plus Ashburner *et al.* (1984), Pélandakis *et al.* (1991), and Pélandakis and Solignac (1993), hypothesized relationships among the subgroups of the *melanogaster* group (Figure 2). Using characters from morphology, metaphase chromosomes and polytene chromosome banding, Ashburner *et al.* (1984) concentrated on relationships of species within the *melanogaster* subgroup; however, they did postulate a general hypothesis of relationships among the species subgroups (Figure 2).

Using 28S rRNA, Pélandakis *et al.* (1991) and Pélandakis and Solignac (1993) studied the subgenus *Sophophora* and the genus *Drosophila*. Both studies found paraphyly within the genus *Drosophila*, the subgenus *Sophophora*, and the *melanogaster* group. The segment that contains the *melanogaster* group is redrawn here (Figure 2). None of the six hypotheses of relationships among the subgroups are in complete agreement (Figure 2). The early studies of Hsu (1949) and Okada (1954), however, did agree that the *suzukii* subgroup is basal to two main lineages: (1) *melanogaster*, *takahashii* and *ficusphila* and (2) *montium*, *ananassae* and *nipponica*. Bock and Wheeler (1972) only agree with Hsu (1949) and Okada (1954) with respect to the affinity between *ananassae* and *montium*. The later studies (Bock and Wheeler, 1972; Ashburner *et al.*, 1984; Pélandakis *et al.*, 1991; Pélandakis and Solignac, 1993) generally agree that the *ananassae*, the *montium* and all other subgroups – *takahashii*, *suzukii*, *ficusphila*, *melanogaster*, *elegans* and *eugracilis* seem distinct.

The *melanogaster* species group has a total of 174 species and 3 subspecies unequally distributed among 12 subgroups, except five species which are *incertae sedis* (Appendix A). Ever since work began on the *melanogaster* group the number of species has increased at an astounding rate, especially within the last 50 years (Figure 3). One of these subgroups - the *melanogaster* subgroup contains the laboratory paradigm species, *D. melanogaster* (DeSalle and Carew, 1992; Ashburner *et al.*, 1984; Lachaise *et al.*, 1988).

Given the reasonable assumption that extinction rates among subgroups is probably equivalent, then the much higher number of species in an apparently monophyletic *montium* subgroup would indicate far greater speciation for this subgroup. Since ages of the subgroups are not known, it is unknown if such differences are due to *rates* of speciation.

The *montium* subgroup, by far the largest subgroup, contains 87 taxa (Appendix A). (Gupta and Sundaran, [1990] described *D. maggulae* and placed it in the *montium* subgroup; however, they said it is similar to *D. atripex* Bock and Wheeler, 1972 which is a species in the *ananasase* subgroup. According to their illustrations *D. maggulae* looks like a member of the *ananassae* subgroup). Approximately half of these flies have been placed into the species complexes of *auraria, bakoue, bocqueti, jambulina, kikkawai, nikananu,* and *serrata* (Table 1). The species complexes within the *montium* subgroup have been established on the basis of morphology of male genitalia and sex combs, as well as hybridization and biochemical analyses (e.g., Tsacas and Chassagnard, 1992; Ayala, 1965 a & b; Kim *et al.*, 1993). The majority of papers concerning the *montium* subgroup are descriptions of new species or local faunal analyses (e.g., Shyamala and Ranganath, 1990; Chassagnard *et al.*, 1997; Singh and Dash, 1998; Singh and Gupta, 1979; De and Gupta, 1996). The numerous investigations of the *montium* subgroup contain relatively small sample sizes due to the enormous size of the group and the reliance of species in culture for biochemical analyses (e.g., largest number of species, 29, in Ohnishi and Watanabe, 1984; 9 species in Konstantopoulou *et al.*, 1997; 6 species in Pissios and Scouras, 1993; 18 species in Kim *et al.*, 1993). Morphological, hybridization, karyological and biochemical data have been analyzed using an evolutionary taxonomy or a phenetic approach (e.g., Tsacas and Chassagnard, 1992; Kim *et al.*, 1989 [used asymmetric mating preference model of Watanabe and Kitagawa, 1979 and 1981]; Shyamala and Ranganath, 1989; Kim *et al.*, 1993). Two biochemical papers (Pissios and Scouras, 1993; Nikolaidis and Scouras, 1996) included a parsimony analyses and then ignored the results, and one morphological paper (Chassagnard *et al.*, 1997) described the relationship among three species approximating a

cladistic approach. Relationships among the species are poorly understood (Lemeunier *et al.*, 1986; Drosopoulou and Scouras, 1995). A revision for this subgroup is sorely lacking.

This dissertation focuses on relationships within the subgenus *Sophophora* at the level of the *melanogaster* species group, and also focuses on species relationships in the *montium* subgroup. A cladistic analysis of molecular data is employed. A phylogeny is generated to examine relationships among subgroups of the *melanogaster* group and among species of the *montium* subgroup. The first section of the dissertation is an exploration into the utility of different gene regions for phylogenetic analyses. The second section analyzes relationships among species subgroups of the *melanogaster* group. The third section is an analysis of species relationships in the *montium* subgroup.

CHAPTER 1

Evaluating a Gene Region's Phylogenetic Utility

CONTENTS FOR CHAPTER 1

INTRODUCTION

Can molecular biologists evaluate the character information of a gene region prior to phylogenetic analysis? Morphologists through an examination of specimens or literature can a priori evaluate different characters or characters systems for variability applicable to the taxonomic level of phylogenetic inquiry. DNA sequence information is virtually unknown for most species; therefore, molecular biologists must first generate the sequence in order to explore its character potential. Perhaps some aspect of the sequence data, such as nucleotide base content, or amount of transitions, transversions or codon position changes, indicates a gene region's phylogenetic utility. These aspects of DNA sequence were tested for predictability using 5 gene regions for *Drosophila* species within the *melanogaster* species group. The gene regions were chosen to broadly survey the genome. The five gene regions chosen were *alcohol dehydrogenase* (*Adh*), *cytochrome oxidase* II (*co ii*), *hunchback* (*hb*), 28S ribosomal DNA and 16S ribosomal DNA.

These five gene regions can be grouped according to the portion of the *Drosphila* genome represented: three gene regions are nuclear – *Adh*, *hb*, 28S, and two genes are mitochonrial – *co ii* and 16S. The nuclear genes are from different chromosomes in *Drosophila melanogaster* (*Adh* from 35B3, *hb* from 85A8--10 and 28S from h26--32). Even though this study includes *D. melanogaster* and other taxonomically close species, the metaphase chromosome structure differs enough to have allowed Ashburner *et al.* (1984) to recognize and divide the *melanogaster* species group into three major lineages: the *ananassae* subgroup, the *montium* subgroup, and the *melanogaster* + Oriental subgroups. (Oriental subgroups are the *elegans, eugracilis, ficusphila, suzukii*, and *takahashii* subgroups.) My study does not provide conclusive evidence (e.g., in situ hybridzation) as to the exact chromosomal location of the nuclear genes in each species, but this study will assume that these gene regions are independent. My study includes two mitochondrial gene regions with assumed character independence although some have argued that sampling

more than one mitochondrial gene is redundant because of the lack of recombination in the mitochondrion.

The five gene regions sampled can also be grouped by the general structural types of protein coding (*Adh*, *co ii*, and *hb*) and ribosomal (28S and 16S). The protein coding genes are single copy. Even though the ribosomal genes are in redundant stretches along chromosomes, these repeats are quite homogeneous (like a single gene) as a result of concerted evolution.

Adh, *co ii*, *hb*, 28S, and 16S each have different functions in the organism. Many Drosophilids, especially within the *melanogaster* species group, live and feed on decaying fruit (Bock, 1980). *Alcohol dehydrogenase* (*Adh*) codes for the enzyme that detoxifies the alcohol on these substrates thus ensuring the survival of larvae and adults (Geer *et al.*, 1985). *Adh* has been used in many taxonomic investigations of *Drosophila* especially at the species and population level. *Hunchback* (*hb*) is a developmental gene that initially acts as a maternal morphogen establishing the anterior - posterior axis in the fertilized egg and later as a gap gene in early embryonic development (Tautz *et al.*, 1987). First studied *in D. melanogaster*, *hb* plays the same role across a broad range of species in the genus *Drosophila* (DeSalle pers. comm.). *Cytochrome oxidase* II is one of the mitochondrial cytochrome transport molecules. The 3' end was found to be the most variable part of the gene in other studies (Baker and DeSalle, 1997; Baker *et al.*, 1998). The 28S gene is part of the small nuclear ribosomal gene. Fragments from the D1 and D2 domains of 28S gene, corresponding to those used in phylogenetic studies of Pélandakis *et al.* (1991) and Pélandakis and Solignac (1993), were sequenced. The 16S gene is part of the ribosome for the mitochondria. The portion of the gene sequenced corresponds to the fragment used in the study of the *Drosphila repleta* species group. These DNA regions chosen also facilitate comparisons with previous studies (e.g., Baker and DeSalle, 1997; Thomas and Hunt, 1993; Beckenbach *et al.*, 1993).

METHODS

Fly Stocks

Flies were obtained from the National *Drosophila* Species Resource Center at Bowling Green and D. LaChaise (Table 2). Cultures were maintained on cornmeal-agar-yeast medium in $^1/_2$ pt milk bottles in a 21°C incubator on a 12-hour light: 12-hour dark cycle.

Taxon Sampling

A total of 24 taxa were used in this study (Table 3). The ingroup contains representative taxa from seven of the twelve subgroups within the *melanogaster* group. Outgroup taxa consist of five species from the *obscura* group - two from each of the subgroups of *obscura*, and *pseudoobscura* and one from the *affinis* subgroup (Barrio *et al.*, 1994).

The Gene Regions

For *Adh* a 290 bp region spanning the second exon was sequenced. This fragment corresponds to the one used in previous taxonomic investigations (e.g., Thomas and Hunt, 1993; DeSalle, 1992). In this study a fragment of the *hb* coding region 5' end containing Box B (Tautz *et al.*, 1987) was sequenced and analyzed. The length of this fragment varied from 436 –513 bp in length. *co ii* was sequenced for 385 bp at the 3' end. For 28S a portion of D1 181-183 bp long and D2 312 -314 bp long was sequenced. A portion of the 16S gene region 424-426 bp long was sequenced.

Molecular Techniques

Single fly DNA extractions were performed (DeSalle *et al.*,1993). DNA was PCR amplified using PE *taq* polymerase with primers developed for use on Drosophilid taxa in (Appendix B). PCR products were purified using Gene Clean II (Bio 101). The *Adh* fragment was cloned using Invitrogen's TA cloning kit for many of the taxa. Gene regions were sequenced in both directions by either manual or automated sequencing methods. All of the 16S and most of the *Adh*, *co ii*, and 28S sequences were generated manually. The

remainder including all the *hb* sequences were done using an ABI 373 automated sequencer. Manual sequencing of double stranded PCR products and clones was done using ^{35}S and United States Biochemical's Sequenase version 2.0 DNA sequencing kit according to manufacturers instructions. Automated sequencing of double stranded PCR product was accomplished according to ABI Prism DNA sequencing kit, purified by sephadex columns and run using Applied Biosystems 373A machine and DNA sequence protocols. Sequences were checked and corrected using Sequencher 3.0 (Gene Codes Corp.) sequence analysis software. Most of the DNA sequence was generated by this study with the exceptions noted (Table 4).

Character Assignment

All gene region sequence ends were trimmed to start at the same position from the furthest right and left primers. The *Adh* and *co ii* DNA sequences contained no indels (insertions or deletions); therefore, alignment was straightforward. The 16S gene region sequence length varied by 2 bp. The 28S gene region was divided into two parts corresponding D1 and D2 fragments of Pélandakis and Solignac (1993). Length variation for 28S also was minimal (2 bp D1 and 2 bp D2). For both 16S and 28S gene regions multiple alignments varying the internal (gap) cost to change cost were performed using MALIGN (Wheeler and Gladstein, version 2.7). Alignment ambiguous sites were removed from the nucleotide sequence prior to phylogenetic analyses (Gatsey *et al.*, 1994). The *hb* gene region sequence length varied by 57 bp making the alignment more complicated. Homology assessment for *hb* required conversion of the nucleotide bases to amino acid sequence. Multiple alignments varying the gap penalty were run using the Clustal method in MEGALGN (DNASTAR, version 1.02). Sites that were alignment ambiguous were removed. The remaining sequence was returned to nucleotide bases prior to phylogenetic analyses.

Molecular Character Assessment

The number of phylogenetically informative characters contributed by the data sets (all 5 gene regions, *Adh*, *co ii*, *hb*, 28S, 16S, coding genes [*Adh*, *co ii*, *hb*], ribosomal genes [28S, 16S] nuclear genes [*Adh*, *hb*, 28S], and mitochondrial genes [*co ii*, 16S]) were compared. Each of the separate gene regions (i.e., *Adh*, *co ii*, *hb*, 28S, and 16S), as well as all 5 gene regions combined, were examined for biases in total nucleotide base composition. Transition or transversion bias was evaluated using saturation plots for phylogenetically informative characters of the gene regions *Adh*, *co ii*, *hb*, 28S, 16S, ribosomal genes (16S + 28S), and all 5 gene regions combined. Saturation plots were also generated to examine biases in codon positon (1^{st}, 2^{nd}, 1^{st} + 2^{nd}, and 3^{rd} positions) for the phylogenetically informative characters of *Adh*, *co ii* and *hb* gene regions separate as well as combined.

Phylogenetic Analyses

Phylogenetic hypotheses were generated for all gene region combinations including a simultaneous analysis of all five gene regions (Kluge, 1989). All the trees were rooted with the five outgroup species chosen from the sister taxon, the *obscura* group. Only informative characters were used to generate trees and tree statistics. PAUP 4.0 (Swofford, 2000) was used to generate all phylogenies. Heuristic tree searches were performed with random addition of taxa, TBR branch swapping and repeated 20 times. The characters were given an equal weight of 1 and run unordered. When there was more than one most equally parsimonious tree, a strict consensus tree was generated and presented as the phylogenetic hypothesis for that analysis. Successive weighting trees / successive approximations analyses employing the iterative method of Farris (1969; Carpenter, 1988) based on the rescaled character c.i. values were generated for the simultaneous analysis tree .

Character Support at Nodes

To describe character support for nodes in the simultaneous analysis strict consensus tree, Bremer support values (Bremer, 1988; 1994) and bootstrap values (Felsenstein, 1985) were calculated (Autodecay ver. 2.9.8, Eriksson, 1998). Bootstrap

analyses employed 1000 replicates with each replicate containing twenty heuristic searches with random addition of taxa and TBR branch swapping.

Comparison of Phylogenetic Hypotheses

Tree topologies (usually strict consensus trees) for all the various data combinations were compared to the nodes resulting in the strict consensus tree from the simultaneous analysis. All trees were evaluated and compared for CI, RI, ingroup resolution, and monophyly of traditional taxonomic groups.

RESULTS

Character Assignment

All the parameter files listed in Appendix C were tested on the 16S and 28S data, however, only certain parameters worked for each data set. Perhaps this result is due to the minimal sequence length variation or constraints of the MALIGN program. The 16S gene region produced an alignment for parameters P6, P8, and P10. All three parameters yielded the exact same alignment for 16S (length = 55). D1 of the 28S gene region produced alignments for parameter files P2, P4, P6, P8 and P10. These 5 parameters yielded the exact same alignment for D1 of 28S (length = 4). The D2 of 28S yielded two equally most parsimonious alignments (length = 98) for P1 parameter file. There were 2 alignment ambiguous sites between these two alignments which were removed prior to the phylogenetic analyses (Appendix C). The resulting alignments of 426 bp for 16S and 495 bp for 28S (183 bp D1 and 312 bp D2) were placed in the matrix for phylogenetic analyses.

Alignment parameters for the *hb* gene region varied as follows: (1) the gap length penalty was set at a value of 10; (2) the amino acid change cost was according to the PAM250 residue weight table (Dayhoff, 1978); and (3) the gap penalty value varied from 8 to 30. Three stretches of amino acid sequence (13, 8 and 18 amino acids long, respectively) were considered alignment ambiguous (Appendix C). These alignment ambiguous sites

were removed and the amino acids converted to nucleotides leaving 435 bp of DNA sequence for *hb* gene region in the PAUP matrix.

Molecular Character Assessment

The number of phylogenetically informative characters contributed by the protein coding gene regions (*Adh*, *co ii*, *hb*) separate as well as combined is approximately 25% (Table 5). The ribosomal gene regions (28S and 16S) each provide very few phylogenetically informative characters (8% to 4%) (Table 5). Due to the presence of ribosomal gene regions in the nuclear and mitochondrial data sets, their phylogenetic informative character contribution is only 19% to 13% respectively. For each of the separate gene regions (*Adh*, *co ii*, *hb*, 28S and 16S) as well as all 5 gene regions combined the relative nucleotide base composition varied between all characters and the phylogenetically informative characters (Figure 4). Phylogenetically informative regions need to be explored because these are used to build trees. The mitochondrial gene regions (*co ii* and 16S) were strongly A, T biased as has been observed in other insect mitochondria (Clary and Wolstenholme, 1985; DeSalle *et al.*, 1987; Liu and Beckenbach, 1992). The 28S (nuclear ribosomal) gene region has a bias in favor of A and against C. For the other nuclear gene regions: *Adh* has a bias in favor of C, and *hb* has a bias in favor of both C and G. All 5 gene regions combined are biased in favor of T. Despite nucleotide base composition biases, there was no saturation exhibited for either transitions or transversions for the gene regions *Adh*, *co ii*, *hb*, 28S, 16S, ribosomal genes (16S + 28S), and all 5 gene regions combined (Appendix D). Third codon position sites by far provide the most phylogenetically informative characters (Table 6). Second codon positions provide so few phylogenetically informative characters that saturation plots could not be made for the *Adh* and *co ii* data. In all instances (1^{st}, 2^{nd}, $1^{st} + 2^{nd}$, and 3^{rd} position) there was no saturation exhibited for the gene regions *Adh*, *co ii* and *hb* separate as well as combined (Appendix D).

Phylogenetic Analyses

All the data combinations yielded more than one equally most parsimonious tree with the exception of *co ii* + *hb* +16S + 28S. Strict consensus trees are presented with tree statistics in Appendix E. The simultaneous analysis resulted in 8 equally most parsimonious trees. These eight trees are well resolved and they vary with respect to hypotheses of relationships for the *montium* species subgroup (Figure 5). Successive weighting / successive approximations analysis was performed for the simultaneous analysis. A single tree resulted but it did not match any of the eight most equally parsimonious trees (Figure 6).

Character Support at Nodes

Bremer and Bootstrap values are present for each node of the simultaneous analysis strict consensus tree (Figure 7). Out of a total of 17 nodes in the simultaneous analysis strict consensus tree 6 nodes had Bremer values less than 3 (Figure 7). Only nodes 12 and 14 had bootstrap values less than 50% (Figure 7 and 8).

Comparison of Phylogenetic Hypotheses

A comparison of the CI, RI and ingroup resolution from the strict consensus trees for the different data combinations (a total of 31 combinations) is in Table 7. Node numbers 1, 2, 4, 5, 8, 9, 13, 15, 16, and 17 are frequently present (in 20 or more trees) (Figure 8 and Appendix E).

DISCUSSION

Aspects such as bias in nucleotide base content, saturation of transitions, transversions or 1st, 2nd, 3rd, 1st and 2nd codon positions that could lead to inaccurate primary homology statements and may reflect the information content of a gene region were examined. The phylogenetically informative characters are the only ones which need to be examined because they produce the tree topology and their relative composition may be different from all characters included (Figure 4).

The nucleotide base bias seen in the 16S data is exaggerated when just examining the phylogenetically informative characters. The large number of adenine (A) nucleotide bases may be a reflection of small sample size (only 16 phylogenetically informative characters). Despite the gene regions sampled having nucleotide base biases, these biases were not enough to cause the sites to become saturated.

Gene Regions and Tree Structure

The phylogenetic signal contained in a single or various combinations of gene regions was compared for tree resolution. The simultaneous analysis strict consensus tree is well resolved except for relationships within the *montium* species subgroup. The 16S gene region has very little structure and does not have ingroup monophyly. Most of the characters in the 16S data have very few state changes and few steps, many character c.i.'s = 1.00. In other words, each individual charcter produces very little structure and what exists is uncontradicted. The other ribosomal gene region 28S also has a relatively low number of phylogenetically informative characters (42) especially in relation to the total number of characters (495 bp). The 28S gene region in my study has ingroup monophyly and contains 13 ingroup nodes – the same amount of ingroup resolution as the simultaneous analysis. In contrast, Pélandakis *et al.* (1991) and Pélandakis and Solignac (1993) using the same gene region found the *melanogaster* species group to be paraphyletic. It is important to note that Pélandakis *et al.* (1991) and Pélandakis and Solignac (1993) had more and different taxa from my study. Previous, studies have demonstrated that taxon sampling is important to resulting tree topology (Lecointre *et al.*, 1993; Graybeal, 1998; Hillis, 1998; Poe, 1998). Of the single gene region trees, the mitochondrial gene regions (*co ii* and 16S) each had little resolution but the signal contained in each did produce different tree topologies (Appendix E), despite being "linked' on the mitochondrion. *Adh* + *co ii* + *hb* + 28S produced a strict consensus tree with the exact same topology as the 5 gene regions simultaneous analysis strict consensus tree. The sixteen phylogenetically informative characters contributed by the 16S data did not alter tree topology. Also the

number of equally most parsimonious trees and RI were the same between the data sets and the CI only differed negligibly, by 0.002. The incongruence length difference (Mickevich and Farris, 1981) between the *Adh* + *co* ii + *hb* + 28S and 16S tree lengths versus all 5 gene regions is only 6 extra steps resulting from the character conflict among the data sets.

In general, the mean ingroup resolution increases with the number of data sets combined (for one data set mean = 10.8, for two data sets mean = 11.3, for three data sets mean = 12.5, for four data sets mean = 14.6). However, the range of values for ingroup resolution does overlap. The amount of ingroup resolution for the simultaneous analysis strict consensus tree does not follow this trend (ingroup resolution = 13). The simultaneous strict consensus tree has less resolution than the average of four gene regions combined. The data set co ii + hb + 16S + 28S yielded one most parsimonious tree and had a fully resolved ingroup; this tree had the most structure.

The CI of a tree from a single or combination data set does not seem to have predictive value. The tree resulting from the 16 S data has the best ensemble consistency index. This is also the tree with very little resolution, the fewest phylogenetically informative characters (16) and the lack of ingroup monophyly. The CI just indicates that the characters agree with the topology of the tree, and 16S has very little topology and very few characters to conflict with.

Nodes were tested for consistency across the various data set trees and compared to the Bremer and bootstrap values. Each of the 17 nodes in the simultaneous analysis strict consensus tree are seen in an average of 11 trees. There are six nodes (nodes 3, 6, 10, 11, 12, and 14) that appear in 12 or less various gene region combinations. All of these six nodes have Bremer values less than 3. Two of these nodes (node 12 and 14) have bootstrap values less than 50%. Nodes seen in fewer of the data combination trees do seem to have the lower Bremer and bootstrap values.

Node 14 which has low values for both Bremer and bootstrap unites the *montium* subgroup with the *melanogaster* + Oriental subgroups lineage. Ashburner *et al.* (1984),

Pélandakis *et al.* (1991) and Pélandakis and Solignac (1993) were unable to resolve relationships among the 3 major lineages. Pélandakis *et al.*, (1991) attributed this lack of resolution to an almost simultaneous splitting of the 3 lineages not affording enough time to accumulate many characters. Presence of the three major lineages was explored in the strict consensus trees (and single most parsimonious tree for *co ii* + *hb* + 16S + 28S) of the various data combinations. Twenty one of the 31 data combinations produced trees containing the three major lineages of the *ananassae* subgroup, the *montium* subgroup, and the *melanogaster* + Oriental subgroups. Of these 21 trees: 12 have the relationship the *montium* subgroup with the *melanogaster* + Oriental subgroups (including the simultaneous analysis tree), one has the relationship *melanogaster* + Oriental subgroups with the *ananassae* subgroup, two have the *ananassae* subgroup with the *montium* subgroup as sister, and six trees have the relationship between the three clades as unresolved. The various data combinations exhibit all possible relationships for the three lineages. Clearly the resulting hypothesis of relationships depends on the data; this observation may have caused Ashburner *et al.* (1984) to say they could not resolve the nodes and Pélandakis *et al.* (1991) to say that there are very few characters at the node for the three lineages.

Taxonomy

Many of the traditional taxonomic groups were confirmed by this study (Figures 7 and 8). Monophyly was established for the *melanogaster* species group (i.e., node 4) in all data combinations except 16S. The *takahashii* subgroup (node 6) was monophyletic in 7 out of the 31 data combinations. The *melanogaster* species subgroup (node 9) was monophyletic in all but four data sets (i.e., *co ii*, *hb*, 16S, and *hb* + 16S). The *yakuba* complex of the *melanogaster* species subgroup was seen in all tree reconstructions. The *ananassae* species subgroup (node 17) was monophyletic in all but one tree (i.e., *co ii*). The *montium* species subgroup (node 15) was seen in all but one tree (i.e., 16S) The other subgroups of *eugracilis* and *ficusphila* had only one representative; therefore, monophyly could not be

tested. The *suzukii* species subgroup had two representatives and was not monophyletic in any of the data combination reconstructions. All the nodes referring to monophyly of the traditional groups appeared in 20 or more reconstructions. Therefore, the following nodes appeared often enough that further studies are warranted to determine if these affiliations need to be established. Node 5 seen in all but the 16S tree may reflect a complex within the *takahashii* species subgroup. Node 1 which represents the pseudoobsscura subgroup within the obscura group. My cladogram has the obscura subgroup paraphyletic (node 2). Node 16 which unites the *montium* subgroup representatives *D. diplacantha*, *nikananu* and *seguyi* is seen in twenty reconstructions. There is no previous complex affiliation uniting these three species (in fact they are from different complexes); however, they are the only *montium* species subgroup representatives from Africa.

IN CONCLUSION

Can a gene region's utility be a priori evaluated?

Characters should not be given quality distinctions (e.g., good or bad), but gene regions must hold sufficient variation to pursue sequencing it for a particular taxonomic investigation. Of the various aspects of the DNA sequence data explored, the only indicator which is not a priori is the number of phylogenetically informative characters. To determine the number of phylogenetically informative characters within a gene region, DNA sequences must be obtained for a small representative sample encompassing the taxonomic categories to be studied. Because recognition of characters depends upon alignment, the number of phylogenetically informative characters in a stretch of DNA sequence relies upon the taxa sampled. Visually inspecting aligned sequences can indicate if the stretch of DNA is almost invariant or if alignment ambiguous sites can be reduced or resolved with more taxa sampled, thus providing more useable character information.

Variability is the key. Gene regions that seem quite variable should be pursued. This concept is not new to systematics. Morphologists visually survey an organism for

character systems that seem variable at the level of study. Too much variability usually comes down to noise or unalignable sequence (Davis and Nixon, 1992).

Previously sequenced gene regions in a data bank such as GenBank can be surveyed for taxa in the investigation. Differing phylogenetic histories require that one surveys gene regions for the taxa under investigation. Taxonomic categories are not equivalent, especially the levels of species group, species subgroup and species complex within Drosophilds. As illustrated in this study, 28S provides more resolution compared to Pélandakis *et al.*, (1991) and Pélandakis and Solignac (1993). The 16S gene region used in this study is the same portion used for the study of the *Drosophila repleta* group. This study finds 16S invariant and most probably misleading; whereas, the *repleta* study finds 16S to be a source of numerous characters (R. Baker, pers. comm.). Therefore, predictions can not be made from previous studies of equivalent or different taoxonomic categories.

For Future Study

In my study, the 28S gene region resulted in *melanogaster* species group monophyly, in contrast to Pélandakis *et al.*, (1991) and Pélandakis and Solignac (1993), thus illustrating the effect of taxon sampling. I conclude, however, that 28S is a poor gene because it has very few phylogenetically informative characters. In my study 16S exhibits no ingroup monophyly, very little resolution, and a large nucleotide base bias. In addition, 16S yields even fewer phylogentically informative characters for approximately the same length of sequence as 28S. Therefore, 16S is a poor gene region to pursue sequencing for more taxa.

CHAPTER 2

Phylogenetic Relationships Among Subgroups of the *melanogaster* Species Group

CONTENTS FOR CHAPTER 2

Page

INTRODUCTION

Fruit flies, or members of the family Drosophilidae, comprise as of the year 2000 over 3,300 species (Wheeler, 1981; Wheeler, 1986; Grimaldi, pers. comm.). One of 57 genera (*sensu* Grimaldi, 1990), the genus *Drosophila* itself contains nearly one third of all species in the family (Wheeler, 1981; Wheeler, 1986; Grimaldi, pers. comm.). This genus is divided into 12 subgenera (Grimaldi, 1990). Some subgenera are further subdivided into species groups. The subgenus, *Sophophora* contains the species most familiar to all students of biology, *Drosophila melanogaster* Meigen. As the eukaryotic model organism, *D. melanogaster* and its close relatives have been subjects for studies in virtually all fields of biology. Yet, very little is known about the phylogenetic relationships of species within *Sophophora*, and within its largest and most intensively studied species group, the *melanogaster* species group.

Presently, the *melanogaster* group contains 174 species and 3 subspecies, and as such represents one of the largest radiations within the genus *Drosophila*. Many of these species are narrowly endemic; however, five species – *melanogaster*, *simulans*, *malerkotliana*, *ananassae*, and *kikkawai* are cosmopolitan commensals of humans (Lemeunier *et al.*, 1986). The majority of species of the *melanogaster* group occur in the Oriental region (113 species, 64 % of the total) (Bock and Wheeler, 1972; Throckmorton, 1975; Bock, 1980; Ashburner *et al.*, 1984; Lemeunier *et al.*, 1986; Toda, 1991), which is considered its center of origin. The *melanogaster* group is then thought to have radiated into the Australasian, Afrotropical and east Palearctic Regions. "Centers of Origin" thought in biogeography has permeated evolutionary studies on *Drosophila* (e.g., Throckmorton, 1975 and other refs.), but now this concept has been largely superseded by vicariance and cladistic biogeography (Nelson and Platnick, 1981). "Centers of Origin" has traditionally been described as those areas with the highest diversity of species or other taxa. Cladistic biogeography, however, relies on phylogenetic relationships of organisms and any geographic patterns revealed from this information.

Morphological and biochemical data have unequivocally established that the sister group to the *melanogaster* group is the *obscura* group (Throckmorton, 1975; Powell and DeSalle, 1995). Morphological evidence rests partly on the shared possession of a sex comb in the males of both groups, although it is not well developed in some species. The *obscura* group contains approximately 55 species (Gleason *et al.*, 1997) adapted to cool temperate forest conditions, thus, having primarily a Holarctic distribution and also found at high elevations in the northern regions of the neotropics, afrotropics, as well as China, and Taiwan (Wheeler, 1981; Takamori and Okada, 1983; Cariou *et al.*, 1988; Watanabe *et al.*, 1996; Watanabe and Sperlich, 1997). Species of the *obscura* group are often difficult to distinguish on the basis of morphological characters, so most of the subgroup diagnoses are based on molecular data (e.g., allozymes and DNA sequences), and morphological characters which concur are added (Lakovaara and Saura, 1982; Cariou *et al.*, 1988; Barrio, *et al.*, 1994). The *obscura* group is presently divided into five subgroups: *obscura*, *affinis*, *pseudoobscura*, *microlabis* and *subobscura* (Sturtevant 1942; Lakovaara and Saura, 1982; Cariou *et al.*, 1988; Barrio *et al.*, 1994).

The *melanogaster* group is comprised of 12 subgroups of disparate sizes: *ananassae*, *denticulata*, *elegans*, *eugracilis*, *ficusphila*, *flavohirta*, *longissima*, *melanogaster*, *montium*, *rhopaloa*, *suzukii*, and *takahashii* with five species *incertae sedis* (Hsu, 1949; Okada, 1954; Bock and Wheeler, 1972; Tsacas, 1979, 1980; Bock, 1980; Okada, 1984; Toda, 1991; Lemeunier *et al.*, 1986). The subgroup diagnoses are primarily based on characters from male genitalia and sex combs, but internal female genitalia, and setae on the palp, oral region and head are also used (Hsu, 1949; Okada, 1954; Bock and Wheeler, 1972; Bock, 1980; Toda, 1991). Bock (1980) and Toda (1991) provided the most recent subgroup diagnoses.

Most species within the *melanogaster* group are easily collected at fruit baits, so their natural habits are hardly explored. However, since many of these species are culturable on fruit-breeding medium and have short generation times, these species are ideal

experimental subjects. Bock (1980) indicated that the primitive ecological condition for the *melanogaster* group is adult feeding and larval breeding on decaying fruits. Derived habits for the *melanogaster* group are flower breeding, as seen in *D. elegans*, *D. sahdryii*, and *D. flavohirta*, and human commensalism, as seen in the cosmopolitan species.

Many species from the *melanogaster* group have been reported from single localities. It is unclear, however, if this is real phenomenon or an artifact of collecting.

Metaphase karyotypes are traditionally characterized for many species of the *melanogaster* group (e.g., Clayton and Wheeler, 1975; Baimai, 1980; Shyamala and Rangnath, 1989; see Lemeunier *et al.*, 1986 review). Metaphase chromosomes have been used, in conjunction with other evidence, to help identify new species (e.g., *D. watanabei*), to hypothesize relationships within (e.g., *ananassae* subgroup, *erecepeae* complex) and among subgroups (e.g., *melanogaster* group) (Ashburner *et al.*, 1984; Gupta and Gupta, 1992; Lemeunier *et al.*, 1997). Below is a brief review of each subgroup. The taxonomic history and subgroup affiliations are summarized in Appendix A.

The *ananassae* subgroup

The *ananassae* subgroup contains 23 species and 2 subspecies, most of them divided among 3 complexes – *ananassae*, *bipectinata*, and *erecepeae* – and some of which have no affiliation (Appendix A; Lemeunier, *et al.*, 1997). Each species of the *ananassae* subgroup possesses a surstylus with two sets of teeth (Toda, 1991). All, but *D. varians* Bock and Wheeler, have a differentiated cercal clasper (Toda, 1991). These species, however, vary in characters of the sex comb (absent to present in either transverse, oblique or longitudinal row[s]) and the aedeagus (Toda, 1991). The complexes are primarily defined by adeagal structure (Lemeunier *et al.*, 1997). Despite lacking the cercal clasper and possessing a cercus with bristles similar to species of the *suzukii* subgroup, *D. varians* Bock and Wheeler is included within the *ananassae* subgroup based on chromosomal homology (Bock and Wheeler, 1972). The *ananassae* subgroup is thought to have

originated in southeast Asia; its distribution is Oriental, Afrotropical, Neotropical, eastern Palearctic, and Australasian (Lemeunier *et al.*, 1986). *Drosophila ananassae* has been collected in flowers and *D. moneri* breeds facultatively in rotting flowers of *Hibiscus tiliaceus* (McEvey *et al.*, 1987). This is one of the more studied subgroups, as based on allozymes, metaphase chromosomes, polytene chromosomes, and hybridization studies (see Lemeunier *et al.*, 1986; Lemeunier *et al.*, 1997 and references therein).

The *denticulata* subgroup

The *denticulata* subgroup contains 3 species (Appendix A; Lemeunier *et al.*, 1986). Two species, *D. pseudodenticulata* Takada and Momma and *D. microdenticulata* Panigrahy and Gupta, are known from single localities in western Malaysia and India, respectively (Lemeunier *et al.*, 1986). The third species, *D. denticulata* Bock and Wheeler, is found throughout northern Queensland, Australia, New Guinea and the Philippines (Bock, 1976; Lemeunier *et al.*, 1986). Characters of male genitalia such as finger-like anterior parameters, and a sex comb consisting of a few large teeth (Bock, 1980) are shared by members of this group.

The *elegans* subgroup

The *elegans* subgroup contains 4 species (Appendix A). Three species are known from single localities: *D. subelegans* Okada from Central Province, Sri Lanka and *D. sahydrii* Prakash and Reddy and *D. neoelegans* Gupta and Singh each from single localities in India (Lemeunier *et al.*, 1986; Okada, 1988). *Drosophila elegans* Bock and Wheeler has the widest distribution being reported from the Ryukyu Islands, the Philippines, New Guinea and Myanmar (Toda, 1991). *D. elegans* and *D. sahyadrii* breed in flowers of *Ipomoea* (Convolvulaceae) (Bock 1980; Okada and Carson 1982). Diagnosis for this subgroup includes an apical black patch on the wing of males as well as sex comb of males arranged in short transverse rows on 3 proximal tarsal segments of fore leg (Toda, 1991).

The *eugracilis* subgroup

The species of the monotypic *eugracilis* subgroup is a common fly encountered in the Oriental and Australasian regions (Lemeunier *et al.*, 1986; Appendix A). This subgroup's diagnosis includes a male sex comb of two teeth on the first tarsal segment and a sexually dimorphic abdomen which in the male is apically black and sharply truncate without protruding genitalia (Bock, 1980; Toda, 1991). Studies have presented hypotheses for the position of the *eugracilis* subgroup within the *melanogaster* group. Okada (1964) and Bock and Wheeler (1972) proposed that the *eugracilis* subgroup was related to the *ficusphila*, *suzukii* and *takahashii* subgroups based on the small hooked bristles of the mid-tibiae and secondary tarsal segment of males. Other hypotheses regarding the position of the *eugracilis* subgroup are similar (see below).

The *ficusphila* subgroup

The *ficusphila* subgroup contains 6 species (Appendix A). The *ficusphila* subgroup is distributed in the Oriental, Australasian and eastern Palearctic regions (Lemeunier *et al.*, 1986). *Drosophila ficusphila* Kikkawa and Peng is distributed across a relatively large area that includes Japan, Korea, the Ryukyu Islands, Taiwan, southern China, eastern India, Java, and the Andaman Island, Nicobar Islands, Australia, Myanmar (Toda, 1991). Whereas, the other five of the species have a more restricted distribution: *D. smithersi* Bock from Queensland, Australia; *D. gorokaensis* Okada and Carson from New Guinea; *D. flavicauda* Toda from Myanmar; and *D. kanaka* and *D. levii* from New Caldonia (Tsacas and Chassagnard, 1988; Toda, 1991). Ecology of this subgroup is not well known. *Drosophila ficusphila* Kikkawa and Peng derives its name from an association with *Ficus* sp. fruits; however, this association has not been confirmed by subsequent collections (Lemeunier *et al.*, 1986). *Drosophila gorokaensis* Okada and Carson was found emerging from the spadix of *Anydrium* (Araceae) (Okada and Carson, 1982). Tsacas and Chassagnard (1988) hypothesized based on distribution and morphology that *D. kanaka* and *D. levii* have a close affinity. The unique double longitudinal sex comb on the first and second tarsal

segments makes *ficusphila* subgroup species readily identifiable. The subgroup diagnosis also includes male characters such as an apical protuberance of the foreleg's second tarsal segment, an elongate, narrow cercus, constricted in the middle with longer bristles on the upper part than on the lower part plus an adeagal apodeme longer than the adeagus (Toda, 1991).

The *flavohirta* subgroup

The *flavohirta* subgroup was established by Bock (1980) for the single species, *D. flavohirta* Malloch, 1924 (Appendix A). This species is associated with *Eucalyptus* flowers and possesses unique body coloration, possibly for camouflage. Also *D. flavohirta* does not possess sex combs. Bock's (1980) inclusion of *D. flavohirta* in the *melanogaster* group is based on male genitalic characters. The subgroup's origin is clearly Australia but it has also been found associated with *Eucalyptus* introduced into Africa (Lemeunier *et al.*, 1986). The accidental introduction of *D. flavohirta* to southern Africa and Madagascar is a major problem for *Eucalyptus* in those areas (McEvey *et al.*, 1989).

The *longissima* subgroup

Toda (1991) recently established the *longissima* subgroup. This subgroup contains *D. longissima* Okada and Carson, previously of the *montium* subgroup, and *D. myamaungi* Toda. Both species possess the "montium-like" sex comb. Toda's (1991) diagnosis includes: simple surstylus without teeth, a very long, thick-tube adeagus and a small, arched hypandrium. These species are located in Papua New Guinea and Myanmar (Lemeunier *et al.*, 1986; Toda, 1991).

The *melanogaster* subgroup

The *melanogaster* subgroup contains 8 species with a uniquely restricted Afrotropical distribution with the exception of *D. melanogaster* Meigen, 1830 and *D. simulans* Sturtevant, 1919 which are human commensals with a worldwide distribution (Appendix A; Lemeunier *et al.*, 1986). Diagnosis for the *melanogaster* subgroup includes such characters as: several prominent bristles on the palpus, short, oblique sex comb located

distally on the first tarsal segment, and a surstylus with two sets of teeth – a medial or ventromedial row or cluster of pointed teeth and lateral row of more blunt, darker teeth (Toda, 1991).

Drosophila erecta and *D. orena* inhabit montane forests with *D. erecta* found on west coast, equatorial Africa (e.g., Ivory Coast, Guinea, Cameroon, etc.) and *D. orena* from a single collection on Mt. Lefo, west Cameroon (Lachaise *et al.*, 1988). Both species are believed to be host-plant specialists with *D. erecta* specializing on fruits of *Pandanus* (Pandanaceae) (Lachaise *et al.*, 1988). *Drosophila yakuba* and *D. teissieri* are host-plant generalists with a more widespread African distribution (Lachaise *et al.*, 1988). *Drosophila yakuba* primarily inhabits the savannas, whereas, *D. teissieri* primarily inhabits the forests (Lachaise *et al.*, 1988). *Drosophila sechellia*, endemic to the Seychelles Islands, is a host-plant specialist on *Morinda* (Rubiaceae) fruits; however, this may be by default since it is the only suitable resource available (Lachaise *et al.*, 1988). *Drosophila mauritiana*, endemic to the Island of Mauritius, is a domestic species whose native host-plant relationship is undeterminable since it exploits introduced fruits (Lachaise *et al.*, 1988).

Data from morphology, metaphase and polytene chromosome, allozyme, mtDNA, hydridization, courtship behavior and others has been used to establish three species complexes: *melanogaster*, *yakuba*, and *erecta*. A hypothesis of relationships for species in this subgroup is: (*erecta*, *orena* (*yakuba*, *teissieri* (*melanogaster* (*sechellia*, *maritiana*, *simulans*)))) (Lachaise *et al.*, 1988). After multiple types of data (e.g., DNA-DNA hybridization, chromosomal inversions, allozymes, etc.) and numerous genes sequenced Caccone *et al.* (1996) conclude that *D. sechellia* and *D. mauritiana* are sister taxa and that the similarities are not due to parallel evolution as argued by Coyne and Kreitman (1986). Despite numerous investigations (e.g., Lachaise *et al.*, 1988; Coyne and Kreitman, 1986; Hey and Kliman, 1993; Caccone *et al.*, 1988; Caccone *et al.*, 1996) the trichotomy of *sechellia*, *maritiana*, and *simulans* has yet to be convincingly resolved.

The *montium* subgroup

Largest of all the subgroups, *montium* contains 87 species (Appendix A). Members of this large speciose subgroup are found in the Oriental, Australasian, Neotropical, Afrotropical, and eastern Palearctic regions (Lemeunier *et al.*, 1986). *Drosophila kikkawai,* with its widespread distribution, is the sole Neotropical representative. The majority of species are Asian; therefore, the Oriental region is thought to be the origin for this subgroup (Lemeunier *et al.*, 1986). Members of the *montium* subgroup possess a distinctive large longitudinal sex comb along the entire length of the first and second tarsal segments (exception: members of the *nikananu* complex as well as *D. exiguitata* and *D. paraviaristata.* Also with the establishment of the *longissima* subgroup, now another subgroup possesses this distinguishing character [Toda, 1991]). Other diagnostic characters include: male external genitalia with surstylus and cercal clasper, the cercal clasper usually with very large bristles, and clearly defined apical bands on the abdomen in both sexes. Further investigation needs to be done as to the placement of *D. exiguitata* and *D. paraviaristata.* The male genitalia of *D. exiguitata* is somewhat unique, however, Bock (1980) felt it should be retained within the *montium* subgroup. Whereas, external genitalia of *D. paraviaristata* lacks a cercal clasper and was only provisionally retained in the *montium* subgroup by Bock (1980).

Approximately half of the total species have been placed within one of seven species complexes - *bocqueti, kikkawai, nikananu, serrata, bakoue, auraria, jambulina* (Bock and Wheeler 1972; Tsacas, 1984; Ohnishi and Watanabe, 1984; Lemeunier et al., 1986; Kimura, 1987; Kim *et al.*, 1989;Tsacas and Chassagnard, 1992; Gupta and Gupta, 1992; Kim *et al.*, 1993; Chassagnard *et al.*, 1997). These complexes were established using morphology (primarily male sex combs and external genitalia), hybridization and allozymes (e.g., Tsacas and Chassagnard, 1992; Ayala, 1965 a & b; Kim *et al.*, 1993). *D. barbarae,* originally included in the *kikkawai* complex (Tsacas and David, 1978), currently is a member of the *jambulina* complex based on allozyme and hybridization studies (Ohnishi and Watanabe,

1984; Kim *et al.*, 1989; Kim *et al.*, 1993). Despite numerous investigations, relationships are not well resolved (Lemeunier *et al.*, 1986; Drosopoulou and Scouras, 1995). This could be attributed to the relatively small sample sizes due to the enormous size of the group and the reliance of species in culture for biochemical analyses. A phylogenetic study using 49 exemplar species (24 *montium* species) and *cytochrome oxidase* ii, *Alcohol dehydrogenase* and *hunchback* gene regions was recently performed (Schawaroch, 2000).

The *rhopaloa* subgroup

A second subgroup recently established by Toda (1991), the *rhopaloa* subgroup, contains a total of 5 species (Appendix A; Toda 1991). The *rhopaloa* subgroup comprises species previously from the *suzukii* (1), *takahashii* (1), *montium* (2) subgroups and one new species – *D. fuyamai* Toda. Although unique male genitalic characters (e.g., including cercus with a single large tooth and hypandrium with 2 pairs of long processes in addition to anterior and posterior parameres [Toda, 1991]) establish this subgroup, the species vary in sex comb and wing characters; therefore, this could be an unstable association. The *rhopaloa* subgroup occurs in the Oriental region (Lemeunier *et al.*, 1986; Toda, 1991).

The *suzukii* subgroup

The *suzukii* subgroup contains 16 species and 1 subspecies with a distribution that includes the Oriental, eastern Palearctic, Afrotropical, and Australasian regions (Lemeunier *et al.*, 1986; Appendix A). Members vary with respect to morphology of sex-comb, aedeagus and anterior paramere (Bock and Wheeler, 1972; Bock, 1980; Toda, 1991). This miscellaneous collection of species has been united on the basis of male genitalic structures such as, surstylus with several sets of distinctly different teeth, cercus with lower bristles differentiated from upper bristles, and large posterior paramere (Toda, 1991). The monophyly of this subgroup has been questioned (Bock and Wheeler, 1972; Bock, 1980; Toda, 1991).

The *takahashii* subgroup

The *takahashii* subgroup contains 13 species (Appendix A), and is one of the better studied subgroups, based on allozymes, hybridization, metaphase and polytene chromosomes, and climatic adaptations (Parkash *et al.*, 1994; Kimura *et al.*, 1994; see also Lemeunier *et al.*, 1986). The *takahashii* subgroup is believed to have originated in India because this area exhibits the greatest endemism, although distribution of the subgroup also includes the Oriental, Australasian, eastern Palearctic regions (Lemeunier *et al.*, 1986). Hybridization tests suggest affinity of *paralutea*, *pseudotakahashii*, *lutescens*, *takahashii* and *trilutea* (Watanabe and Kawanishi, 1983; Bock and Wheeler, 1972; Lemeunier *et al.*, 1986). Toda (1991), who gave the latest diagnosis for the *takahashii* subgroup, noted that females alone cannot be identified to species. Toda's (1991) diagnosis includes: a single prominent bristle apically located on the palpus, sex combs arranged in short transverse rows on 2 proximal tarsal segments of fore leg, surstylus with comb of primary teeth ventrolaterally, row of secondary teeth dorsolaterally and several medial bristles and a pair of long or moderate submedian spines relatively close to each other on caudal margin of the hypandrium. Bock (1980) hypothesized that the *takahashii* subgroup was closest to the *melanogaster* subgroup. There have been other placements for the *takahashii* subgroup within the *melanogaster* group (see below).

Relationships among the subgroups have been hypothesized using morphological, chromosomal and 28S rRNA sequence data (Hsu 1949; Okada, 1954; Bock and Wheeler, 1972; Ashburner *et al.*, 1984; Pélandakis *et al.*, 1991 and Pélandakis and Solignac, 1993). The early studies of Hsu (1949) and Okada (1954) were done when less than 12% of the current species for the *melanogaster* group were known. Interestingly, both studies used male external genitalic characters, however, Hsu only used epandrial characters, whereas, Okada only used phallic characters. Despite different data sets and methods of analysis (i.e., evolutionary taxonomy for Hsu and phenetics for Okada), both agreed on the

relationships among species (i.e., *suzukii* is the most basal subgroup giving rise to the *melanogaster - takahashii – ficusphila* lineage and the *ananassae - montium - nipponica* lineage [*nipponica* is no longer part of the *melanogaster* species group see Okada, 1984]).

Bock and Wheeler's (1972) monograph on the *melanogaster* group presented two general conclusions on relationships among subgroups. First, based on the presence of both surstyli and cercal claspers, the *ananassae* and *montium* subgroups form a separate lineage from the other subgroups. Second, within the remaining subgroups a cluster of more closely related subgroups (i.e., *eugracilis, ficusphila, suzukii* and *takahashii*) could be recognized due to the presence of small hooked bristles on the mid-tibiae of the males and other characters from male genitalia. Using characters from morphology, metaphase chromosomes and polytene chromosome banding, Ashburner *et al.* (1984) concentrated on relationships of species within the *melanogaster* subgroup; however, they did recognize three distinct lineages within the *melanogaster* group: (1) *ananassae* subgroup, (2) *montium* subgroup and (3) the *elegans, eugracilis, ficusphila, melanogaster, suzukii, takahashii* subgroups. Ashburner *et al.* (1984) were not sure how these lineages were interrelated or where to place the remaining subgroups. Using 28S rRNA sequences, Pélandakis *et al.* (1991) and Pélandakis and Solignac (1993) studied the subgenus *Sophophora* and the genus *Drosophila*, respectively. Both studies proposed extensive paraphyly within the genus *Drosophila*, the subgenus *Sophophora*, and the *melanogaster* species group. With respect to the *melanogaster* group, they concluded with the use of parsimony and neighbor joining methods that 3 lineages exist: (1) the *obscura* and *fima* groups are allied with the *ananassae* subgroup, (2) there is a *montium* subgroup lineage, and (3) the *melanogaster* subgroups links with an Oriental subgroups lineage. (Oriental subgroups are the *elegans, eugracilis, ficusphila, suzukii*, and *takahashii* subgroups.) Pélandakis *et al.* (1991) concluded that the relationship among the three lineages remains unresolved and that the lineages proposed are in agreement with Ashburner *et al.* (1984) if

the placement of the *obscura* and *fima* groups is ignored. Placement of the *dentissima*, *flavohirita*, *longissima* and *rhopaloa* subgroups has not yet been addressed.

While the above-proposed relationships are largely in conflict, there are, however, some areas of general agreement. Bock and Wheeler (1972) agreed with Hsu (1949) and Okada (1954) with respect to the affinity between *ananassae* and *montium*. Bock and Wheeler (1972), Ashburner *et al.* (1984), Pélandakis *et al.* (1991) and Pélandakis and Solignac (1993) generally agreed that the *ananassae*, the *montium* and all other subgroups – *elegans*, *eugracilis*, *ficusphila*, *melanogaster*, *suzukii* and *takahashii* seemed distinct. Plus the *takahashii*, *suzukii*, *ficusphila* and *elegans* subgroups may have a closer affinity (Bock and Wheeler, 1972).

Starch gel electrophoretic and hybridization studies were used to determine relationships among eight species within the *melanogaster* group: four from the *montium* subgroup, one from the *suzukii* subgroup, one from the *takahashii* subgroup and two from the *melanogaster* subgroup (Kim and Lee, 1991; Kim *et al.*, 1992; Lee *et al.*, 1993; Lee *et al.*, 1994). The resulting UPGMA dendrogram yield two clades: a *montium* subgroup and *melanogaster* + Oriental subgroups. The phylogenetic analyses in Nigro *et al.* (1991) study of the *melanogaster* + Oriental subgroups lineage were questionable. This study compared mt DNA sequences for representatives *D. eugracilis*, *D. takahashii* and from the *melanogaster* subgroup. The scheme of relationships presented: *D. eugracilis* (*D. takahashii* (*melanogaster* subgroup)). This hypothesis is in agreement with Bock (1980) who felt that the *takahashii* subgroup was sister to the *melanogaster* subgroup.

Despite all the above research, relationships among the subgroups are still poorly understood. As early as 1972, Bock and Wheeler felt that a combination of morphological and biochemical data would be necessary to resolve the relationships within the *melanogaster* group. Morphology thus far has provided approximately 40 characters, hardly enough to resolve relationships among 174 species in the *melanogaster* group. Neither the Ashburner *et al.* (1984) study combining morphological and chromosomal data

nor the Pélandakis *et al.* (1991) and Pélandakis and Solignac (1993) studies employing 28SrRNA sequence data could resolve the relationship among the three lineages of *ananassae*, *montium* and *melanogaster* plus the Oriental subgroups. Pélandakis *et al.* (1991) concluded that these lineages evolved almost simultaneously. Although the D3 expansion region of 28S rDNA has provided character information for resolving relationships among the holometabolous insect orders (see Whiting *et al.*, 1997), the D1 and D2 regions of 28S rRNA are extremely invariant for taxa within the *melanogaster* species group (see figures in Pélandakis and Solignac, 1993; Schawaroch, 2000). As a result Pélandakis *et al.* (1991) consensus of equally most parsimonious trees yielded no resolution. Since neighbor-joining always produces a single resolved tree because the method of tree construction and the criterion for choosing a tree are conflated (Swofford *et al.*, 1996), Pélandakis *et al.* (1991) presented a neighbor-joining tree with bootstrap support values for the nodes. Even the bootstrap value at the node supporting the 3 lineages was very low (bootstrap value = 6). Pélandakis *et al.* (1991) suggested that more sequence data would be need.

In an attempt to maximize informative characters, the present study employs DNA sequences from three gene regions – two nuclear, *Alcohol dehydrogenase* (*Adh*) and *hunchback* (*hb*) and one mitochondrial, *cytochrome oxidase* ii (*co* ii). This study is the first to employ a cladistic analysis of this large complicated group, based on molecular data. In comparison to the previous molecular studies, this study's sample is much larger – 43 species representing 8 subgroups within the *melanogaster* group.

METHODS

Fly Stocks

Flies were obtained from the National *Drosophila* Species Resource Center at Bowling Green and D. LaChaise (Table 8). Molecular work was based on taxa from cultured species. Cultures were housed in a 21°C incubator on a 12-hour light: 12-hour dark cycle. Cultures were maintained on cornmeal-agar-yeast medium in $^1/_2$ pt milk bottles.

Taxon Sampling

A total of 49 taxa were used in this study (Table 9). The ingroup contains representative taxa from eight of the twelve subgroups within the *melanogaster* group. The majority of taxa (24 species) for this study were sampled from the *montium* subgroup. The remaining 19 species were from other subgroups in the *melanogaster* group, however, three subgroups were represented by only one taxon. Outgroup taxa consist of six species from the *obscura* group - two from each of the subgroups of *obscura*, *pseudoobscura* and *affinis* in the *obscura* group (Barrio *et al.*, 1994).

The Gene Regions

Three gene regions were used in this study, two nuclear (290 bp region of *Alcohol dehydrogenase* (*Adh*) and *hunchback* (*hb*)) and one mitochondrial (384 bp *cytochrome oxidase ii* (*co ii*)). Most drosophilids live upon and are attracted to the yeast growing on rotting fruit (Borror *et al.*, 1989), *Adh* is the enzyme which detoxifies alcohol thus insuring their survival (Geer *et al.*, 1985). A 290 bp region, which spans the second exon of *Adh*, was sequenced. A second nuclear gene region *hb* is a developmental gene and is expressed at various stages of development in *Drosophila*. It initially acts as a maternal morphogen establishing the anterior - posterior axis in the fertilized egg and later as a gap gene in early embryonic development (Tautz *et al.*, 1987). In this study a fragment of the *hb* coding region 5' end containing Box B (Tautz *et al.*, 1987) was sequenced and analyzed. The mitochondrial cytochrome transport molecule *co ii* was sequenced for 385 bp at the 3' end. This end was found to be the most variable part of the gene in other studies (Baker and DeSalle, 1997; Baker *et al.*, 1998). These DNA regions were chosen to facilitate comparisons with previous studies (e.g., Baker and DeSalle, 1997; Thomas and Hunt, 1993; Beckenbach *et al.*, 1993).

Molecular Techniques

Genomic DNA was prepared using single fly preps (DeSalle *et al.*, 1993). Fly DNA was PCR amplified using PE *taq* polymerase with primers developed for use on

Drosophilid taxa in (Appendix B). PCR products were purified using Gene Clean II (Bio 101). For approximately half of the taxa, the *Adh* fragment was cloned using Invitrogen's TA cloning kit. Gene regions were sequenced in both directions by either manual or automated sequencing methods. Most of the *Adh* and *co ii* sequences were generated manually; whereas, the *hb* sequences were done using an ABI 373 automated sequencer. Manual sequencing of double stranded PCR products and clones was done using ^{35}S and United States Biochemical's Sequenase version 2.0 DNA sequencing kit according to manufacturers instructions. Automated sequencing of double stranded PCR product was accomplished according to ABI Prism DNA sequencing kit, purified by sephadex columns and run on Applied Biosystems 373A DNA sequence protocols. Sequences were checked and corrected using Sequencher 3.0 (Gene Codes Corp.) sequence analysis software. Most of the DNA sequence was generated by this study with the exceptions noted (Table 10).

Character Assignment

All gene region sequence ends were trimmed to start at the same position from the furthest right and left primers. The *Adh* and *co ii* DNA sequences contained no indels (insertions or deletions); therefore, alignment was straightforward. Homology assessment for *hb* was more complicated plus gaps in the *hb* sequence were coded as a combination of "missing' and 5th state depending upon the alignment context. A thorough discussion of the alignment and gap coding for the *hb* gene region is in Appendix F.

Molecular Character Assessment

Previous studies of eukaryotes have reported bias in total nucleotide base composition for transitions, transversions (e.g., DeSalle *et al.*, 1987), and codon positions (e.g., Graybeal, 1993). Such biases would interfere with putative homology statements. Therefore, in this study character data was described with respect to the base composition, and the amount of transitions and transversions, as well as 1st, 2nd, and 3rd position changes in saturation plots. Saturation plots are an a priori attempt to estimate the unobservable changes in nucleotide sites or homoplasy in the sequence data. Distance models that correct

for biases are subsequently applied in likelihood analyses. Phylogenetically informative transitions, transversions and 1st, 2nd and 3rd position codon sites were examined for each of the three single data partitions as well as the total combined data set (i.e., *Adh*, *hb*, *co ii* and *Adh* + *hb* + *co ii*). Each model used here allowed for progressively more hidden variability within the sequence. The Jukes Cantor model (1969) assumes equal base frequencies and equal probabilty for all substitutions. The Kimura 2 parameter model (1980) assumes equal nucleotide base frequencies and allows the frequency of transitions and transversions to occur at different rates. The Felsenstein (1981) model (i.e., F81) allows for unequal nucleotide base frequencies but maintains that all substitutions are equally likely. The Hasagawa *et al.* model (1985) allows for both unequal nucleotide base frequencies as well as different frequency of occurrence for transitions and transversions. The LogDet model by Lockhart *et al.* (1994) allows for different pairs of sequences to have different relative base compositions whereas all other models assume the change is uniform across all pairwise comparisons. Saturation curves made with the uncorrected distance or uncorrected p-distance (i.e., a pairwise comparison where the total number of differences is divided by the total number of available sites) illustrate the equally weighted characters as they were used for my parsimony analyses.

Some authors (Bull *et al.*, 1993; Huelsenbeck *et al.*, 1994; Lutzoni, 1997; Shaffer *et al.*, 1997) are proponents of quantifying the variability among data partitions and using that as the criterion to determine if partitions should be combined in a simultaneous analysis. To quantify the congruence between each of the data partitions and the combined analysis tree, the incongruence length difference (ILD) (Mickevitch and Farris, 1981) was calculated for phylogenetically informative characters. ILD values were tested for significance (Farris *et al.*, 1994, 1995) using the partition-homogeneity test for 111 iterations with ten random addition TBR searches in PAUP 4.0 (Swofford, 2000). The resulting CI and RI estimated and actual values were compared. Estimated CI and RI values were calculated according to the formula: CI = partition 1 CI (% characters to combined analysis) + partition 2 CI (%

characters to combined analysis) + for the number of partitions in the combined analysis. To calculate the estimated RI, substitute RI for CI in the preceding formula.

Phylogenetic Analyses

Phylogenetic hypotheses were generated for all three gene regions separately and combined (i.e., *co ii*, *Adh*, *hb*, *Adh* + *hb*, *Adh* + *co ii*, and *hb* + *co ii*) to examine the signal contained within each partition. Finally, a combined analysis of all three genes for all the taxa was generated (Kluge, 1989). This combined analysis establishes a phylogenetic hypothesis for the *montium* subgroup as well as higher level relationships of the subgroups within the *melanogaster* group. The tree generated by the combined analysis allows a test of the monophyly of the *melanogaster* group and the *montium* subgroup. All the trees were rooted with the six outgroup species chosen from the sister taxon, the *obscura* group. Only informative characters were used to generate trees and tree statistics. PAUP 4.0 (Swofford, 2000) was used to generate all phylogenies. Heuristic tree searches were performed with random addition of taxa, TBR branch swapping and repeated 20 times. The characters were given an equal weight of 1 and run unordered.

The amount of character conflict in the *Adh* data resulted in the generation of number of trees beyond the memory capacity of PAUP 4.0. Therefore, *Adh* data was analyzed in three ways: (1) a single heuristic search was allowed to reach the maximum 14300 trees allowed by PAUP, then TBR branch swapping was performed on those trees; (2) a series of heuristic searches was run on PAUP (100 replicates with random addition of taxa, maxtrees=200 and TBR branch swapping conducted on all 200 tree topologies to determine the shortest [most parsimonious] topology. This search was done in an effort to sample a wider range of the tree space and to avoid being stuck on a local minimum or island); and (3) a series of heuristic searches (max*) using the program NONA (Goloboff, 1998) (1000 replicates [mult * 1000] with random addition of taxa, at the end of each search a single tree was saved and TBR branch swapping was done on each of those 1000 trees to select the most parsimonious topology). Because PAUP and NONA conduct TBR

swapping at different times searching of tree space is different. However the real crux of the matter is the amb - and amb + settings. PAUP 4.0 is designed to run amb - searches but the usual PAUP default and my searches were performed with amb + setting. The amb + setting results in tremendously more trees being saved during TBR branch swapping. NONA uses amb - therefore saving less trees during TBR swapping. The amb - command is more rigorous because it collapses branches if one of the optimizations does not support that branch. This difference between the two programs has been seen to produce different results if the data sets are homoplasious.

When there was more than one most equally parsimonious tree a strict consensus tree was generated and presented as phylogenetic hypotheses for that analysis. Successive weighting trees/ successive approximations analyses employing the iterative method of Farris (1969; Carpenter, 1988) based on the rescaled character c.i. values were generated for comparison to the separate phylogenetic hypotheses. When the successive weighting tree resulted in more than one tree a strict consensus was used for comparative purposes.

Character Support at Nodes

Support for the final hypothesis (i.e., combined analyses) was evaluated in three ways: (1) support for nodes using Bremer and bootstrap values, (2) individual data set contributions using PBS and counting shared nodes, and (3) exploring individual character support at controversial nodes using MacClade. To describe character support for nodes in the combined analysis, Bremer support values (Bremer, 1988; 1994) and bootstrap values (Felsenstein, 1985) were calculated (Autodecay ver. 2.9.8, Eriksson, 1998). Bootstrap analyses employed 1000 replicates with each replicate containing ten heuristic searches with random addition of taxa and TBR branch swapping. To quantify each of the individual gene region contributions to the combined analysis, partitioned Bremer support (PBS) (Baker and DeSalle, 1997) values were calculated.

RESULTS

Molecular Character Assessment

The number of informative characters contributed by each data set for phylogeny reconstruction is listed in Table 11. All the gene regions sampled in this study are protein coding. The relative character contribution for each codon position is listed in Table 12. Using phylogenetically informative characters, the transition and transversion changes for nucleotide base pair positions were presented in the context of tree topology (Figure 9 and 10) as well as measured using pairwise distance calculations in saturation plots. Transition and transversion saturation plots were made for the each individual gene region and the total combination (i.e., *Adh*, *hb*, *co ii*, *Adh* + *co ii* + *hb*) (Appendix G). Saturation plots were also made for codon positions (Appendix G). Second codon position saturation plots were not possible for *Adh* and *co ii* gene regions due to a lack of characters (refer to Table 12). The nucleotide compositions for the each of the gene regions comparing all character versus only phylogenetically informative characters are in Figure 11.

ILD calculations were made for all possible data combinations (Figure 12). A significance test was run comparing the three parts (*Adh*, *co ii* and *hb*) to the whole (simultaneous analysis). The partitions were found to be significantly different ($p = 0.009$ out of 111 iterations) (Figure 12). The PAUP 4.0 partition-homogeneity test used to determine if estimated samples of parts are significantly different to the whole, required 20 days using a fully dedicated IMAC computer. Estimated and actual CI and RI values were compared in Figure 12.

Phylogenetic Analyses

The total data set was subdivided into six separate analyses: *Adh*, *co ii* (mitochondrial), *hb*, *Adh* + *co ii*, *Adh* + *hb* (nuclear), and *hb* + *co ii*. Each separate and partial combined data analyses yielded more than one equally most parsimonious tree; therefore, strict consensus trees are presented with tree statistics in Appendix H. All three methods used to analyze the *Adh* data set produced a strict consensus tree with the same tree

topology and the same most parsimonious tree statistics. Tree statistics reported (i.e., number of equally most parsimonious trees) are from the NONA runs. The *Adh + co ii* analysis resulted in the fewest number of equally most parsimonious trees (four); each tree is pictured in Appendix H. Successive weighting / successive approximations analysis was performed for each of the six data sets (i.e., *Adh*, *co ii* (mitochondrial), *hb*, *Adh + co ii*, *Adh + hb* (nuclear), and *hb + co ii*. In all instances, except *Adh + hb*, more than one successive approximations tree was obtained, therefore the strict consensus tree was presented (Appendix H). None of successive weight tree topologies was contained within its respective original 'most parsimonious trees' topologies. A table was made for all the trees comparing CI, RI, resolution (i.e., number of nodes within the ingroup), monophyly of species subgroups, and presence of the 3 basal clades (Table 13). A simultaneous analysis (total evidence) of all three gene regions (*Adh + hb + co ii*) resulted in a single, most parsimonious tree (Figure 13).

Character Support at Nodes

Bremer, Bootstrap, and partitioned Bremer support values were calculated and compared for each node of the total evidence single most parsimonious cladogram (Figure 13 and Table 14). The total evidence most parsimonious cladogram nodes that are also present in the single and two data partitions combined cladograms are listed in Table 14. Total nodes matching the total evidence most parsimonious cladogram and each of the various data partitions strict consensus cladogram are listed in Table 15. Nodes were considered equivalent if they contained the same taxa in any order. Only nodes present in the strict consensus cladogram were counted; even though, one of the equivalently most parsimonious cladograms may have contained structure for that node.

DISCUSSION

This discussion will first describe the molecular data in general and then discuss the single, most parsimonious cladogram resulting from a total evidence analysis in a phylogenetic context.

DESCRIBING MOLECULAR CHARACTER ASSESSMENT, TREES AND NODE SUPPORT

A. MOLECULAR CHARACTER ASSESSMENT

Only approximately 30% of all the nucleotides sequenced were considered informative to phylogeny reconstruction of the *melanogaster* species group (Table 11). All the DNA regions sequenced coded for particular genes. As expected, the third position sites contributed the greatest number of phylogenetically informative characters (Table 12). The first and second codon positions also contributed character information with the second position having the least contribution.

Base Composition

The overall nucleotide base composition for each gene region as well as just phylogenetically informative sites is depicted in Figure 11. The percent nucleotide base composition for nuclear data combined or separate (i.e., *Adh* + *hb*, *Adh*, *hb*) varies between all sites and only phylogenetically informative sites (from the greatest to the least: C, A, G, T versus C, G, T, A). This change in relative composition could be result of higher level phenomenon e.g., codon bias, 3-D ergonomics, etc.. Nuclear regions combined or separate have a bias in favor of the C nucleotide base. As in other insect mitochondrial studies (Clary and Wolstenholme, 1985; DeSalle *et al.*, 1987; Liu and Beckenbach, 1992), the *co ii* gene region has a strong A and T bias with relatively little C and G nucleotide base content. Interestingly, in light of the combined versus separate data analyses controversy, the combined data percent nucleotide base content is almost equally distributed among the four nucleotide bases.

Saturation

Transitions and Transversions

Total transitions outnumber transversions with the *hb* data being the most extreme, *co ii* the least extreme, and both *Adh* and all 3 genes combined having the same ratio (Figure 9). In most instances, among the transitions TC and CT changes are the more prevalent

especially within *co ii* TC 69% and *Adh* CT 55% (Figure 10). Among types of transversions, *co ii* data is very biased favoring AT and TA changes (49% and 50% respectively), little CA changes (0.5%), and no GC and CG changes (undefined) (Figure 10). This can be attributed to the nucleotide base bias (Figure 11). Although the other gene regions are not as extreme in regards to types of transitions, biases can be noted, such as: *hb* favors CA and GT changes, *Adh* favors CA, GC and CG changes and all three gene regions combined favor AT and TA changes (Figure 10).

Saturation Curves. - Generally, as the various distance models allow for more biases in base changes, there seems to be greater tendency for each data set to exhibit saturation. Within the *Adh* data transition and transversion saturation begins with the HKY85 model and becomes more pronounced in the LogDet model (Appendix G). The *hb* data exhibits saturation for the transitions and transversions beginning with the Jukes Cantor model and increasing with each model thereafter (Appendix G). Within the *co ii* data both transitions and transversions seem to become saturated with F81 and Kimura 2 parameter models and are highly saturated with the HKY85 and LogDet models (Appendix G). The combined (all 3 gene regions) data, which was used for phylogenetic interpretations, exhibits slight saturation for transversions in HKY85 and LogDet models (Appendix G).

The "saturation" seen in some graphs is an artifact resulting from misapplication of the models. The distance models allow for various unobservable changes in the data and are used to straighten the leveling off curve of a graphed line (the saturated line). Interestingly all gene regions, separate and combined, using uncorrected p-distance do not exhibit saturation for either the transitions or transversions and do not require corrections to be made (Figure 14 and 15). My phylogenetic analyses involve the combined (all 3 gene regions) data and an uncorrected p-distance model neither of which are saturated.

1st, 2nd, and 3rd positions

Saturation curves for codon positions were based on informative characters only. Second positions yield very few characters and were not plotted for *Adh* and *co ii* gene regions (Table 12 and Appendix G). The first positions have very few characters (Appendix G). Both the first and second position curves each show a cloud of scattered points meaning no saturation and few characters. The third position codon sites provide the greatest number of characters (Table 12). For both the *co ii* and *Adh* data sets saturation appears in the HKY85 and Log Det models (Appendix G). The *hb* 3rd positions are saturated in the Kimura 2 parameter, HKY85 and LogDet models (Appendix G). The third position for all three gene regions combined are not saturated (Appendix G).

As seen in the transition and transversion saturation plots, the "saturation" seen in some third codon position graphs is an artifact resulting from misapplication of the models. For all gene regions, separate and combined, the uncorrected p-distance does not exhibit saturation or require corrections to be made. My phylogenetic analyses involve the combined (all 3 gene regions) data and an uncorrected p-distance model neither of which are saturated (Figure 16).

Incongruence Length Difference (ILD)

Combining data partitions increases character conflict as reflected in the ILD values (Figure 12). This increased character conflict (homoplasy) is also reflected in the combined CI and RI values being lower than the estimated CI and RI values (Figure 12). The CI and RI values of the separate partition cladograms are ranked from greatest to least *hb* > *Adh* > *co ii*. A separately analyzed data partition, however, provides information such as resolution and character internal consistency (CI) on that data partitions' cladogram. In order to understand the contribution of a data partition within a combined analysis, the partition must be examined in the context of the combined analysis cladogram. Therefore, each data partition (*Adh*, *co ii* and *hb*) was individually placed on the simultaneous analysis cladogram. The CI and RI values of each partition on the simultaneous analysis cladogram

can be ranked from greatest to least *hb* > *Adh* > *co ii*. (Actually the same as resulted from separately analyzed partitions). Ensemble character consistency for each nuclear gene region (*Adh* and *hb*) is greater on the simultaneous analysis topology than the total combined data set CI (Figure 12). The mitochondrial gene region (*co ii*) emsemble character consistency on the simultaneous analysis cladogram topology is lower than the combined data set CI (Figure 12).

My simultaneous analysis matrix contains heterogeneous data partitions as determined using the PAUP partition-homogeneity test ($p = 0.009$). Bull *et al.* (1993) are opposed to combining heterogeneous data partitions because the heterogeneity may either (1) indicate radically different evolutionary history, such as horizontal transfer, or (2) produce an erroneous cladogram topology. The prior agreement approach has problems, for example, in the case of two data partitions: which one is "correct" and which is "erroneous"? More data partitions may not make the decision any easier. Baker and DeSalle's (1997) data contained eight gene region partitions which were heterogeneous to the point that no two partitions could be combined. By not including all the data, resolution could be lost especially if data partitions contribute information for different levels of the analysis (e.g., Hillis, 1987). Studies have demonstrated that simultaneous analyses provide greater resolution (Remsen and DeSalle, 1998; Miller *et al.*, 1997; Olmstead and Sweere, 1994). In addition, combining heterogeneous data sets provides increased support for nodes as inferred from Bremer values (Remsen and DeSalle, 1998; Baker and DeSalle, 1997).

Unless there is evidence to the contrary (e.g., horizontal transfer of transposable elements) all attributes of an organism have a shared history. Sources of evidence (i.e., characters) should be varied, thereby negating the problem of single character or gene phylogenies (or homogenized data). Corroboration between various sources of characters (e.g., mtDNA, genomic DNA, morphology [head, internal and external genitalia, etc.]) including heterogeneous ones will most probably produce the most robust hypotheses.

These reasons have contributed to my decision to use a simultaneous analysis of all three gene regions for my hypothesis of relationships.

B. TREES

Data Partitions

The *Adh* data were difficult to analyze and required an different analysis program. This difficulty cannot solely be due to (1) a large number of equally most parsimonious trees, because *hb* had more trees or (2) a large amount of homoplasy because *co ii* had a lower CI. The problem may be attributable to a small number of informative characters relative to the number of taxa. This ratio combined with a lack of resolution (low number of nodes for the ingroup) caused the program to swap many equally possible topologies extending the length and taxing the memory of the cladogram search.

.In general, as the number of gene regions combined in the analysis increased so did the resolution, with the exceptions of *co ii* and *hb + co ii* (Appendix H and Table 13). The simultaneous analysis of two or more data sets had all but the *suzukii* subgroup monophyletic. The *Adh* and *hb* gene regions each had no monophyly for the *suzukii* and *takahashii* subgroups, and *co ii* had these plus the *ananassae* subgroup not monophyletic. The 3 basal clades (i.e., *ananassae* subgroup, *montium* subgroup, and *melanogaster* + Oriental subgroups) appeared in 4 different analyses each presenting one of the four possible combinations.

The number of taxa rather than the number of characters effects the CI value (Sanderson and Donoghue, 1989; Meier *et al.*, 1991; Baker *et al.*, 1998). Based on the regression line for numerous morphological and molecular data sets, Sanderson and Donoghue (1989) derived a formula for estimating the CI (i.e., CI + 0.90-0.022[number of taxa]+0.000213[number of taxa]^2). For my sample of 49 taxa, the CI should equal 0.333. The CI values for *Adh + co ii*, *hb + co ii*, and *Adh + hb + co ii* are very close to Sanderson and Donoghue's (1989) estimate (Table 6). The *hb* CI could be inflated because the data matrix contained missing values (Sanderson and Donoghue, 1989). The low CI value for

co ii reflects greater homoplasy which may be the result of A and T nucleotide base (character) bias. The nuclear gene regions by themselves (i.e., *Adh* and *hb*) and combined (i.e., *Adh* + *hb*) have elevated CI values.

The RI does directly correspond with the CI but not with the amount of resolution (Table 8). RI reflects the amount of synapomorphy, especially the number of deep internal nodes (Farris, 1989; Siebert, 1992). The definition of deep internal nodes seems subjective, especially in a well resolved tree. As defined here, resolution includes all nodes for the ingroup not just the deep internal nodes. The resolution and the RI values do not have a correlation (Table 13). The CI seems to be a better comparison value than RI (Goloboff, 1991).

Successive weighting analyses provided trees with more resolution (Appendix H). The resolution, however, was not simply adding to the previously existing nodes but some nodes actually changed. Successive weighting trees retained the previous monophyly or lack of monophyly for the subgroups that was exhibited in the initial strict consensus tree with the exception of the *Adh* + *co ii* successive weighting tree where monophyly decreased when the *takahashii* subgroup became paraphyletic. The three basal clades if previously present, remained; however, the relationships in the *hb* + *co ii* successive weighting tree became resolved.

Total Evidence Analysis

Interestingly, the total evidence analysis resulted in a single, most parsimonious cladogram (Figure 13). The total evidence cladogram is well resolved and the resolution much greater than other analyses (single and two gene regions) (Table 13). The total evidence tree has CI and RI values lower than *Adh*, *hb*, or *Adh* + *hb* trees (Table 13). The CI value, however, is appropriate for the number of taxa (Sanderson and Donoghue, 1989). The total evidence tree had a high level of monophyly for the traditionally established taxonomic groups (Table 13). The three basal clades of the *ananassae*, *montium* and

melanogaster + Oriental subgroups were present (also in *hb*, *Adh* + *hb*, and *hb* + *co ii* cladograms).

Previous studies have demonstrated that combined analysis cladograms produce different even unpredictable results from the separate analyses (Barrett *et al.*, 1991). A total evidence analysis employs less extraneous hypotheses to character data (Kluge, 1989). My total evidence cladogram had the best resolution, high level of monophyly, and an appropriate CI for the number of taxa. It seems better to maximally summarize all character evidence simultaneously than to try and combine various tree topology components. For these reasons, I have used the total evidence cladogram as a basis for my phylogenetic analyses.

C. NODE SUPPORT

Bremer, Bootstrap, Partitioned Bremer Support

The BS and B values do not always exhibit a direct correlation (see nodes 2, 4, 19, 24, 28, 38, 39 and 44) (Figures 13 and Table 13). A node containing a Bremer support about 3 or better and bootstrap value greater than 50% is considered well supported. The partitioned Bremer support values present the relative contributions of each separate data partition to the individual nodes Bremer values (Table 13). As determined by PBS analysis, nodes 25, 31, 39 and 44 are supported by a single data set, while the other two data sets contribute zero (25 and 31- *co ii*, 39 – *Adh* and 44 – *hb*). The most parsimonious tree topology was contradicted most often by *Adh* data (17), than *co ii* (11) and than *hb* (9). There were 6 occurrences when a single data set (*hb*) was contradicted by two negative PBS values (from *co ii* and *Adh*) at a single node. Based on PBS values, *hb* seems most favorable to nodes in the total evidence most parsimonious cladogram despite the difference in cladogram topologies.

Cladogram nodes among the various separate and partial combined analyses were compared to the total evidence cladogram (Table 13). Besides new nodes appearing in the total evidence tree (e.g., 7, 31, 37 and 46), support for tree nodes can not always be predicted

either. Although a node may be present in all six different data combinations (i.e., 5, 6, 11, 18, 29, 32, 40, 41, 42, and 45) it did not always have 100% bootstrap value (i.e., 11, 41, 42, and 45) or all positive PBS values (e.g., 11, 29, and 41). The total number of tree nodes in common between the total evidence tree and each individual gene region is inversely related to the total number of negative PBS values. Therefore, *Adh* shares 19 nodes, *co ii* 24 nodes and *hb* 25 nodes with the total evidence tree (Table 13).

PHYLOGENETIC DISCUSSION

Phylogenetic hypotheses are constrained by the taxa sampled. The importance of taxon choice and its ability to influence cladogram structure has been demonstrated (Lecointre *et al.*, 1993; Graybeal, 1998; Hillis, 1998; Poe, 1998). To obtain an adequate DNA sample for small species, such as drosophilids, the whole specimen must be sacrificed. Fortunately, a number of drosophild species are maintained in laboratory cultures. These cultures, however, limit the scope of the molecular studies. Most, but not all, of the available stocks have been included in my study. Forty-three species have been sampled from the *melanogaster* group which is only 25% of the currently known species. However, this is the greatest sampling of representative taxa within the *melanogaster* group for any biochemical investigation (almost sampling Bock and Wheeler's 1972 monograph where most of these cultures were established). With the caveat above in mind, I would like to examine the phylogenetic hypotheses derived from my most parsimonious cladogram, which is the most thorough hypothesis to date (Figure 17).

Testing Monophyly and Relationships Within Subgroups

The *melanogaster* group

A necessary prerequisite for phylogenetic investigations is that the group of interest must be "natural" or monophyletic which is defined by shared derived features (synapomorphies) (Hennig, 1966). Putative homologies must undergo the test of cladogram construction in order to demonstrate that they are actually synapomorphies (Patterson 1982; DePinna, 1991). All of the literature concerning the *melanogaster* group

has not employed cladogram construction, but, in their own way, researchers have proposed putative synapomorphies defining the groups (subgroups etc); it is these characters on which I will be focusing much of my discussion. Since the establishment of the *melanogaster* group in 1942 by A. H. Sturtevant, morphological characters such as body coloration, presence of male sex combs, and characters from male genitalia, female internal genitalia, and setae on the palp, oral region and head, have been used to support or indicate the group's monophyly. In the most parsimonious cladogram, monophyly was supported for the *melanogaster* group (BS = 36; B = 100%). This demonstrates that besides morphology many of the DNA characters sampled strongly support a monophyletic *melanogaster* group. These findings are in contrast to Pélandakis *et al.* (1991) and Pélandakis and Solignac (1993) who found the *melanogaster* group to be paraphyletic.

My analysis included representative taxa from 8 of the 12 subgroups within the *melanogaster* group. Since the *elegans* and *ficusphila* subgroups were each represented by a single taxon, in this study, monophyly could not be tested. Monophyly also was untestable for the *eugracilis* subgroup, which is monotypic. Of the remaining five subgroups, four were found to be monophyletic: *ananassae*, *melanogaster*, *montium* and *takahashii*; however, monophyly was not supported for the *suzukii* subgroup.

The *ananassae* subgroup

The *ananassae* subgroup has been characterized by the presence of a cercal clasper and a surstylus with two sets of teeth. For members of this subgroup, *D. varians* is morphologically unique - lacking the cercal clasper and possessing a cercal plate with bristles similar to species of the *suzukii* subgroup. *Drosophila varians* has been included within the *ananassae* subgroup based on chromosomal homology (Bock and Wheeler, 1972). In this study six species were chosen to represent the *ananassae* subgroup three from the *ananassae* complex (*D. ananassae*, *pallidosa* and *phaeopleura*), one from the *bipectinata* complex (*D. malerkotliana*), one from the *ercepeae* complex (*D. ercepeae*) and one unassigned species, *D. varians*. My most parsimonious cladogram supported

monophyly for the *ananassae* subgroup with very large Bremer and Bootstrap values (BS = 9, and B = 98%). Within the *ananassae* subgroup clade, representatives of the *ananassae* complex species form a cluster. Since the remaining species complexes have only single representatives it can not be determined if they are natural groups. It is interesting to note that despite previous studies questioning its inclusion, the morphologicallly aberrant *D. varians* does turn out to be nested within the *ananassae* subgroup.

The *melanogaster* subgroup

The *melanogaster* subgroup has been well investigated and much data has been used to indicate its monophyly (Lachaise *et al.*, 1988). This study sampled the familiar *D. melanogaster* and the two species of the *yakuba* complex *D. yakuba* and *D. teissieri*. The *melanogaster* subgroup forms a well supported clade (BS = 8 and B = 90%) located in a relatively more derived position within various Oriental species and subgroups. The *yakuba* complex has even greater Bremer and Bootstrap values (B = 16 and B = 100%). With respect to the species sampled, my cladogram is in agreement with the current hypothesis of relationships within the *melanogaster* subgroup.

The *montium* subgroup

The *montium* subgroup is monophyletic with very large Bremer and Bootstrap values (BS = 14, and B = 99%) at the basal node. A monophyletic *montium* subgroup is in agreement with numerous morphological and biochemical studies that have indicated putative synapomorphies (e.g, Bock and Wheeler, 1972; Ashburner *et al.*, 1984; Scouras, 1995). These finding are in contrast to Tsacas and David (1978) who felt that due to its enormous size and distribution the *montium* subgroup could not be monophyletic. These criteria are not applicable in phylogenetic studies that define monophyletic groups. Toda (1991) removed species (i.e., *D. rhopaloa, palmata* and *longissima*) from the *montium* subgroup when he established the *rhopaloa* and *longissima* subgroups. Unfortunately, none of these species or other representatives of these subgroups were included in the

present study. One way to assess this classification in lieu of sequencing is to assess diagnostic value of morphological characters (determined by total evidence of the analyzed taxa). The two taxa in the *longissima* subgroup do, however, possess a sex comb identical to ones in the *montium* subgroup. Since this morphological character corroborates the molecular phylogeny presented here, the *longissima* subgroup may either be a complex within *montium* or a subgroup sister to the *montium* subgroup. Many morphological features of the *rhopaloa* subgroup are variable and its monophyly seems questionable (to me) at this time. A thorough discussion of the complexes and relationships within this large complicated subgroup can be found in Schawaroch (2000). It can be noted that: (1) *D. barbarae* is not a member of a clade which contains either *jambulina* or *kikkawai* complex representatives, and (2) much of the resolution within the *montium* subgroup is supported by Bremer values less than or equal to 2 and Bootstrap values less than 50%.

The *takahashii* subgroup

My most parsimonious cladogram supports monophyly for the *takahashii* subgroup, even though Bremer and Bootstrap values are relatively low (BS = 1 and B = 43%). The division of the *takahashii* subgroup into two clades may indicate the presence of species complexes: *paralutea* + *prostipennis* and *takahashii* + *lutescens*. This differs from the hypothesized affinity of *D. lutescens*, *paralutea*, *pseudotakahashii*, *takahashii* and *trilutea* based on hybrization tests (Watanabe and Kawanishi, 1983; Bock and Wheeler, 1972; Lemeunier *et al.*, 1986). A morphological feature that seems to correspond with this division seen in my cladogram is the number of rows of male sex comb in the second tarsal segment (see data presented in Table by Bock and Wheeler, 1972 p.13). Although included within the *melanogaster* + Oriental subgroups clade, the *takahashii* subgroup is not the sister taxon to the *melanogaster* subgroup as Bock (1980) hypothesized.

The *suzukii* subgroup

Members of this subgroup apparently exhibit the greatest range in morphological characters, particularly for sex comb, phallic and periphallic structures. The putative

synapomorphies for this subgroup are generalized male genitalic characters, such as surstylus with several sets of distinctly different teeth, cercus with lower bristles differentiated from upper bristles, and large posterior paramere (Toda, 1991). For these reasons monophyly of the *suzukii* subgroup has been questioned (Bock and Wheeler, 1972; Bock, 1980; Toda, 1991). In my most parsimonious cladogram the *suzukii* subgroup was polyphyletic. The *suzukii* subgroup representatives in this study exhibit the complete range in variation with respect to the sex combs and molecularly-based clades seem to fall out along sex comb morphology. *Drosophila mimetica* is sister to the *takahashii* clade (BS = 3; B = 59%), and all taxa have similar sex combs – a transverse row each on the first and second tarsal segments. *Drosophila biarmipes* is the most basal member in a clade containing the *melanogaster* and *eugracilis* subgroups (BS = 2; B = 40%). All taxa have a sex comb located on the first tarsal segment, however, the *D. biarmpes* and the *melanogaster* subgroup sex comb is in an oblique orientation, whereas, the *eugracilis* sex comb contains two teeth of either oblique or longitudinal orientation. *Drosophila lucipennis* forms a clade with *D. elegans* (BS = 12; B = 99%). *Drosophila lucipennis* has no sex comb, in contrast to *D. elegans* whose sex comb is a series of transverse rows along the first 3 tarsal segments. When Bock and Wheeler established the *elegans* subgroup for the species *D. elegans*, they mentioned that *D. elegans* "...bears some superficial resemblance to several of the members of the *suzukii* subgroup but differs substantially in the structure of the male genitalia" (p.27-28). Only the *D. lucipennis* + *D. elegans* clade has high Bremer and Bootstrap values. Low Bremer and Bootstrap values were also seen in relationships between the clades containing *D. mimetica* and *D. biarmipes* (BS = 2; B = 40%) and the sister to this clade which contains *D. lucipennis* (BS = 3; B = 63%). Further work on the status and/or redefinition of the *suzukii* subgroup is indicated.

Each of the three clades: *D mimetica - takahashii* subgroup, *D.biarmpes-eugracilis-melanogaster* subgroups, and *D. lucipennis - elegans* subgroup are supported by 5, 7 and 17 unambiguous characters (Figure 18). Most of the character ci and ri's are not very

large. It is interesting to note that the *D mimetica - takahashii* subgroup clade and *D. lucipennis - elegans* subgroup clade are each supported by a character with a c.i. and r.i. value of 1, or a distinctive synapomorphy in support of that clade.

Exploring alternative topologies. - Alternative topologies/hypotheses were explored to estimate the cost of creating a monophyletic *suzukii* subgroup as well as to determine if a strong affinity exists with another clade. The shortest monophyletic combination for each of the 3 previous *suzukii* taxa locations are presented in Figure 19. Support for a monophyletic *suzukii* subgroup would minimally increase the most parsimonious cladogram length by 31 steps , decrease the number of unambiguous characters support at the node, and reduce the CI and RI values. All three tree topologies (A, B, and C) have similar length, CI and RI values; however, topology "C" is inferior because it does not possess any unambiguous synapomorphy supporting a *suzukii* subgroup clade. Of the three alternative hypotheses/topologies the least costly arrangement places the *suzukii* subgroup sister to the *takahashii* subgroup. But, of course, according to my data a polyphyletic *suzukii* subgroup is best and in keeping with previous hypotheses.

Relationships Among Clades

The 3 lineages

In my most parsimonious cladogram the *melanogaster* group is subdivided into three major clades: the *ananassae* subgroup, the *montium* subgroup, and the *melanogaster* + Oriental subgroups. The presence of these three lineages is in agreement with Ashburner *et al.* (1984). Ashburner *et al.* (1984), however, did not hypothesize any relationships among these lineages. My hypothesis also agrees with Pélandakis *et al.* (1991) who interpret their findings to support the three lineages of Ashburner *et al.* 1984 with the exception of the *obscura* and *fima* groups within the *melanogaster* group. To explain the weak support for the relationship among the three lineages, Pélandakis *et al.* (1991) concluded that these lineages evolved almost simultaneously. Their result, however, most likely is attributable to the lack of variation within the 28S rDNA gene region at this

phylogenetic level. My most parsimonious cladogram hypothesized the relationship for the three lineages as: (*ananassae*, *montium*) *melanogaster* + Oriental subgroups.

The *ananassae* and *montium* subgroups are sister taxa

In my most parsimonius cladogram the *ananassae* and *montium* subgroups are sister taxa. The node supporting this relationship has low Bremer and Bootstrap values (BS 1 and B = 35%). Despite the weak support, the sister relationship between the *ananassae* and *montium* subgroup also has a morphological basis. The very early studies of Hsu (1949) and Okada (1954) noticed an affinity between the *ananassae* and *montium* subgroups. Bock and Wheeler (1972) hypothesized that the *ananassae* and *montium* subgroups formed a separate lineage based on the putative synapomorphies of a surstylus and cercal claspers. Ashburner *et al.* (1984) hypothesized no resolution among the three lineages. Pélandakis *et al.* (1991) and even Pélandakis and Solignac (1993) presented the relationship: subgenus *Drosophila* (*melanogaster* + oriental (*montium*, *ananassae* + *fima*+ *obscura*)). Despite the paraphyly and the lack of characters *ananassae* and *montium* are still seen as having a greater affinity.

According to the partitioned Bremer support values, *hb* data lends support to an *ananassae* + *montium* clade whereas *Adh* and *co ii* each do not (i.e., *hb* = 2.16, *Adh* = -0.33 and *co ii* = -0.83). The *ananassae* + *montium* subgroups node contains 8 unambiguous characters (Figure 20). Character 612 has a ci and ri value of 1.0 and character 121 has a ci = 0.5 and ri = 0.9, while the remaining characters have lower values. Character 612 is diagnostic for the *ananassae* + *montium* subgroups sister relationship.

Exploring alternative hypotheses. - The *ananassae* + *melanogaster* + Oriental clade has 7 unambiguous characters supporting its node (Figure 21). This tree is only one step longer than my most parsimonious cladogram. Yet, the data seems less supportive for this configuration since there are no individual characters with a ci or ri value of 1. A *montium* + *melanogaster* + Oriental clade hypothesis has even less support. The tree is 4 steps longer and there are only 4 unambiguous characters at the node, each with c.i. values <

0.5 (Figure 22). It is interesting to explore the feasibility of the alternative hypotheses/topologies, especially considering the controversy at this node. However, my original, most parsimonious cladogram is the best summary of all the character information contained in my data, and the corroborating morphological characters "presence of a surstylus and a cercal clasper" support an *ananassae* + *montium* subgroups sister relationship.

Relationships within the *melanogaster* + Oriental subgroups

The *melanogaster* + Oriental subgroups clade seems well supported with a Bremer value of 3 and a bootstrap value of 85%. The relationships of the subgroups within this clade can be summarized as: ((((*melanogaster*, *eugracilis*) *suzukii*) (*takahashii*, *suzukii*)) *elegans*, *suzukii*)) *ficusphila*. The support for these relationships varies but they have relatively low Bremer (< 3) and Bootstrap (< 50%) values. My cladogram disagrees with previous hypotheses.

The early studies of Hsu (1949) and Okada (1954) placed the *suzukii* subgroup at the base of the *melanogaster* group. Using my most parsimonious cladogram, even within the *melanogaster* + Oriental clade, none of the three representative *suzukii* subgroup taxa are basal. Bock (1980) felt the *takahashii* subgroup was closest to the *melanogaster* subgroup. The studies of Kim and Lee (1991), Kim *et al.* (1992), Lee *et al.* (1993), and Lee *et al.* (1994) at the subgroup level hypothesized a hierarchy of (*melanogaster*, *takahashii*) *suzukii*. Nigro *et al.* (1991) present a scheme of relationships for the *melanogaster* + Oriental subgroups as: *D. eugracilis* (*D. takahashii* (*melanogaster* subgroup)). Okada (1964) and Bock and Wheeler (1972) each felt that due to the putative synapomorphies of hooked setae on the mid- tibiae of males and other characters of the male genitalia, the *eugracilis*, *ficusphila*, *suzukii* and *takahashii* subgroups have a closer affinity. My most parsimonious cladogram does not support this hypothesis because the *melanogaster* and *elegans* subgroups are contained within that clade.

Mapping the hooked seta character on the most parsimonious cladogram (i.e., *eugracilis*, *ficusphila*, *suzukii*, and *takahashii* representatives are coded for presence, all other taxa absence) increases the length of the tree by three steps and does not affect the CI and RI values (now has 343 informative characters versus the usual 342 therefore L= 1543, CI= 0.349 RI= 0.667 see Figure 23). According to the most parsimonious cladogram, the presence of hooked setae on the mid-tibia evolved once at the base of the *melanogaster* + Oriental subgroups clade and was lost twice: (1) at the node for the *melanogaster* subgroup clade and (2) at the terminal for the *elegans* subgroup.

The subdivisions within the *melanogaster* + Oriental subgroups clade parallels sex comb morphology changes (see discussion for *suzukii* subgroup).

CHAPTER 3

Species Relationships in the *Drosophila montium* Subgroup

CONTENTS FOR CHAPTER 3

INTRODUCTION

Besides being the most diverse subgroup of the *melanogaster* group, the *montium* subgroup is additionally interesting for the following reasons. Species within the *montium* subgroup possess Balbiani rings (Mavragani-Tsipidou *et al.*, 1994; Drosopoulou and Scouras, 1995; Scouras, 1995) which have been observed thus far only in the fly family, Chronomidae. Balbini rings are found in chromosomal regions containing a high number of reverse tandem duplications; although the evolutionary origin of Balbiani rings is unknown transposable elements may be involved (Scouras, 1995). The *montium* subgroup also provides a model system for the study of the transmission (horizontal versus vertical) of the transposable element, *hobo*. Within the genus *Drosophila*, *hobo* and *hobo*-like sequences are restricted to the two species subgroups, *melanogaster* and *montium* (Daniels *et al.*, 1990). For species within the *melanogaster* group (i.e., *D. melanogaster*, *D. simulans* and *D. mauritiana*) *hobo* has been hypothesized to be horizontally transferred (Simmons, 1992). Studies of the position of ß-tubulin genes within the *montium* subgroup seem to corroborate Ohno's (1970) theory for the evolution of a gene family (Mavragani-Tsipidou *et al.*, 1994; Drosopoulou and Scouras, 1995; Scouras, 1995). The closely related species of the *serrata* compex have been used to study speciation processes (Ayala 1965a; 1965b; Baimai, 1970a; 1970b). Kimura (1987) has studied the adaptive radiation of the closely related species of the *auraria* complex. Many of these conclusions are conjecture because a rigorous phylogeny of the *montium* subgroup is necessary for the understanding and interpretation of these phenomena.

The genus *Drosophila* Fallén, as presently defined with 12 subgenera, contains approximately 1,000 species (Wheeler, 1981; Wheeler, 1986; Grimaldi, 1990). The subgenus *Sophophora* Sturtevant, being of exceptional biological interest, since it comprises *D. melanogaster* and relatives. *Sophophora* contains seven named species groups and some unclassified species (the exact composition of *Sophophora* varies between authors; see Lemeunier *et al.*, 1986, and Scouras, 1995, versus Ashburner, 1989). The *melanogaster*

species group contains 174 species and 3 subspecies and is the largest species group in the genus *Drosophila* (Toda, 1991; Appendix A). These species, with the exception of 5 species that are *incertae sedis*, are unequally distributed among 12 subgroups. The 12 subgroups are *ananassae*, *denticulata*, *elegans*, *eugracilis*, *ficusphila*, *flavohirta*, *longissima*, *melanogaster*, *montium*, *rhopaloa*, *suzukii* and *takahashii* (Hsu, 1949; Okada, 1954; Bock and Wheeler, 1972; Tsacas, 1979, 1980; Bock, 1980; Okada, 1984; Toda, 1991) (Figure 1).

Relationships among the subgroups have been hypothesized using morphological, chromosome and 28S rRNA sequence data (Hsu, 1949; Okada, 1954; Bock and Wheeler, 1972; Ashburner *et al.*, 1984; Pélandakis *et al.*, 1991; Pélandakis and Solignac, 1993) (Figure 2). Proposed relationships in these studies are largely in conflict; although, areas of general agreement can be pointed out. The two earlier studies of Hsu (1949) and Okada (1954) agreed that the basal lineage is the *suzukii* subgroup and that the more derived subgroups can be divided into two lineages (1) *melanogaster*, *takahashii* and *ficusphila* and (2) *montium*, *ananassae* and *nipponica*. At the time of Bock and Wheeler's (1972) monograph the known species of the *melanogaster* group increased more than fourfold; thereby giving current researchers better understanding of the diversity of the group. A consensus of the three most recent hypotheses (Bock and Wheeler, 1972; Ashburner *et al.*, 1984: Pélandakis *et al.*, 1991; Pélandakis and Solignac, 1993) results in three separate divisions: the *ananassae*; the *montium*; and all other subgroups – *elegans*, *eugracilis*, *ficusphila*, *melanogaster*, *suzukii* and *takahashii*. Within the last grouping the *elegans*, *ficusphila*, *suzukii* and *takahashii* subgroups may have a closer affinity (Bock and Wheeler, 1972). The remaining subgroups – *denticulata*, *flavohirita*, *longissima* and *rhopaloa* – are unplaced.

More than half the species of the *melanogaster* group are in the *montium* subgroup. The first species, *D. montium* de Meijere, was described in 1916 and numerous new species were described primarily since the 1950's, amounting to the current 87 species in the

montium subgroup (Figure 24). Representatives of the *montium* subgroup are distributed in all biogeographic regions except Antarctica, although the sole Neotropical representative is the widespread, introduced species, *D. kikkawai*. The highest concentration of species is in the Oriental region, especially southeast Asia (Lemeunier *et al*, 1986). Therefore, this region has been considered the "center of origin" for the *montium* subgroup (Lemeunier *et al.*, 1986). The Oriental Region has also been proposed as the "center of origin" for the *melanogaster* group as well as the family Drosophilidae (Bock and Wheeler, 1972; Throckmorton, 1975; Bock, 1980; Ashburner *et al.*, 1984; Lemeunier *et al.*, 1986; Toda, 1991). The "center of origin" concept has been replaced with vicariance and cladistic biogeography (Nelson and Platnick, 1981) which relies on the phylogenetic relationships of organisms and geographic patterns rather than patterns of species diversity.

The sister group to the *montium* subgroup is unknown, despite all the research on classification of *melanogaster* subgroups. Pélandakis *et al.* (1991) found that the branching order for the *montium* subgroup with respect to the other two lineages (i.e., *ananassae* and remaining subgroups) changed depending on the clustering method and/or the outgroups chosen. Pélandakis *et al.* (1991) attributed this to the short node between these three lineages, hypothesizing that the groups must have arisen almost simultaneously. At the next higher taxonomic level morphological and biochemical data have established the sister to the *melanogaster* group as the *obscura* group (Throckmorton, 1975; Powell and DeSalle, 1995).

Approximately half of the species within the *montium* subgroup have been placed into seven species complexes – *auraria*, *bakoue*, *bocqueti*, *jambulina*, *kikkawai*, *nikananu* and *serrata* (Table 1).

The *auraria* complex

This well studied complex is endemic to southeast Asia. Bock and Wheeler (1972) created a complex of four species closely related to *D. auraria*. These species were initially

placed together based on extreme morphological similarity. This affinity has also been demonstrated using data from starch gel electrophoresis, 2-dimensional gel electrophoresis, RFLP of mtDNA, hybridizations and mating asymmetry tests (Ohnishi, Kim and Watanabei, 1983; Ohnishi, Kawanishi and Watanabei, 1983; Ohnishi and Watanabe, 1984; Kim *et al.*, 1989; Kim *et al.*, 1993). Kimura (1987) included *D. auraria*, *biauraria*, *quadraria*, *subauraria* and *D. triauraria* in the complex and *D. rufa*, *D. asahinai* and *D. lacteicornis* as close relatives. Lemeunier *et al.* (1986) thought further study might add *D. trapezifrons* to the complex. The latest definition for the *auraria* complex includes eight species: *D., auraria*, *biauraria*, *lacteicornis*, *quadraria*, *rufa*, *subauraria*, *triauraria*, and *yuwanensis* (Kim *et al.*, 1993). Various investigations of the *auraria* complex concluded that the species are very closely related (Drosopoulou and Scouras, 1995 and references therein). Studies, however, disagree about the relationships among the species of the *auraria* complex, perhaps due to lack of sufficient resolution. Due to a low number of characters, some studies question the validity of the species such as Kimura (1987) who treated *D. quadraria* is a geographic race of *D. triauraria*. Kimura (1987) proposed that members of the *auraria* complex originated in forest habitats, evolved to live in fields, and later became human commensals. This scenario agrees with Bock's (1980) conclusion that most derived ecological condition for the *melanogaster* species group was human commensalism.

The *bakoue* complex

The *bakoue* complex contains ten Afrotropical species: *D. bakoue*, *curta*, *greeni*, *malagassaya*, *seguyi*, *seguyiana*, *tsacasi*, *vulcana*, and two undescribed species, spp. *C* and *D* (Chassagnard *et al.*, 1997). This complex was defined by Rafael (1984) based on species so similar that she performed hybridization tests to demonstrate that they were indeed distinct species. In their recent review of the *bakoue* complex, Chassagnard *et al.* (1997) defined the *seguyi* cluster to contain the species: *D. seguyi*, *seguyiana*, *curta* and a

sp. *D*. The position of sp. *D* is unclear, but the relationship of the three others is (*seguyi*, *seguyiana*), *curta* (Chassagnard *et al.*, 1997).

The *bocqueti* complex

This complex was first established by Tsacas in 1979 and later updated in 1984 to contain three species. Two are from the African mainland (*D. bocqueti* and *D. burlai*) and one species from Grand Comore Island (*D. chauvacae*). These species possess the unique morphological character of fusion of the anterior parameres with expansion of the novasternum (Tsacas, 1984). Because this morphological character seems derived almost to the point of loss, Tsacas (1979; 1984) hypothesized that this complex is the most derived within the *montium* subgroup, although discerning this actually depends on recognition of various synapomorphies, not just one or a few autapomorphies.

The *jambulina* complex

Data from metaphase chromosomes, starch gel electrophoresis, 2-dimensional gel electrophoresis, RFLP of mtDNA, hybridization (asymmetric mating/crosses), and morphology have established the *jambulina* complex (Watanabe *et al.*, 1982; Ohnishi and Watanabe, 1984; Kim *et al.*, 1989; Gupta and Gupta, 1992; Kim *et al.*, 1993). This Asian complex contains the species: *D. jambulina*, *punjabiensis*, *watanabei* (previously referred to as *punjabiensis*-like) and *barbarae* (Ohnishi and Watanabe, 1984; Kim *et al.*, 1989; Gupta and Gupta, 1992; Kim *et al.*, 1993). *Drosophila barbarae* was initially placed in the *kikkawai* complex by Tsacas and David (1978) based on morphology. The dendrogram constructed by Ohnishi and Watanabe (1984) placed *D. barbarae* in the *jambulina* complex rather than the *kikkawai* complex. Starch gel electrophoresis, 2-dimensional gel electrophoresis, hybridization (asymmetric mating/crosses), and RFLP of mtDNA data has confirmed *D. barbarae*'s placement within the *jambulina* complex; however, the hypothesized relationship among the four species of the group has varied (Ohnishi and Watanabe, 1984; Kim *et al.*, 1989; Kim *et al.*, 1993).

The *kikkawai* complex

This is the largest complex with representatives in the Afrotropical and Oriental regions, plus the widespread species, *D. kikkawai*. Bock and Wheeler (1972) noted an affinity between *D. kikkawai* and many species, and in 1978 Tsacas and David clearly defined the complex: long, longitudinal sex combs on the 1st and 2nd tarsal segments, two teeth on the cercal clasper, and slender posterior parameres (Tsacas and David, 1978). Initially the *kikkawai* complex contained 9 species – *D. barbarae*, *brevina*, *diplacantha*, *kikkawai*, *leontia*, *lini*, *pennae*, and tentatively *mysorensis* and *anomelani* (Tsacas and David, 1978). Lemeunier *et al.* (1986) expanded the definition of this complex to include *D. bocki* and possibly *D. sampagensis* and *D. cauverii*. Starch gel electrophoresis, 2-dimensional gel electrophoresis, RFLP of mtDNA, and hybridization (asymmetric mating) studies by Ohnishi and Watanabe (1984), Kim *et al.* (1989), and Kim *et al.* (1993) included seven species in the *kikkawai* complex - *D. barbarae*, *bocki*, *kikkawai*, *leontia*, *lini*, *pennae*, plus another, *D. lini*-like species. These three studies found that the *kikkawai* complex species that were tested shared a common node on their distance tree, except *D. barbarae* which grouped with the *jambulina* complex. Thus, these papers do not consider *D. barbarae* a member of the *kikkawai* complex. Hypotheses in the studies above disagreed about relationships among the species of the *kikkawai* complex. Tsacas and David (1978) did not believe that *D. montium*, which had been mistaken with *D. kikkawai*, should be included in the *kikkawai* complex. A relatively new species, *D. cryptica* De and Gupta which has a very close affinity to *D. diplacantha*, may need to be included in this complex.

The *nikananu* complex

When Tsacas (1981) described the species *D. xanthia* and *D. phylae* he noted an affinity with *D. nikananu* and in 1984 defined the *nikananu* complex based on the short male sex combs. Tsacas and Chassagnard (1992) formally defined the *nikananu* complex based primarily on the uniquely short male sex combs (tarsal segment 1 and 0-3 teeth on the 2nd tarsal segment) as well as a protuberance on the surstylus. This complex contains

the endemic African species *D. dictena, nikananu, phylae, xanthia, doussoui*, and two undescribed species, sp. aff. *nikananu* and sp. E aff. with *phylae* (Tsacas and Chassagnard, 1992; Chassagnard, 1991; Bock and Wheeler, 1972; Chassagnard *et al.*, 1997). The endemic Indian species, *D. gundensis* Prakash and Reddy has a similar sex comb but needs further examination of other characters (Tsacas and Chassagnard, 1992). Due to the similarity of the sex combs to those of the *melanogaster* subgroup, the *nikananu* complex has been proposed as most basal within the *montium* subgroup. "linking" the *montium* with the *melanogaster* subgroup (Tsacas, 1979; Tsacas, 1984; Tsacas and Chassagnard, 1992).

The *serrata* complex

The *serrata* complex contains the Australasian species of *D. serrata, D. birchii* and *D. dominicana*. Initially these species were considered populational variants of *D. serrata*. Dobzhansky and Mather (1961) subdivided southern and northern populations into the subspecies *D. serrata serrata* and *D. serrata birchii*. Using morphological characters from the male external genitalia and hybridization data, Ayala (1965a; 1965b) elevated the subspecies *D. serrata birchii* to a species and established the species *D. dominicana* from a population on Madang, New Guinea. During the hybridization experiments, Ayala (1965b) noted that populations *D. serrata* and *D. birchii* were quite variable in their extent of sexual isolation (also that within and across populations of *D. serrata* female abdominal pigmentation varied). Using hybridization, metaphase and salivary gland chromosome maps for populations of *D. birchii*, Baimai (1970a; 1970b) confirmed that *D. birchii* has variable and distinct populations. Further investigation might yield cryptic species.

Laboratories in Japan, Greece and France have been researching relationships within the *montium* subgroup as a whole. The Japanese have primarily focused on the relationships within and among the *jambulina, kikkawai* and *auraria* complexes. The Greek investigations have sampled *montium* species across a broad geographic range.

Their studies have included up to eleven species and sometimes multiple populations. These species are *D. auraria*, *bicornuta*, *birchii*, *diplacantha*, *jambulina*, *kikkawai*, *quadraria*, *seguyi*, *serrata*, *triauraria* and *vulcana* (stock 3116.11 UT and 3120.5 UT actually is *D. watanabei* Gupta and Gupta, 1992 instead of *D. jambulina* as indicated in their study) (Drosopoulou and Scouras, 1995). The French research has focused on the drosophilids in Africa, and established the *bakoue*, *bocqueti*, *kikkawai*, and *nikananu* species complexes.

In 1984, Ohnishi and Watanabe conducted the largest investigation of the *montium* subgroup by sampling 29 species. As a result, the *jambulina* complex was established, which now included *D. barbarae*, and the *jambulina* complex was found to be closely related to the *kikkawai* and the *auraria* complexes. These complexes are related to each other in the following arrangement: (*jambulina*, *kikkawai*) *auraria* (Ohnishi and Watanabe, 1984; Kim *et al.*, 1993). Relationships of the species within the complexes have varied across these and other studies (cf., above paragraphs on the individual complexes).

Tsacas (1979) felt the *montium* subgroup was the most basal in the *melanogaster* group. Investigations in the Scouras laboratory have indicated that the *montium* subgroup is relatively more basal than the *melanogaster* subgroup. However, these studies view the *montium* subgroup as "transitory" or "linking" the *obscura* group to the *melanogaster* subgroup. This is based on the presence of Balbiani rings (Mavragani-Tsipidou *et al.*, 1994; Drosopoulou and Scouras, 1995; Scouras, 1995). Genomic locations for ß-tubulin genes are dispersed in *D. melanogaster* but are varied from clustered to dispersed positions within species of the *montium* subgroup (Mavragani-Tsipidou *et al.*, 1994; Drosopoulou and Scouras, 1995; Scouras, 1995). Also, there is greater sequence similarity within the *montium* subgroup for the *hsp*70 and *hsp*68 (heat shock protein) genes as compared to the *melanogaster* complex (Drosopoulou *et al.*, 1996; Scouras, 1995). The *hsp*70 gene is found in single copy in the *repleta* group, *obscura* group and *montium* subgroup; whereas, it is a duplicate copy in the *melanogaster* subgroup (Drosopoulou *et al.*, 1996).

The *montium* subgroup was hypothesized to have radiated from a center of origin in southeast Asia southwesterly to its present Australasian, Indian and African locations (see reviews by Lemeunier *et al.*, 1986; Scouras, 1995). The species radiation pattern, proposed through a "center of origin" concept, is also confirmed by characters in RFLP of mtDNA, differentiation and then elimination of Balbiani rings (BR1 and BR2), electrophoretic patterns of "small" *hsp* genes, and clustered and then dispersed chromosomal locations for ß-tubulin genes (Pissios and Scouras, 1993; Nikolaidis and Scouras, 1996; Mavragani-Tsipidou *et al.*, 1994; Konstantopoulou *et al.*, 1997; Drosopoulou and Scouras, 1995). The African region may be a place of secondary *montium* species differentiation since various studies have found that either *D. vulcana* or *D. seguyi* emerge as separate, unique lineages (Pardali *et al.*, 1996). Tsacas and Lachaise (1974) and Tsacas (1979; 1984), however, believe that the *montium* subgroup has had multiple separate introductions into Africa because complexes with representatives in Africa do not form a separate clade. Rather, these complexes have been hypothesized to be the most primitive, providing a close link to the *melanogaster* subgroup (i.e., *nikananu* complex), the most derived (i.e., *bocqueti* complex), and an ancient link between the Oriental and African regions (i.e., *kikkawai* complex) (Tsacas 1979; Tsacas and Lachaise, 1974; Tsacas 1984).

Tsacas and David (1978) concluded that due to the large size and broad geographic range, the *montium* subgroup must be polyphyletic, although these are not criteria applied in phylogenetic studies since monophyletic groups are defined on the basis of synapomorphies and such groups can be both speciose and widespread. Other researchers, however, (Kim *et al.*, 1993; Drosopoulou and Scouras, 1995; Nikoladis and Scouras, 1996) indicated that the species of the *montium* subgroup are very closely related (Nikolaidis and Scouras, 1996). Investigations from Scouras' laboratory have found putative synapomorphies for monophyly of the subgroup such as Balbiani rings, reverse tandem chromosome duplications, darkening anterior spiracles in late third instar larvae,

electrophoretic patterns of "small" *hsps*, mt DNA size, and features of metaphase chromosomes (Scouras, 1995).

Phylogeny of the *montium* subgroup at present is abscent despite recognition of species complexes and numerous investigations. A revision of the *montium* subgroup, as well as its relation to the other subgroups of the *melanogaster* group, is needed to understand relationships of species and complexes, and biogeography (Rafael, 1984; Tsacas and David, 1978; Tsacas, 1979; Lemeunier *et al.*, 1986). Resolution for such a speciose group requires numerous characters as is readily provided with biochemical data; however, such studies are limited by stock availability (Lemeunier *et al.*, 1986).

Literature on the *montium* subgroup consists primarily of descriptions of new species or local faunal analyses (e.g., De and Gupta, 1996; Singh and Gupta, 1979; Muniyappa *et al.*, 1981; Shyamala and Ranganath, 1990; Singh and Dash, 1998; Chassagnard *et al.*, 1997). Tsacas (1979) and Grimaldi (1990; 1991) have noted that only a limited understanding of relationships can be obtained through faunal analyses. Morphological, hybridization (mating/crosses) and biochemical studies (e.g., metaphase chromosomes, *in situ* hybridizations and electrophoresis) have used a descriptive or evolutionary taxonomy approach to determine relationships within the *montium* subgroup (Tsacas and Chassagnard, 1992; Chassagnard *et al.*, 1997; Rafael, 1984; Tsacas, 1975; Tsacas and David, 1978; Tsacas, 1984; Tsacas and Lachaise, 1974; Shyamala and Ranganath, 1989; Baimai, 1980; Drosopoulou and Scouras, 1995; Konstantopoulou *et al.*, 1997; Pardali *et al.*, 1996; Drosopoulou *et al.*, 1996; Mavragani-Tsipidou *et al.*, 1994; Kimura, 1987). Relationships have also been determined using allozyme and RFLP of mtDNA data (Kalanti-Makri *et al.*, 1985; Triantaphyllidis *et al.*, 1978; Ohnishi and Watanabe, 1984; Ohnishi, Kim and Watanabei, 1983; Ohnishi, Kawanishi and Watanabei, 1983; Tsakas and Tsacas, 1984; Shyamala and Ranganath, 1990; Watanabe *et al.*, 1982; Kim *et al.*, 1993). These studies, however, produced distance data, which was analyzed

phenetically using UPGMA (Sneath and Sokal, 1973) and Neighbor-Joining (Saitou and Nei, 1987).

Biochemical and other studies employing phenetic distance have problems accessing homology for the reconstruction of phylogeny (Avise, 1994; Murphy *et al.*, 1996; Swofford *et al.*, 1996). Allozymes can have hidden heterogeneity; therefore, even after tree construction, the homology remains questionable (Murphy *et al.*, 1996). In the distance analyses above, the RFLPs of mtDNA are coded as presence or absence of restriction fragment lengths. Avise (1994) and Swofford *et al.* (1996) noted that this type of coding violates the assumption of independence of characters. Phenetic analyses compare general, overall similarity in order to reconstruct phylogenies (Farris, 1979). Highly autapomorphic taxa, for example, are artificially inflated in rank, contrary to phylogenetic position conversely, taxa can also share a plesiomorphic resemblance. Therefore, phenetics may not yield monophyletic groups (Farris, 1979). UPGMA and neighbor-joining always produce a single, resolved tree, but the method of tree construction and criterion for choosing such a tree are conflated (Swofford *et al.*, 1996). By comparing the percent difference of overall similarity, UPGMA and neighbor joining lose the individual character information (Avise, 1994; Swofford *et al.*, 1996). Therefore, one is unable to infer the evolution of characters and to calculate the character support at nodes.

Some studies tried to expand their approach to analyses of the data. The hybridization study of Kim *et al.* (1989) applied an asymmetrical mating preference criterion, according to a model by Watanabe and Kawanishi (1979; 1981). Pissios and Scouras (1993) and Nikolaidis and Scouras (1996) coded their mtDNA RFLP data as both distance and discrete characters, generating trees using both phenetic and parsimony methods. Each study, however, ignored the parsimony results and discussed the tree produced through phenetic analyses. Chassagnard *et al.* (1997) approximated a cladistic approach when describing the relationship among three species.

The present investigation tested the monophyly and relationships within the *montium* subgroup, and it contains one of the largest samples of species studied in the group to date. This is also the first time a cladistic analysis of DNA sequences has been used to resolve relationships within the *montium* subgroup.

METHODS

The characters were aligned and gap coded as in chapter 2. The phylogeny reconstruction and node support are based on the analyses in chapter 2, except that this chapter focuses on the the *montium* subgroup. Characters were optimized at the nodes using ACCTRAN which is logically superior to DELTRAN because ACCTRAN maintains initial putative homology statements until tree construction demonstrates the statements to be false (DePinna, 1991). Node support was evaluated and compared among the actual characters and the commonly used inference statistics of Bremer support, and partitioned Bremer support (PBS) (Schawaroch, in prep). The *montium* subgroup portion of the tree was examined for amino acid character evolution at nodes for each of the individual gene regions (i.e., *Adh*, *co* ii and *hb*).

RESULTS

A list of synapomorphies for the *montium* subgroup nodes is in Table 16. Bremer, partitioned Bremer and bootstrap node support values for the *montium* subgroup are compared in Table 17. The number of actual characters at each node is presented for each individual gene region facilitating comparisons with PBS values (Table 18). Amino acid changes at each node are presented in Table 16. A single, most parsimonious cladogram resulted from the simultaneous analysis of the three gene regions (see Chapter 2). The phylogeny is presented in two ways: (1) in total showing the location of the *montium* subgroup within the *melanogaster* group (Figure 17) and (2) a close up of the *montium* subgroup portion compared to the traditional taxonomic affiliations (Figure 25).

DISCUSSION

Comparing Inference Statistics to Actual Characters at Nodes

Discrepancies among inference statistics (bootstrap, Bremer, partitioned Bremer support) as well as in comparison to nodes from separate gene region analyses has been discussed in Chapter 2. The applicability of bootstrap analysis to evaluating topology of cladograms has been the subject of many papers (e.g., Kluge and Wolf, 1993; Sanderson, 1995) and will not be discussed here. Instead, this section compares node support as defined by the characters and the PBS values. Most nodes are supported by characters from all three gene regions (Table 18). Yet, some nodes are exclusively supported by characters from a single gene region (*co ii* = 23, 25, 31, 38 and *hb* = 34). The PBS values (Table 17) are in general agreement: (a) for nodes 23, 25, 31, and 38 because *co ii* provides greater than zero PBS values while *Adh* and *hb* PBS values are zero or negative and (b) for node 34 where *hb* provides the only positive PBS value. Some nodes have character support from two gene regions (*Adh* and *hb* = 27, 33, 39; *co ii* and *hb* = 36, 37). The PBS values agree for node 27 where both *Adh* and *hb* values are greater than zero (*co ii* = 0) and nodes 36 and 37 where values are positive from *hb* and *co ii* and negative from *Adh*. PBS values, however, do not reflect character support at node 33 which has a positive value from *hb* and negative values each from *Adh* and *co ii* and at node 39 which has a positive value from *Adh* and zero values from each *hb* and *co ii*. Nodes 24, 26 28, 29, 30, 34, 35, 41, 43, and 44 have character support from all three gene regions but PBS values are negative or zero for one or more gene regions.

In total, *co ii* provides the greatest amount, while *Adh* and *hb* provide equal amounts of character support (synapomorphies) summed for all nodes of the *montium* subgroup. In contrast, gene region contribution for total *montium* subgroup as estimated by summing PBS values ranks support from greatest to least *hb* > *co ii* > *Adh*. The same pattern appears when the total number of PBS positive and zero node values are summed per gene region.

A more extensive discussion about the discrepancy between characters at a node and the inference statistics (Bremer and partitioned Bremer support) values for that particular node is in Schawaroch (in prep). Since this discrepancy exists I feel character (i.e., synapomorphy) data for nodes should be reported (Table 16).

Amino Acid Changes

Within the list of synapomorphies (Table 16) the nucleotide changes have been converted into their subsequent amino acids. Many of the amino acid changes were silent substitutions (165 silent substitutions out of 192 knowable amino acid changes at *montium* subgroup nodes [one change unknowable because partially in the primer]). The mitochondrial gene region had less non- silent changes than the nuclear gene regions (*co* ii = 7% versus *Adh* = 19% and *hb* = 19%), perhaps reflecting greater conservation in the mitochondrion. These silent substitutions would not have been detected as change and a cladistic analysis on the amino acid level may have yielded a different tree topology. Besides amino acids having less sensitivity (actual change is undetected), they also provide less characters because some codons had more than one position change (e.g., nodes 29, 40 and 45). Sampling among the amino acids at the nodes, there seems to be codon bias; however, this has not been explored at this time. To determine if there is any codon bias for clades and within gene regions, all the nucleotide sequence will have to be translated, codons scored for each position, and compared. This is beyond the scope of the present chapter but would be useful for future study.

PHYLOGENETIC DISCUSSION

Testing Monophyly of the *montium* Subgroup

Simultaneous analysis of the three gene regions (*Adh*, *hb* and *co ii*) yielded a single, well resolved most parsimonious cladogram (L=1540 steps, CI =0.349) (Figure 17). In this cladogram the *montium* subgroup, largest of all the subgroups and containing more than half the species within the *melanogaster* group, is monophyletic. This result contradicts Tsacas and David (1978) who hypothesized a polyphyletic *montium* subgroup based on

evolutionary taxonomic and faunal treatments. The criteria of size and distribution are not applicable in phylogenetic studies which define monophyletic (natural) groups on the basis of synapomorphies. A monophyletic *montium* subgroup is, however, in agreement with numerous morphological and biochemical studies (e.g, Bock and Wheeler, 1972; Ashburner *et al.*, 1984; Scouras, 1995). Morphological characters that corroborate *montium* subgroup monophyly include a distinctive large longitudinal sex comb along the entire length of the first and second tarsal segments (exception: members of the *nikananu* complex, as well as *D. exiguitata* and *D. paraviaristata*. With the establishment of the *longissima* subgroup, there is now another subgroup possessing this distinguishing character [Toda, 1991]). Other diagnostic characters include: male external genitalia with surstylus and cercal clasper, the cercal clasper usually with very large bristles, and clearly defined apical bands on the abdomen in both sexes. Further investigation needs to be done as to the placement of *D. exiguitata* and *D. paraviaristata*. The male genitalia of *D. exiguitata* is somewhat unique; however, Bock (1980) felt it should be retained within the *montium* subgroup. Whereas, external genitalia of *D. paraviaristata* lacks a cercal clasper and was only provisionally retained in the *montium* subgroup by Bock (1980). Other putative synapomorphies indicating monophyly of the *montium* subgroup are Balbiani rings, reverse tandem chromosome duplications, darkening anterior spiracles in late third instar larvae, electrophoretic patterns of "small" *hsps*, mt DNA size, and features of metaphase chromosomes (Scouras, 1995).

Many studies state that the *montium* subgroup species are closely related (Kim *et al.*, 1993; Drosopoulou and Scouras, 1995; Nikolaids and Scouras, 1996). My analyses do not infer the degree of relatedness, yet the node for the *montium* subgroup clade is very well supported by numerous characters from all three gene regions (26) and the large Bremer and bootstrap values (14 and 99% respectively).

Approximately half of the total 87 species within the *montium* subgroup have been given complex affiliations. Monophyly of the complexes was tested with the exceptions of

the *bocqueti* complex (no representatives sampled) and the *nikananu* complex (represented by a single species).

Testing Monophyly of the Traditional Species Complexes

The *auraria* complex

The species are valid since each taxon posesses its own set of autapomorphies. The *auraria* complex forms a monophyletic clade in a relatively basal position within the *montium* subgroup (Figure 25). The *auraria* complex node is well supported by a total of 10 characters from all three gene regions, Bremer value of 6 and bootstrap value of 95% (Table 17 and 18). The hypothesis of relationships for the five representative taxa sampled are: *rufa* (*biauraria* (*auraria* (*quadraria*, *triauraria*))). Previous studies (e.g., Bock and Wheeler, 1972; Kim *et al.*, 1992; Kimura, 1987) have noted that four species are closely related: *D. auraria*, *biauraria*, *triauraria* and *quadraria*. The degree of relatedness is indeterminable with my investigation but the node for the clade relating the four taxa is well supported by a total of 22 characters from all three genes regions, a Bremer value of 22 and a bootstrap value of 100% (Table 17 and 18). For the internal relationships, node 39 relies on 3 characters from the *Adh* and *hb* gene regions and node 38 is supported by two characters from the *co ii* gene region, thus demonstrating that the choice of data directly effects hypotheses (Table 18). Mapping species habitat of Kimura (1987) on my phylogeny maintains forest habitat as basal and places fields and human commensalism as derived.

The *bakoue* complex

Four species represented the *bakoue* complex (i.e., *D. greeni*, *seguyi*, *tsacasi* and *vulcana*). Despite the extreme morphological similarity which required Rafael (1984) to use mating tests to confirm species, the *bakoue* complex forms a paraphyletic clade due to the inclusion of *D. diplacantha* (*kikkawaii* complex) and *D. nikananu* (*nikananu* complex) (Figure 25). The support for this paraphyletic clade (BS =2, B = 34%, and characters =2) as well as relationships within have low Bremer and bootstrap values (Table 17 and 18).

Node 23 (ie., *greeni* and *seguyi*) and node 25 [ie., (*vulcana*, *nikananu*) *diplacantha*)] have characters only from *co ii* gene region. Node 27 (basal node creating the clade) has characters only from *Adh* and *hb* gene regions. Two *bakoue* complex representatives *D. greeni* and *D. seguyi* are sister taxa with relatively higher Bremer and bootstrap values (BS = 3, B = 58%). Interestingly this paraphyletic *bakoue* complex clade contains all the African representatives used in my taxon sample (see below).

The *jambulina* complex

Drosophila punjabiensis and *D. watanabei* are sister taxa (BS=14, B=100%, and 18 characters from all 3 gene regions) (Table 17 and 18; Figure 25). This relationship is not surprising since *D. watanabei* was only recently described from stocks initially thought to be *D. punjabiensis* (Wantanabe *et al.*, 1982: Gupta and Gupta, 1992). *Drosophila barbarae*, included in the *jambulina* complex by biochemical studies (e.g., Ohnishi and Watanabe, 1984), does not form a clade with the other *jambulina* complex representatives. *Drosophila punjabiensis* and *D. watanabei* are part of a larger clade that includes *D. serrata* (the *serrata* complex), however, this node is not well supported (BS = 2, B = 33%, and 8 characters from *Adh* and *hb*) and more characters may easily change this hypothesis.

The *kikkawai* complex

In my cladogram the *kikkawai* complex is polyphyletic (Figure 25). *Drosophila kikkawai* and *D. lini* are sister taxa supported by 12 characters from all three gene regions and Bremer and bootstrap values of 8 and 93%, respectively (Table 17 and 18). *Drosophila kikkawai* and *D. lini* are part of a larger clade that includes *D. orosa*, but this clade is not as well supported (BS = 2, B = 27, and 4 characters from *co ii* and *hb*). In my cladogram, *D. diplacantha* is not with the other *kikkawai* complex representatives, but rather nested within a clade containing other African species. *Drosophila barbarae*, included in the *kikkawai* complex by Tsacas and David (1978) based on morphology, instead forms a clade with *D. birchii* and *D. mayri*. These disparate positions for the *kikkawai* complex representatives are well supported in the instance of *D. barbarae*. Even though, Tsacas and David (1978)

had well defined characters for the *kikkawai* complex, their study and my cladogram do not agree. This discrepancy may be attributed to (1) less taxa known at the time of Tsacas and David (1978), (2) morphological characters being pleismorphic, or (3) my taxon sampling.

The *serrata* complex

Two of the three taxa in this complex were sampled (i.e., *D. serrata* and *D. birchii*). Initially the species of *D. serrata* and *D. birchii* were considered populational variants, then subspecies and finally distinct species (Dobzhansky and Mather, 1961; Ayala, 1965a; 1965b). This taxonomic history alone indicates that these two species share many characters. The biogeographic range for this complex has been relatively thoroughly sampled and the three representatives are all Australasian (see Ayala, 1965 a & b and Lemeunier *et al.*, 1986). In my cladogram, however, the *serrata* complex is polyphyletic (Figure 25). *Drosophila serrata's* basal position in the clade containing *D. punjabiensis* and *D. watanabei* is not very well supported (BS = 2, B = 33%, and 8 characters from *Adh* and *hb*), and may easily be subject to change with more character information (Table 17 and 18). Yet *D. birchii* does seem nested in a well-supported clade (see discussion below). Previous studies such as, 2 dimensional gel protein electrophoresis (Ohnishi and Watanabe, 1984), RFLP of mt DNA (Pissios and Scouras, 1993; Nikolaidis and Scouras, 1996), and chromsomal location of ß-tubulin genes (Drosopoulou and Scouras, 1995) also question the sister relationship of *D. serrata* and *D. birchii*. This complex needs further study.

Potentially New Species Complexes

The species affiliations (mentioned below) are well supported. *At this point, these complexes are tentative because a revision needs to be done for complete species sampling.* In relation to previous studies my taxon sample is large, but it is only approximately 28% of the known species. Therefore, not all clades in my phylogeny are given a (new) complex designation.

The "*parvula* complex" (lineage F)

Drosophila parvula and *D. kanapiae* form their own separate clade in a basal position within the *montium* subgroup (Figure 25). The node supporting the sister relationship of these two species has high Bremer value (7), bootstrap value (97%) and numerous characters (15 characters from all 3 gene regions) (Table 17 and 18). Two of these characters, 483 and 765, each have a c.i. = 1 indicating unique support for the clade without exhibiting any homoplasy (Table 16). Bock and Wheeler (1972) described these two species and noted that *D. parvula* and *D. kanapiae* possessed similar posterior parameters. These posterior parameters are morphologically distinct for species within the *montium* subgroup. Further investigation (e.g., a revision) may confirm the existence of this complex.

The "*birchii* complex" (lineage B)

This complex contains four species with the following relationship: *bicornuta* (*birchii* (*barbarae*, *mayri*)) (Figure 25). The most basal node (30) forming this complex has Bremer and bootstrap values of 4 and 68% respectively (Table 17). The next inner node (29) has much higher values (BS = 15 and B= 100%) supporting the relationship of three taxa. Characters from all 3 gene regions support each of the nodes (node 30 = 10 characters, node 29 = 21 characters, node 28 = 6 characters) (Table 18). Each of these nodes is supported by some characters which do not exhibit homoplasy (characters 534 and 1036 for node 30; characters 36 and 232 for node 29; characters 761 for node 28) (Table 16). *Drosophila bicornuta* and *D. mayri* have no previous complex affiliations. *Drosophila barbarae's* complex designation has varied from the *kikkawai* complex based on morphology (Tsacas and David, 1978) to the *jambulina* complex based on biochemical and hybridization studies (e.g., Ohnishi and Watanabe, 1984). *Drosphila birchii* ,based on morphology and hybridization data, has been a member of the *serrata* complex (Ayala, 1965a; 1965b). Perhaps if future studies corroborate the *serrata* complex it will be sister group to the (*barbarae*, *mayri*) clade. Mitochondrial DNA RFLP studies (Pissios and

Scouras, 1993; Nikolaidis and Scouras, 1996) place *D. bicornuta* and *D. birchii* as sister taxa, whereas Drosopoulou and Scouras' (1995) ß-tubulin gene study places *D. bicornuta* closer to *D. serrata* than to *D. birchii*. Ohnishi and Watanabe (1984) included all four taxa in their protein electrophoresis study and only *D. bicornuta* and *D. mayri* formed a clade (actually sister taxa). This study is the first to propose a complex affiliation for these species which seems probable and should be investigated with future studies.

Relationships Among Complexes and Species

Although the relationships among species of the *montium* subgroup are fully resolved, many nodes relating clades have low Bremer support (≤2) and bootstrap values (<50%).

Tsacas (1979; 1984) and Tsacas and Chassagnard (1992) placed *D. nikananu* as most basal within the *montium* subgroup, a "link" to the *melanogaster* subgroup. My analyses place *D. baimaii* as most basal within the *montium* subgroup clade (Figure 25). Bock and Wheeler (1972) noted that *D. baimaii* is unique among *montium* subgroup species due to the lack of enlarged medial bristles on the cercal clasper, anterior parameters resembling ones in the *takahashii* and *suzukii* subgroups, and a blind process of the aedeagus similar to the *eugracilis* subgroup.

Ohnishi and Watanabe (1984) and Kim *et al.* (1993) hypothesized the relationship for 3 complexes as follows: *auraria* (*jambulina* , *kikkawai*). Generally, my study is in agreement by placing representatives of the *auraria* complex in a basal position to the *kikkawai* and *jambulina* complex representatives.

Examining a total of eleven *montium* subgroup representatives, Scouras' laboratory proposed evolutionary scenarios for chromosomal and gene structures in the absence of a cladogram. The interrelationships proposed had a general biogeographic component - progressing southwesterly from Southeast Asia to Australia, India and then Africa (see Scouras, 1995). My cladogram has representatives from the Asian, Australian and Indian

regions throughout; however, a clade containing all the representatives from the African region has a more derived position.

Biogeography

The area cladogram for species in my phylogeny is presented in Figure 26. The outgroup (the *melanogaster* species group) has both Africa and southeast Asia represented. The basal area for the *montium* subgroup is southeast Asia, which is also the most species diverse area traditionally hypothesized to be the "center of origin". There have been two colonizations of Australia, once in lineage C (*D. serrata*) and once in lineage B (*D. birchii*) (Figures 25 and 26). The separate colonizations of Australia result from *serrata* complex polyphyly in my phylogeny (for further discussion see above). In my study, all five representatives of African species form a single clade: *tsacasi* ((*greeni*, *seguyi*) (*diplacantha* (*vulcana*, *nikananu*))) despite different complex affiliations. This clade is located in a derived position within the *montium* subgroup. This agrees with studies from Scouras' laboratory that hypothesize Africa as a place of secondary species differentiation, especially with reference to *D. seguyi* and *D. vulcana* (Pardali *et al.*, 1996). But a single colonization of Africa disagrees with Tsacas and Lachaise (1974) and Tsacas (1979; 1984) who hypothesize the *montium* subgroup to have multiple separate introductions to Africa because three complexes with African representatives seem disconnected (i.e., *bocqueti* complex most derived, *nikananu* complex most basal and *kikkawai* complex "links" African and Oriental species). A general progression across areas from the basal to derived clades concurs with Scouras' laboratory which concluded, based on various studies of chromosome composition and structure, that the *montium* subgroup radiated from southeast Asia in a generally southwesterly direction to Australia, then India and finally to Africa (Scouras, 1995).

The biogeography hypotheses presented here are extremely tentative for numerous reasons. First, a revision is needed to place all species in a phylogentic context to have a better representation of area relationships. Second, distributions of many species are known

from single localities and species diversity may be an artifactual due to the standard fruit-bait collecting technique. Third, biogeographic hypotheses are plagued with questions regarding effects of possible extinction events, (e.g, previous widespread species now in small ecological refugia), and the ability of *Drosophila* to disperse, thus not easily isolated on land-masses.

The *montium* Subgroup's Position Within The *melanogaster* Group

Tsacas (1979) hypothesized the *montium* subgroup to be the most basal within the *melanogaster* group. Many studies from Scouras' laboratory hypothesized that within the *melanogaster* species group the *montium* subgroup was in a more basal position than the *melanogaster* subgroup (Scouras, 1995). My cladogram does not suppport either of the above hypotheses, but rather presents three major clades for the *melanogaster* species group (Figure 17). The first, most basal, clade contains the *melanogaster* and Oriental subgroups. This clade is sister to a clade that contains both the *ananassae* and *montium* subgroups (Figure 17).

IN CONCLUSION

As previously noted, much more extensive sampling is needed before any real understanding of relationships or further questions such as biogeography can be understood. There are a few more *montium* species available in cultures, however, the representative sampling of the subgroup would still be small. A revision is needed especially at this time with so many new species since Bock and Wheeler's (1972) monograph and such reliance of representatives of this subgroup for ecological and evolutionary studies (e.g., Kimura, 1987; Scouras, 1995). It is interesting how my results concur with previous morphology. Plus some studies whose methods of analysis were questionable (e.g., phenetics and scenarios) also had agreement with my cladogram.

CONCLUSION

The phylogeny generated from this dissertation research is the most thorough to date for the following reasons: (1) it is one of the largest taxon samples for the *melanogaster* species group, (2) it uses DNA sequence from three independent gene regions and (3) it employs a cladistic analysis of the data.

This research demonstrated that both molecules and morphology provide useful characters and corroborative evidence. During my research, morphological and molecular sequence both indicated the presence of mislabeled stocks from the National *Drosophila* Species Resource Center at Bowling Green. Morphological characters were found to correspond with clades resulting from the simultaneous analysis of the three gene regions. The molecular analyses indicated the presence of two complexes within the *takahashii* subgroup. This division is an original hypothesis of relationships for the *takahashii* even though it corresponds to a previously known sex comb difference. The *suzukii* subgroup representatives, *D. biarmipes* and *D. mimetica* each affiliated with different clades, but within each clade taxa possessed similar sex comb structure. My phylogeny resolves the three majors clades of the *melanogaster* species group by placing the *ananassae* + *montium* subgroups as sister taxa. This sister relationship corresponds to the putative morphological homology - presence of both surstyli and cercal claspers. Presence of hooked setae on the mid-tibia may be a synapomorphy for the *melanogaster* + Oriental subgroups lineage even though it was lost twice - once in the *melanogaster* subgroup and once in the *elegans* subgroup.

Future research

Future work could explore the possibility of codon bias corresponding with cladogensis. Hypotheses such as the evolution of gene families, chromosomal structural changes, and the evolution of transposable elements have been proposed in absence of a phylogeny. It would be interesting to test these hypotheses to the phylogenetic hypothesis of relationships derived from my study. My research indicated that the literature of the *melanogaster* species group was fraught with problems such as poor descriptions,

inadequate diagrams and inconsistent nomenclature that can only be solved by a morphological revision of the *melanogaster* species group.

TABLES

Table 1. List of the known species in the *montium* subgroup. Approximately half of these species have been placed in closer affiliations called species complexes. These species complexes have been established using a variety of data, including morphology, hybridization, and allozymes. Note that *D. barbarae* is placed in both the *kikkawai* and *jambulina* complexes. Based on morphological data from male sex combs, secondary claspers, and posterior parameters, Tsacas and David (1978) included *D. barbarae* in the *kikkawai* complex. More recent investigations, using allozymes and hybridization data, resulted in the placement of *D. barbarae* within the *jambulina* complex (Ohnishi and Watanabe, 1984; Kim *et al.*, 1989; Kim *et al.*, 1993).

complex	species
	agumbensis Prakash & Reddy, 1978
	artecarina Takada & Momma, 1975
	asahinai Okada, 1964
	austroheptica Tsaur & Lin, 1991
	baimaii Bock & Wheeler, 1972
	bhagamandalensis Muniyappa, Reddy & Krishnamurthy, 1981
	bicornuta Bock & Wheeler, 1972
	brahmagiriensis Muniyappa, Reddy & Krishnamurthy, 1981
	constricta Chen, Shao & Fan, 1988
	cryptica De & Gupta, 1996
	davidi Tsacas,1975
	eupyga Tsacas, 1981
	exiguitata Takada, Momma & Shima, 1973
	flavopleuralis Takada, Momma & Shima, 1973
	gangotrii Muniyappa & Reddy, 1981
	ifestia Tsacas, 1984
	kanapiae Bock & Wheeler, 1972
	khaoyana Bock & Wheeler, 1972
	kinabaluana Takada, Momma & Shima, 1973
	lacteicornis Okada, 1965
	longipectinata Takada, Momma & Shima, 1973
	madikerii Muniyappa & Reddy, 1981
	mayri Mather & Dobzhansky, 1962
	megapyga Tsacas, 1981
	montium de Meijere, 1916
	nagarholensis Prakash & Reddy, 1980
	neobaimai Singh & Dash, 1998
	neokhaoyana Singh & Dash, 1998
	neotrapezifrons Ranganath, Krishnamurthy & Hedge, 1983
	nigrialata Takada, Momma & Shima, 1973
	nigropleuralis Takada, Momma & Shima, 1973
	orosa Bock & Wheeler, 1972
	paraviaristata Takada, Momma & Shima, 1973
	parvula Bock & Wheeler, 1972
	pectinifera Wheeler & Takada, 1964
	penicillipennis Takada, Momma & Shima, 1973
	pseudobaimaii Takada, Momma & Shima, 1973
	pseudomayri Baimai, 1970
	rhombura Okada & Carson, 1983
	serrula Tsacas, 1984
	suborosa Kumar & Gupta, 1992
	tani Cheng & Okada, 1985
	tricombata Singh & Gupta, 1979
	truncata Okada, 1964
	yuwanensis Kim & Okada, 1988

complex	species
auraria	*auraria* Peng,1937
	biauraria Bock & Wheeler, 1972
	lacteicornis Okada, 1965
	quadraria Bock & Wheeler, 1972
	rufa Kikkawa & Peng, 1938
	subauraria Kimura, 1983
	trapezifrons Okada, 1966
	triauraria Bock & Wheeler, 1972
bakoue	*bakoue* Tsacas & Lachaise, 1974
	sp. *C* in Chassagnard, Tsacas & Lachaise,1997
	curta Chassagnard & Tsacas, 1997
	sp. *D* in Chassagnard, Tsacas & Lachaise,1997
	greeni Bock & Wheeler, 1972
	malagassya Tsacas & Rafael, 1982
	seguyi Smart, 1945
	seguyiana Chassagnard & Tsacas,1997
	tsacasi Bock & Wheeler, 1972
	vulcana Graber, 1957
bocqueti	*bocqueti* Tsacas & Lachaise, 1974
	burlai Tsacas & Lachaise, 1974
	chauvacae Tsacas, 1984
jambulina	*barbarae* Bock & Wheeler, 1972
	jambulina Parshad & Paika, 1964
	punjabiensis Parshad & Paika, 1964
	watanabei Gupta & Gupta, 1992
kikkawai	*anomelani* Reddy & Krishnamurthy, 1973
	barbarae Bock & Wheeler, 1972
	bocki Baimai, 1979
	brevina Wheeler, 1981
	cauverii Muniyappa & Prakash, 1982
	diplacantha Tsacas & David, 1978
	kikkawai Burla, 1954
	leontia Tsacas & David, 1978
	lini Bock & Wheeler, 1972
	mysorensis Reddy & Krishnamurthy, 1970
	pennae Bock & Wheeler, 1972
	sampagensis Muniyappa & Reddy, 1980
nikananu	*dictena* Tsacas & Chassagnard, 1992
	dossoui Chassagnard, 1991
	sp. *E* in Chassagnard, Tsacas & Lachaise,1997
	gundensis Prakash & Reddy, 1977
	nikananu Burla, 1954
	phyale Tsacas, 1981
	xanthia Tsacas, 1981
serrata	*birchii* Dobzhansky & Mather, 1961
	dominicana Ayala, 1965
	serrata Malloch, 1927

Table 2. Species with culture/stock numbers for all of the stocks used in this study. All species were obtained from the National *Drosophila* Species Resource Center at Bowling Green with the exception of *D. teissieri* Brazzaville isofemale line 16 which was a gift from D. LaChaise to G. Simmons. All species identifications were confirmed based on male gentalic dissections. *Drosophila ficusphila* was originally mislabled by the National *Drosophila* Species Resource Center at Bowling Green to be *D. pennae* 14028-0631.0. Representatives of all taxa sampled have been placed in the collections at the American Museum of Natural History.

Species	Culture/Stock
D. ambigua	14011-0091.0
D. bifasciata	14012-0181.0
D. pseudoobscura	14011-0121.0
D. persimilis	14011-0111.0
D. affinis	14012-0141.0
D. eugracilis	14026-0451.0
D. ficusphila	misID 14028-0631.0
D. lucipennis	14023-0331.0
D. mimetica	14023-0341.0
D. ananassae	14024-0371.0
D. varians	14024-0431.0
D. takahashii	14022-0311.5
D. paralutea	14022-0281.0
D. prostipennis	14022-0291.0
D. teissieri	gift D. LaChaise
D. yakuba	14021-0261.0
D. bicornuta	14028-0511.0
D. dipacantha	14028-0586.0
D. punjabiensis	14028-0641.0
D. seguyi	14028-0671.0
D. nikananu	14028-0601.0
D. kikkawai	14028-0561.3
D. serrata	14028-0681.0

Table 3. Traditional taxonomic grouping (species group and species subgroup) for species used in this study. The species subgroups for *obscura* are based on relationships proposed by Barrio *et al.*, 1992.

species group	species subgroup	species
obscura	*obscura*	*D. ambigua*
		D. bifasciata
	pseudoobscura	*D. pseudoobscura*
		D. persimilis
	affinis	*D. affinis*
melanogaster	*eugracilis*	*D. eugracilis*
	ficusphila	*D. ficusphila*
	suzukii	*D. lucipennis*
		D. mimetica
	melanogaster	*D. melanogaster*
		D. teissieri
		D. yakuba
	takahashii	*D. takahashii*
		D. paralutea
		D. prostipennis
	ananassae	*D. ananassae*
		D. varians
	montium	*D. bicornuta*
		D. dipacantha
		D. kikkawai
		D. nikananu
		D. punjabiensis
		D. seguyi
		D. serrata

Table 4. Source of gene regions employed in this study. *** are regions sequenced for this study. See text for details. Publ. are previously published sequences obtained from Genbank (accession numbers in parentheses) with the exception of the 28S D1 domain which was entered from the Pélandakis and Solignac (1993) reference. ambigua: Adh (X54813), co ii (M95145); bifasciata: co ii (M95147); pseudoobscura: Adh (M60989), co ii (M95150), 28S (X71203), 16S (M93993); persimilis: Adh (M60997), co ii (M95143); affinis: co ii (M95140), 28S (X71207); eugracilis: 28S (X71175); ficusphila: 28S (X71181); mimetica: 28S (X71179); melanogaster; Adh (M11290), co ii (AF200828), hb (Y00274), 28S D1 (X71158) and D2 (X71159), 16S (AF200828); teissieri; Adh (X54118), 28S (X71169); yakuba: Adh (X57376), co ii (X00924), 28S (X71167), 16S (X03240); takahashii: 28S (X71177); ananassae: 28S (X71197); varians: 28S (X71199); kikkawai: 28S (X71185); serrata: 28S (X71189)

species	*Adh*	*co ii*	*hb*	*28S*	*16S*
D. ambigua	Publ.	Publ.	***	***	***
D. bifasciata	***	Publ.	***	***	***
D. pseudoobscura	Publ.	Publ.	***	Publ.	Publ.
D. persimilis	Publ.	Publ.	***	***	***
D. affinis	***	Publ.	***	Publ.	***
D. eugracilis	***	***	***	Publ.	***
D. ficusphila	***	***	***	Publ.	***
D. lucipennis	***	***	***	***	***
D. mimetica	***	***	***	Publ.	***
D. melanogaster	Publ.	Publ.	Publ.	Publ.	Publ.
D. teissieri	Publ.	***	***	Publ.	***
D. yakuba	Publ.	Publ.	***	Publ.	Publ.
D. takahashii	***	***	***	Publ.	***
D. paralutea	***	***	***	***	***
D. prostipennis	***	***	***	***	***
D. ananassae	***	***	***	Publ.	***
D. varians	***	***	***	Publ.	***
D. bicornuta	***	***	***	***	***
D. dipacantha	***	***	***	***	***
D. kikkawai	***	***	***	Publ.	***
D. nikananu	***	***	***	***	***
D. punjabiensis	***	***	***	***	***
D. seguyi	***	***	***	***	***
D. serrata	***	***	***	Publ.	***

Table 5. The total versus the number of phylogentically informative characters for each gene region separately and in various combinations.

Data	Total Characters	Only Informative Characters	Percent
all 5 gene regions	2030	335	17
Adh	290	69	24
co ii	384	90	24
hb	435	118	27
28S	495	42	8
16S	426	16	4
coding genes (*Adh*, *co* ii, *hb*)	1109	277	25
ribosomal genes (28S, 16S)	921	58	6
nuclear genes(*Adh*, *hb*, 28S)	1220	229	19
mitochondrial genes (*co* ii, 16S)	810	106	13

Table 6. The total and the number of phylogentically informative characters for codon positions of each protein coding gene region, separate and various combinations.

	1st position changes			2nd position changes			3rd position changes		
Data	Total Characters	Only Informative Characters	%	Total Characters	Only Informative Characters	%	Total Characters	Only Informative Characters	%
Adh + hb + co ii	370	38	10	369	14	4	370	225	61
Adh	97	11	11	96	4	4	97	54	56
hb	145	12	8	145	9	6	145	97	67
Adh + hb (nuclear)	242	23	10	241	13	5	242	151	62
co ii (mitochondrial)	128	15	12	128	1	0.80	128	74	58

Table 7. Tree statistics of CI, RI, resolution, and monophyly are compared. * Resolution defined here means the total number of nodes for the ingroup on the simultaneous analysis most parsimonious cladogram (e.g., this would be nodes 5-17).

	CI	RI	resolution*
Adh	0.440	0.650	14
co ii	0.393	0.554	5
hb	0.547	0.759	11
28S	0.584	0.816	13
16S	0.606	0.764	
Adh + *co* ii	0.407	0.588	14
Adh + *hb*	0.485	0.698	11
Adh + 28S	0.457	0.680	15
Adh + 16S	0.539	0.639	6
co ii + *hb*	0.458	0.658	10
co ii + 28S	0.425	0.624	10
co ii + 16S	0.403	0.560	11
hb + 28S	0.535	0.755	11
hb + 16S	0.544	0.752	12
16S + 28S	0.562	0.781	13
Adh + *co* ii + *hb*	0.444	0.642	11
Adh + *co* ii + 28S	0.422	0.620	16
Adh + *co* ii + 16S	0.410	0.587	13
Adh + *hb* + 28S	0.488	0.706	15
Adh + *hb* + 16S	0.486	0.695	11
Adh + 16S + 28S	0.459	0.673	10
co ii + *hb* + 28S	0.466	0.675	11
co ii + *hb* + 16S	0.461	0.657	10
co ii + 16S + 28S	0.433	0.626	13
hb + 16S + 28S	0.536	0.751	15
Adh + *co* ii + *hb* + 28S	0.452	0.658	13
Adh + *co* ii + *hb* + 16S	0.447	0.643	16
Adh + *co* ii + 16S + 28S	0.425	0.618	12
Adh + *hb* + 16S + 28S	0.489	0.703	15
co ii + *hb* + 16S + 28S **	0.469	0.674	17
Adh + *co* ii + *hb* + 16S + 28S	0.454	0.658	13

Table 8. Species with culture/stock numbers for all of the stocks used in this study. All species were obtained from the National *Drosophila* Species Resource Center at Bowling Green with the exception of *D. teissieri* Brazzaville isofemale line 16 which was a gift from D. LaChaise to G. Simmons. All species identifications were confirmed based on male gentalic dissections. The following three species were originally mislabled by the National *Drosophila* Species Resource Center at Bowling Green. *D. ficusphila* was incorrectly labeled as *D. pennae* 14028-0631.0. *D. ercepeae* was incorrectly labeled as *D. greeni* 14028-0712.0. *D. greeni* was incorrectly labeled as *D. ercepeae* 14024-0432.0. Representatives of all taxa sampled have been placed in the collections at the American Museum of Natural History. The species, *D. rajasekari* (14023-0361.3) was ordered from the stock center; however, *D. rajarsekari* Reddy and Krishnamurthy, 1968 and *D. raychaudhurii* Gupta, 1969 were made junior synonyms of *D. biarmipes* Malloch, 1924 by Bock 1980. *D. jambulina* Parshad and Paika, 1964 has been found to be an Indian endemic species. Collections made in Indochina (Thailand and Cambodia, e.g, *D. jambulina* 14028-0531.1) are actually *D. watanabei* Gupta and Gupta, 1992.

Species	Culture/Stock
D. ambigua	14011-0091.0
D. bifasciata	14012-0181.0
D. pseudoobscura	14011-0121.0
D. persimilis	14011-0111.0
D. affinis	14012-0141.0
D. tolteca	14012-0210.0
D. elegans	14027-0461.0
D. eugracilis	14026-0451.0
D. ficusphila	misID 14028-0631.0
D. lucipennis	14023-0331.0
D. mimetica	14023-0341.0
D. biarmipes	14023-0361.3
D. teissieri	gift D. LaChaise
D. yakuba	14021-0261.0
D. takahashii	14022-0311.5
D. lutescens	14022-0271.0
D. paralutea	14022-0281.0
D. prostipennis	14022-0291.0
D. ananassae	14024-0371.0
D. ercepeae	misID 14028-0712.0
D. m. malerkotiana	14024-0391.0
D. pallidosa	14024-0433.0
D. phaeopleura	14024-0434.0
D. varians	14024-0431.0
D. auraria	14028-0471.1
D. barbarae	14028-0491.2
D. baimaii	14028-0481.1
D. biauraria	14028-0501.0
D. bicornuta	14028-0511.0
D. birchii	14028-0521.0
D. dipacantha	14028-0586.0
D. greeni	misID 14024-0432.0
D. watanabei	14028-0531.1
D. kanapiae	14028-0541.0
D. kikkawai	14028-0561.3
D. lini	14028-0581.0
D. mayri	14028-0591.0
D. nikananu	14028-0601.0
D. orosa	14028-0611.0
D. parvula	14028-0621.0
D. punjabiensis	14028-0641.0
D. rufa	14028-0661.0
D. quadraria	14028-0651.0
D. seguyi	14028-0671.0
D. serrata	14028-0681.0
D. triauraria	14028-0691.0
D. tsacasi	14028-0701.0
D. vulcana	14028-0711.0

Table 9. Traditional taxonomic grouping (species group and species subgroup) for species used in this study. The species subgroups for *obscura* are based on relationships proposed by Barrio *et al.*, 1992.

species group	species subgroup	species
obscura	*obscura*	*D. ambigua*
		D. bifasciata
	pseudoobscura	*D. pseudoobscura*
		D. persimilis
	affinis	*D. affinis*
		D. tolteca
melanogaster	*elegans*	*D. elegans*
	eugracilis	*D. eugracilis*
	ficusphila	*D. ficusphila*
	suzukii	*D. lucipennis*
		D. mimetica
		D. biarmipes
	melanogaster	*D. melanogaster*
		D. teissieri
		D. yakuba
	takahashii	*D. takahashii*
		D. lutescens
		D. paralutea
		D. prostipennis
	ananassae	*D. ananassae*
		D. ercepeae
		D. m. malerkotiana
		D. pallidosa
		D. phaeopleura
		D. varians
	montium	*D. auraria*
		D. baimaii
		D. barbarae
		D. biauraria
		D. bicornuta
		D. birchii
		D. dipacantha
		D. greeni
		D. watanabei
		D. kanapiae
		D. kikkawai
		D. lini
		D. mayri
		D. nikananu
		D. orosa
		D. parvula
		D. punjabiensis
		D. quadraria
		D. rufa
		D. seguyi
		D. serrata
		D. triauraria
		D. tsacasi
		D. vulcana

Table 10. Source of gene regions employed in this study. *** are regions sequenced for this study. See text for details. Publ. are previously published sequences obtained from Genbank (accession numbers in parentheses) with the exception of the 28S D1 domain which was entered from the Pélandakis and Solignac (1993) reference. ambigua: co ii (M95145), Adh (X54813); bifasciata: co ii (M95147); pseudoobscura: co ii (M95150), Adh (M60989); persimilis: co ii (M95143), Adh (M60997); affinis: co ii (M95140); tolteca: co ii (M95147); melanogaster; co ii (AF200828), Adh (M11290), hb (Y00274); teissieri; Adh (X54118); yakuba: co ii (X00924), Adh (X57376).

species	co II	Adh	hb
D. ambigua	Publ.	Publ.	•••
D. bifasciata	Publ.	•••	•••
D. pseudoobscura	Publ.	Publ.	•••
D. persimilis	Publ.	Publ.	•••
D. affinis	Publ.	•••	•••
D. tolteca	Publ.	•••	•••
D. elegans	•••	•••	•••
D. eugracilis	•••	•••	•••
D. ficusphila	•••	•••	•••
D. lucipennis	•••	•••	•••
D. mimetica	•••	•••	•••
D. biarmipes	•••	•••	•••
D. melanogaster	Publ.	Publ.	Publ.
D. teissieri	•••	Publ.	•••
D. yakuba	Publ.	Publ.	•••
D. takahashii	•••	•••	•••
D. lutescens	•••	•••	•••
D. paralutea	•••	•••	•••
D. prostipennis	•••	•••	•••
D. ananassae	•••	•••	•••
D. ercepeae	•••	•••	•••
D. m. malerkotiana	•••	•••	•••
D. pallidosa	•••	•••	•••
D. phaeopleura	•••	•••	•••
D. varians	•••	•••	•••
D. auraria	•••	•••	•••
D. baimaii	•••	•••	•••
D. barbarae	•••	•••	•••
D. biauraria	•••	•••	•••
D. bicornuta	•••	•••	•••
D. birchii	•••	•••	•••
D. dipacantha	•••	•••	•••
D. greeni	•••	•••	•••
D. watanabei	•••	•••	•••
D. kanapiae	•••	•••	•••
D. kikkawai	•••	•••	•••
D. lini	•••	•••	•••
D. mayri	•••	•••	•••
D. nikananu	•••	•••	•••
D. orosa	•••	•••	•••
D. parvula	•••	•••	•••
D. punjabiensis	•••	•••	•••
D. quadraria	•••	•••	•••
D. rufa	•••	•••	•••
D. seguyi	•••	•••	•••
D. serrata	•••	•••	•••
D. triauraria	•••	•••	•••
D. tsacasi	•••	•••	•••
D. vulcana	•••	•••	•••

Table 11. The total versus the number of phylogentically informative characters for each gene region, separate as well as various combinations.

Data	Total Characters	Only Informative Characters
all 3 gene regions	1108	342
Adh	290	92
hb	434	138
nuclear (*Adh* + *hb*)	724	230
co ii (mitochondrial)	384	112

Table 12. The total and the number of phylogentically informative characters for codon positions of each gene region.

Data	1st Position changes		2nd Position changes		3rd Position changes	
	Total Characters	Only Informative Characters	Total Characters	Only Informative Characters	Total Characters	Only Informative Characters
all 3 gene regions	369	52	368	21	369	267
Adh	97	18	96	8	97	66
hb	144	16	144	12	144	108
nuclear (*Adh* + *hb*)	241	34	240	20	241	174
co ii (mitochondrial)	128	18	128	1	128	93

Table 13. Tree statistics of CI, RI, resolution, and monophyly are compared. * Resolution defined here means the total number of nodes for the ingroup on the simultaneous analysis most parsimonious cladogram (e.g., this would be nodes 6-46).

	Adh	hb	co ii	Adh + hb	Adh + co ii	hb + co ii	Adh + hb + co ii
CI	0.379	0.423	0.296	0.39	0.323	0.348	0.349
RI	0.702	0.744	0.612	0.71	0.64	0.668	0.667
resolution*	26	26	34	34	36	31	41
Monophyly:							
ananassae	yes	yes	no	yes	yes	yes	yes
melanogaster	yes	yes	yes	yes	yes	yes	yes
montium	yes	yes	yes	yes	yes	yes	yes
suzukii	no	no	no	no	no	no	no
takahashii	no	no	no	yes	yes	yes	yes
3 clades?	no	yes	no	yes	no	yes	yes
If yes,relationship		an (mont + mel/or)		mont(an + mel/or)		unresolved	mel/or(an + mont)

Table 14. Nodes were compared using Bremer, Bootstrap, and partitioned Bremer support values. Data partitions were divided into single gene (i.e., *Adh*, *hb*, and *co ii*) and double gene regions (i.e., *Adh* + *hb*, *adh* + *co ii* and *hb* + *co ii*) and each node was examined to see if it appeared in the total evidence tree for each type of partition.

Node Number	Partitioned Bremer Support Value (PBS)			Bremer Value	Bootstrap Value (%)	Gene Region combination with Node		
	hb	*co ii*	*Adh*			single	double	Total
1	-1	12	4	15	100	2	3	5
2	-8.33	7.67	2.67	2	75	2	1	3
3	-0.33	4.67	0.67	5	88	1	2	3
4	3.67	-3.33	0.67	1	63	1	1	2
5	25.25	5.88	4.88	36	100	3	3	6
6	8	1	1	10	100	3	3	6
7	1.63	-0.25	-0.38	1	25	0	0	0
8	0.8	0.4	-0.2	1	43	0	2	2
9	0.67	1	1.33	3	59	1	2	3
10	8.5	7.5	0	16	100	2	3	5
11	-0.33	3.5	4.83	8	90	3	3	6
12	-3.5	4	1.5	2	49	0	1	1
13	-2.8	3.2	1.6	2	40	1	1	2
14	10.25	2.5	-0.75	12	99	2	2	4
15	-0.67	2.67	0	2	40	1	0	1
16	3	-1	1	3	63	1	2	3
17	8.5	3.75	-9.25	3	85	1	2	3
18	2	5	2	9	100	3	3	6
19	0	2	-1	1	54	1	2	3
20	6.67	5.67	-7.33	5	97	2	3	5
21	3.5	2.5	-4	2	48	1	2	3
22	9.6	-5.1	4.5	9	98	2	3	5
23	0	3.5	-0.5	3	57	1	2	3
24	-0.33	1.33	1	2	59	1	2	3
25	0	1	0	1	42	1	2	3
26	2	-0.5	0.5	2	46	0	2	2
27	1.33	0	0.67	2	34	0	3	3
28	0	1	1	2	66	2	3	5
29	8.92	-1.33	7.42	15	100	3	3	6
30	7.27	1.68	-4.95	4	68	2	1	3
31	0	1	0	1	13	0	0	0
32	4	7	3	14	100	3	3	6
33	5.33	-2.33	-1	2	33	1	0	1
34	4	-0.33	-1.67	2	17	1	0	1
35	6	3	-1	8	93	2	3	5
36	2.64	0.27	-0.91	2	27	1	1	2
37	2.31	0.85	-1.15	2	13	0	0	0
38	-1	2	0	1	66	1	2	3
39	0	0	2	2	89	1	2	3
40	5	13	4	22	100	3	3	6
41	9.5	-3	-0.5	6	95	3	3	6
42	3	2.5	1.5	7	97	3	3	6
43	7.5	-3	-2.5	2	27	1	1	2
44	1	0	0	1	63	2	1	3
45	3	6.61	4.39	14	99	3	3	6
46	2.17	-0.83	-0.33	1	35	0	0	0

Table 15. Total nodes were counted for each data partition to see whether it contained nodes that also appeared in the total evidence tree.

	Adh	hb	co ii	Adh + hb	Adh + co ii	hb + co ii
outgroup	2	2	2	3	2	2
ingroup	16	22	21	25	24	28
subtotal	18	24	23	28	26	30
Node 5	1	1	1	1	1	1
Total	19	25	24	29	27	31

Table 16. List of synapomorphies for the *montium* subgroup as derived from the *melanogaster* species group phylogeny. Characters and their respective c.i. values are reported. The nucleotide base changes and the resulting amino acid changes are included. Although three gene regions (*Adh, co ii, hb*) were used for the phylogentic reconstruction of relationships, all do not contribute characters to every node. Some instances there is more than one codon position change for a single amino acid; therefore, the number of nucleotide changes do not always result in an equivalent amount of amino acid changes. Double line arrows indicate unambiguous character changes and single line arrows indicate characters changes optimized at the node using ACCTRAN.

	Adh				*co ii*				*hb*			
Node	Character	c.i.	Nucleotide Change	Amino Acid Change	Character	c.i.	Nucleotide Change	Amino Acid Change	Character	c.i.	Nucleotide Change	Amino Acid Change
23					522	0.167	A ==> T	Ala ==> Ala				
					531	0.250	T ==> C	Met ==> Ile				
					612	0.167	T ==> C	Asn ==> Asn				
					621	0.250	T ==> A	Gly ==> Gly				
24	151	0.125	C ==> T	Asp ==> Asp	354	0.300	A --> T	Gly --> Gly	1013	0.500	A --> C	Pro --> Pro
					450	0.125	T ==> C	Phe ==> Phe				
25					366	0.111	T ==> C	Try ==> Try				
					576	0.273	T --> A	Thr --> Thr				
					615	0.333	T ==> A	Arg ==> Arg				
26	247	0.333	G --> C	Pro --> Pro	354	0.300	T --> A	Gly --> Gly	941	0.250	T ==> C	Gly ==> Gly
									968	0.667	T ==> C	Gly ==> Gly
27	172	0.111	T ==> C	Ile ==> Ile					740	0.500	G ==> A	Gln ==> Gln
28	172	0.111	T ==> C	Ile ==> Ile	381	0.500	T ==> C	Try ==> Try	722	0.429	G --> C	Gln --> His
					591	0.100	T ==> C	Asn ==> Asn	746	0.667	G --> T	Gln --> His
									761	1.000	T ==> C	His ==> His
29	36	1.000	T ==> A	Phe ==> Try	292	0.200	T ==> C	Leu ==> Leu	677	0.100	T --> C	Ser --> Ser
	85	0.500	G ==> A	Leu ==> Leu	294	0.200	A ==> T	Leu ==> Leu	842	0.111	A ==> G	Gln ==> Gln
	104	0.600	C ==> T	Leu ==> Phe	300	0.667	A --> G	Pro --> Pro	887	0.167	G ==> A	Gln ==> Gln
	113	0.167	G ==> A	Val ==> Ile	330	0.250	A ==> T	His ==> His	965	0.300	C ==> T	Pro ==> Pro
	115	0.250	G ==> T	Thr ==> Thr	357	0.200	T ==> C	Val ==> Met	968	0.667	T ==> C	Gly ==> Gly
	202	0.600	G ==> A	Lys ==> Ile	478	0.200	G ==> A	Arg --> Arg				
	231	0.500	A ==> T		480	0.375	A ==> T					
	232	1.000	G ==> C		585	0.333	A --> G					

	Adh				co ii				hb			
Node	Character	c.i.	Nucleotide Change	Amino Acid Change	Character	c.i.	Nucleotide Change	Amino Acid Change	Character	c.i.	Nucleotide Change	Amino Acid Change
30	43	0.429	G ==> A	Pro ==> Pro	423	0.222	T ==> A	Pro ==> Pro	827	0.333	C ==> T	Asp ==> Asp
	151	0.125	C ==> T	Asp ==> Asp	486	0.091	T ==> A	Pro ==> Pro	857	0.429	G ==> T	Thr ==> Thr
					522	0.167	A ==> T	Ala ==> Ala	1036	1.000	C ==> A	Thr ==> Asn
					528	0.167	T ==> A	Val ==> Val				
					543	1.000	A ==> T	Thr ==> Thr				
31					436	0.200	T ==> G	Ser ==> Ala				
32	106	0.250	C ==> T	Leu ==> Leu	357	0.200	T ==> C	His ==> His	860	0.500	G ==> C	Leu ==> Leu
	115	0.250	G ==> C	Val ==> Val	366	0.111	T ==> C	Try ==> Try	890	0.333	C ==> T	Asp ==> Asp
	139	0.375	C ==> A	Gly ==> Gly	372	0.333	T ==> C	Ser ==> Ser	1035	0.333	A ==> G	Thr ==> Ala
	215	0.400	C ==> A	Leu ==> Met	411	0.286	A --> T	Ser --> Ser	1050	0.167	C --> T	Leu --> Leu
	220	0.667	C ==> G	Asp ==> Glu	444	0.400	T ==> C	Asp ==> Asp	1061	0.429	C ==> T	Leu ==> Leu
					481	0.250	T ==> C	Leu ==> Leu				
					582	0.231	T ==> A	Gly ==> Gly				
					612	0.167	T ==> C	Asn ==> Asn				
33	1	0.200	G ==> A	PRIMER					677	0.100	T ==> C	Ser ==> Ser
	130	0.333	C ==> T	Asn ==> Asn					725	0.167	C ==> G	His ==> Gln
	148	0.125	C ==> T	Asp ==> Asp								
	190	0.250	C --> T	Gly --> Gly								
	244	0.222	C --> A	Gly --> Gly								
	250	0.200	C ==> T	Gly ==> Gly								
34									1013	0.500	C ==> A	Pro ==> Pro
									1028	0.222	G ==> A	Thr ==> Thr
									1050	0.167	T --> C	Leu --> Leu

	Adh				*co ii*				*hb*			
Node	Character	c.i.	Nucleotide Change	Amino Acid Change	Character	c.i.	Nucleotide Change	Amino Acid Change	Character	c.i.	Nucleotide Change	Amino Acid Change
35	190	0.250	C ==> T	Gly ==> Gly	411	0.286	A --> T	Ser --> Ser	710	1.000	C ==> A	Leu ==> Leu
					486	0.091	T ==> A	Pro ==> Pro	947	0.200	C ==> T	Phe ==> Phe
					519	0.214	A ==> T	Ala ==> Ala				
					546	0.200	T ==> C	Met ==> Ile				
					558	0.333	A ==> T	Gly ==> Gly				
					585	0.333	A ==> T	Arg ==> Arg				
					618	0.286	T ==> A	Pro ==> Pro				
					621	0.250	T --> A	Gly --> Gly				
					633	0.222	T ==> A	Gly ==> Gly				
36					438	0.333	T ==> A	Ser ==> Ser	734	0.429	G ==> A	Gln ==> Gln
					522	0.167	A ==> T	Ala ==> Ala				
					591	0.100	T --> C	Asn --> Asn				
37					354	0.300	A ==> T	Gly ==> Gly	827	0.333	T ==> C	Asp ==> Asp
					384	0.250	T --> A	Ser --> Ser				
					411	0.286	T --> A	Ser --> Ser				
					465	0.250	A --> T	Val --> Val				
38					375	0.333	T ==> C	Try ==> Try				
					513	0.222	A ==> G	Val ==> Val				
39	193	1.000	G ==> C	Leu ==> Leu					1004	0.143	C --> T	Asn --> Asn
	215	0.400	C ==> A	Leu ==> Met								

	Adh				co ii				hb			
Node	Character	c.i.	Nucleotide Change	Amino Acid Change	Character	c.i.	Nucleotide Change	Amino Acid Change	Character	c.i.	Nucleotide Change	Amino Acid Change
40	59	1.000	C ==> A	Pro ==> Lys	298	0.250	T ==> C	Leu ==> Leu	765	1.000	T ==> C	Ser ==> Pro
	60	1.000	C ==> A	Gln ==> Glu	327	0.500	A ==> G	Glu ==> Glu	1019	0.250	A ==> G	Pro ==> Pro
	155	1.000	C ==> G	Ile ==> Ile	372	0.333	T ==> C	Ser ==> Ser	1034	0.600	A ==> C	Thr ==> Thr
	268	0.500	T ==> C		387	0.500	T ==> C	Asp ==> Asp	1037	0.300	T ==> C	Thr ==> Thr
					414	1.000	T ==> C	Try ==> Try	1088	0.429	C ==> A	Pro ==> Pro
					450	0.125	T ==> C	Phe ==> Phe				
					486	0.091	T ==> A	Pro ==> Pro				
					531	0.250	T ==> C	Met ==> Ile				
					534	0.250	T ==> C	His ==> His				
					552	0.250	T ==> C	Ala ==> Ala				
					591	0.100	T ==> C	Asn ==> Asn				
					654	0.250	A ==> G	Gly ==> Gly				
					660	0.500	T ==> C	Asn ==> Asn				
41	106	0.250	C --> G	Leu --> Leu	433	0.333	T ==> C	Leu ==> Leu	908	0.333	A ==> C	Ala ==> Ala
	169	1.000	C ==> A	Thr ==> Thr	480	0.375	A --> T	Val ==> Val	911	0.600	C ==> G	Ala ==> Ala
	172	0.111	T ==> C	Ile ==> Ile	567	0.167	T ==> A	Val ==> Val	995	0.250	T ==> C	Try ==> Try
	241	0.333	C ==> A	Gly ==> Gly								
42	29	0.143	G ==> A	Val ==> Ile	481	0.250	T ==> C	Leu ==> Leu	692	0.250	T ==> C	Ser ==> Ser
	160	0.167	C ==> T	Ile ==> Ile	483	1.000	A ==> T	Ala ==> Ala	722	0.429	G --> C	Gln --> His
	196	0.400	C ==> T	Val ==> Val	519	0.214	A ==> T	Ser ==> Ser	734	0.429	G ==> A	Gln ==> Gln
	244	0.222	C --> A	Gly --> Gly	537	0.200	A --> T		765	1.000	T ==> A	Ser ==> Thr
									842	0.111	A --> G	Gln --> Gln
									1061	0.429	C ==> T	Leu ==> Leu
									1070	0.750	C --> A	Pro --> Pro

	Adh					*co ii*					*hb*				
Node	Character	c.i.	Nucleotide Change	Amino Acid Change		Character	c.i.	Nucleotide Change	Amino Acid Change		Character	c.i.	Nucleotide Change	Amino Acid Change	
43	43	0.429	A --> G	Pro -->	Pro	582	0.231	A ==> T	Gly ==>	Gly	1019	0.250	C ==> A	Pro ==>	Pro
	151	0.125	T ==> C	Asp ==>	Asp						1035	0.333	G ==> A	Ala ==>	Thr
44	43	0.429	C --> A	Pro -->	Pro	384	0.250	A --> T	Ser -->	Ser	1010	0.300	A ==> T	Arg ==>	Arg
	61	0.333	C --> G	Pro -->	Pro	438	0.333	A ==> T	Ser ==>	Ser	1049	0.143	A ==> G	Lys ==>	Lys
	271	0.400	C ==> A	Gly ==>	Gly	480	0.375	T --> A	Val -->	Val					
						486	0.091	A --> T	Pro -->	Pro					
						528	0.167	A --> T	Val -->	Val					
						615	0.333	A --> T	Arg -->	Arg					
						633	0.222	A --> T	Gly -->	Gly					
45	19	0.600	C ==> G	Pro ==>	Pro	436	0.200	G ==> T	Ala ==>	Ser	677	0.100	C --> T	Ser -->	Ser
	40	0.333	T ==> C	Try ==>	Try	454	0.250	T ==> C	Leu ==>	Leu	691	0.500	A ==> G	Asn ==>	Ser
	106	0.250	G --> C	Leu -->	Leu	456	0.600	A ==> T	Ser ==>	Ser	701	0.250	A --> G	Glu -->	Glu
	154	0.333	C --> T	His -->	His	537	0.200	T --> A	Pro -->	Pro	791	0.667	C ==> T	Thr ==>	Thr
	190	0.250	T --> C	Gly -->	Gly	549	0.500	A ==> T	Thr ==>	Thr	792	0.500	C ==> A	Leu ==>	Met
	202	0.600	C ==> G	Thr ==>	Thr	597	0.200	T ==> A			806	0.429	A ==> G	Thr ==>	Thr
	205	0.429	C ==> G	Thr ==>	Thr						842	0.111	G --> A	Gln -->	Gln
	244	0.222	A --> C	Gly -->	Gly						1037	0.300	G --> T	Ala -->	Ala
	247	0.333	A ==> G	Pro ==>	Pro										
	256	0.250	C ==> T	Ile ==>	Ile										
	274	0.667	C ==> G	Ser ==>	Ser										
	280	0.250	T ==> G	Thr ==>	Thr										

Table 17. Comparison table of various inference statistics values for the nodes contained within the *montium* subgroup. Bremer support also referred to as decay indices for each node (Bremer, 1988; 1994). Partitioned Bremer support (PBS) gives a single data partition's (*co ii*, *Adh*, or *hb)* Bremer value for the support at node of the simultaneous analysis tree. PBS values can be positive (complementary support for node), negative (contradictory support for existence of node) or zero (no support either way for node) (Baker and DeSalle. 1997). Bootstrap values are percentages resulting from 1000 replicates of an artificial matrix consisting of my data resampled with replacement.

Node	Partitioned Bremer Support			Bremer value	bootstrap value
	hb	co ii	Adh		
23	0	3.5	-0.5	3	57
24	-0.33	1.33	1	2	59
25	0	1	0	1	42
26	2	-0.5	0.5	2	46
27	1.33	0	0.67	2	34
28	0	1	1	2	66
29	8.92	-1.33	7.42	15	100
30	7.27	1.68	-4.95	4	68
31	0	1	0	1	13
32	4	7	3	14	100
33	5.33	-2.33	-1	2	33
34	4	-0.33	-1.67	2	17
35	6	3	-1	8	93
36	2.64	0.27	-0.91	2	27
37	2.31	0.85	-1.15	2	13
38	-1	2	0	1	66
39	0	0	2	2	89
40	5	13	4	22	100
41	9.5	-3	-0.5	6	95
42	3	2.5	1.5	7	97
43	7.5	-3	-2.5	2	27
44	1	0	0	1	63
45	3	6.61	4.39	14	99

Table 18. The total number of unambiguous and ambiguous characters contributed by each gene region for each node of the *montium* subgroup are listed here. Characters were optimized on the simultaneous analyses most parsimonious cladogram using ACCTRAN.

Node	Characters			Total
	Adh	*co ii*	*hb*	
23	0	4	0	4
24	1	2	1	4
25	0	3	0	3
26	1	1	2	4
27	1	0	1	2
28	1	2	3	6
29	8	8	5	21
30	2	5	3	10
31	0	1	0	1
32	5	8	5	18
33	6	0	2	8
34	0	0	3	3
35	1	9	2	12
36	0	3	1	4
37	0	4	1	5
38	0	2	0	2
39	2	0	1	3
40	4	13	5	22
41	4	3	3	10
42	4	4	7	15
43	2	1	2	5
44	3	7	2	12
45	12	6	8	26
Total	57	86	57	200

FIGURES

Figure 1. The taxonomic placement of the *melanogaster* group and the *montium* subgroup within the subgenus *Sophophora* into seven named species groups. Notice that the seventh species group is either *dispar* Mather, 1955 (cf. Lemeunier *et al.*, 1986; Scouras, 1995) or *populi* Throckmorton, 1975 (cf. Ashburner, 1989). The *melanogaster* group is further subdivided into its respective species subgroups. *D. melanogaster* is a member of the *melanogaster* subgroup. Note the uneven distribution of species among the subgroups with the *montium* subgroup being the most speciose. Also contained within the *melanogaster* group are 5 taxa *incertae sedis*.

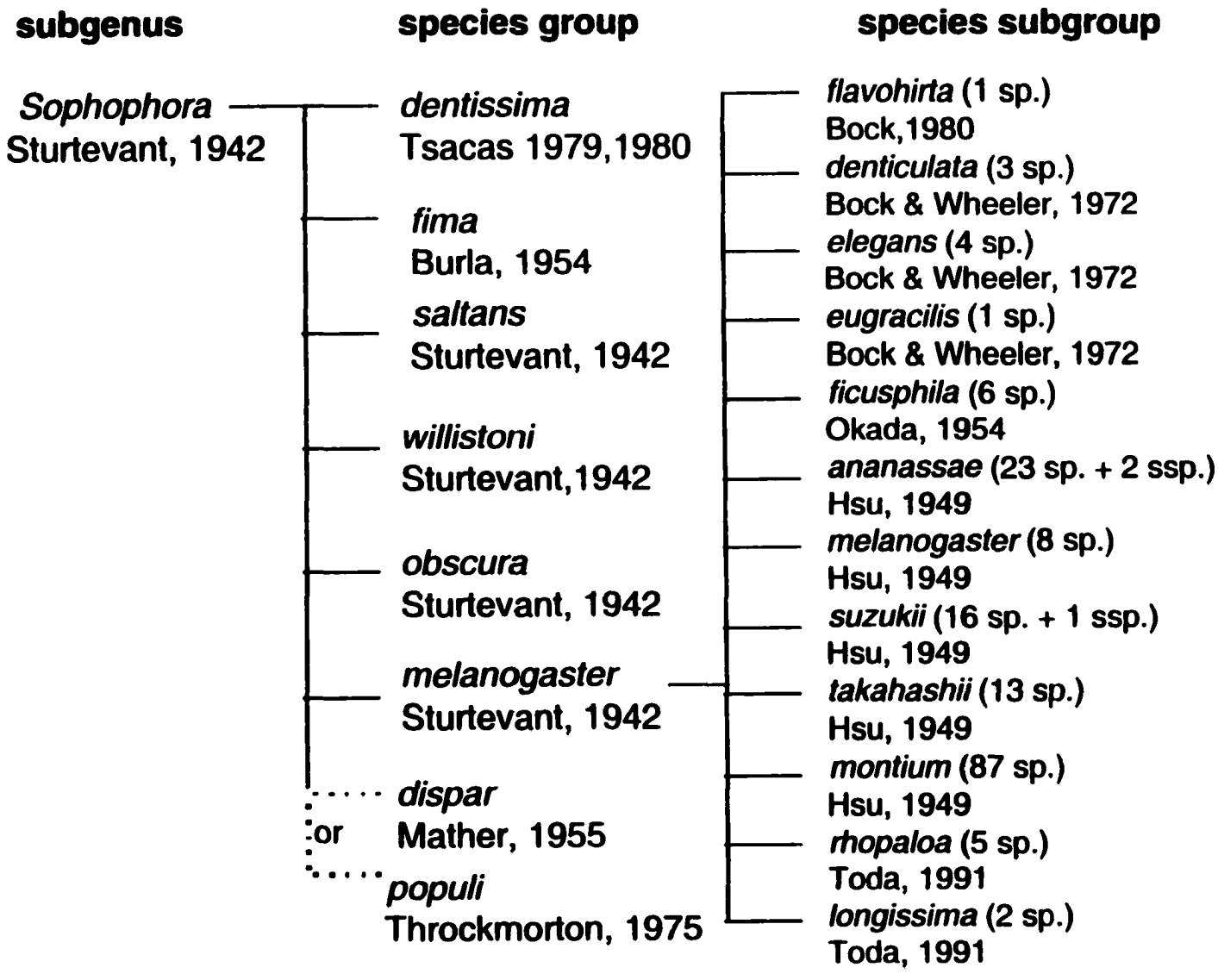
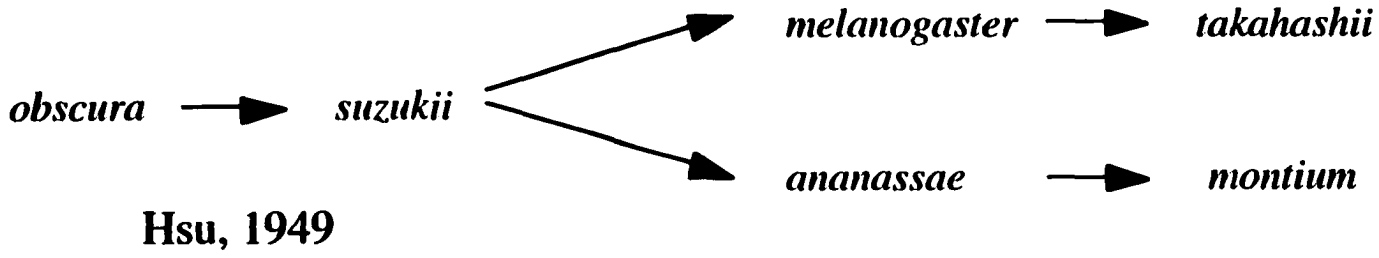
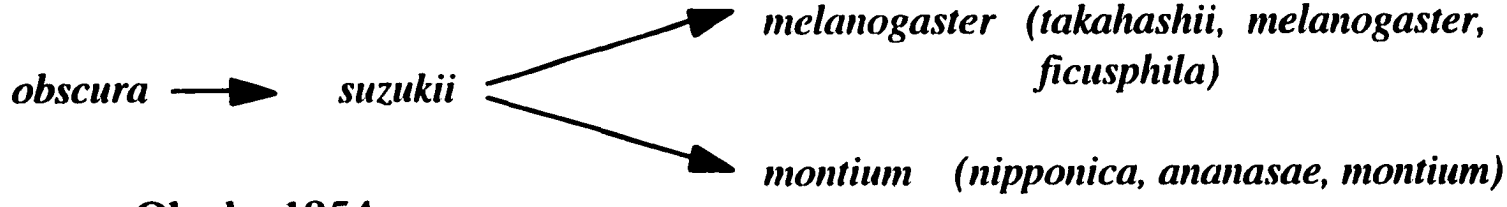


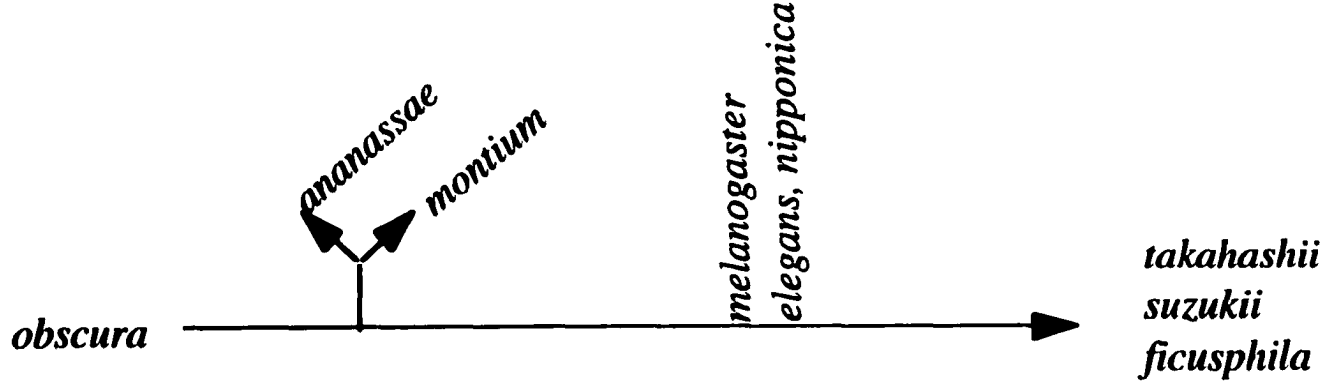
Figure 2. Hypotheses proposed for relationships among subgroups of the *melanogaster* group. Hsu's (1949) hypothesis is based on male periphallic structures. Okada (1954) used a phenetic analysis of principally male phallic characters. Bock and Wheeler (1972) added 30 species to the *melanogaster* group. They proposed the relationships for the subgroups of the *melanogaster* group based primarily from characters of the male sex combs and external genitalia, but also abdominal banding, malpighian tubules, testes color, wing color and female spermatheaca, and ventral receptacle. Ashburner *et al.* (1984) used characters from male external genitalia and sex combs, plus metaphase chromosome karyotypes and polytene chromosome banding pattern, for the *melanogaster, ananassae, montium, suzukii, ficusphila, elegans, eugracilis* and *takahashii* subgroups. Pélandakis *et al.* (1991) as well as Pélandakis and Solignac (1993) sequenced the D1 and D2 domains of 28S rRNA (approximately 550bp) to examine relationships in the subgenus *Sophophora* and the genus *Drosophila*. Only the portion that overlaps my study is redrawn here. Pélandakis *et al.* (1991) and Pélandakis and Solignac (1993) concluded that the *melanogaster* group is paraphyletic.



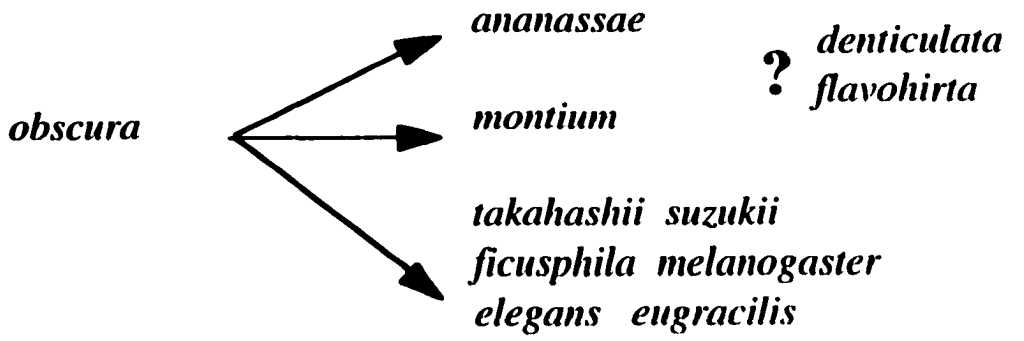
Hsu, 1949



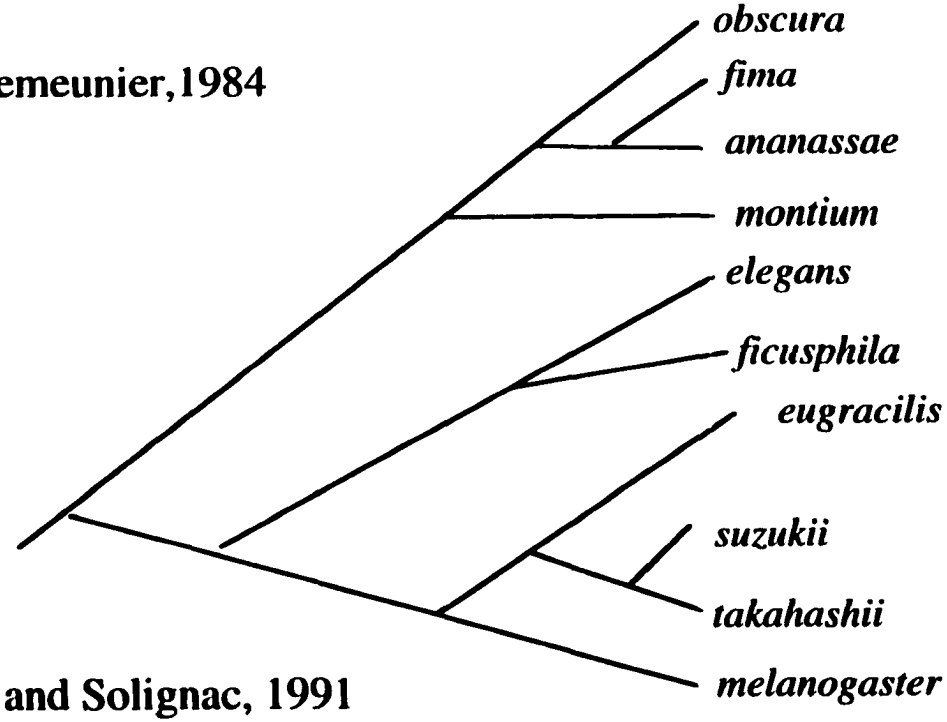
Okada, 1954



Bock and Wheeler, 1972



Ashburner, Bodmer and Lemeunier, 1984



Pélandakis, Higgins and Solignac, 1991
Pélandakis and Solignac, 1993

Figure 3. The cumulative number of species described for the *melanogaster* group. The bars depict the number of new species described each year. The line shows the cumulative number of known species to date. Notice the apparent lack of interest in these small flies in the early 20th century, prior to the widespread use of *D. melanogaster* for genetics. The "spike" at 1972 results from Bock and Wheeler's (1972) monograph. Lack of an asumptote indicates more species are to be found and described.

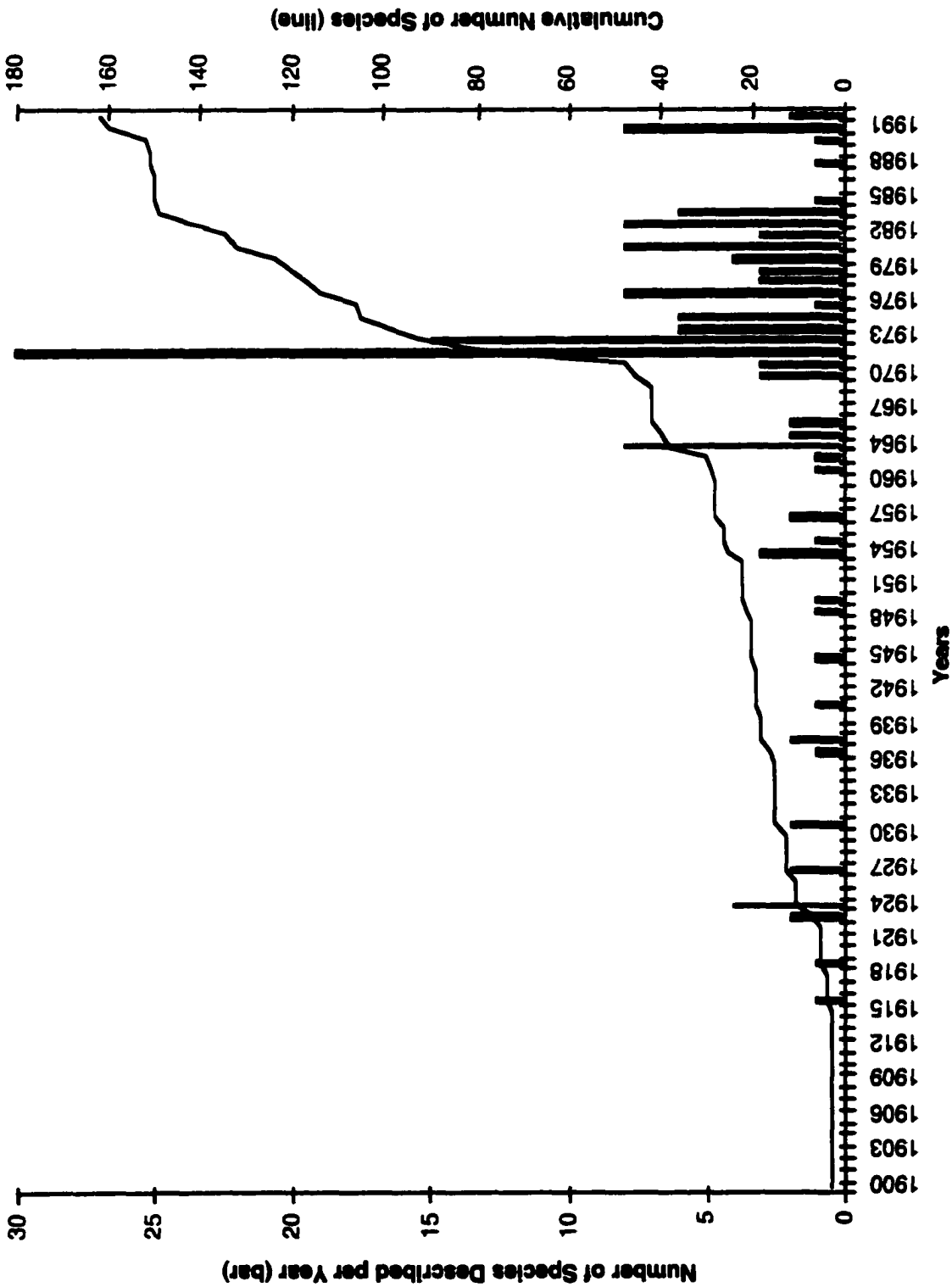


Figure 4. The percent composition for each nucleotide base all characters and phylogenetically informative characters. Comparing the five individual gene regions and all five gene regions combined.

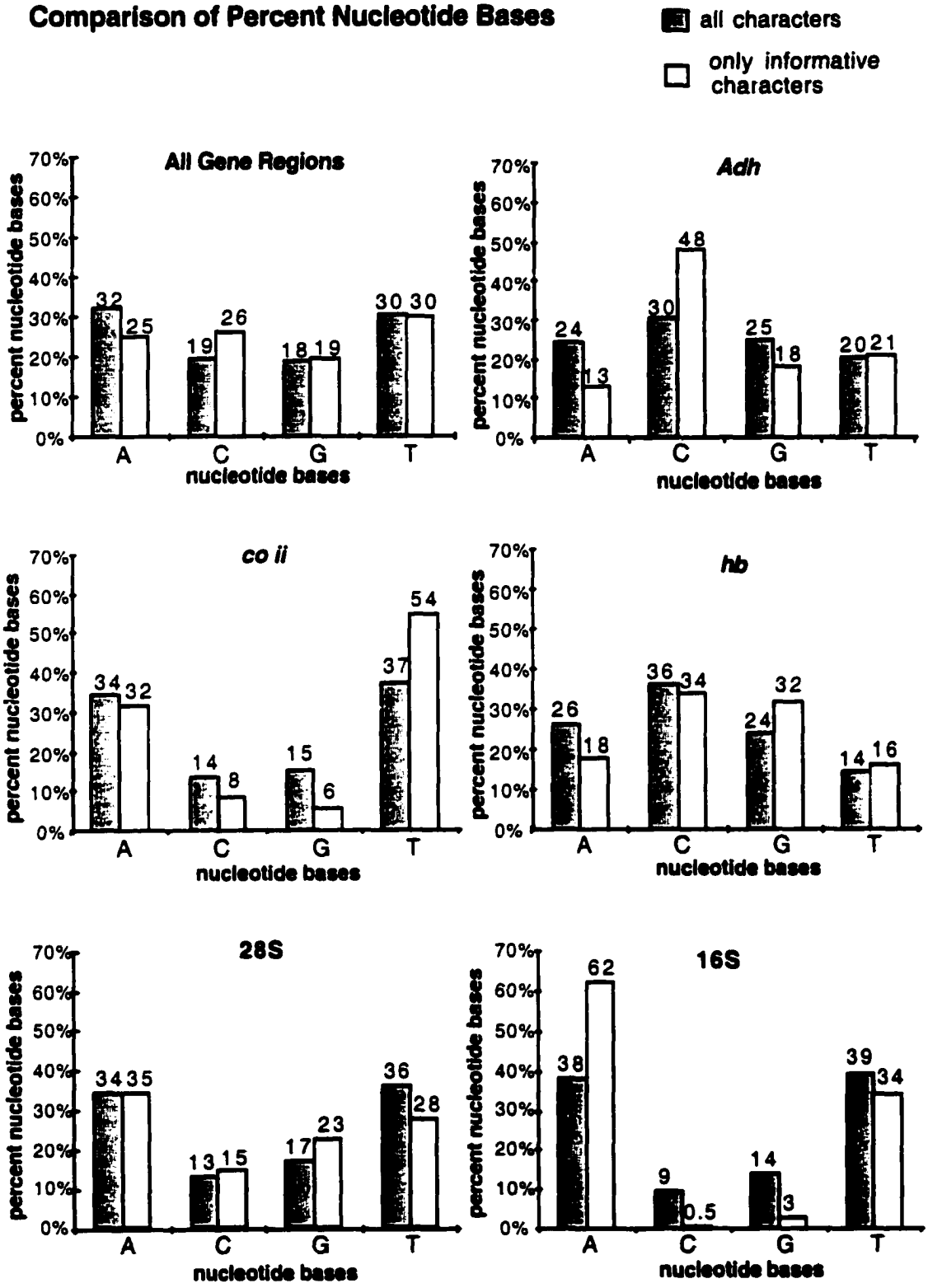


Figure 5. These are the eight equally most parsimonious cladograms resulting from a simultaneous analysis of the data (i.e., *Adh* + *co* ii + *hb* + 16S + 28S). All nodes are the same except for varying degrees of resolution within the *montium* species subgroup clade.

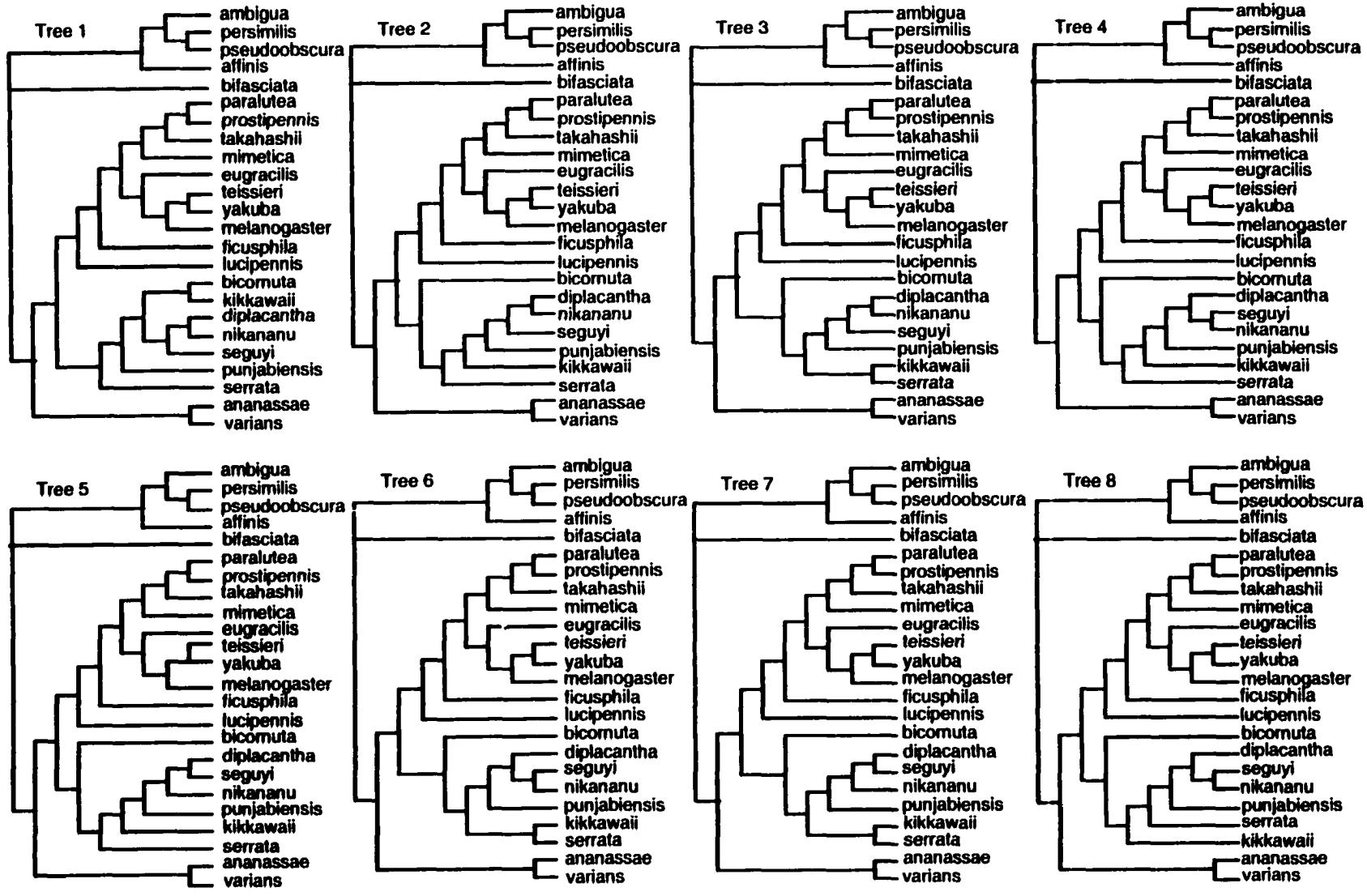


Figure 6. There was only one cladogram produced as a result of successive weighting. The relationships scheme does not exactly match any of the equally most prarsimonious cladograms from the simultaneous analysis.

Successive Weighting Tree

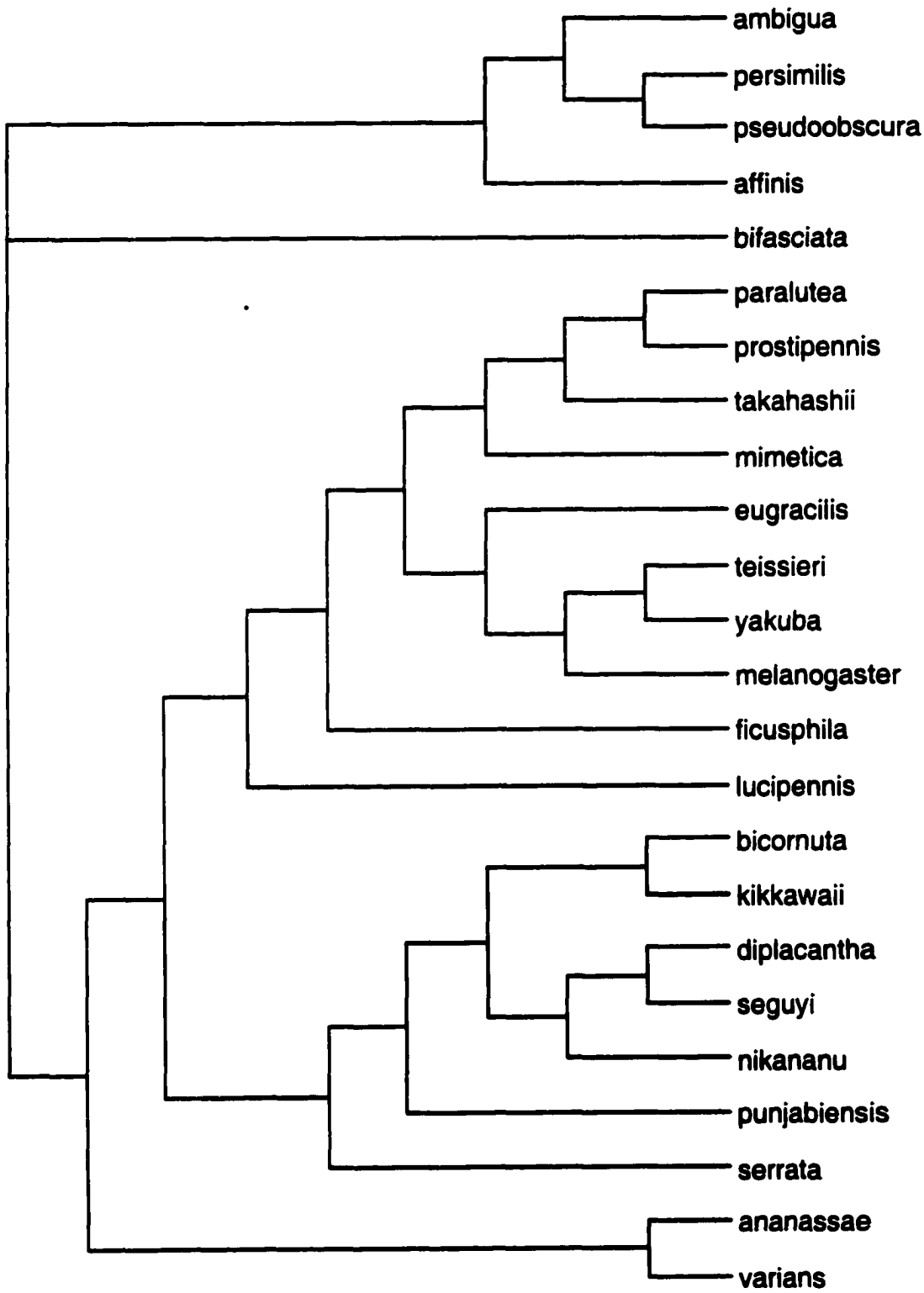


Figure 7. Molecular phylogenetic reconstruction of the *melanogaster* species group. This is the strict consensus cladogram resulting from a simultaneous analysis of the gene regions *Adh, co ii, hb* , 28S and 16S (L = 1035, CI = 0.45 and RI = 0.66). Numbers at nodes above the line are the Bremer support values (Bremer, 1998; 1994) and below the line are the bootstrap values (Felsenstein, 1985). Most previously established groups are supported except for the *obscura* and *suzukii* species subgroups. The traditional species groups and subgroups are abbreviated as follows: obs. = obscura, psd. = pseudoobscura, aff. = affinis, mel. = melanogaster, tak. = takahashii, suz. = suzukii, eleg. = elegans, eug. = eugracilis, ana. = ananassae, ficus. = ficusphila, and mont. = montium.

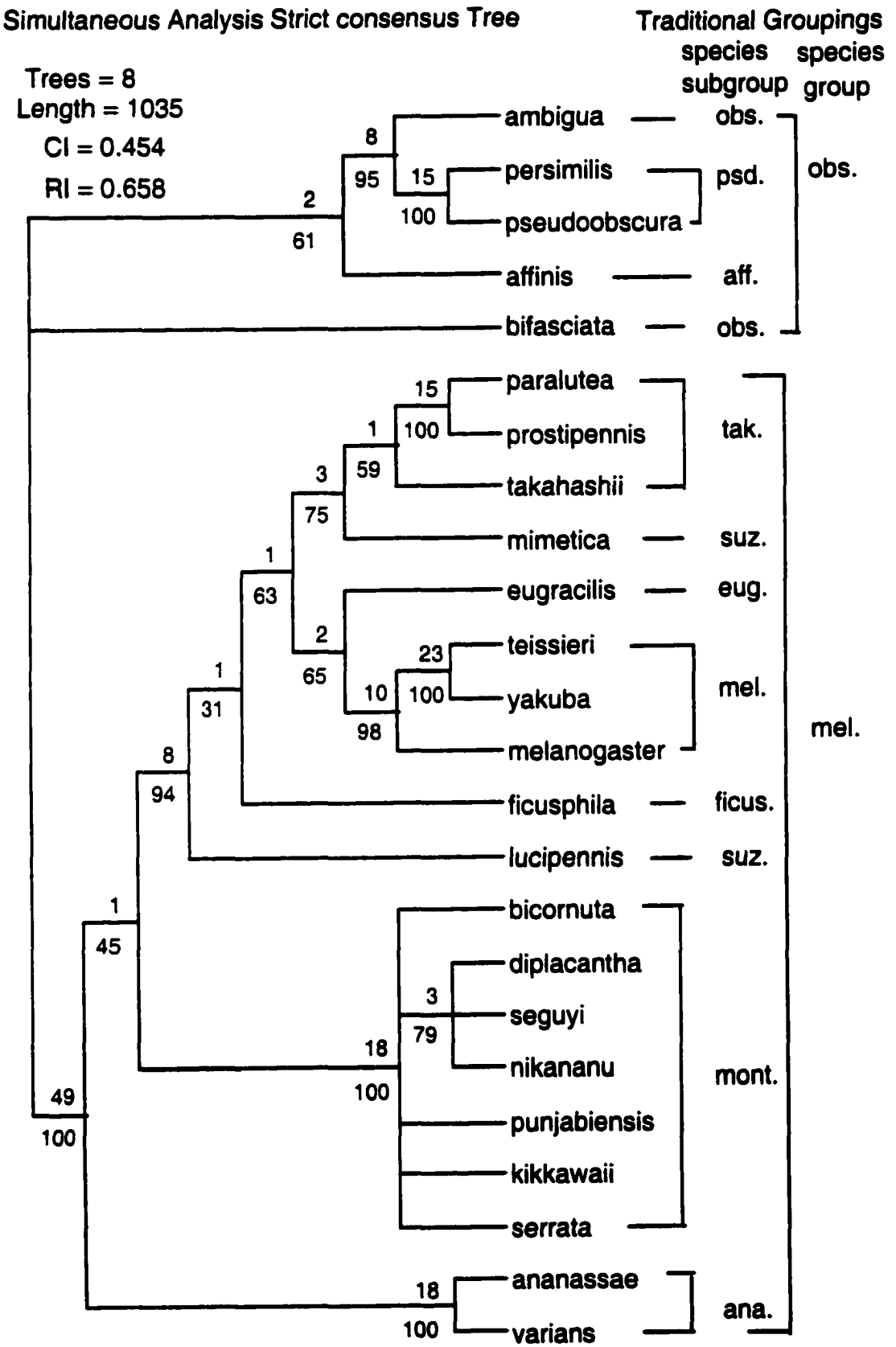


Figure 8. Node numbers for the strict consensus cladogram resulting from the simultaneous analysis of the five gene regions.

Simultaneous Analysis Strict Consensus Tree

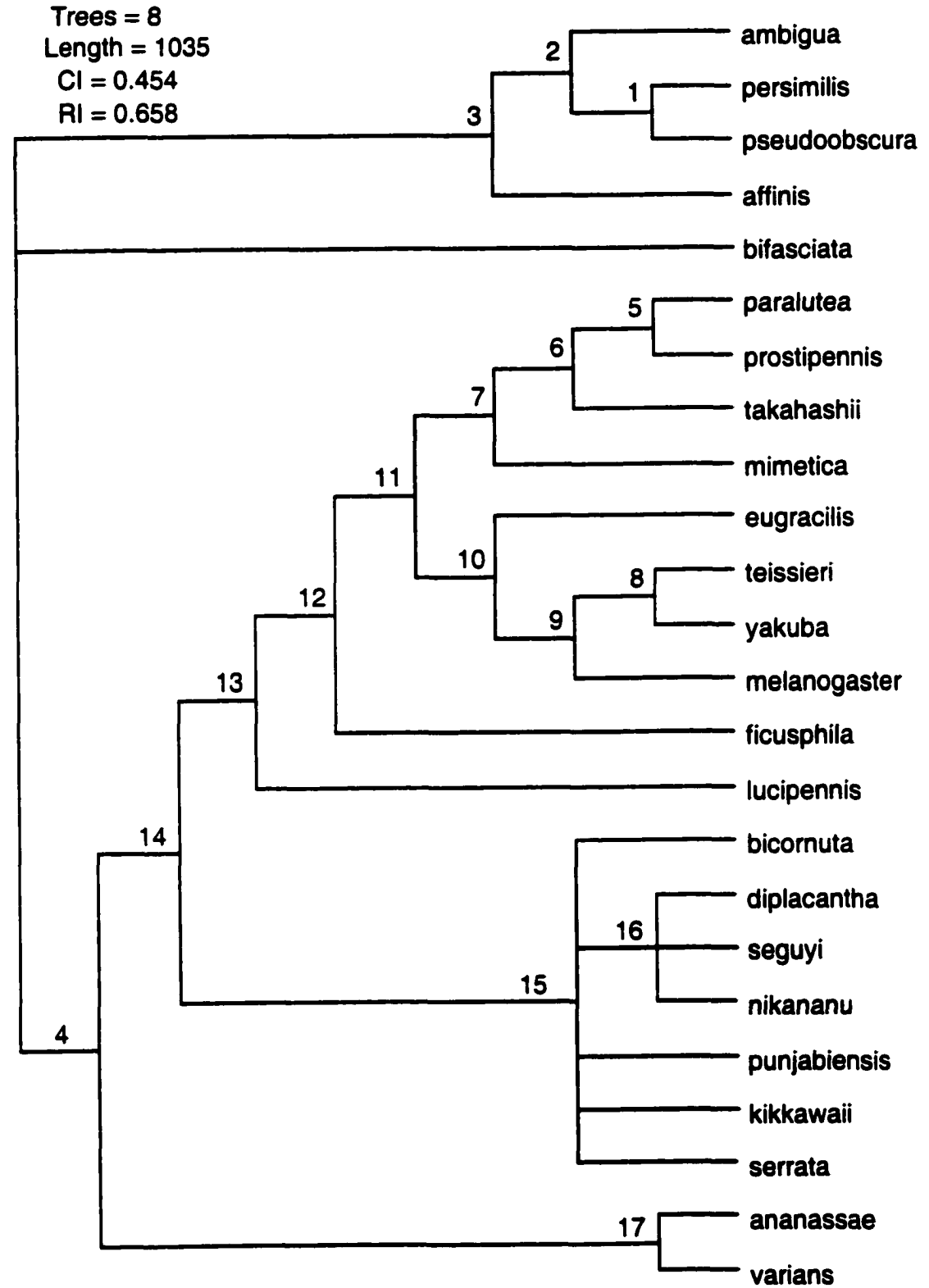


Figure 9. The total percent of transitions and transversions for the phylogentically informative characters for the 3 separate gene regions as well as the total evidence data set.

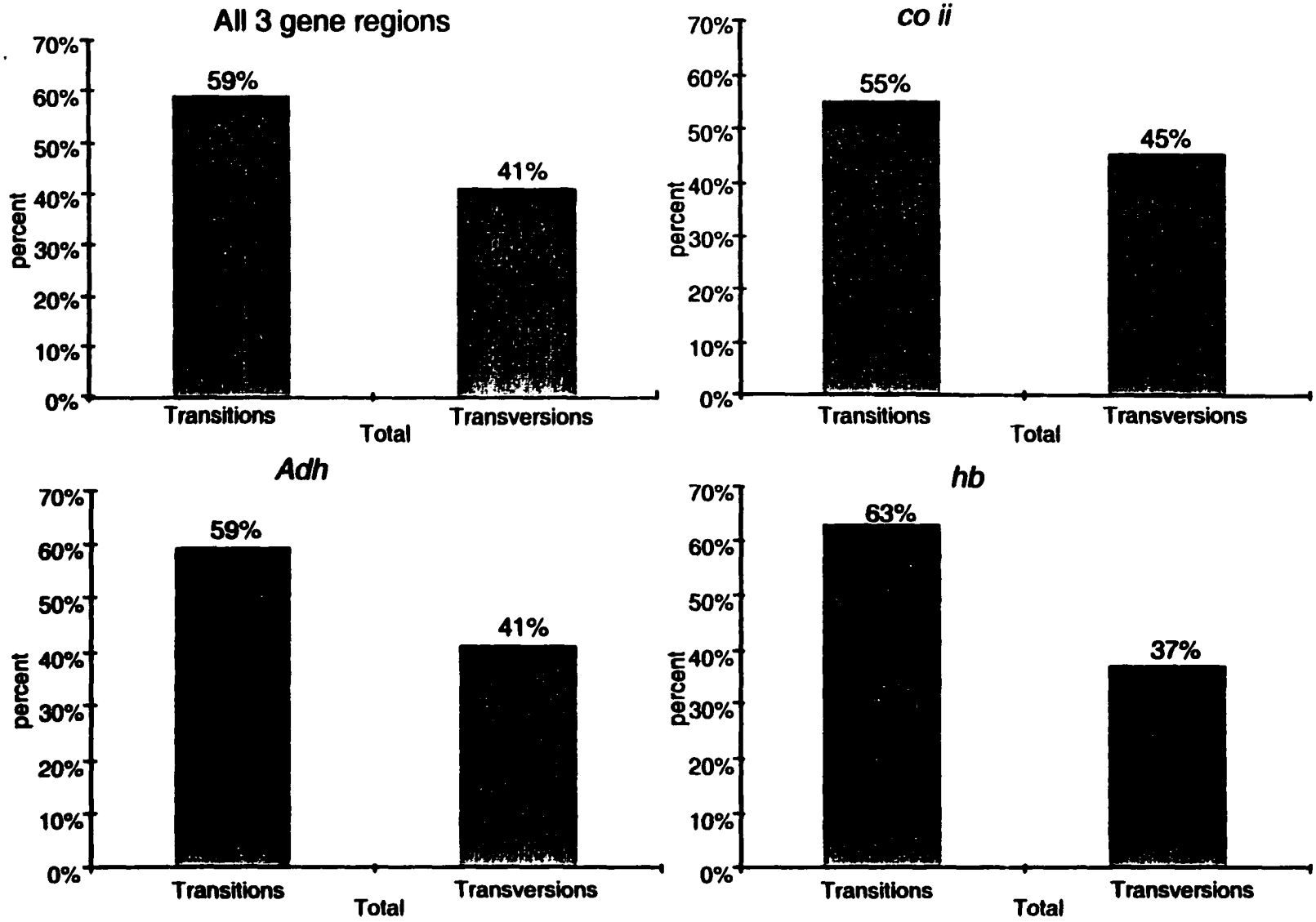
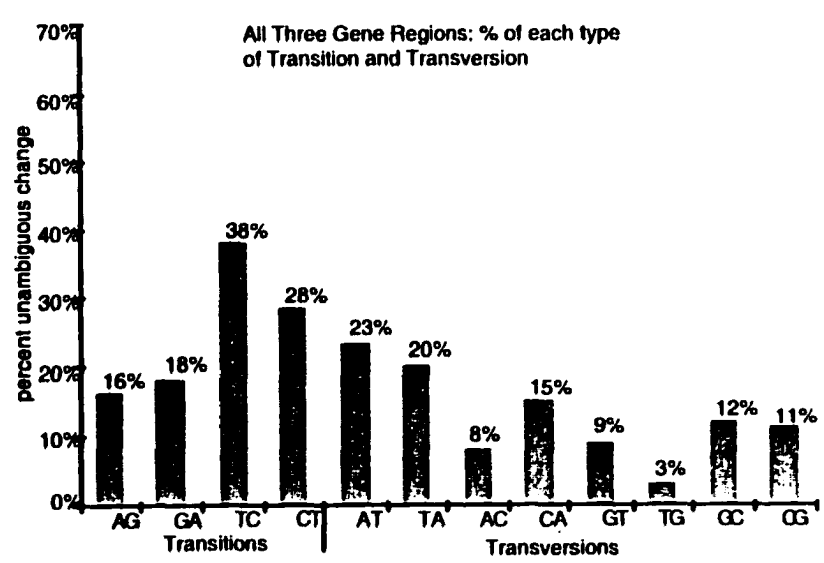
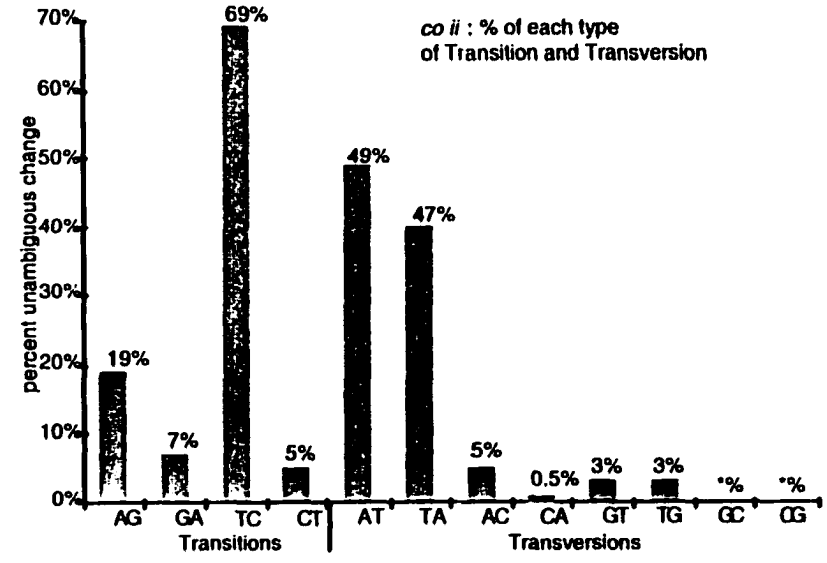
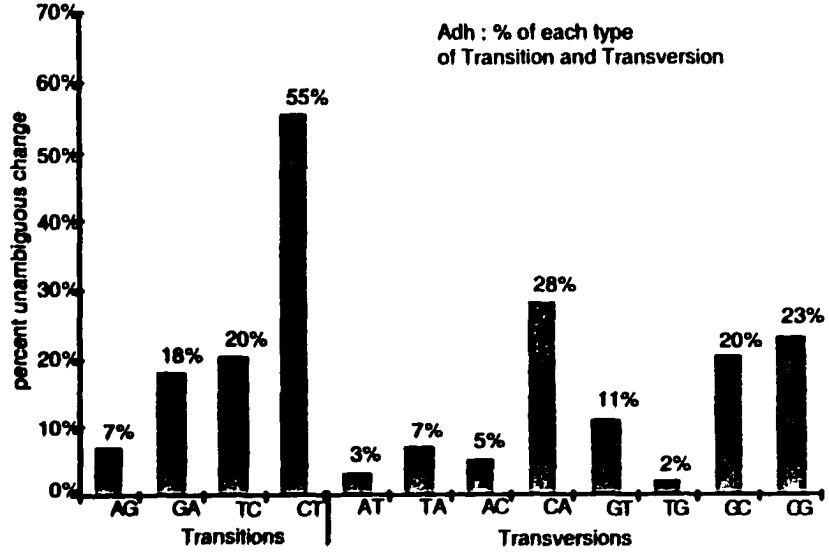


Figure 10. The percent of each type of transition and transversion for the phylogenetically informative characters for the 3 separate gene regions as well as the total evidence data set.







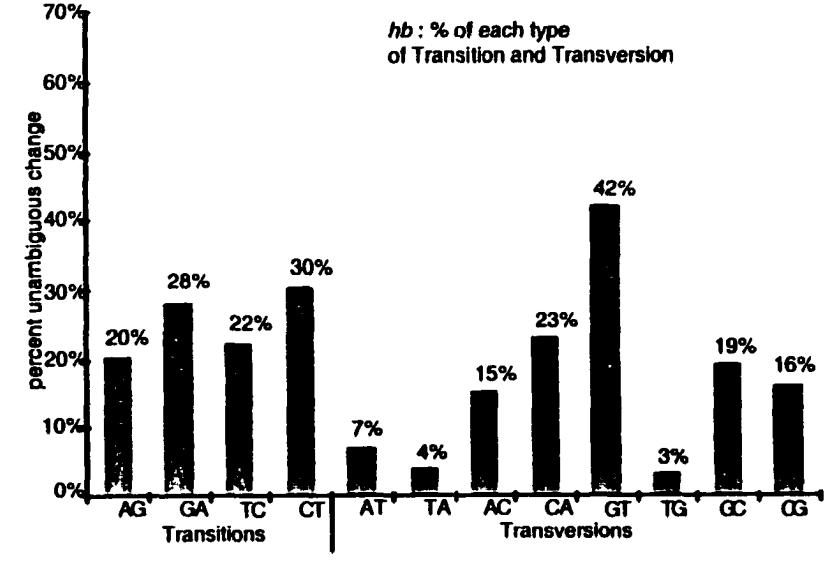


Figure 11. The percent composition for each nucleotide base all characters and phylogenetically informative characters. Comparing the three individual gene regions, the two nuclear genes combined and the total evidence (all 3 gene regions combined).

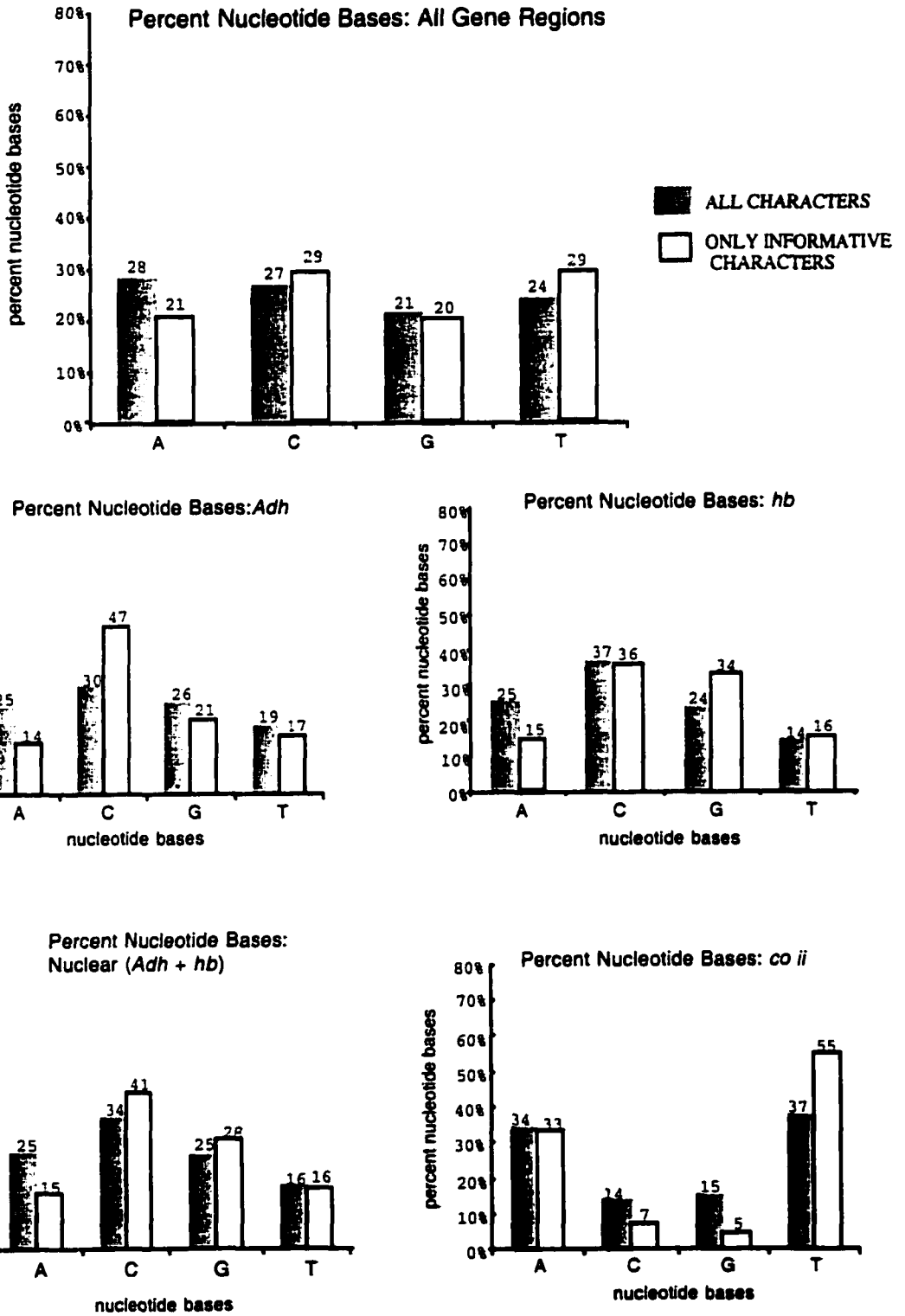


Figure 12. The lengths table presents the sum of the tree lengths for each data partition in comparison to the combined analysis tree length and gives the resulting ILD value. For the total combined data analysis (*Adh* + *co* ii + *hb*) estimated CI and RI were calculated according to the formula presented in the methods section; these estimates were compared to the actual total combined analysis tree's CI and RI. Using MacClade each data partition was placed on the simultaneous analysis tree to determine that data partitions characters contribution to the length, CI and RI within the context of the simultaneous analysis.

Lengths

	sum of parts	combined analysis	ILD*
Adh + co ii	925	948	23
Adh + hb	934	969	35
co ii + hb	1083	1112	29
Adh + co ii + hb	1471	1540	69

* is the amount of added steps (homoplasy)

	CI		RI	
	estimated	combined	estimated	combined
Adh + co ii + hb	0.370	0.349	0.689	0.667

Actual Contribution on the Most Parsimonious Tree

	Length	CI	RI
Adh	414	0.355	0.670
co ii	556	0.286	0.592
hb	570	0.405	0.724
TOTAL	1540	0.349	0.667

Figure 13. Node numbers for the single most parsimonious cladogram resulting from the simultaneous analysis of the 3 gene regions.

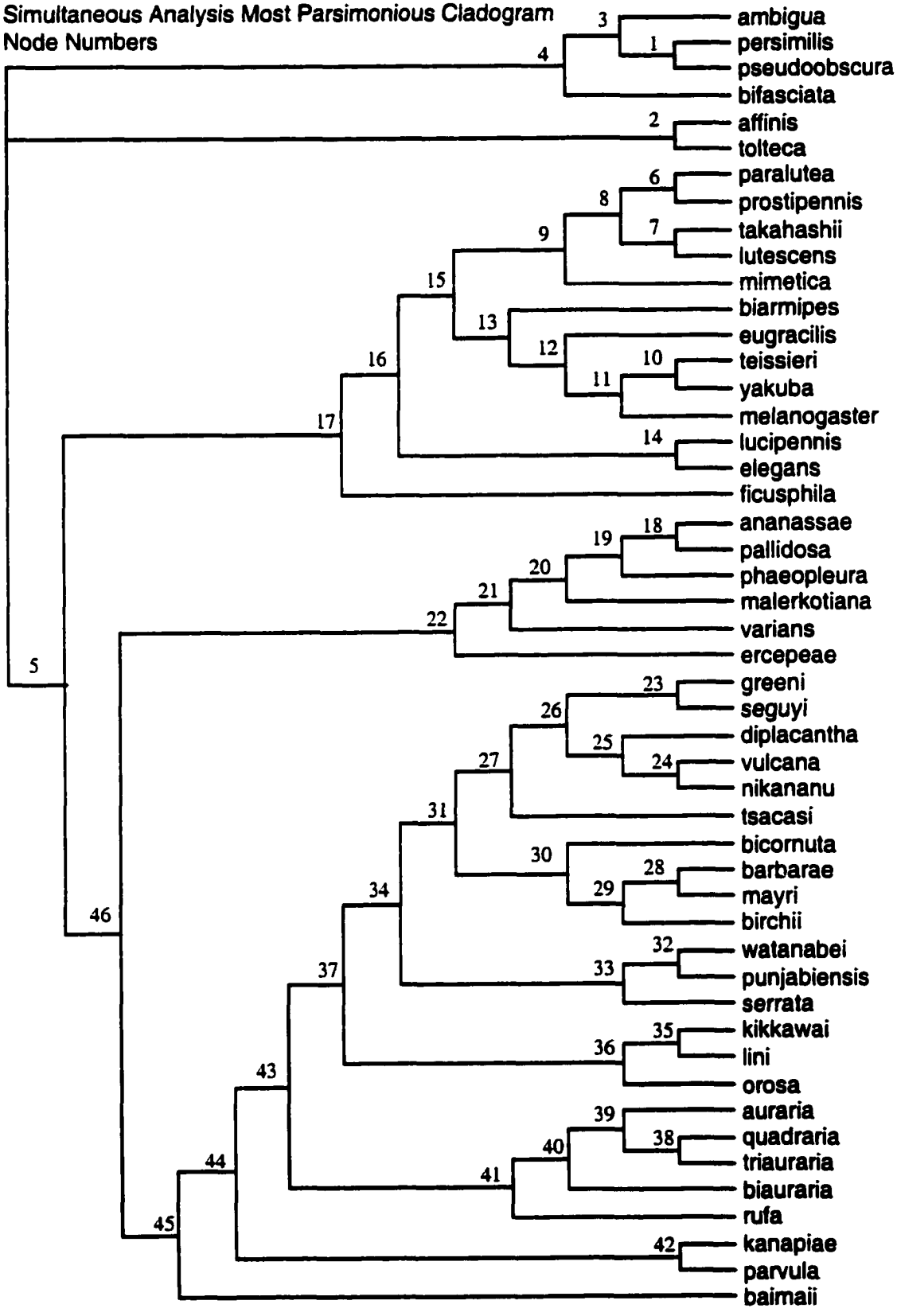


Figure 14. Transitions plotted for the uncorrected p pairwise distance for each separate gene region as well as all 3 gene regions combined. Comparisions include only the phylogenetically informative characters.

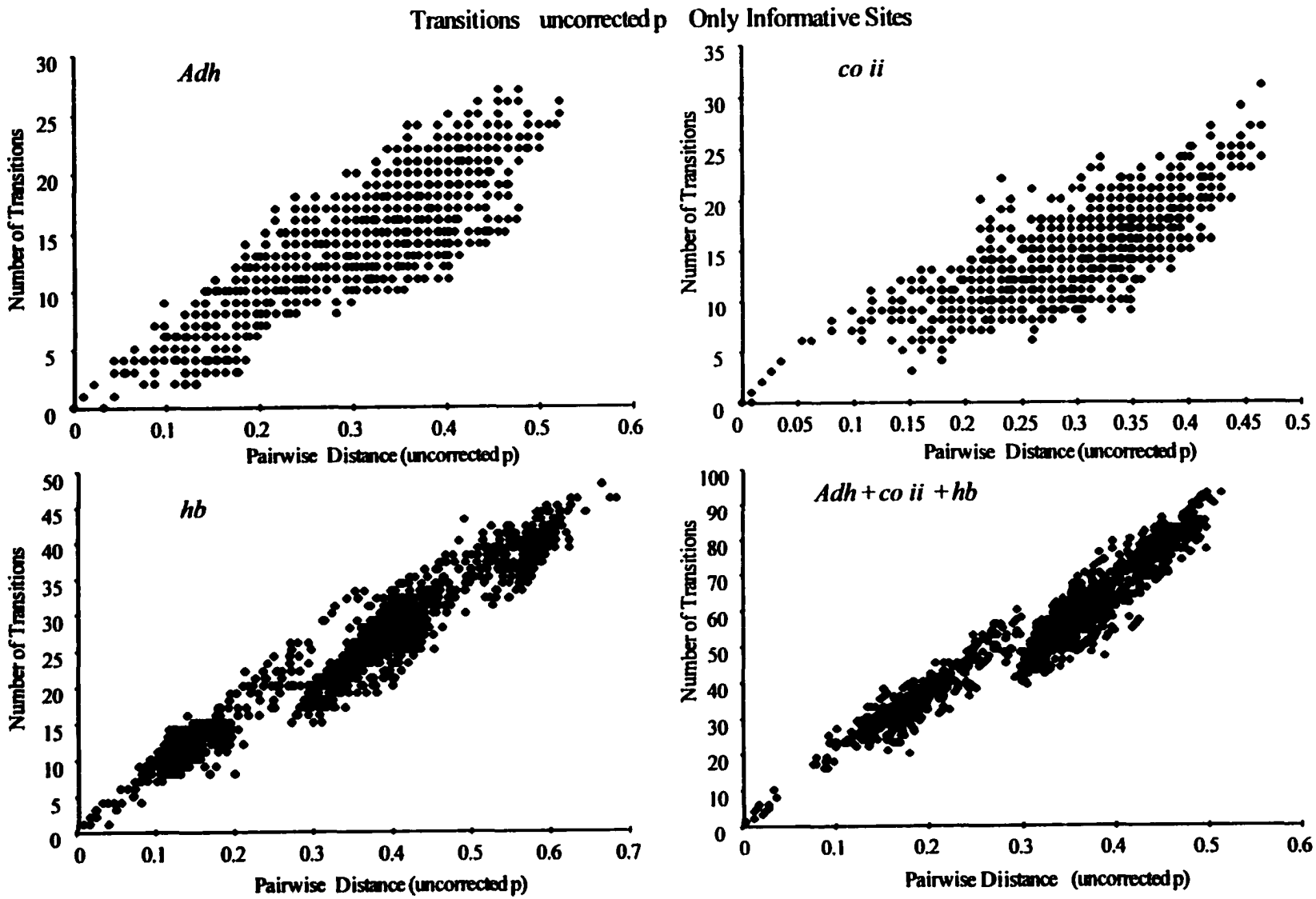


Figure 15. Transversions plotted for the uncorrected p pairwise distance for each separate gene region as well as all 3 gene regions combined. Comparisons include only the phylogenetically informative characters.

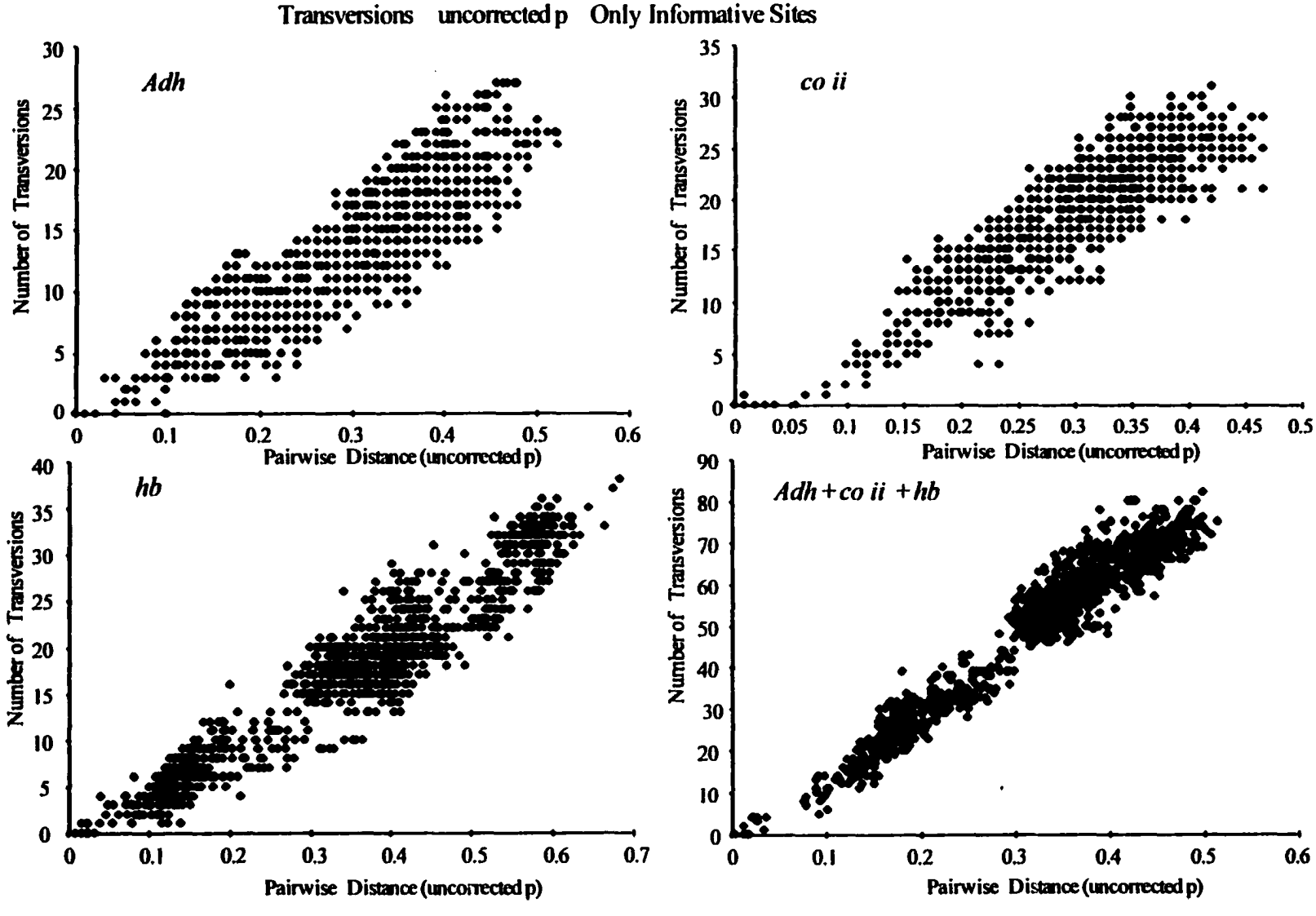
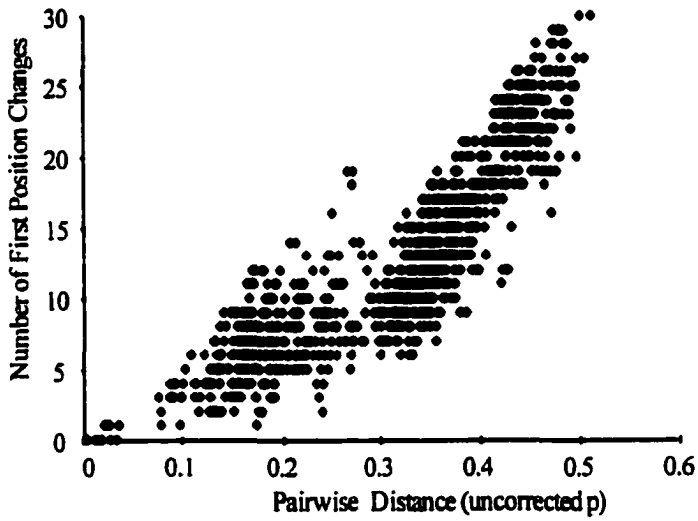


Figure 16. The first second and third codon postions plotted against the uncorrected p pairwise distance for the 3 gene regions combined. Comparisons include only the phylogenetically informative characters.

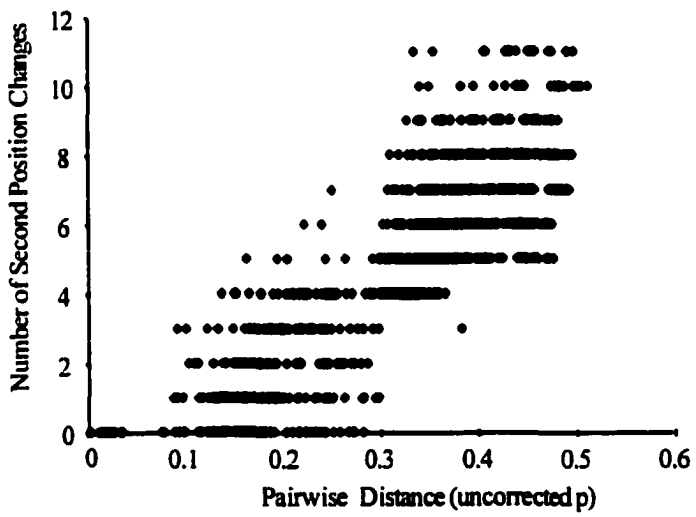
Adh + *co ii* + *hb* Number of Condon Position Changes

Only Informative Sites

1st position



2nd position



3rd position

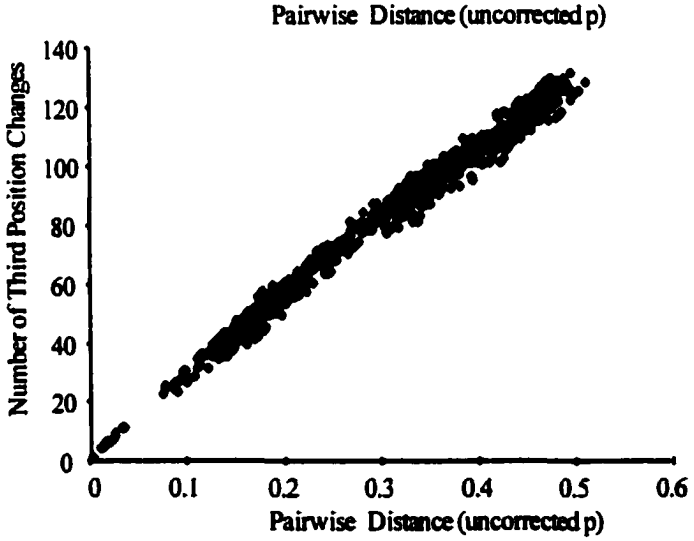


Figure 17. Molecular phylogenetic reconstruction of the *melanogaster* species group. This is the most parsimonious tree resulting from a simultaneous analysis of the gene regions *co ii*, *Adh*, *hb* (L= 1540, CI = .35). Numbers at nodes above the line are the Bremer support values (Bremer, 1998; 1994) and below the line are the bootstrap values (Felsenstein, 1985). Most previously established groups are supported except for *obscura* and *suzukii* species subgroups. The traditional species groups and subgroups are abbreviated as follows: obs. = obscura, psd. = pseudoobscura, aff. = affinis, mel. = melanogaster, tak. = takahashii, suz. = suzukii, eleg. = elegans. eug. = eugracilis, ana. = ananassae, ficus. = ficusphila, and mont. = montium.

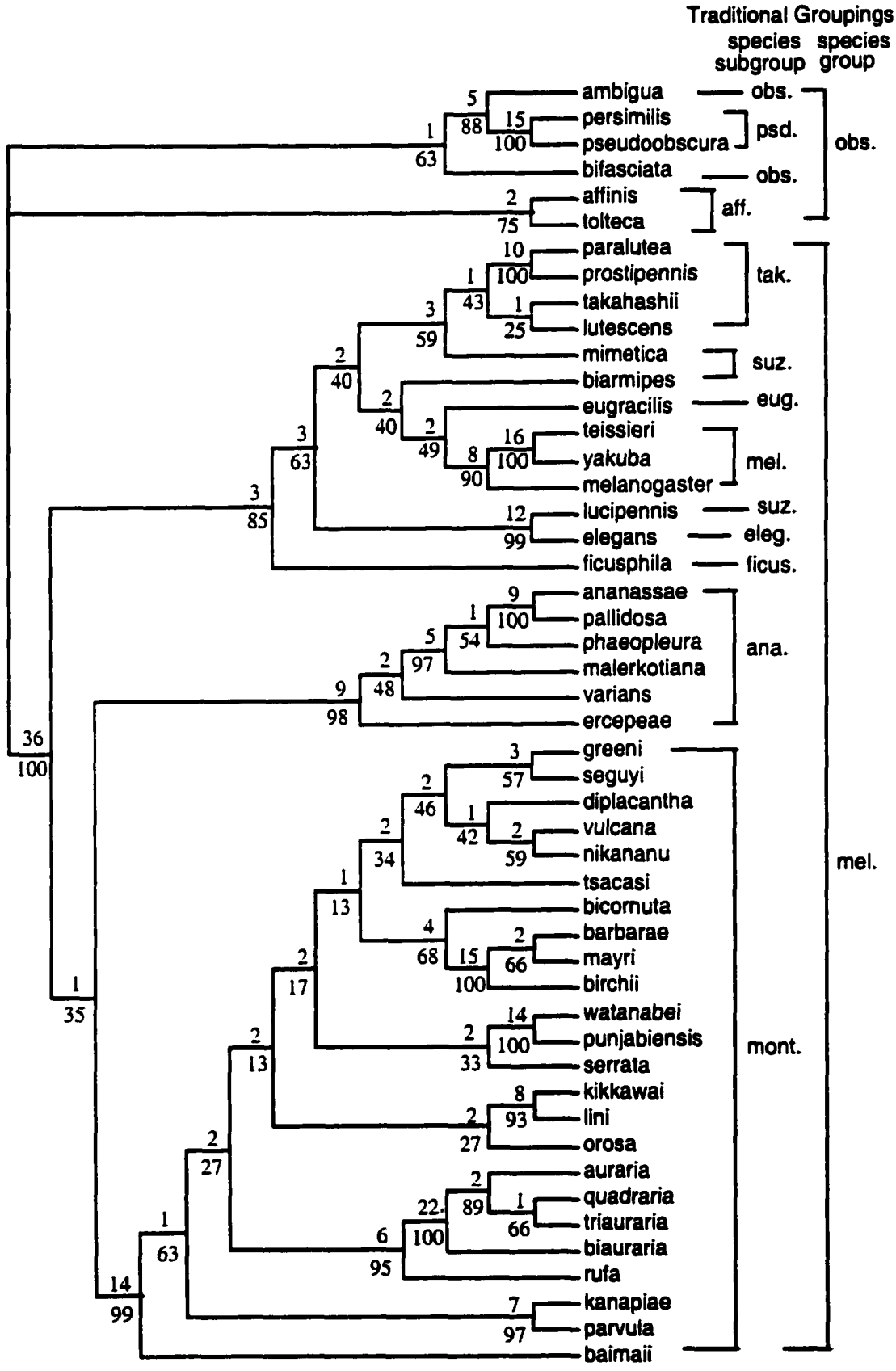
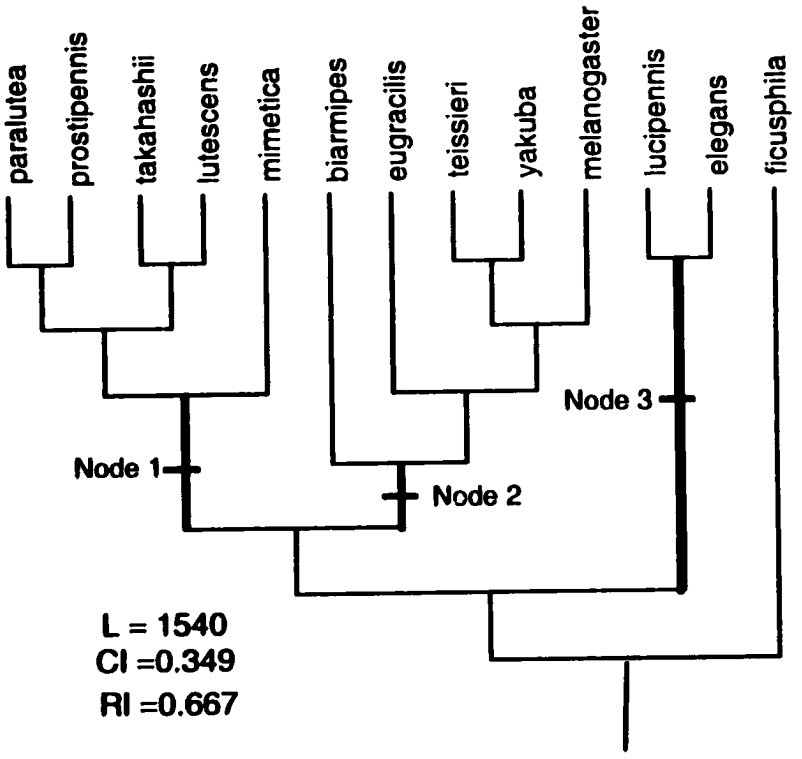


Figure 18. This is a partial view of the most parsimonious cladogram resulting from the simultaneous analysis. This region depicts the *melanogaster* + Oriental subgroups clade. Nodes 1, 2, and 3 support clades containing a *suzukii* subgroup representive. Thorough descriptions of each unambiguous character supporting these three nodes are listed.



Node 1

character	states	steps	ci	ri
52	2	1	1	1
82	3	7	0.286	0.5
148	2	8	0.125	0.563
214	2	9	0.111	0.467
426	2	6	0.167	0.545

Node 2

character	states	steps	ci	ri
40	2	3	0.333	0.889
101	2	3	0.333	0.778
208	4	11	0.273	0.5
366	2	9	0.111	0.385
546	3	10	0.2	0.467
558	3	6	0.333	0.636
770	4	5	0.6	0.5

Node 3

character	states	steps	ci	ri
187	4	6	0.5	0.625
205	4	7	0.429	0.789
259	3	3	0.667	0.833
280	3	8	0.25	0.714
330	3	8	0.25	0.5
436	3	10	0.2	0.652
481	2	4	0.25	0.5
615	3	6	0.333	0.789
794	4	11	0.273	0.429
856	3	4	0.5	0.818
887	2	6	0.167	0.444
908	3	6	0.333	0.714
950	2	3	0.333	0.5
1019	4	12	0.25	0.625
1028	3	9	0.222	0.682
1031	4	4	0.75	0.909
1091	2	1	1	1

Figure 19. Attempts were made to evaluate the cost of creating a monophyletic *suzukii* subgroup. The most parsimonious arrangement for each of the three taxa in each previous location on the cladogram is depicted.

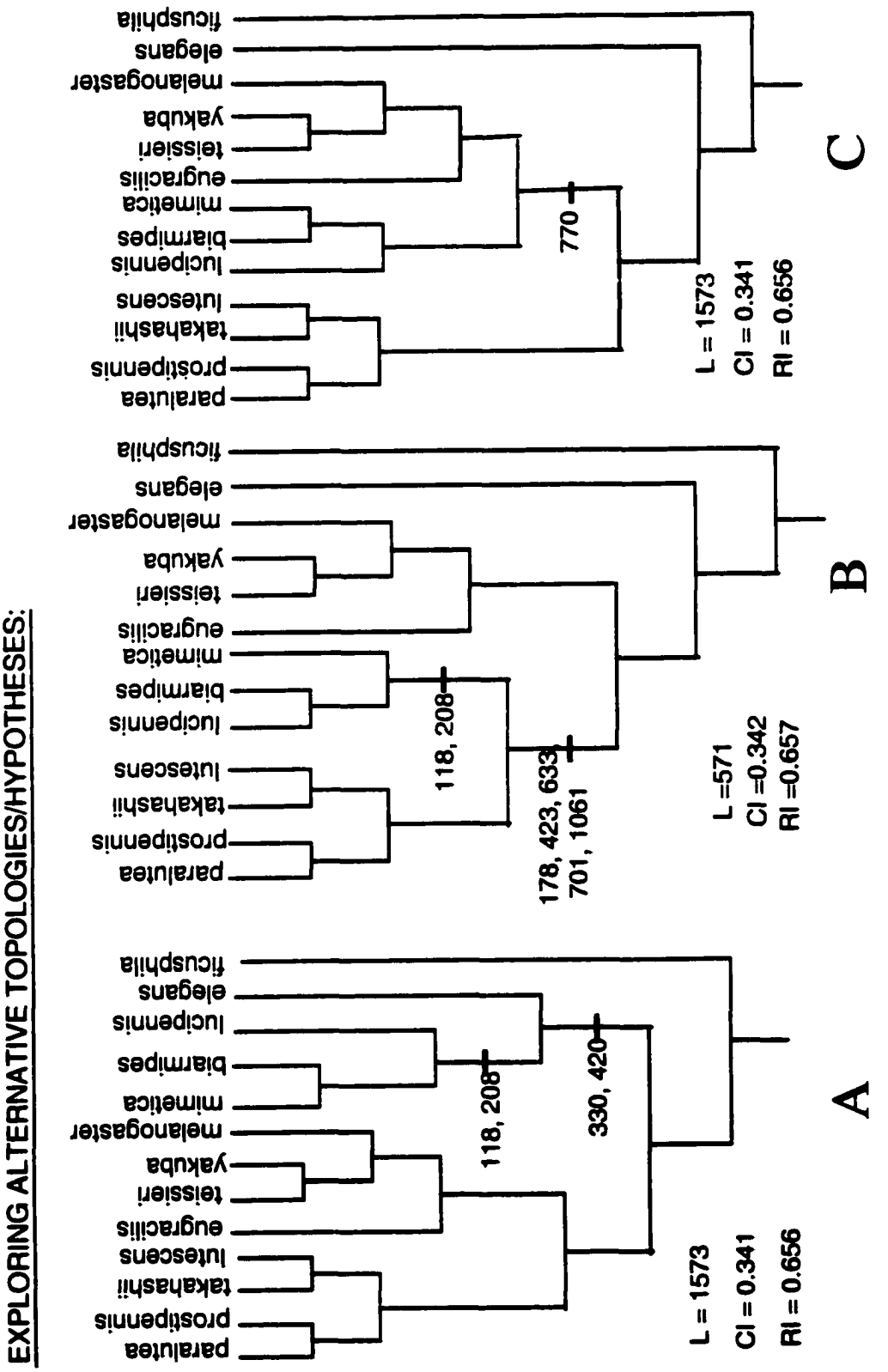
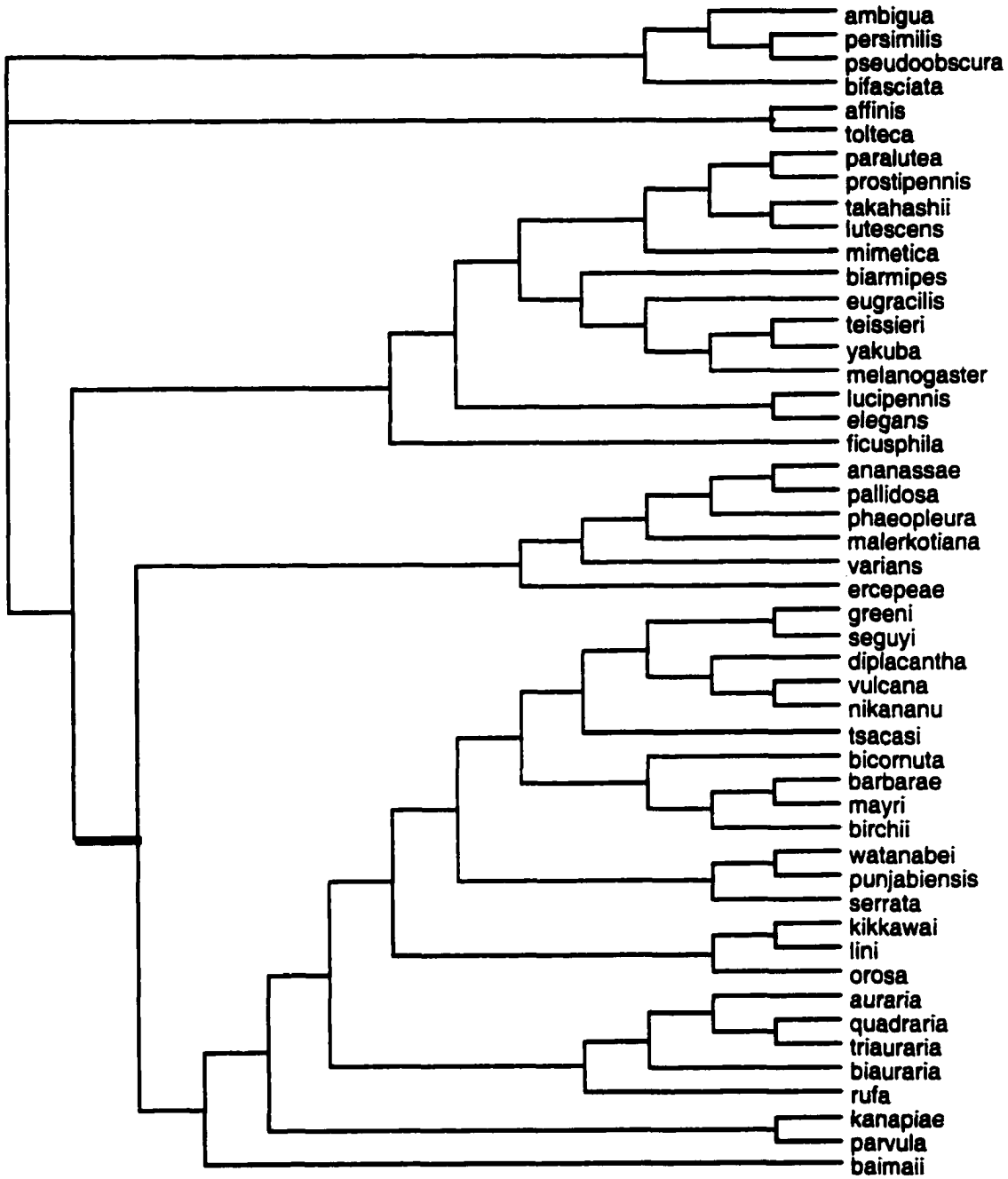


Figure 20. This is the simultaneous analysis most parsimonious cladogram. The node supporting the *ananassae* + *montium* subgroups sister group relationship is thoroughly described for each unambiguous character. (Done using MacClade)



Tree

L = 1540

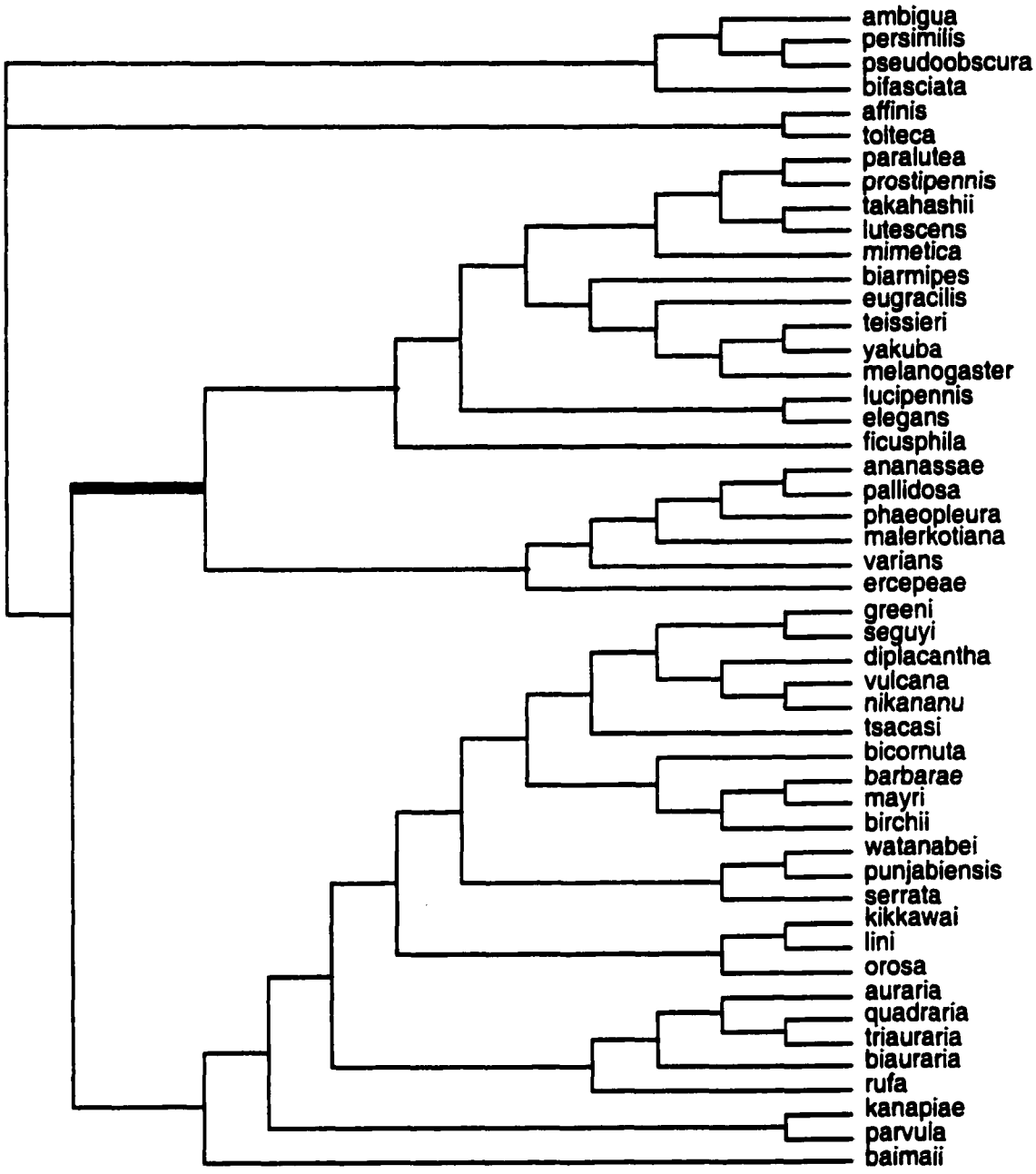
CI = 0.349

RI = 0.667

ananassae + *montium* clade

Character	states	steps	ci	ri
115	3	8	0.25	0.76
121	3	4	0.5	0.9
522	3	12	0.167	0.524
544	2	6	0.167	0.667
612	2	6	0.167	0.762
695	3	2	1	1
696	2	3	0.333	0.714
725	3	12	0.167	0.545

Figure 21. This is the simultaneous analysis most parsimonious cladogram with a change in the basal node creating an *ananassae* + *melanogaster* + Oriental subgroups clade. The node supporting this sister group relationship is thoroughly described for each unambiguous character. Tree statistics are also presented for this topology. (Done using MacClade).

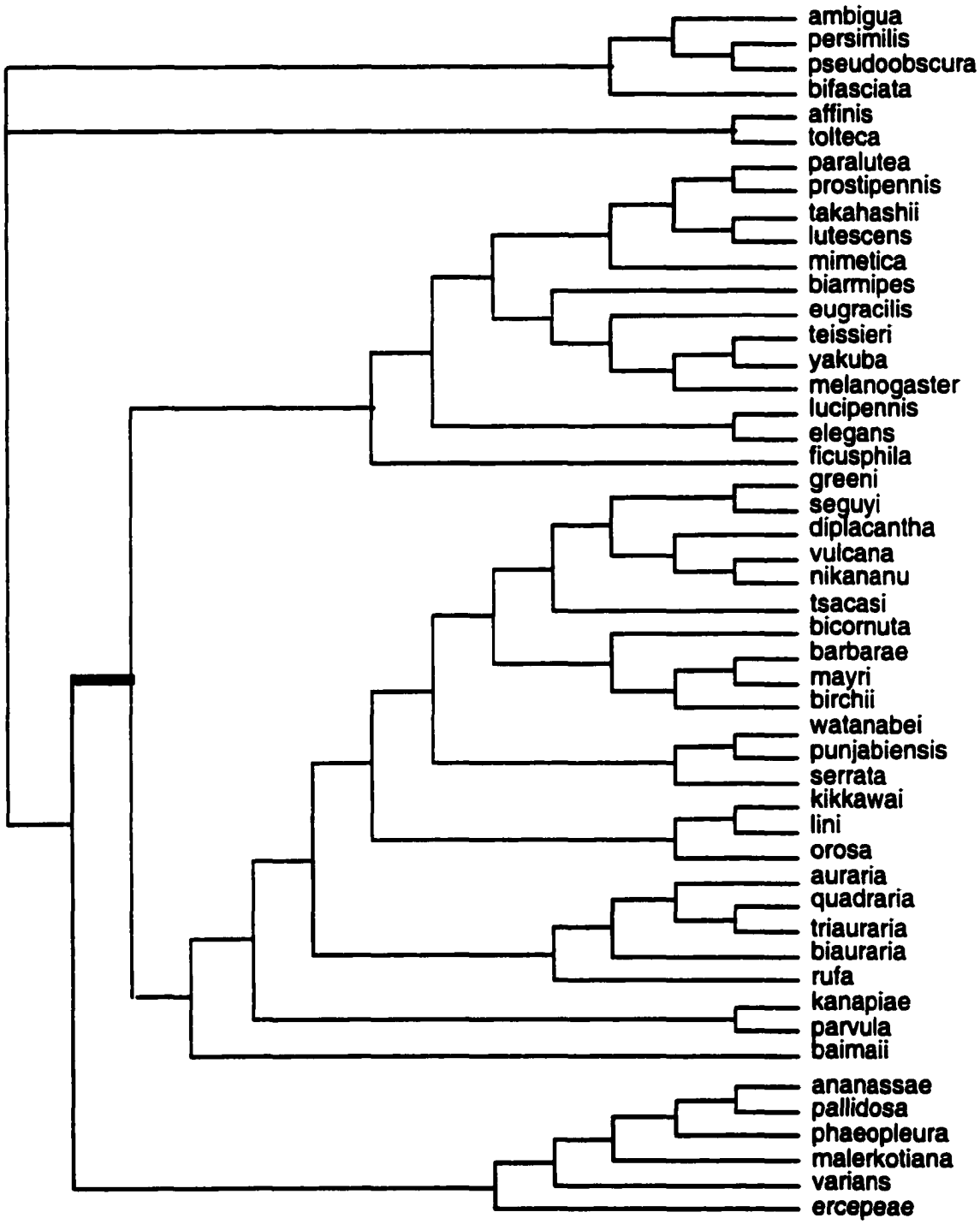


ananassae + *melanogaster* + Oriental Clade

Tree
L = 1541
CI = 0.348
RI = 0.667

Character	states	steps	ci	ri
154	2	2	0.5	0.947
190	4	11	0.273	0.579
244	3	8	0.25	0.727
537	2	4	0.25	0.85
677	2	9	0.111	0.619
701	2	3	0.333	0.833
842	2	8	0.125	0.696

Figure 22. This is the simultaneous analysis most parsimonious cladogram with a change in the basal node creating a *montium* + *melanogaster* + Oriental subgroups clade. The node supporting this sister group relationship is thoroughly described for each unambiguous character. Tree statistics are also presented for this topology. (Done using MacClade).



Tree
L = 1544
CI = 0.348
RI = 0.666

montium + *melanogaster* + Oriental clade

Character	states	steps	ci	ri
447	3	5	0.4	0.625
687	2	2	0.5	0.917
719	2	4	0.25	0.7
1050	2	5	0.2	0.826

Figure 23. The mid-tibiae hooked bristle character was mapped onto the simultaneous analysis most parsimonious cladogram. Resulting tree statistics are also presented.

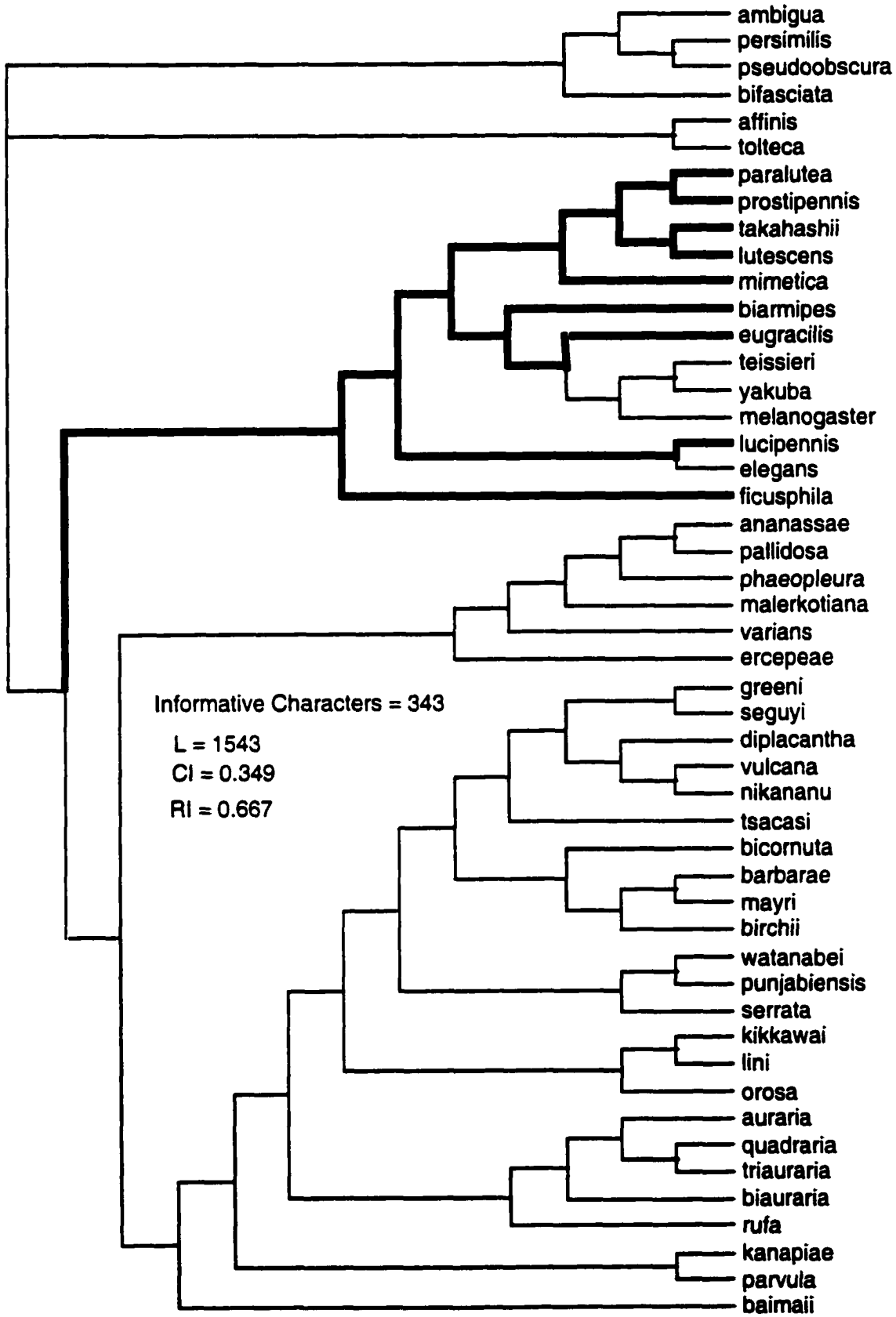


Figure 24. The cumulative number of species described for the *montium* subgroup. The bars depict the number of new species described each year. The line shows the cumulative number of known species to date. In 1972, Bock and Wheeler nearly doubled the size of the *montium* subgroup. Containing 87 species, this is the largest subgroup within the *melanogaster* group with new species currently being described (pers. comm. M. J. Toda).

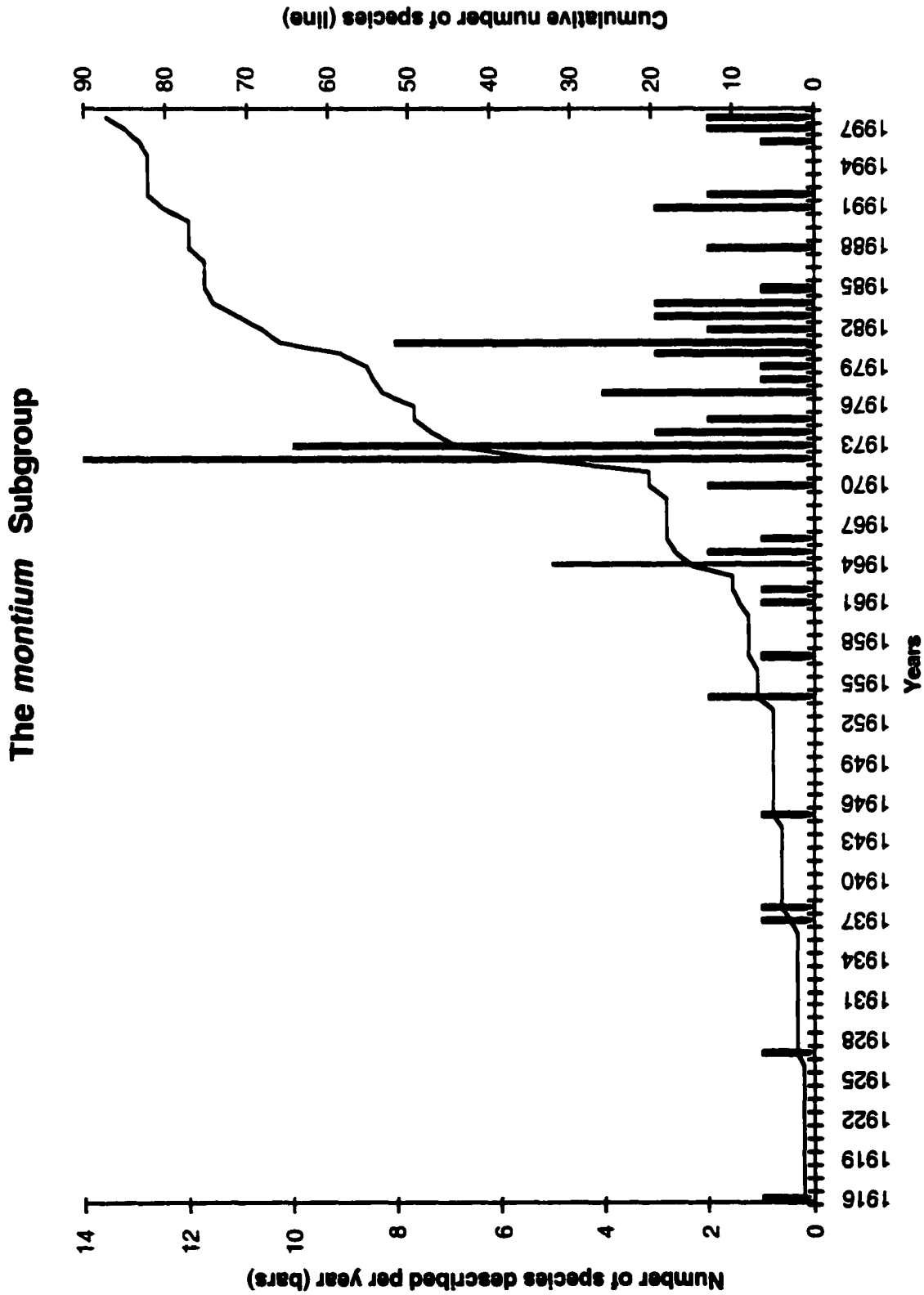


Figure 25. Close up of the *montium* subgroup clade from the reconstruction of the *melanogaster* species group. This is the most parsimonious tree resulting from a simultaneous analysis of the gene regions *Adh*, *co ii*, *hb* (L= 1540, CI = .35). Numbers at the nodes are labels referred to by other figures and in the text. Representatives were sampled from six of the seven species complexes. Only the *auraria* and possibly *jambulina* complexes are coming out as monophyletic. Not all the species sampled have been assigned to a species complex. Lineages A – F are non-official taxonomic designations for the circumscribed clades.

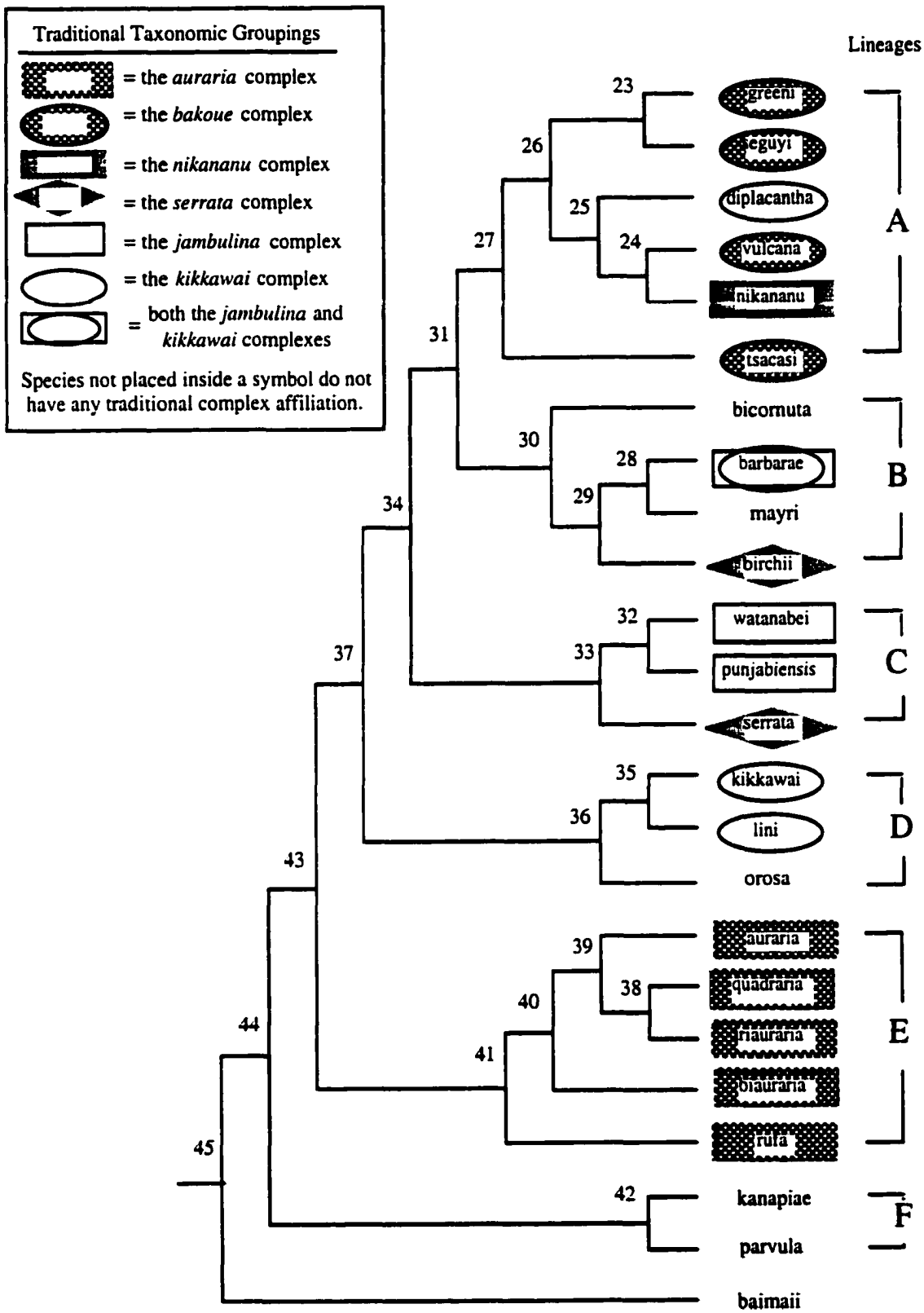
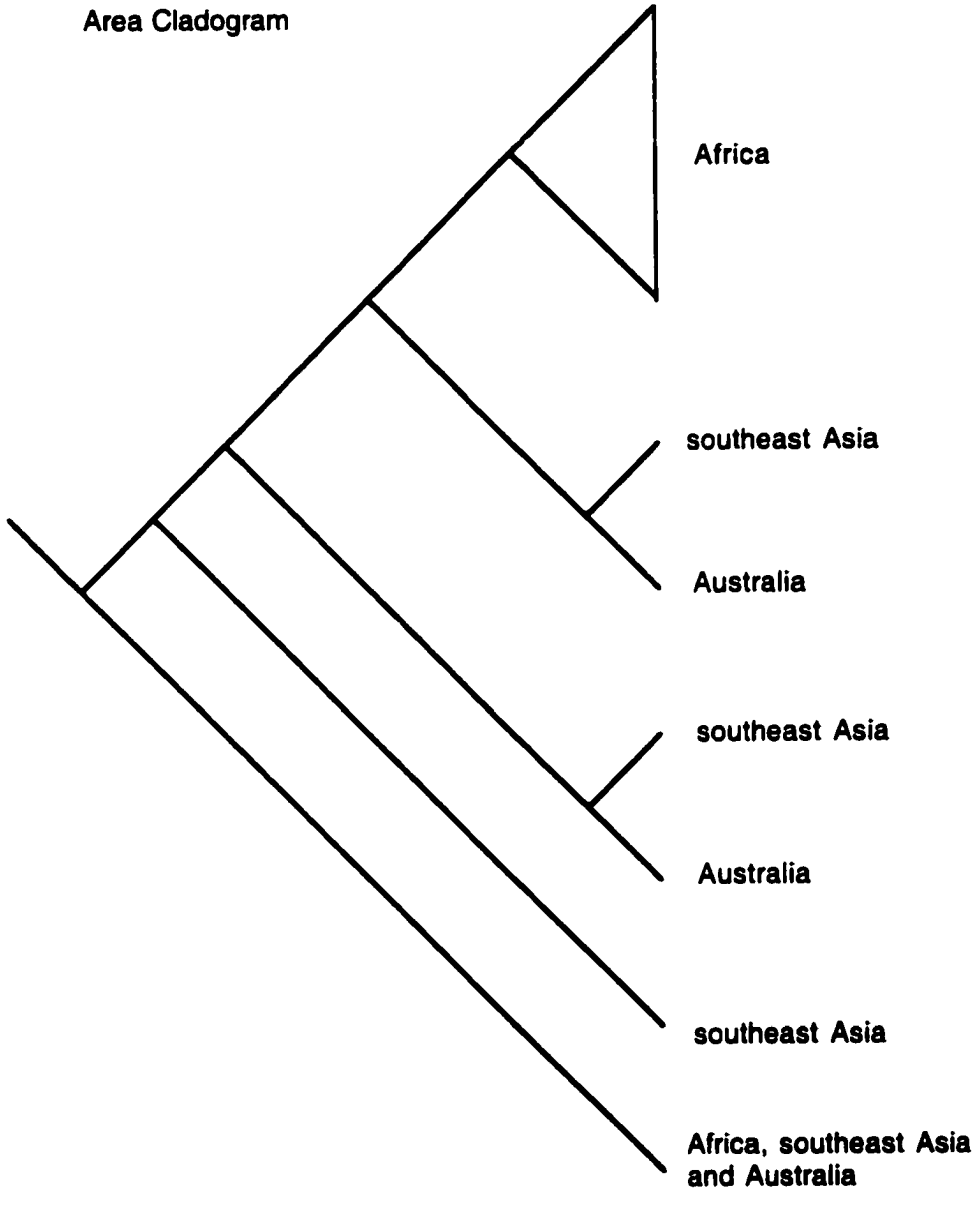


Figure 26. This is an area cladogram for the *montium* subgroup representavies in my study. The outgroup, representatives from the *melanogaster* species group have distributions in Africa, southeast Asia and Australia. The areas correspond to the lineages depicted in figure 5 are as follows: lineages D-F are from southeast Asia (*D. kikkawai* has a widespread distribution), lineage C and D each depict a separate colonization of Australia (for lineage C - *D. serrata* and for lineage B - *D. birchii*), and lineage A is the Africa clade. Distributions for my taxa were from Wheeler, 1981 and 1986; Lemeunier *et al.*, 1986)



APPENDICES

APPENDIX A: Nomenclatural Comments

Table 1 is a current listing of the *melanogaster* group species with their subgroup affiliation. The *nipponica* subgroup has been removed by Okada (1984) and the *dentissima* subgroup has been elevated to species group status by Tsacas (1979; 1980). At present, there are a total of 12 named subgroups and 5 species *incertae sedis* (see revisions of the *melanogaster* species group: Hsu, 1949; Okada, 1954; Bock and Wheeler, 1972; Bock, 1980 and partial revision by Toda, 1991, plus Lemeunier *et al.*, 1986 for listing of the *incertae sedis* species).

Table 2 lists the valid names for all 174 species and 3 subspecies of the *melanogaster* group. The first eight names, however, are species from the *obscura* group pertinent to this dissertation. The flies of the *melanogaster* group are listed in alphabetical order with their synonyms where applicable.

Nota Bene:

** This list is primarily compiled using lists in Wheeler's catalogue (1981; 1986). Other important compilations are Bock and Wheeler (1972) and Bock (1980). Bock and Wheeler (1972); Bock (1980) and Lemeunier *et al.* (1986). (I got the list of unclassified species from Lemeunier *et al.*, 1986.)

†† These references had a publication date discrepancy between Wheeler (1981; 1986) and Lemeunier *et al.* (1986) and Zoological Record on-line catalogue. *D. ercepeae* Tsacas & David 1974/5 publication year differs between Lemeunier *et al.* (1986) and Wheeler (1981) respectively. Tsacas & David 1977/8 was an end of the year issue which I felt needed checking. Publication years for these articles where checked in their original sources and listed here. (See below for details.)

When Bock (1971) designated subspecies for *malerkotliana* and for *pseudoananassae*, he referred to populations containing the originally described species as *malerkotliana malerkotliana* and *pseudoananassae pseudoananassae*, respectively. Hence, other papers may also contain these designations for *malerkotliana* and *pseudoananassae*.

The synonymies listed for *D. melanogaster*, Meigen, 1830 were from Wheeler (1981), however, Bock and Wheeler (1972) and Bock (1980) list more species. It is not clear as to who or when the synonymies were confirmed by examining types.

The species name *D. constricta* was used twice for two species: *D. constricta* Okada & Carson, 1983 and *D. constricta* Chen, Shao & Fan, 1988 (in Chen *et al.*, 1988). Okada and Carson, 1983 have priority on usage of the name according to nomenclatural rules in the Bulletin of Zoological

Nomemclature. From the original species descriptions these two species seem very different. Therefore, Okada and Carson 's 1983 species keeps the name *D. constricta* and a new name must be found for Chen, Shao and Fan's species.

Gupta and Sundaran, (1990) described *D. maggulae* and placed it in the *montium* subgroup; however, they said it is similar to *D. atripex* Bock & Wheeler, 1972 which is a species in the *ananasase* subgroup. According to their illustrations *D. maggulae* looks like a member of the *ananassae* subgroup. Therefore, *D maggulae* has been removed from the *montium* subgroup and listed with the *ananassae* subgroup.

One author is referred by two ways in the literature – G. Seereema Reddy and G. S. Reddy; here only G. S. Reddy is used.

References with publication year discrepancy:

Gupta, J. P. and B. K. Singh. 1978. Two new and two unrecorded Indian species of *Drosophila* (Dipt.) from Kurseong, Darjeeling. Entomologist's Monthly Magazine. 113(1977): 71-78. [Note: On front cover of journal it says "published 18 April, 1978".]

Muniyappa, N. and G. S. Reddy. 1981. *Drosophila madikerii* sp. nov. from Coorg District (Western Ghats) Karnataka, India (Diptera: Drosphilidae). Oriental Insects. 14(1980): 499-502.[Note: Says on bottom of reference "published January 1981" and also in the end of year index gives the actual publication date as: "published January 15, 1981".]

Muniyappa, N. and G. S. Reddy. 1981. Description of a new species *Drosophila gangotrii* (Diptera: Drosophilidae) from south India. Journal of the Bombay Natural History Society. 77(1980): 486-490. [Note: Date of publication April 27, 1981.]

Singh, B. K. and J. P. Gupta. 1979. Two new and two unrecorded species of the genus *Drosophila* Fallen (Diptera: Drosophilidae) from Shillong, Meghalaya, India. Proceedings of the Zoological Society (Calcutta). 30(1977): 31-38. [Note: On back cover it says "Issued December 1979".]

Tsacas, L. 1981. Les groupes d'espèces du sous-genre *Sophophora* Sturtevant (Diptera, Drosophilidae, *Drosophila*) et le rôle du functionnement des génitalia mâles dans la définition des taxons supraspécifiques. Bulletin de la Societe Zoologique de France. 105(1980): 529-543. [Note: Date of publication March 2, 1981.]

Tsacas, L. and M.-T. Chassagnard. 1992. Le complexe *Drosophila nikananu*: description d'une nouvelle espèce africaine et analyse de quelques caractères morphologiques du groupe *melanogaster* (Diptera, Drosophilidae). Nouvelle Revue d'Entomologie (N. S.). 8(1991): 385-398. [Note: On journal back cover it says: "paru le 27 mars 1992" and inside the back page it says "Dépôt légal :mars 1992".]

Tsacas, L. and J. David. 1975: Les Drosophilidae (Diptera) de l'ile de la Reunion et de l'ile Maurice. 1. Deux nouvelles especes du genre *Drosophila*. Bulletin Mensuel de la Societe Linneenne de Lyon. 44: 134-143. The year is 1975. Lemeunier, *et al.*, 1986 had a misprint.

Tsacas, L. and J. David. 1978. Systematics and biogeography of the *Drosophila kikkawai*-complex, with descriptions of new species (Diptera, Drosophilidae). Annales de la Societe Entomologique de France (N. S.). 13(1977): 675-693. [Note: Date of publication February 28, 1978.]

New species since Wheeler 1986 catalogue:

Chassagnard, M.-T. 1991. Une nouvelle espèce afrotropicale du sous-groupe *Drosophila montium* groupe *melanogaster* (Diptera, Drosophilidae). Revue Francaise d'Entomologie (N. S.). 13: 119-122.

Chassagnard, M.-T. and N. Groseille. 1992. *Drosophila* (*Sophophora*) *ochrogaster* n. sp. du sous-groupe *ananassae* de Nouvelle Caledonie (Diptera, Drosophilidae). Revue Francaise d'Entomologie (N. S.). 14: 63-68.

Chassagnard, M.-T.and L. Tsacas and D. Lachaise. 1997. Drosophilidae (Diptera) of Malawi. Annals of the Natal Museum. 38: 61-131.

Chen, H. 1988. A new species of *Drosophila* (*Sophophora*) from China (Diptera: Drosophilidae). Entomotaxonomia. 10: 193-195.

Chen, H.-Z., Z. Shao, Z.-d. Fan and T. Okada. 1988. A new and a newly recorded species of *Drosophila* (*Sophophora*) (Diptera, Drosophilidae) from China. Kontyú. 56: 839-842.

Cheng, H.-Z.; Okada, T. 1985. A new species of *Drosophila* (*Sophophora*) from China (Diptera, Drosophilidae). Kontyú. 53(1): 202-203.

De, A. and J. P. Gupta. 1996. Records of drosophilid species from west Bengal with description of one new and two previously unrecorded species from India (Insecta: Diptera: Drosophilidae). Senckenbergiana biologica 76: 129-133.

Gupta, J. P. and A. K. Sundaran. 1990. Further record of two new and one known species of *Drosophila* (Diptera: Insecta) from Karnataka, India. Proceedings of the Zoological Society (Calcutta). 43: 31-35.

Gupta, J. P. and K. K. Gupta. 1992. *Drosophila watanabei*, a new species of the *montium* subgroup of the *melanogaster* species group of the subgenus *Sophophora* (Diptera: Drosophilidae). Oriental Insects. 26: 201-205.

Kim, B. K. and T. Okada. 1988. A new species of the *Drosophila montium* species subgroup (Diptera, Drosophilidae) from Amami-oshima Island. Proceedings of the Japanese Society of Systematic Zoology. No. 38: 57-61.

Kumar, A. and J. P. Gupta. 1988. Additions to the drosophilid fauna of northeast India (Diptera). Annales de la Societe Entomologique de France (N. S.). 24: 337-342.

Kumar, A. and J. P. Gupta. 1992. Descriptions of three new species of Drosophilidae (Diptera) from India. Oriental Insects. 26: 207-212.

Lemeunier, F., S. Aulard, M. Arienti, J.-M. Jallon, M.-L. Cariou, and L. Tsacas. 1997. The *ercepeae* complex: new cases of insular speciation within the *Drosophila ananassae* species subgroup (*melanogaster* group) and descriptions of two new species (Diptera: Drosophilidae). Annals of the Entomological Society of America. 90: 28-42.

McEvey, S. F., J.-R. David and L. Tsacas. 1987. The *Drosophila ananassae* complex with description of a new species from French Polynesia (Diptera: Drosophilidae). Annales de la Societe Entomologique de France (N. S.). 23: 377-385.

Okada, T. 1988. Family Drosophilidae (Diptera) from the Lund University Ceylon Expedition in 1962 and Borneo Collections in 1978-1979. Entomologica Scandinavica Supplement. No. 30: 111-151.

Singh, B. K. and S. Dash. 1994. Record of further new species from Uttarakhand region, India. Drosophila Information Service. 75: 60-61.

Singh, B. K. and S. Dash. 1998. Drosophilidae of Kumaun Region, India with the description of four new species (Insecta: Diptera). Proceedings of the Zoological Society (Calcutta). 51: 45-56.

Tsacas, L. and M.-T. Chassagnard. 1988. Dipteres Drosophilidae de Nouvelle-Caledonie. 1. *Drosophila*: sous genres *Drosophila* et *Sophophora*. Memoires du Museum National d'Historie Naturelle Serie A Zoologie. 142: 143-154.

Tsacas, L. and M.-T. Chassagnard. 1992. Le complexe *Drosophila nikananu*: description d'une nouvelle espèce africaine et analyse de quelques caractères morphologiques du groupe *melanogaster* (Diptera, Drosophilidae). Nouvelle Revue d'Entomologie (N. S.). 8(1991): 385-398.

Tsacas, L. and M.-T. Chassagnard. 1995. *Drosophila plagiata* Bezzi nouveau membre du groupe *melanogaster* [Diptera, Drosophilidae]. Revue Francaise d'Entomologie (N. S.). 17:19-22.

Tsaur, S. C. and F. J. Lin. 1991. A new *Drosophila* species in the *montium* Meijere subgroup of the *melanogaster* Meigen species-group in the subgenus *Sophophora* Sturtevant from Taiwan (Diptera: Drosophilidae). Pan-Pacific Entomologist. 67: 24-27.

Zhang, W. and Y. Gan. 1986. Descriptions of eight new species of drosophilid flies from Kunming (Diptera: Drosophilidae). Zoological Research. 7: 351-365.

subgroup	species
ananassae	*ananassae* Doleschall,1858
	andamanensis Gupta & Ray-Chaudhuri, 1970
	atripex Bock & Wheeler, 1972
	bipectinata Duda, 1923
	comorensis Tsacas, 1997
	cornixa Takada, Momma & Shima, 1973
	ercepeae Tsacas & David, 1975
	ironensis Bock & Parsons, 1978
	lachaisei Tsacas, 1984
	maggulae Gupta & Sundaran, 1990
	malerkotliana Parshad & Paika, 1964
	malerkotliana pallens Bock & Wheeler, 1972
	merina Tsacas, 1997
	micropectinata Takada & Momma, 1975
	monieri McEvey & Tsacas, 1987
	nesoetes Bock & Wheeler, 1972
	ochrogaster Chassagnard,1992
	pallidosa Bock & Wheeler, 1972
	parabipectinata Bock, 1971
	pereirai Takada, Momma & Shima, 1973
	phaeopleura Bock & Wheeler, 1972
	pseudoananassae Bock, 1971
	pseudoananassae nigrens Bock & Wheeler, 1972
	vallismaia Tsacas, 1984
	varians Bock & Wheeler, 1972
denticulata	*denticulata* Bock & Wheeler, 1972
	microdenticulata Panigrahy & Gupta, 1983
	pseudodenticulata Takada & Momma, 1975
elegans	*elegans* Bock & Wheeler, 1972
	neoelegans Gupta & Singh, 1978
	sahyadrii Prakash & Reddy, 1979
	subelegans Okada, 1988
eugracilis	*eugracilis* Bock & Wheeler, 1972
ficusphila	*ficusphila* Kikkawa & Peng, 1938
	flavicauda Toda, 1991
	gorokaensis Okada & Carson, 1982
	kanaka Tsacas, 1988
	levii Tsacas, 1988
	smithersi Bock, 1976
flavohirta	*flavohirta* Malloch, 1924
longissima	*longissima* Okada & Carson, 1983
	myamaungi Toda, 1991
melanogaster	*erecta* Tsacas & Lachaise, 1974
	mauritiana Tsacas & David, 1974

subgroup	species
melanogaster	*melanogaster* Meigan, 1830
	orena Tsacas & David, 1978
	sechellia Tsacas & Bächli, 1981
	simulans Sturtevant, 1919
	teissieri Tsacas, 1971
	yakuba Burla, 1954
montium	*agumbensis* Prakash & Reddy, 1978
	anomelani Reddy & Krishnamurthy, 1973
	artecarina Takada & Momma, 1975
	asahinai Okada, 1964
	auraria Peng,1937
	austroheptica Tsaur & Lin, 1991
	baimaii Bock & Wheeler, 1972
	bakoue Tsacas & Lachaise, 1974
	barbarae Bock & Wheeler, 1972
	bhagamandalensis Muniyappa, Reddy & Krishnamurthy, 1981
	biauraria Bock & Wheeler, 1972
	bicornuta Bock & Wheeler, 1972
	birchii Dobzhansky & Mather, 1961
	bocki Baimai, 1979
	bocqueti Tsacas & Lachaise, 1974
	brahmagiriensis Muniyappa, Reddy & Krishnamurthy, 1981
	brevina Wheeler, 1981
	burlai Tsacas & Lachaise, 1974
	cauverii Muniyappa & Prakash, 1982
	chauvacae Tsacas, 1984
	constricta Chen, Shao & Fan, 1988
	cryptica De & Gupta, 1996
	curta Chassagnard & Tsacas, 1997
	davidi Tsacas,1975
	dictena Tsacas & Chassagnard, 1992
	diplacantha Tsacas & David, 1978
	dominicana Ayala, 1965
	dossoui Chassagnard, 1991
	eupyga Tsacas, 1981
	exiguitata Takada, Momma & Shima, 1973
	flavopleuralis Takada, Momma & Shima, 1973
	gangotrii Muniyappa & Reddy, 1981
	greeni Bock & Wheeler, 1972
	gundensis Prakash & Reddy, 1977
	ifestia Tsacas, 1984
	jambulina Parshad & Paika, 1964
	kanapiae Bock & Wheeler, 1972
	khaoyana Bock & Wheeler, 1972
	kikkawai Burla, 1954

subgroup	species
montium	*kinabaluana* Takada, Momma & Shima, 1973
	lacteicornis Okada, 1965
	leontia Tsacas & David, 1978
	lini Bock & Wheeler, 1972
	longipectinata Takada, Momma & Shima, 1973
	madikerii Muniyappa & Reddy, 1981
	malagassya Tsacas & Rafael, 1982
	mayri Mather & Dobzhansky, 1962
	megapyga Tsacas, 1981
	montium de Meijere, 1916
	mysorensis Reddy & Krishnamurthy, 1970
	nagarholensis Prakash & Reddy, 1980
	neobaimai Singh & Dash, 1998
	neokhaoyana Singh & Dash, 1998
	neotrapezifrons Ranganath, Krishnamurthy & Hedge, 1983
	nigrialata Takada, Momma & Shima, 1973
	nigropleuralis Takada, Momma & Shima, 1973
	nikananu Burla, 1954
	orosa Bock & Wheeler, 1972
	paraviaristata Takada, Momma & Shima, 1973
	parvula Bock & Wheeler, 1972
	pectinifera Wheeler & Takada, 1964
	penicillipennis Takada, Momma & Shima, 1973
	pennae Bock & Wheeler, 1972
	phyale Tsacas, 1981
	pseudobaimaii Takada, Momma & Shima, 1973
	pseudomayri Baimai, 1970
	punjabiensis Parshad & Paika, 1964
	quadraria Bock & Wheeler, 1972
	rhombura Okada & Carson, 1983
	rufa Kikkawa & Peng, 1938
	sampagensis Muniyappa & Reddy, 1980
	seguyi Smart, 1945
	seguyiana Chassagnard & Tsacas, 1997
	serrata Malloch, 1927
	serrula Tsacas, 1984
	subauraria Kimura, 1983
	suborosa Kumar & Gupta, 1992
	tani Cheng & Okada, 1985
	trapezifrons Okada, 1966
	triauraria Bock & Wheeler, 1972
	tricombata Singh & Gupta, 1979
	truncata Okada, 1964
	tsacasi Bock & Wheeler, 1972

subgroup	species
montium	*vulcana* Graber, 1957
	watanabei Gupta & Gupta, 1992
	xanthia Tsacas, 1981
	yuwanensis Kim & Okada, 1988
rhopaloa	*fuyamai* Toda, 1991
	kurseongensis Gupta & Singh, 1978
	palmata Takada, Momma & Shima, 1973
	prolongata Singh & Gupta, 1979
	rhopaloa Bock & Wheeler, 1972
suzukii	*apicespinata* Zhang & Gan, 1986
	apodemata Okada & Carson, 1983
	ashburneri Tsacas, 1984
	biarmipes Malloch, 1924
	hypomelana Okada & Carson, 1983
	immacularis Okada, 1966
	lucipennis Lin, 1972
	mimetica Bock & Wheeler, 1972
	nyinyii Toda, 1991
	oshimai Choo & Nakamura, 1973
	plagiata Bezzi, 1908
	pulchrella Tan, Hsu & Sheng, 1949
	siangensis Kumar & Gupta, 1988
	suzukii (Matsumura), 1931
	suzukii indica Parshad & Paika, 1964
	tristipennis Duda, 1924
	unipectinata Duda, 1924
takahashii	*giriensis* Prakash & Reddy, 1977
	jagri Prakash & Reddy, 1979
	liui Chen, 1988
	lutescens Okada, 1975
	nepalensis Okada, 1955
	paralutea Bock & Wheeler, 1972
	prostipennis Lin, 1972
	pseudotakahashii Mather,1957
	pyo Toda, 1991
	retnasabapathyi Takada & Momma, 1975
	takahashii Sturtevant, 1927
	tanorum Okada, 1964
	trilutea Bock & Wheeler, 1972
unclassified	*apectinata* Duda, 1931
	brunettii Ray-Chaudhuri & Mukherjee, 1941
	constricta Okada & Carson, 1983
	illata Walker, 1860
	prashadi Brunetti, 1923

VALID SPECIES NAMES	SYNONYMS & COMMENTS	SOURCE(who did it)
ambigua Pomini, 1940		
persimilis Dobzhansky & Epling, 1944		
pseudoobscura Frolova, 1929		Frolova & Astaurov, 1929
pseudoobscura bogatana Ayala & Dobzhansky, 1974		
affinis Sturtevant, 1916		
affinis iroquois Sturtevant & Dobzhansky, 1936		
bifasciata Pomini,1940 also as bilineata error?		
tolteca Patterson & Mainland, 1944		
ananassae Doleschall,1858		
	caribea Sturtevant, 1916	
	errans Malloch, 1933	
	imparta Walker, 1859	
	similis Lamb, 1914 (preocc.)	
agumbensis Prakash & Reddy, 1978		
andamanensis Gupta & Ray-Chaudhuri, 1970	*andamanensis* Parshad & Singh, 1971	
anomelani Reddy & Krishnamurthy, 1973		
apicespinata Zhang & Gan, 1986		
apectinata Duda, 1931		
apodemata Okada & Carson, 1983		
artecarina Takada & Momma, 1975		
asahinai Okada, 1964		
ashburneri Tsacas, 1984		
atripex Bock & Wheeler, 1972		
auraria Peng,1937		
austroheptica Tsaur & Lin, 1991		
baimaii Bock & Wheeler, 1972		
bakoue Tsacas & Lachaise, 1974		
barbarae Bock & Wheeler, 1972		
bhagamandalensis Muniyappa, Reddy, Krishnamurthy, 1981		
biarmipes Malloch, 1924	*rajasekari* Reddy & Krishnamurthy, 1968	Bock, 1980
	raychaudhurii Gupta, 1969	Bock, 1980
biauraria Bock & Wheeler, 1972		
bicornuta Bock & Wheeler, 1972		

VALID SPECIES NAMES	SYNONYMS & COMMENTS	SOURCE (who did it)
bipectinata Duda, 1923	*szentivanii* Mather & Dobzhansky, 1962	
birchii Dobzhansky & Mather, 1961	was a ssp. of *D. serrata* elevated by Ayala, 1965	[cf. Bock & Wheeler, 1972]
bocki Baimai, 1979		
bocqueti Tsacas & Lachaise, 1974		
brahmagiriensis Muniyappa, Reddy, Krishnamurthy, 1981		
brevina Wheeler, 1981	*brevis* Parshad & Singh, 1971 (preocc.)	Wheeler, 1981
brunettii Ray-Chaudhuri & Mukherjee, 1941		
burlai Tsacas & Lachaise, 1974		
cauverii Muniyappa & Prakash, 1982		
chauvacae Tsacas, 1984		
comorensis Tsacas, 1997		Lemeunier *et al.*,1997
constricta Chen, Shao, Fan, 1988 **		Chen *et al.*,1988
constricta Okada & Carson, 1983 **		
cornixa Takada, Momma & Shima, 1973		
cryptica De & Gupta, 1996		
curta Chassagnard & Tsacas, 1997		Chassagnard *et al.*,1997
davidi Tsacas, 1975		
denticulata Bock & Wheeler, 1972		
dictena Tsacas & Chassagnard, 1992 ††		
diplacantha Tsacas & David, 1978 ††		
dominicana Ayala, 1965		
dossoui Chassagnard, 1991		
elegans Bock & Wheeler, 1972		
ercepeae Tsacas & David, 1975 ††		
erecta Tsacas & Lachaise, 1974		
eugracilis Bock & Wheeler, 1972	*gracilis* (Duda), 1924 (as *Tanygastrella;* preocc.)	Bock & Wheeler, 1972
eupyga Tsacas, 1981		
exiguitata Takada, Momma & Shima, 1973		
ficusphila Kikkawa & Peng, 1938		
flavicauda Toda, 1991		
flavohirta Malloch, 1924		
flavopleuralis Takada, Momma & Shima, 1973		

VALID SPECIES NAMES	SYNONYMS & COMMENTS	SOURCE (who did it)
fuyamai Toda, 1991		
gangotrii Muniyappa & Reddy, 1981 ††		
giriensis Prakash & Reddy, 1977		
gorokaensis Okada & Carson, 1982		
greeni Bock & Wheeler, 1972		
gundensis Prakash & Reddy, 1977		
hypomelana Okada & Carson, 1983		
ifestia Tsacas, 1984		
illata Walker, 1860		
immacularis Okada, 1966		
ironensis Bock & Parsons, 1978		
jagri Prakash & Reddy, 1979		
jambulina Parshad & Paika, 1964		
kanaka Tsacas, 1988		Tsacas & Chassagnard, 1988
kanapiae Bock & Wheeler, 1972		
khaoyana Bock & Wheeler, 1972		
kikkawai Burla, 1954	*atropyga* Duda, 1924 as *montium* var. *xanthopyga* Duda, 1924 as *montium* var.	
kinabaluana Takada, Momma & Shima, 1973		
kurseongensis Gupta & Singh, 1978 ††		
lachaisei Tsacas, 1984		
lacteicornis Okada, 1965		
leontia Tsacas & David, 1978 ††		
levii Tsacas, 1988		Tsacas & Chassagnard, 1988
lini Bock & Wheeler, 1972		
liui Chen, 1988		
longipectinata Takada, Momma & Shima, 1973		
longissima Okada & Carson, 1983		
lucipennis Lin, 1972		Bock & Wheeler, 1972
lutescens Okada, 1975	*lutea* Kikkawa & Peng, 1938 (preocc.) *luteola* Okada, 1974 (preocc.)	[cf. Bock, 1980 explains better than Wheeler, 1981]
malerkotliana Parshad & Paika, 1964		Bock, 1971 refers to as *m.m.*
malerkotliana pallens Bock & Wheeler, 1972	*m. pallida* Bock, 1971	Bock & Wheeler, 1972

VALID SPECIES NAMES	SYNONYMS & COMMENTS	SOURCE (who did it)
madikerii Muniyappa & Reddy, 1981 ††		
maggulae Gupta & Sundaran, 1990	originally *montium* subgr., moved to *ananassae* subgr.	This dissertation. Schawaroch, 2000
malagassya Tsacas & Rafael, 1982		
mauritiana Tsacas & David, 1974		
mayri Mather & Dobzhansky, 1962		
megapyga Tsacas, 1981		
melanogaster Meigen, 1830	*ampelophila* Loew, 1862	[Wheeler, 1981
	approximata Zetterstedt, 1847	possibly others see Bock, 1980
	emulata Ray-Chaudhuri & Mukherjee, 1941	and Bock & Wheeler, 1972]
	fasciata Meigen, 1830	
	nigriventris Macquart, 1843	
merina Tsacas, 1997	prev. described as *"ercepeae* -like" (Da Lage *et al.*,1989)	Lemeunier *et al.*,1997
microdenticulata Panigrahy & Gupta, 1983		
micropectinata Takada & Momma, 1975		
mimetica Bock & Wheeler, 1972		
monieri McEvey & Tsacas, 1987		McEvey *et al.*, 1987
montium de Meijere, 1916		
myamaungi Toda, 1991		
mysorensis Reddy & Krishnamurthy, 1970		
nagarholensis Prakash & Reddy, 1980		
neobaimai Singh & Dash, 1998		
neoelegans Gupta & Singh, 1978 ††		
neokhaoyana Singh & Dash, 1998		
neotrapezifrons Ranganath, Krishnamurthy & Hedge, 1983		
nepalensis Okada, 1955		
nesoetes Bock & Wheeler, 1972		
nigrialata Takada, Momma & Shima, 1973		
nigropleuralis Takada, Momma & Shima, 1973		
nikananu Burla, 1954		
nyinyii Toda, 1991	*lucipennis* Gupta & Singh, 1977 (not Lin, 1972). misdet.	Toda, 1991
ochrogaster Chassagnard, 1992		Chassagnard & Groseille, 1992
opisthomelaina Nolte & Stoch, 1950		
orena Tsacas & David, 1978		

VALID SPECIES NAMES	SYNONYMS & COMMENTS	SOURCE (who did it)
orosa Bock & Wheeler, 1972		
oshimai Choo & Nakamura, 1973		
pallidosa Bock & Wheeler, 1972		
palmata Takada, Momma & Shima, 1973		
parabipectinata Bock, 1971		
paralutea Bock & Wheeler, 1972		
paraviaristata Takada, Momma & Shima, 1973		
parvula Bock & Wheeler, 1972		
pectinifera Wheeler & Takada, 1964		
penicillipennis Takada, Momma & Shima, 1973		
pennae Bock & Wheeler, 1972		
pereirai Takada, Momma & Shima, 1973		
phaeopleura Bock & Wheeler, 1972		
phyale Tsacas, 1981		
plagiata Bezzi, 1908	non-classified subgenus to *Sophophora* subgen., *suzukii* subgroup	Tsacas & Chassagnard 1995
prashadi Brunetti, 1923		
prolongata Singh & Gupta, 1979 ††	originally *suzukii* subgr., moved to *rhopaloa* subgroup	Toda, 1991
prostipennis Lin, 1972		Bock & Wheeler, 1972
pseudoananassae Bock, 1971		[cf. Bock, 1971 refer to *p. pseudoan.*]
pseudoananassae nigrens Bock & Wheeler, 1972	*p. nigra* Bock, 1971 (preocc.)	Bock & Wheeler, 1972
pseudobaimaii Takada, Momma & Shima, 1973		
pseudodenticulata Takada & Momma, 1975		
pseudomayri Baimai, 1970		
pseudotakahashii Mather,1957		
pulchrella Tan, Hsu & Sheng, 1949		
punjabiensis Parshad & Paika, 1964		
pyo Toda, 1991		
quadraria Bock & Wheeler, 1972		
retnasabapathyi Takada & Momma, 1975		
rhombura Okada & Carson, 1983		
rhopaloa Bock & Wheeler, 1972	*coonorensis* Reddy & Krishnamurthy, 1973	Okada & Carson, 1983
rufa Kikkawa & Peng, 1938		
sahyadrii Prakash & Reddy, 1979	originally *suzukii* subgr., moved to *elegans* subgroup	Bock, 1980

VALID SPECIES NAMES	SYNONYMS & COMMENTS	SOURCE (who did it)
sampagensis Muniyappa & Reddy, 1980		
sechellia Tsacas & Bächli, 1981		
seguyi Smart, 1945		
seguyiana Chassagnard & Tsacas, 1997		Chassagnard *et al.*,1997
serrata Malloch, 1927		
serrula Tsacas, 1984		
siangensis Kumar & Gupta, 1988		
simulans Sturtevant, 1919		
smithersi Bock, 1976		
subauraria Kimura, 1983		
subelegans Okada, 1988		
suborosa Kumar & Gupta, 1992		
suzukii (Matsumura), 1931	as (*Leucopheuga*)	Wheeler, 1981
suzukii indica Parshad & Paika, 1964	(as *indicus*)	Wheeler, 1981
takahashii Sturtevant, 1927		
tani Cheng & Okada, 1985		
tanorum Okada, 1964		
teissieri Tsacas, 1971		
trapezifrons Okada, 1966		
triauraria Bock & Wheeler, 1972		
tricombata Singh & Gupta, 1979 ††		
trilutea Bock & Wheeler, 1972		
tristipennis Duda, 1924		
truncata Okada, 1964		
tsacasi Bock & Wheeler, 1972		
unipectinata Duda, 1924		
vallismaia Tsacas, 1984		
varians Bock & Wheeler, 1972		
vulcana Graber, 1957		
watanabei Gupta & Gupta, 1992	as " *punjabiensis* - like" Watanabe et al., 1982	Gupta & Gupta, 1992
xanthia Tsacas, 1981		
yakuba Burla, 1954	prev. desc. as *opisthomelaina* Nolte & Stoch, 1950	[cf. Ashburner, 1989 Text]
yuwanensis Kim & Okada, 1988		

GENE REGION	NAME	SEQUENCE 5' to 3'	DESIGN
Adh	AdhA	tt gas aay ccn gct gch att gcc ga	Baker & DeSalle, 1997 ref. Thomas & Hunt 1993
	AdhB	g agt ava crg gca cyk gdw aga tgg	Baker & DeSalle, 1997 ref. Thomas & Hunt 1993
	AdhL2	tgg gcg gca ttg gny tng aya c	Valerie Schawaroch
	AdhR2	agc cag gar ttg aay ttr tg	Valerie Schawaroch
hb	hbL	gag cag cac aay gcn tgg ta	Baker and DeSalle,1997
	hbR	ggc cat gta ctt cat tcy tc	Baker and DeSalle, 1997
	hb3	ggc gtk ggc tgw gga ctg gg	Rick Baker
	hbL2(int)	ccc agc cag aac gay car aa	Valerie Schawaroch
	hbR2(int)	ccg gca tag aar tgy tgc at	Valerie Schawaroch
co ii	George I	ata cct cga cgt tat tca ga	see Brower, 1994
	Strom	taa ttt gaa cta tyt tac cig c	see Brower, 1994
	Barb I	cca caa att tct gaa cat tga cca	see Brower, 1994
	Eva	gag acc att act tgc ttt cag tca tct	see Brower, 1994
Clones (manual)	PCR left	gct cgg atc cac tag taa c	DeSalle Lab
	PCR right	ctc tag atg cat gct cga g	DeSalle Lab
16S	Kb1	gct gga atg aat ggt tgg	DeSalle Lab (Rick Baker?)
	Kb2	taa tcc aac atc gag gtc gca	DeSalle Lab (Rick Baker?)
	Kb3	tat aat ttt ggg tgt agc cg	DeSalle Lab (Rick Baker?)
	Kb4	aat tta ttg cac taa tct gcc	DeSalle Lab (Rick Baker?)
28S	28SD2	cat ttt ttc cat ata agg aca tt	DeSalle Lab
	28SD2*	ccc gaa ggt atc ctg aat ctt tcg	DeSalle Lab

Protocols

For Cycle Sequencing:

General recipe (but amount of primer and PCR product will vary)
1.5 λ 1/10 primer
3.5 λ DNA
4 λ FS
1 λ water

Program 27 links programs 26 to 24 to 3
26 96°C 1 min.
95°C 15 sec
24 50°C 15 sec
60°C 4 min
for 33 cycles
3 25°C forever

General Recipe for PCR:
5 λ 10X buffer
5 λ dNTP's
0.1 λ Taq
1 λ primer A dilute
1 λ primer B dilute
38 λ sd water
T.V. = 50 λ
(All Primers are 100mM stocks all PCR reactions are for a Total Volume of 50 λ)

KB	°C	Time	KB	°C	Time
Denature	94		Denature	94	1min
Anneal	50 could increase by 3		Anneal	50	1 min 30 sec
Extension	72		Extension	72	2min
No. of cycles	40		No. of cycles	35	
Recipe:			Recipe:		

_ dilute of primer use1 λ each primer 1/10 dilution primer use 0.5 λ in T.V. = 50 λ

KB	°C	Time		°C	Time
Denature	94	1min	Denature	94	1 min
Anneal	54	1min	Anneal	50	1 min 30 sec
Extension	72	1min	Extension	72	2 min
No. of cycles	32X		No. of cycles	38X	
Recipe:			Recipe:		
0.5 λ of 1/10 primer in T.V.=50 λ			use 1 λ of _ dilution of primers		

ADH	°C	Time		ADH (DNA22)	°C	Time
Denature	94	1 min		Denature	94	1min
Anneal	50 +3	1 min	Anneal	57	1min	
Extension	72	1 min 30 sec		Extension	72	1min 30 sec
No. of cycles	37X			No. of cycles	37X	

Recipe: 0.5 λ of each of 1/10 Adh A and B in T.V.=50 λ

ADH New	℃	Time
Denature	94	1 min
Anneal	50	1 min
Extension	72	1 min 30 sec
No. of cycles	30X	

Recipe: 1 λ of 1/10 each primer (L2 & R2)

ADH New	℃	Time
Denature	94	1 min
Anneal	50	1 min
Extension	72	2 min
No. of cycles	35X	

Recipe: 1 λ of 1/10 each primer (L2 & R2)

CO II	℃	Time
Denature	94	1 min
Anneal	54	1 min
Extension	72	1 min
No. of cycles	35X can decrease to 30X too	

Recipe: 0.5 λ of 1/10 each primer T.V.=50 λ

	℃	Time
Denature	94	1 min
Anneal	54	1 min
Extension	72	1 min
No. of cycles	33X	

Recipe: 0.5 λ of 1/10 primer
Also 47 anneal; Also 37 cycles

Amplify with George and Eva sometimes Strom and Eva

HB	℃	Time
Denature	94	1 min
Anneal	52,53,55	1 min
Extension	72	1 min 30 sec
No. of cycles	35X	

Recipe: 1 λ of each 1/10 primer

HB (great)	℃	Time
Denature	94	1 min
Anneal	52	1 min
Extension	72	2 min
No. of cycles	35X	

Recipe: 0.5 λ of each 1/10 primer

HB (some Taxa)	℃	Time
Denature	94	1 min
Anneal	47,48	1 min
Extension	72	1 min
No. of cycles	40X (anneal 46 w/ 37X)	

Recipe: 0.5 λ of each 1/10 primer

HB	℃	Time
Denature	94	1 min
Anneal	46	1 min
Extension	72	2 min
No. of cycles	40X	

Recipe: 2 λ of each 1/10 primer

HB	℃	Time
Denature	94	1 min
Anneal	48	1 min
Extension	72	2 min
No. of cycles	40X	

Recipe: 0.5 ? of 1/10 hbL or R
anneal 46, 50
some need do internal & external PCR to get piece

HB (Rick's protocol)	℃	Time
Denature	94	1 min
Anneal	46	1 min
Extension	72	2 min
No. of cycles	?	

Recipe: 2 microMolar of hb primer

28S	℃	Time
Denature	94	1 min
Anneal	48	1 min
Extension	72	2 min
No. of cycles	35X decrease to 30X	

Recipe: 0 to B primer use 1 λ of 1/10

28S	℃	Time
Denature	94	1 min
Anneal	48	1 min
Extension	72	1 min
No. of cycles	30X	

APPENDIX C: Alignment Parameters and Ambiguous Sites

The parameter files used in MALIGN are listed below. The parameter files varied on the cost factors (internal gaps cost, change cost, leading and trailing gap costs) and on the tree construction methods (quick, build, score 1 and score 2).

The 28S D2 is shown for one of the two equally most parsimonious alignments. The two positions that were alignment ambiguous are at basepair positions 119 and 121 (also see the asterisk below the sequences).

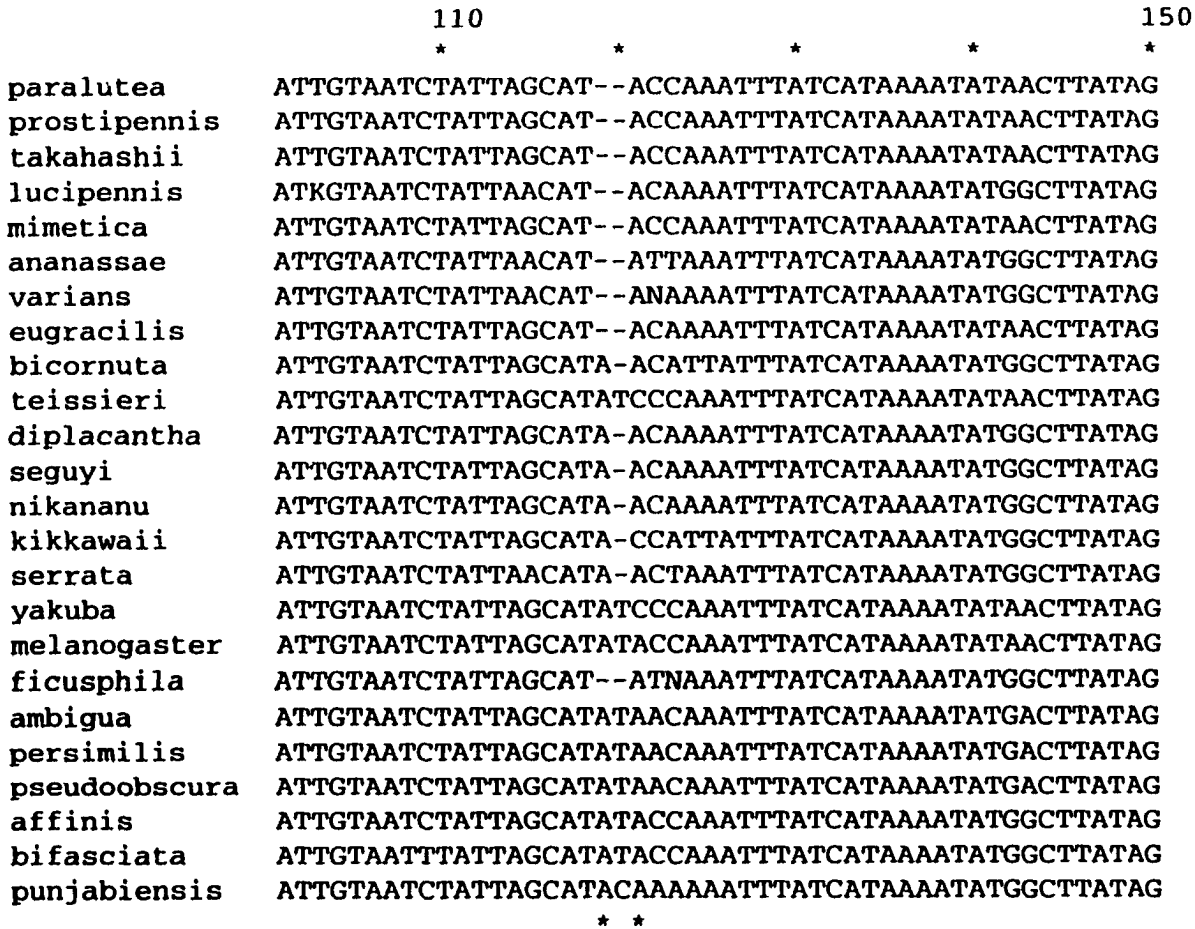
There are three ambiguous regions for the aligned *hb* sequence. They are pictured for both the amino acid as well as the nucleotide base alignments. The alignment shown is using the gap penalty value of 8.

	MALIGN Parameter Files					
	P1	P2	P4	P6	P8	P10
internal	2	2	4	6	8	10
changecost	1	1	1	1	1	1
leading	4	2	4	6	8	10
trailing	4	2	4	6	8	10
paup	--	--	--	--	--	--
quick	build	quick	quick	quick	quick	quick
score 1	2	1	1	1	1	1
keeptrees 100	--	--	--	--	--	--
keepaligns 100	--	--	--	--	--	--
alignswap	--	--	--	--	--	--
treeswap	--	--	--	--	--	--
extragaps 1	--	--	--	--	--	--
inalign	--	--	--	--	--	--
linelength 600	--	--	--	--	--	--
time	--	--	--	--	--	--
phylot	--	--	--	--	--	--
iter	--	--	--	--	--	--

28S D2 Alignment

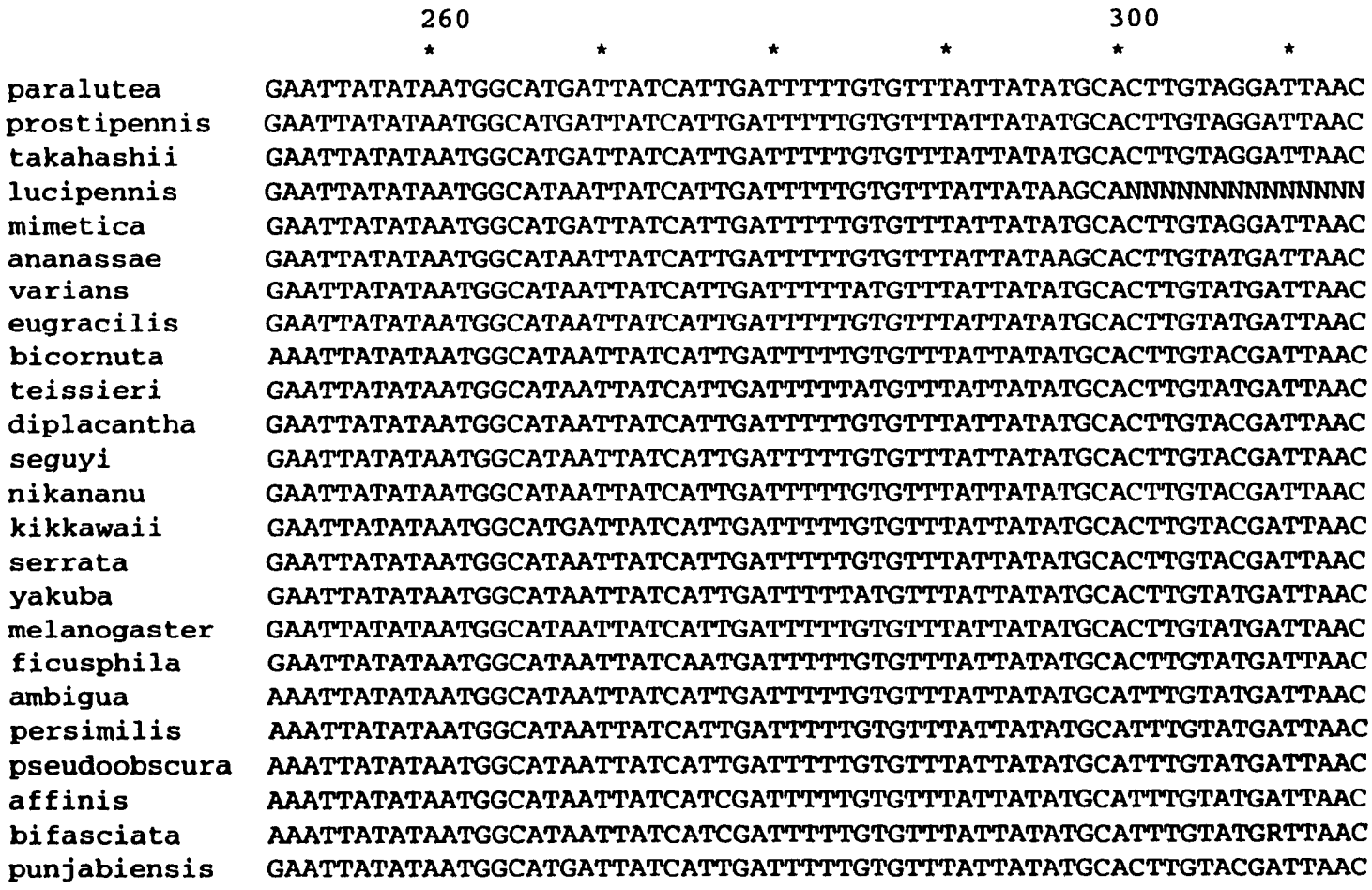
```
                          10                                        50
                          *         *         *         *         *
paralutea         GCCCGATGAACCTGAATATCCGTTATGGAAAATTCATCATTAGAATTGTA
prostipennis      GCCCGATGAACCTGAATATCCGTTATGGAAAATTCATCATTAGAATTGTA
takahashii        GCCCGATGAACCTGAATATCCGTTATGGAAAATTCATCATTAAAATTGTA
lucipennis        GCCCGATGAACCTGAATATCCATTATGGAAAATTCATCATTAAAATTGTA
mimetica          GCCCGATGAACCTGAATATCCGTTATGGAAAATTCATCATTAAAATTGTA
ananassae         GCCCGATGAACCTGAATATCCGTTATGGAAAATTCATCATTAAAGTTATA
varians           GCCCGATGAACCTGAATATCCATTATGGAAAATTCATCATTAGAGTTATA
eugracilis        GCCCGATGAACCTGAATATCCGTTATGGAAAATTCATCATTATAATTGTA
bicornuta         GCCCGATGAACCTGAATATCCGTTATGGAAAATTCATCATTAGAGTTGTA
teissieri         GCCCGATGAACCTGAATATCCGTTATGGAAAATTCATCATTAAAATTGTA
diplacantha       GCCCGATGAACCTGAATATCCGTTATGGAAAATTCATCATTAGAATTGTA
seguyi            GCCCGATGAACCTGAATATCCATTATGGAAAATTCATCATTAGAATTGTA
nikananu          GCCCGATGAACCTGAATATCCGTTATGGAAAATTCATCATTAGAATTGTA
kikkawaii         GCCCGATGAACCTGAATATCCGTTATGGAAAATTCATCATTAGAATTGTA
serrata           GCCCGATGAACCTGAATATCCGTTATGGAAAATTCATCATTAGACTTGTA
yakuba            GCCCGATGAACCTGAATATCCGTTATGGAAAATTCATCATTAAAATTGTA
melanogaster      GCCCGATGAACCTGAATATCCGTTATGGAAAATTCATCATTAAAATTGTA
ficusphila        GCCCGATGAACCTGAATATCCATTATGGAAAATTCATCATTAAAATTGTA
ambigua           GCCCGATGAACCTGAATATCCGTTATGGAAAATTCATCATTAAAGTTGTA
persimilis        GCCCGATGAACCTGAATATCCGTTATGGAAAATTCATCATTAAAGTTGTA
pseudoobscura     GCCCGATGAACCTGAATATCCGTTATGGAAAATTCATCATTAAAGTTGTA
affinis           GCCCGATGAACCTGAATATCCGTTATGGAAAATTCATCATTAAAGTTGTA
bifasciata        GCCCGATGAACCTGAATATCCGTTATGGAAAATTCATCATTAAAGTTGTA
punjabiensis      GCCCGATGAACCTGAATATCCGTTATGGAAAATTCATCATTAGAATTGTA
```

```
                    60                                                  100
                    *         *         *         *         *
paralutea           ATATTTAAACAATATTATGGTAATAGTGTGCATTTTTTCCATATAAGGAC
prostipennis        ATATTTAAACAATATTATGGTAATAGTGTGCATTTTTTCCATATAAGGAC
takahashii          ATATTTAAACAATATTATGATAATAGTGTGCATTTTTTCCATATAAGGAC
lucipennis          ATATTTAAACAATATTATGATAATAGTGTGCATTTTTTCCATATAAGGAC
mimetica            ATATTTAAACAATATTATGAGGATAGTGTGCATTTTTTCCATATAAGGAC
ananassae           ATATTTAAACAATATTATACAAATAATGTGCATTTTTTCCATATAAGGAC
varians             ATATTAAAATAATATTATACAAATAGTGTGCATTTTTTCCATATAAGGAC
eugracilis          ATATTTAAACAATATTATGATAATAGTGTGCATTTTTTCCATATAAGGAC
bicornuta           ATATTTTAACAATATTATGATAATAGTGTGCATTTTTTCCATATAAGGAC
teissieri           ATATTTAAATAATATTATGAGGATAGTGTGCATTTTTTTCCATATAAGGAC
diplacantha         ATGTTTAAATAATATTATGATAATAGTGTGCATTTTTTTCCATATAAGGAC
seguyi              ATGTTTAAACAATATTATGATAATAGTGTGCATTTTTTTCCATATAAGGAC
nikananu            ATATTTAAACAATATTATGATAATAGTGTGCATTTTTTTCCATATAAGGAC
kikkawaii           ATATTTAATAAATATTATGATAATAGTGTGCATTTTTTTCCATATAAGGAC
serrata             GTATTTAAACAATATTATGATAATAGTGTGCATTTTTTTCCATATAAGGAC
yakuba              ATATTTAAATAATATTATAAGGATAGTGTGCATTTTTTTCCATATAAGGAC
melanogaster        ATATTTAAATAATATTATGAGAATAGTGTGCATTTTTTTCCATATAAGGAC
ficusphila          ATATTTAAACAATATTATGATAATAGTGTGCATTTTTTTCCATATAAGGAC
ambigua             ATATTTTAACAATATTATAATAATAGTGTGCATTTTTTTCCATATAAGGAC
persimilis          ATATTTTAACAATATTATAATAATAGTGTGCATTTTTTTCCATATAAGGAC
pseudoobscura       ATATTTTAACAATATTATAATAATAGTGTGCATTTTTTTCCATATAAGGAC
affinis             ATATTTTAACAATATTATAATAATAGTGTGCATTTTTTTCCATATAAGGAC
bifasciata          ATATTTTAACAATATTATAATAATAGTGTGCATTTTTTTCCATATAAGGAC
punjabiensis        ATATTTAAACAATATTATGATGATAGTGTGCATTTTTTTCCATATAAGGAC
```



```
                    160                                                 200
                    *         *         *         *         *
paralutea           TTTATTCAAATTAAATTGCTTGCATTTTAACACAGAATAAATGTTATTAA
prostipennis        TTTATTCAAATTAAATTGCTTGCATTTTAACACAGAATAAATGTTATTAA
takahashii          TTTATTCAAATTAAATTGCTTGCATTTTAACACAGAATAAATGTTATTAA
lucipennis          TTTATTCCAATTAAAATGCTTGCGTTTTAACACAGAATAAATGTCATTAA
mimetica            TTTATTCAAATTAAATTGCTTGCATTTTAACACAGAATAAATGTTATTAA
ananassae           TTTATTCAAATTATTATGCTTGCATTTTAACATAGAATAAATGTCATTAA
varians             TTTATTCCAATTAAATTGCTTGCATTTTAACACAGAATAAATGTCATTAA
eugracilis          TTTATTCCAATTAAATTGCTTGCATTTTAACACAGAATAAATGTTATTAA
bicornuta           TTTATTCAAATTATTATGCTTGCATTTTAACATAGAATAAATGTCATTAA
teissieri           TTTATTCAAATTAAAGTGCTTGCATTTTAACACAGAATAAATGTTATTAA
diplacantha         TTTATTCAAATTATTATGCTTGCATTTTAACATAGAATAAATGTCATTAA
seguyi              TTTATTCAAATTATTATGCTTGCATTTTAACATAGAATAAATGCCATTAA
nikananu            TTTATTCAAATTATTATGCTTGCATTTTAACATAGAATAAATGTCATTAA
kikkawaii           TTTATTCAAATTATTATGCTTGCATTTTAACATAGAATAAATGTCATTAA
serrata             TTTATTCAAATTATTATGCTTGCATTTTAACATAGAATAAATGTCATTAA
yakuba              TTTATTCAAATTAAAGTGCTTGCATTTTAAAACAGAATAAATGTTATTAA
melanogaster        TTTATTCCAATTAAATTGCTTGCATTTTAACACAGAATAAATGTTATTAA
ficusphila          TTTATTTAAATTAAAATGCTTGCGTTTTAACACAAAATAAATGTCATTAA
ambigua             TTTATTCAAATTATTATGCTTGCATTTTAACATAGAATAAATGTCATTAA
persimilis          TTTATTCAAATTATTATGCTTGCATTTTAACATAGAATAAATGTCATTAA
pseudoobscura       TTTATTCAAATTATTATGCTTGCATTTTAACATAGAATAAATGTCATTAA
affinis             TTTATTCAAATTATTATGCTTGCATTTTAACATAGAATAAATGTCATTAA
bifasciata          TTTATTCAAATTATTATGCTTGCATTTTAACATAGAATAAATGTCATTAA
punjabiensis        TTTATTCAAATTATTATGCTTGCATTTTAACATAGAATAAATGTCATTAA
```

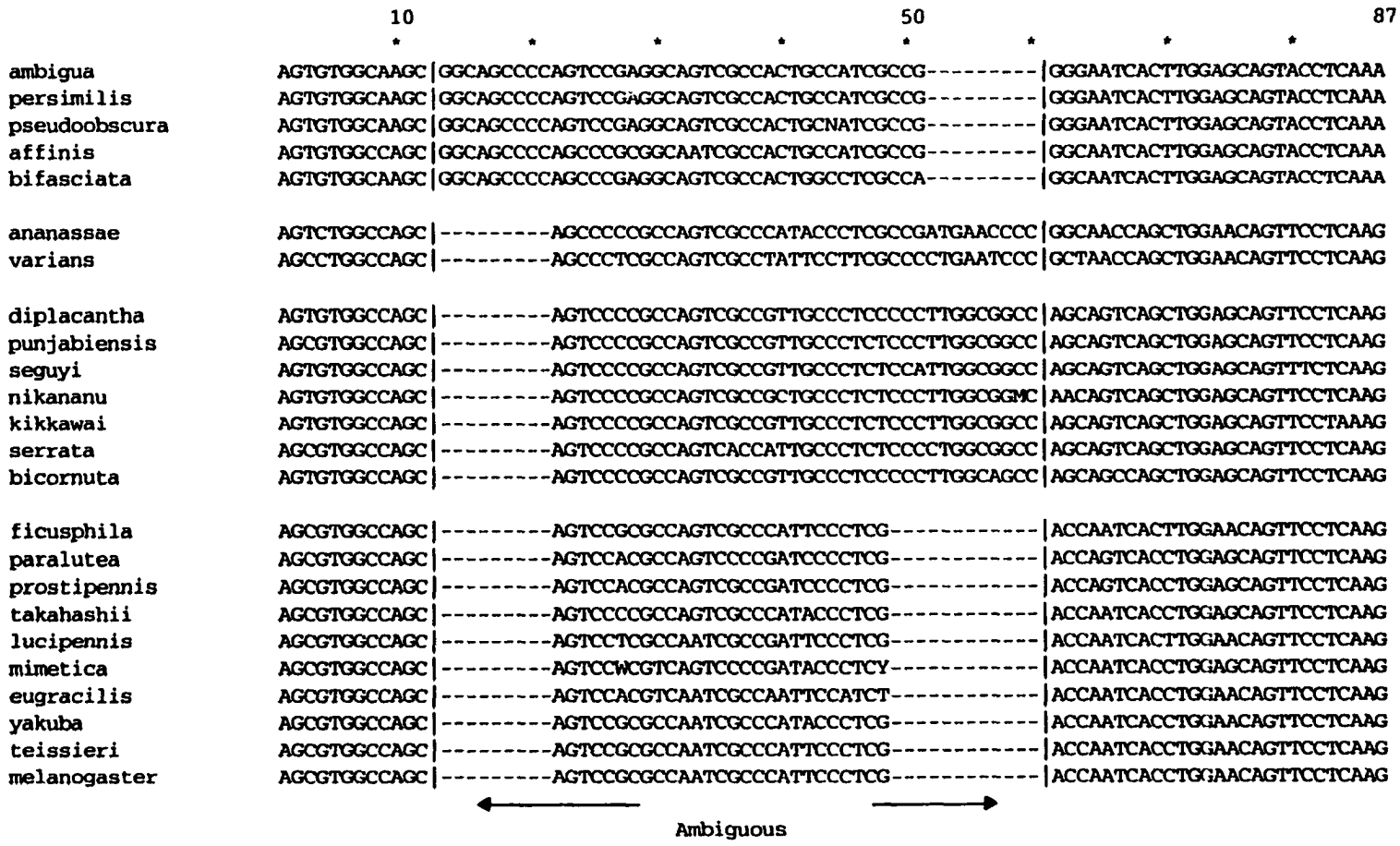
```
                    210                                       250
                    *         *         *         *         *
paralutea           TTTGATAAAGTGCTGATAGATTTATATGAATACAGTGCGTTAATTTTTCG
prostipennis        TTTGATAAAGTGCTGATAGATTTATATGAATACAGTGCGTTAATTTTTCG
takahashii          TTTGATAAAGTGCTGATAGATTTATATGAATACAGTGCGTTAATTTTTCG
lucipennis          TTTGATAAAGTGTTGATAGATTTATAAGAATACAGTGCGTTAATTTTTCG
mimetica            TTTGATAAAGTGTTGATAGATTTATATGAATACAGTGCGTTAATTTTTCG
ananassae           TTTGATAAAGTGTTGATAGACTAATAAGAATACAGTGCGTTAATTTTTCG
varians             TTTGATAAAGTGTTGATAGATTAATAAGAATACAGTGCGTTAATTTTTCG
eugracilis          TTTGATAAAGTGCTGATAGATTTATATGAATACAGTGCGTTAATTTTTCG
bicornuta           TTTGATAAAGTGCTGATAGATTTATATGACTACAGTGCGTTAATTTTTCG
teissieri           TTTGATAAAGTGCTGATAGATTTATATGATTACAGTGCGTTAATTTTTCG
diplacantha         TTTGATAAAGTGCTGATAGATTTATATGACTACAGTGCGTTAATTTTTCG
seguyi              TTTGATAAAGTGCTGATAGATTTATATGACTACAGTGCGTTAATTTTTCG
nikananu            TTTGATAAAGTGCTGATAGATTTATATGACTACAGTGCGTTAATTTTTCG
kikkawaii           TTTGATAAAGTGCTGATAGATTTATATGACTACAGTGCGTTAATTTTTCG
serrata             TTTGATAAAGTGCTGATAGATTTATATGACTACAGTGCGTTAATTTTTCG
yakuba              TTTGATAAAGTGCTGATAGATTTATATGATTACAGTGCGTTAATTTTTCG
melanogaster        TTTGATAAAGTGCTGATAGATTTATATGATTACAGTGCGTTAATTTTTCG
ficusphila          TTTGATAAAGTGCTGATAGATTTATATGATTACAGTGCGTTAATTTTTCG
ambigua             TTTGATAAAGTGTTGATAAATTAAAATTAATACAGTGCGTTAATTTTTCG
persimilis          TTTGATAAAGTGTTGATAAATTAAAATTAATACAGTGCGTTAATTTTTCG
pseudoobscura       TTTGATAAAGTGTTGATAAATTAAAATTAATACAGTGCGTTAATTTTTCG
affinis             TTTGATAAAGTGTTGATAAATTAAAATTAATACAGTGCGTTAATTTTTCG
bifasciata          TTTGATAAAGTGTTGATAAATTAAAATTAATACAGTGCGTTAATTTTTCG
punjabiensis        TTTGATAAAGTGCTGATAGATTTATATGACTACAGTGCGTTAATTTTTCG
```

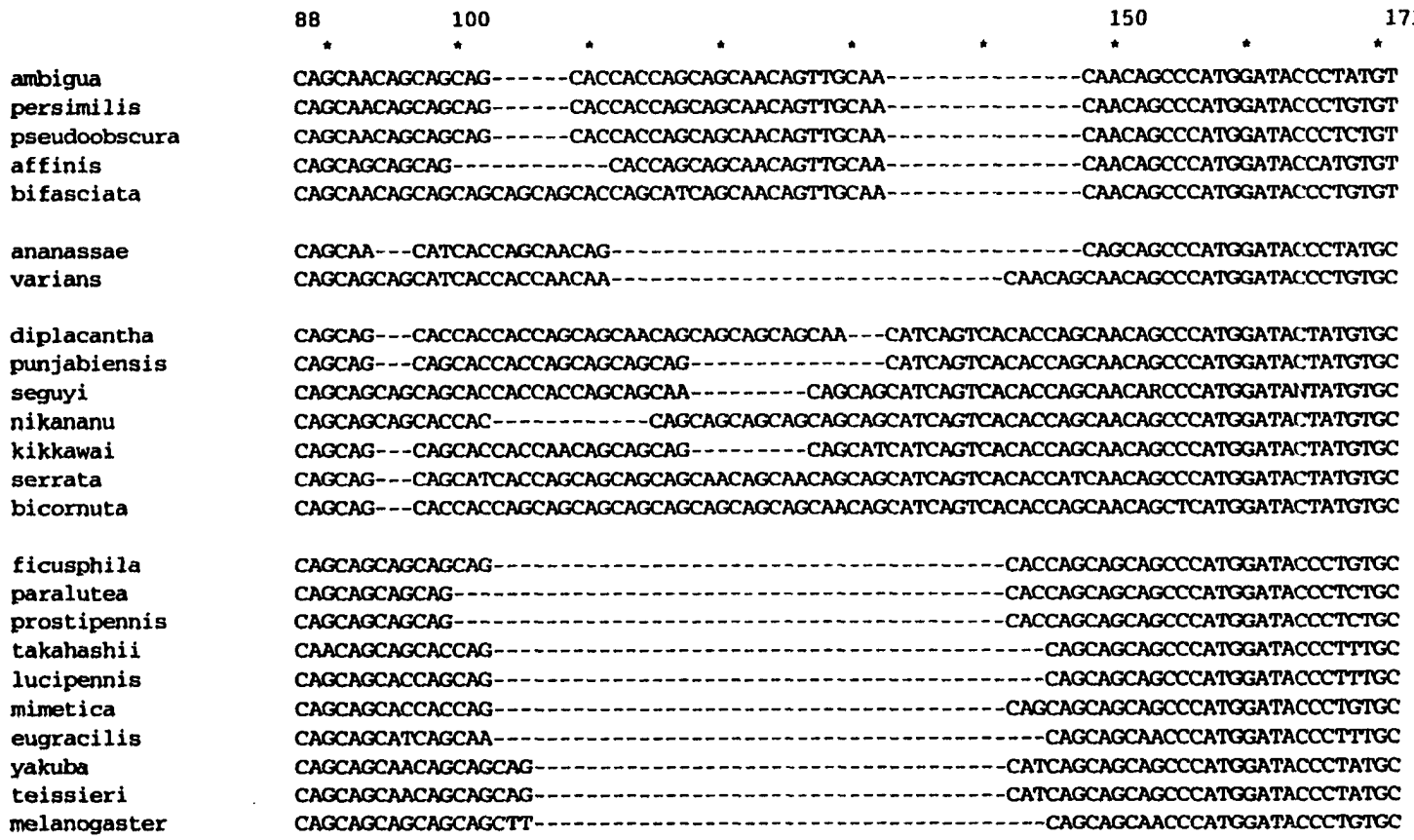


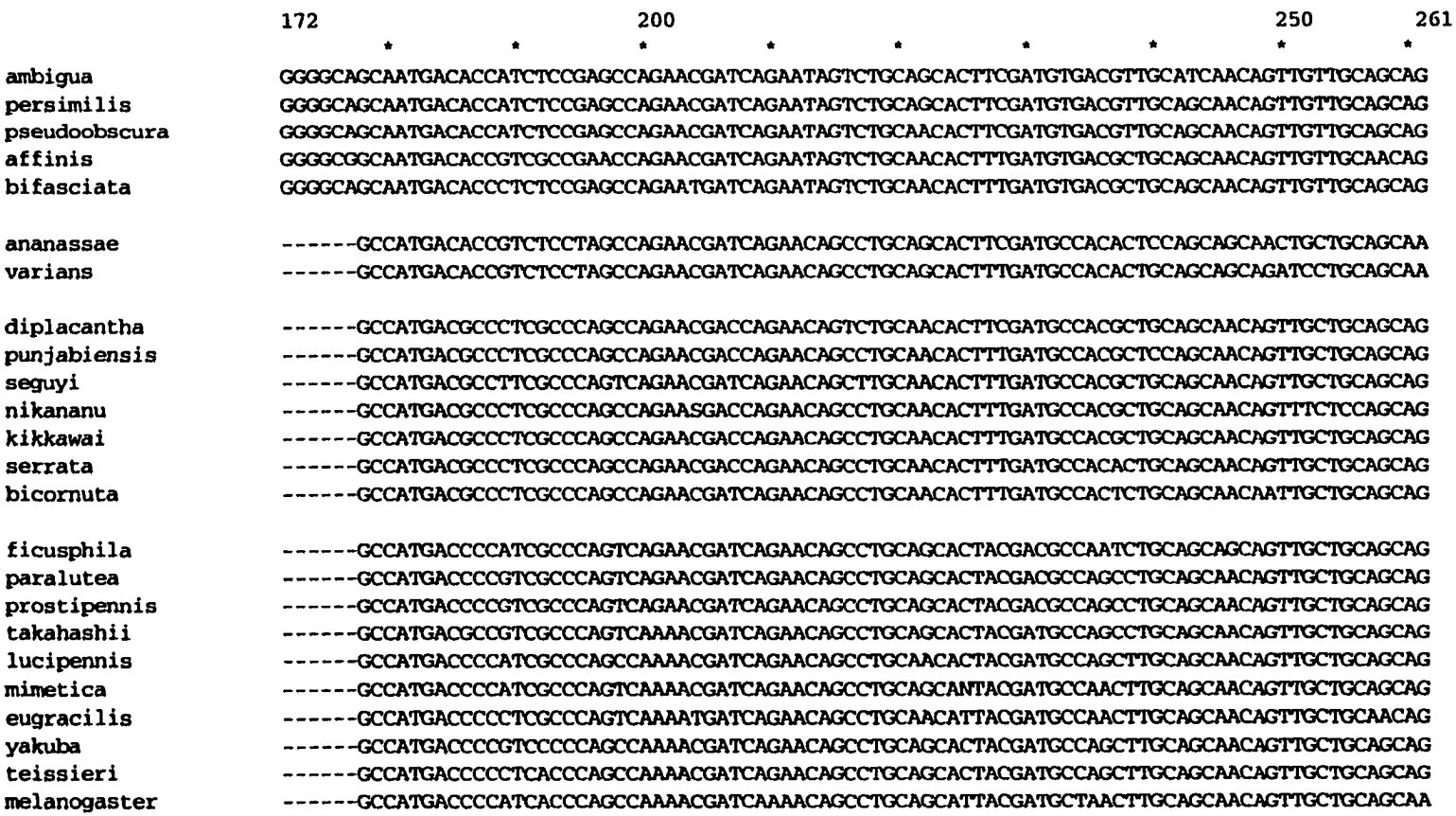
hb amino acid sequence

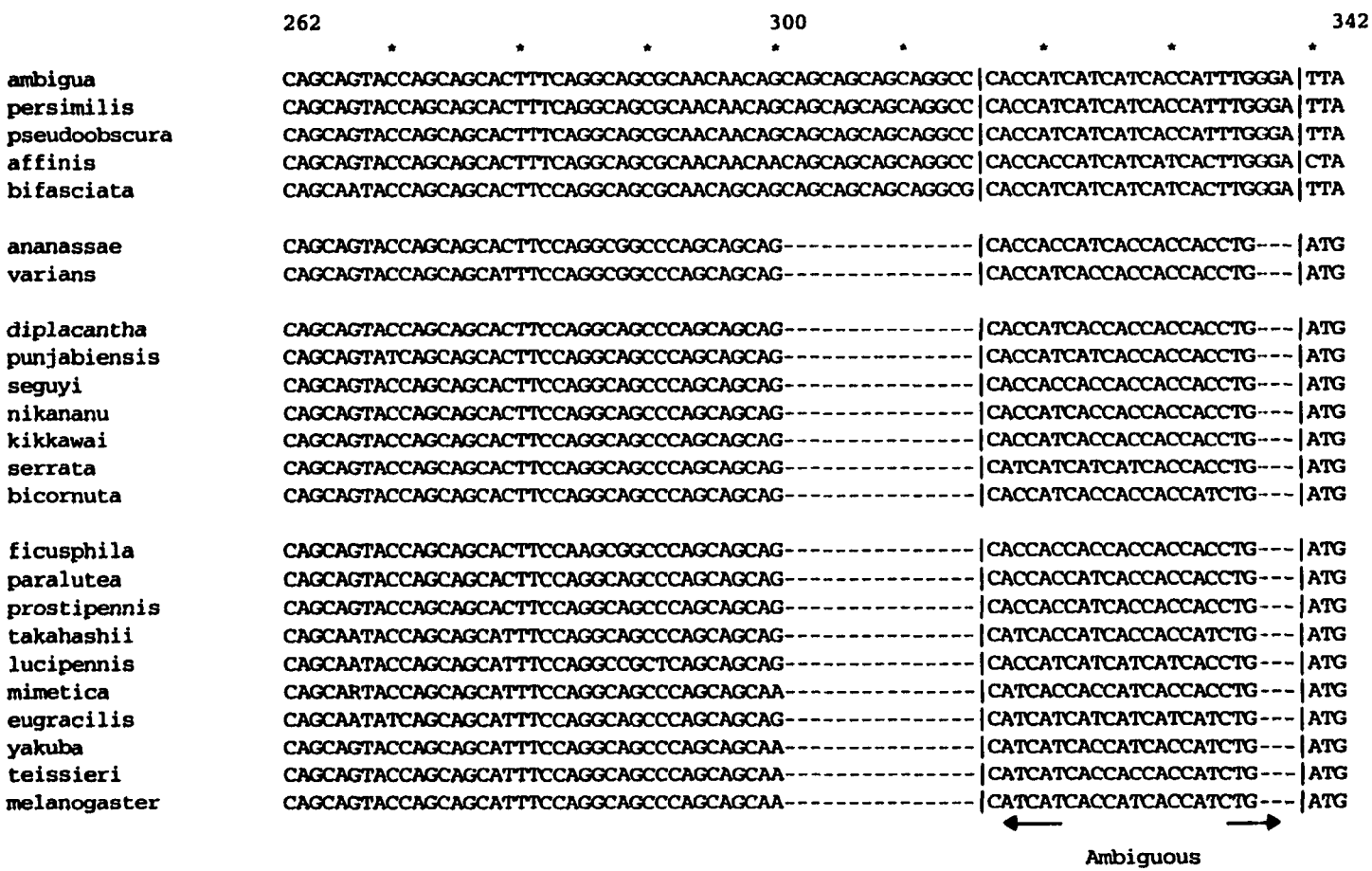
```
                                10                                  50                                  97
                                 *         *         *         *         *         *         *         *         *
ambigua              SVAS|GSPSPRQSPLPSP---|GNHLEQYLKQQQQQ--HHQQQQLQ-----QQPMDTLCGAAMTPSPSQNDQNSLQHFDVTLHQQLLQQQQYQQHFQAA
persimilis           SVAS|GSPSPRQSPLPSP---|GNHLEQYLKQQQQQ--HHQQQQLQ-----QQPMDTLCGAAMTPSPSQNDQNSLQHFDVTLQQQLLQQQQYQQHFQAA
pseudoobscura        SVAS|GSPSPRQSPLXSP---|GNHLEQYLKQQQQQ--HHQQQQLQ-----QQPMDTLCGAAMTPSPSQNDQNSLQHFDVTLQQQLLQQQQYQQHFQAA
affinis              SVAS|GSPSPRQSPLPSP---|GNHLEQYLKQQQQ----HQQQQLQ-----QQPMDTMCGAAMTPSPNQNDQNSLQHFDVTLQQQLLQQQQYQQHFQAA
bifasciata           SVAS|GSPSPRQSPLASP---|GNHLEQYLKQQQQQQQHQHQQQLQ-----QQPMDTLCGAAMTPSPSQNDQNSLQHFDVTLQQQLLQQQQYQQHFQAA

ananassae            SLAS|---SPRQSPIPSPMNP|GNQLEQFLKQQ-HHQQQ-----------QQPMDTLC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
varians              SLAS|---SPRQSPIPSPLNP|ANQLEQFLKQQQHHHQQ----------QQQQPMDTLC--AMTPSPSQNDQNSLQHFDATLQQQILQQQQYQQHFQAA

diplacantha          SVAS|---SPRQSPLPSPLAA|SSQLEQFLKQQ-HHHQQQQQQQQ-HQSHQQQPMDTMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
punjabiensis         SVAS|---SPRQSPLPSPLAA|SSQLEQFLKQQ-QHHQQQQ-----HQSHQQQPMDTMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
seguyi               SVAS|---SPRQSPLPSPLAA|SSQLEQFLKQQQQHHHQQQ---QQHQSHQQQPMDXMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
nikananu             SVAS|---SPRQSPLPSPLAX|NSQLEQFLKQQQHH----QQQQQQHQSHQQQPMDTMC--AMTPSPSQXDQNSLQHFDATLQQQFLQQQQYQQHFQAA
kikkawai             SVAS|---SPRQSPLPSPLAA|SSQLEQFLKQQ-QHHQQQQ---QHHQSHQQQPMDTMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
serrata              SVAS|---SPRQSPLPSPLAA|SSQLEQFLKQQ-QHHQQQQQQQQQQHQSHHQQPMDTMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
bicornuta            SVAS|---SPRQSPLPSPLAA|SSQLEQFLKQQ-HHQQQQQQQQQQQHQSHQQQLMDTMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA

ficusphila           SVAS|---SPRQSPIPS----|TNHLEQFLKQQQQQ------------HQQQPMDTLC--AMTPSPSQNDQNSLQHYDANLQQQLLQQQQYQQHFQAA
paralutea            SVAS|---SPRQSPIPS----|TSHLEQFLKQQQQ-------------HQQQPMDTLC--AMTPSPSQNDQNSLQHYDASLQQQLLQQQQYQQHFQAA
prostipennis         SVAS|---SPRQSPIPS----|TSHLEQFLKQQQQ-------------HQQQPMDTLC--AMTPSPSQNDQNSLQHYDASLQQQLLQQQQYQQHFQAA
takahashii           SVAS|---SPRQSPIPS----|TNHLEQFLKQQQHQ-------------QQQPMDTLC--AMTPSPSQNDQNSLQHYDASLQQQLLQQQQYQQHFQAA
lucipennis           SVAS|---SPRQSPIPS----|TNHLEQFLKQQHQQ-------------QQQPMDTLC--AMTPSPSQNDQNSLQHYDASLQQQLLQQQQYQQHFQAA
mimetica             SVAS|---SPRQSPIPS----|TNHLEQFLKQQHHQ------------QQQQPMDTLC--AMTPSPSQNDQNSLQXYDANLQQQLLQQQQYQQHFQAA
eugracilis           SVAS|---SPRQSPIPS----|TNHLEQFLKQQHQQ-------------QQQPMDTLC--AMTPSPSQNDQNSLQHYDANLQQQLLQQQQYQQHFQAA
yakuba               SVAS|---SPRQSPIPS----|TNHLEQFLKQQQQQ------------HQQQPMDTLC--AMTPSPSQNDQNSLQHYDASLQQQLLQQQQYQQHFQAA
teissieri            SVAS|---SPRQSPIPS----|TNHLEQFLKQQQQQ------------HQQQPMDTLC--AMTPSPSQNDQNSLQHYDASLQQQLLQQQQYQQHFQAA
melanogaster         SVAS|---SPRQSPIPS----|TNHLEQFLKQQQQQL-------------QQQPMDTLC--AMTPSPSQNDQNSLQHYDANLQQQLLQQQQYQQHFQAA

                          ←          →
                          Ambiguous
```

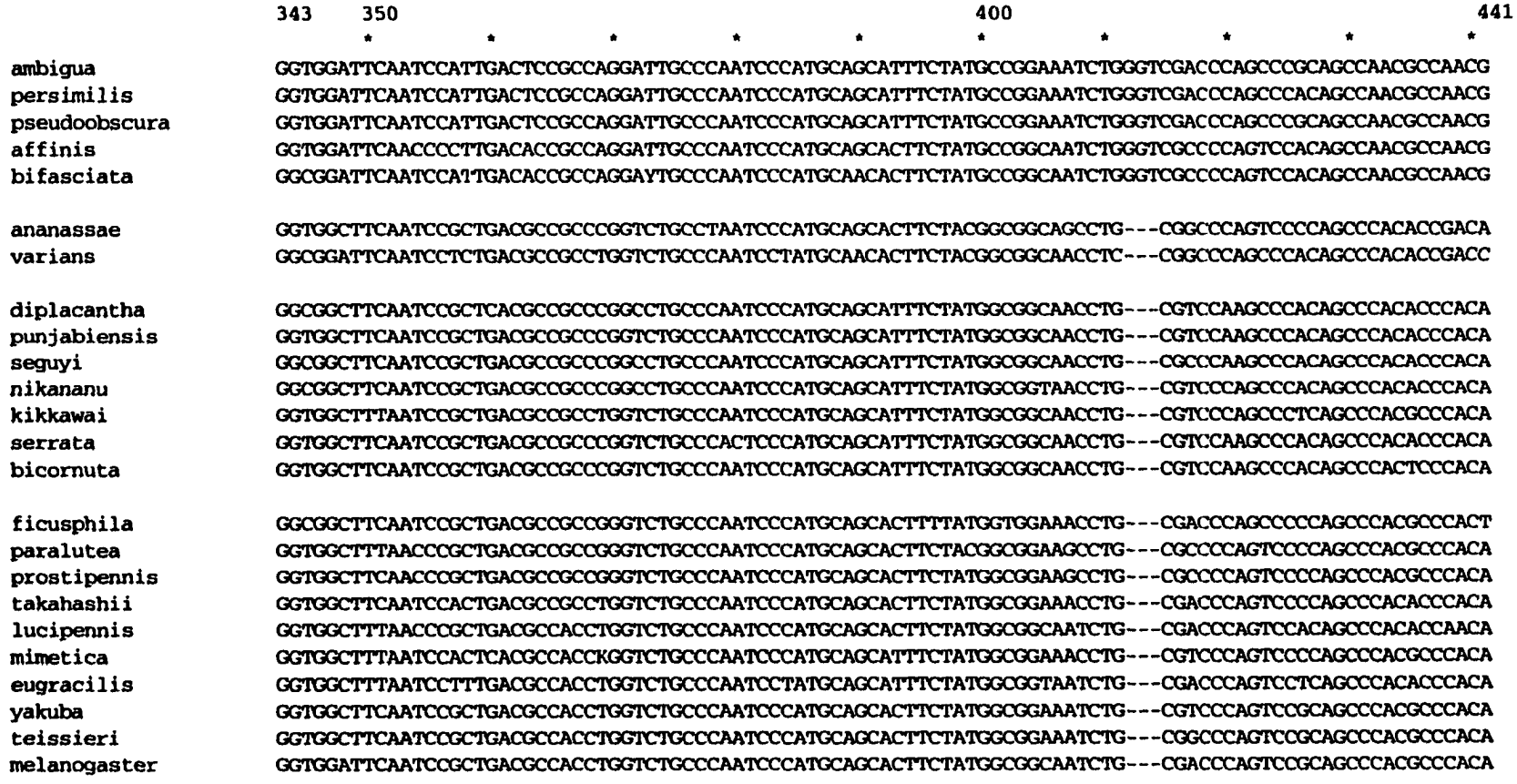
hb nucleotide sequence

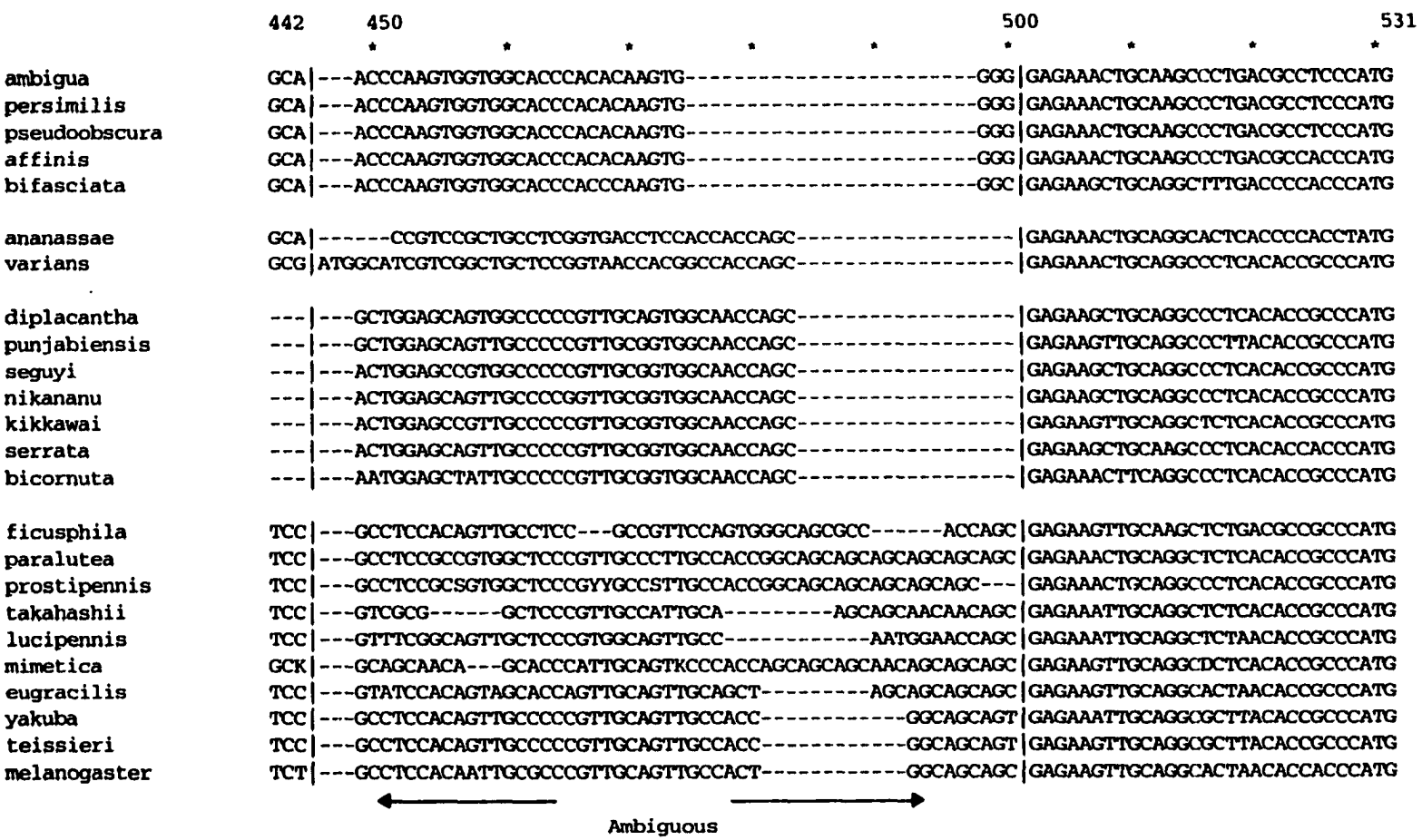












```
                    532                 550                 564
                            *                   *                   *
ambigua             GATGTGACGCCACCAAAGTCACCAGCAAAATCC
persimilis          GATGTGACGCCACCAAAGTCACCAGCGAAATCC
pseudoobscura       GATGTGACGCCACCAAAGTCACCAGCGAAAGCC
affinis             GATGTGACTCCACCAAAGTCACCAGCCAAGTCC
bifasciata          GATGTGACGCCACCAAAGTCACCAGCGAAATCC

ananassae           GACGTGACACCGCCCAAGTCGCCCGCCAAGTCC
varians             GACGTGACGCCGCCCAAGTCGCCCGCCAAGTCC

diplacantha         GATGTGACGCCGCCCAAGTCGCCGGCCAAGTCC
punjabiensis        GATGTGACGCCGCCCAAGTCGCCGGCCAAGTCC
seguyi              GACGTGACGCCGCCCAAGTCGCCGGCCAAGTCC
nikananu            GATGTGACGCCGCCCAAGTCGCCGGCCAAGTCC
kikkawai            GATGTGACGCCGCCCAAGTCGCCGGCCAAGTCC
serrata             GATGTTACGCCGCCCAAGTCGCCGGCCAAGTCC
bicornuta           GATGTGACGCCGCCCAAGTCGCCGGCCAAGTCC

ficusphila          GATGTCACGCCGCCCAAGTCGCCGGCCAAGTCC
paralutea           GATGTCACACCGCCCAAGTCGCCGGCCAAGTCG
prostipennis        GATGTCACGCCGCCCAAGTCGCCGGCCAAGTCG
takahashii          GATGTCACGCCGCCCAAGTCGCCGGCCAAGTCG
lucipennis          GATGTCACACCGCCCAAATCGCCGGCCAAGTCG
mimetica            GATGTCACACCGCCCAAGTCGCCGGCCAAGTCC
eugracilis          GATGTCACACCACCCAAGTCGCCGGCCAAGTCG
yakuba              GATGTCACACCGCCCAAGTCGCCGGCCAAGTCT
teissieri           GATGTCACACCGCCCAAGTCGCCGGCCAAGTCG
melanogaster        GATGTCACACCGCCTAAGTCGCCGGCCAAGTCG
```

APPENDIX D: Saturation Plots (Chapter 1)

Saturation plots were made for nucleotide composition of transitions, transversions, and codon positions as listed below.

Figure 1 (two pages). Total transitions for the informative characters compared to the uncorrected p distance for each of the gene regions separately as well as the ribosomal genes (16S + 28S) combined and all 5 gene regions combined.

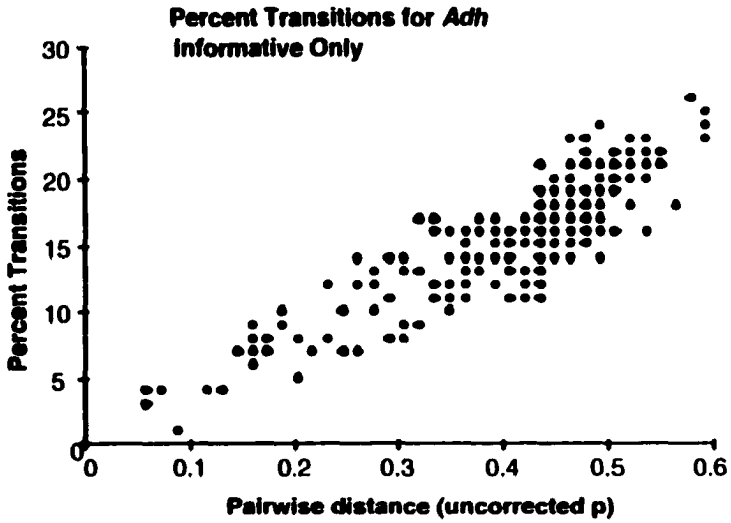
Figure 2 (two pages). Total transversions for the informative characters compared to the uncorrected p distance for each of the gene regions separately as well as the ribosomal genes (16S + 28S) combined and all 5 gene regions combined.

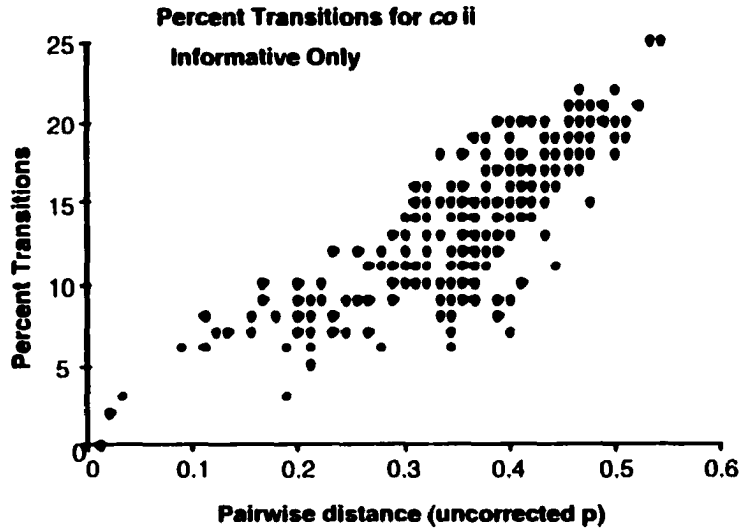
Figure 3. First, first + second and third codon position changes compared to the uncorrected p distance for informative characters within the *Adh* gene region.

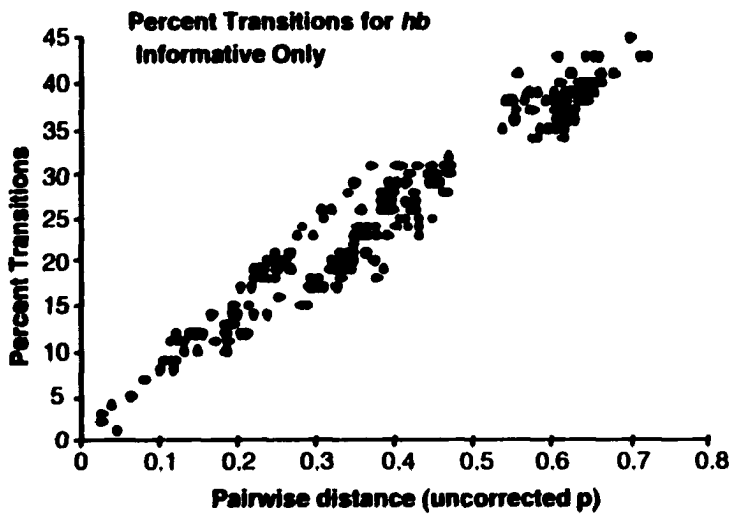
Figure 4. First, first + second and third codon position changes compared to the uncorrected p distance for informative characters within the *co* ii gene region.

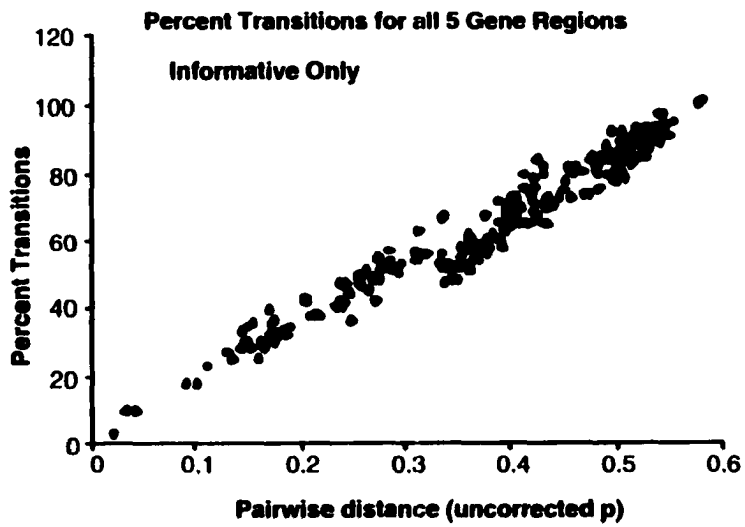
Figure 5. First, second, first + second and third codon position changes compared to the uncorrected p distance for informative characters within the *hb* gene region.

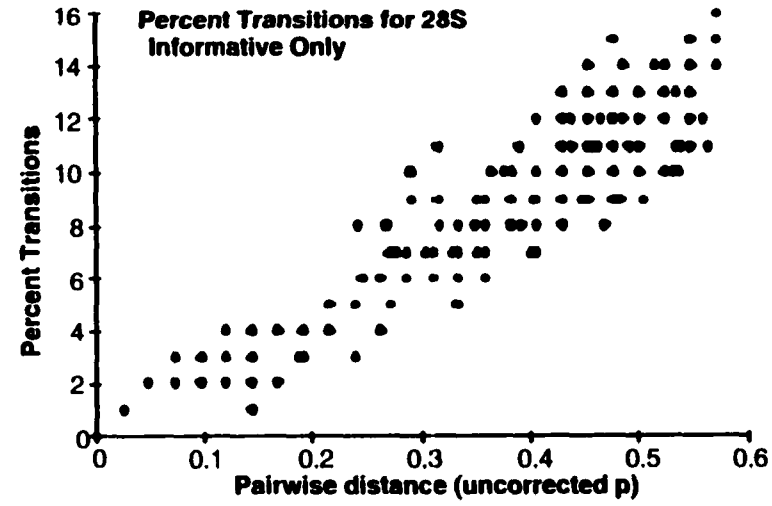
Figure 6. First, second, first + second and third codon position changes compared to the uncorrected p distance for informative characters within the protein coding gene regions of *Adh* + *co* ii + *hb*.

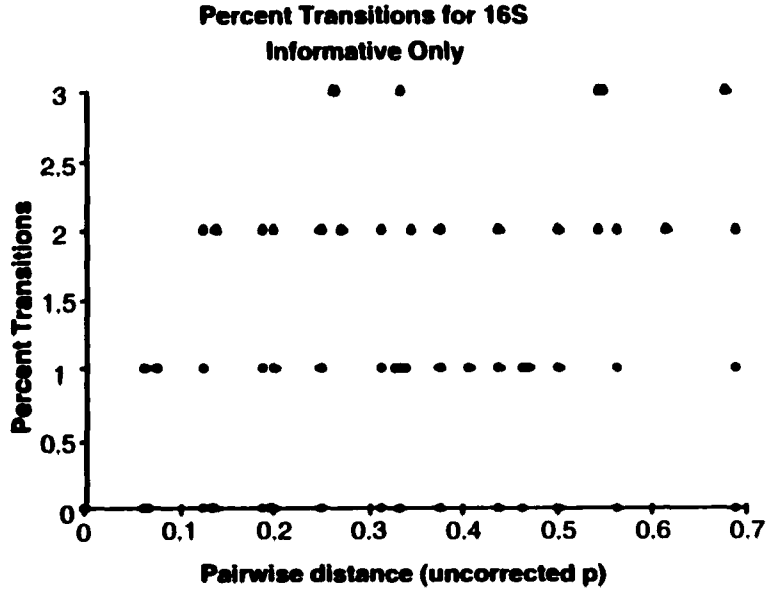


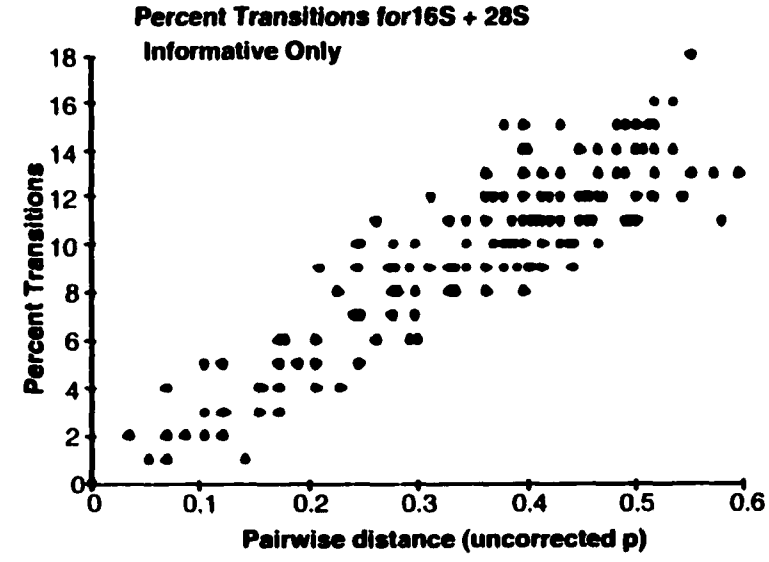


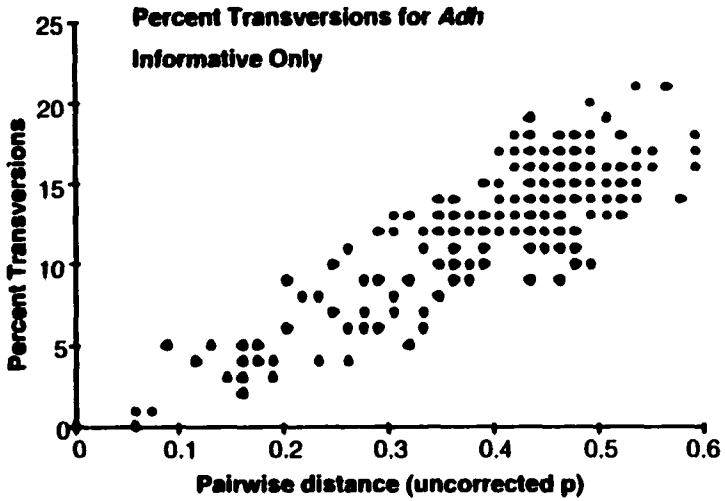


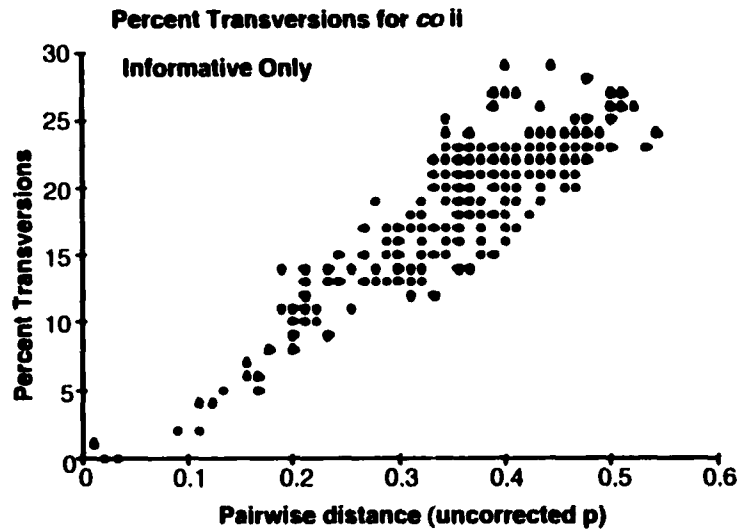


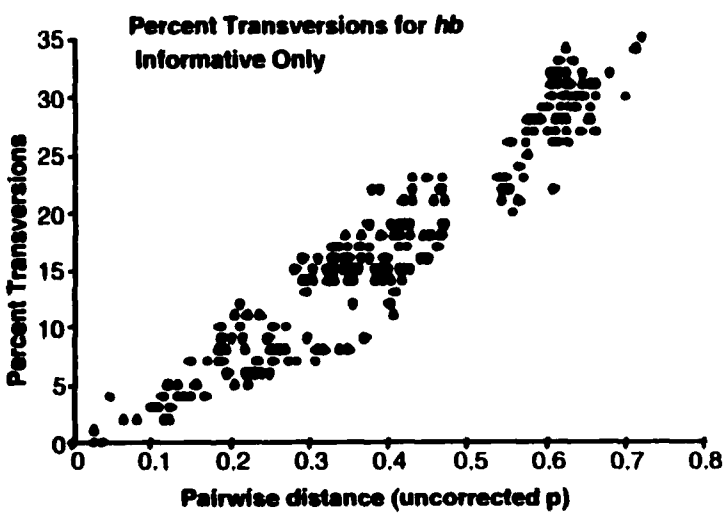


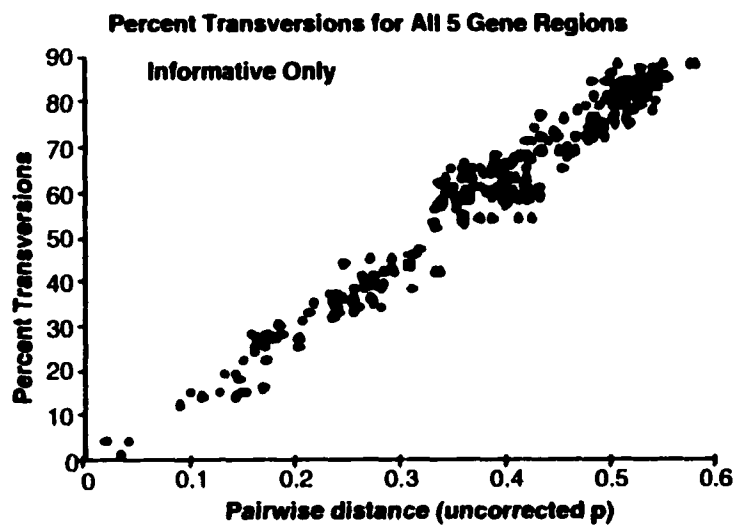


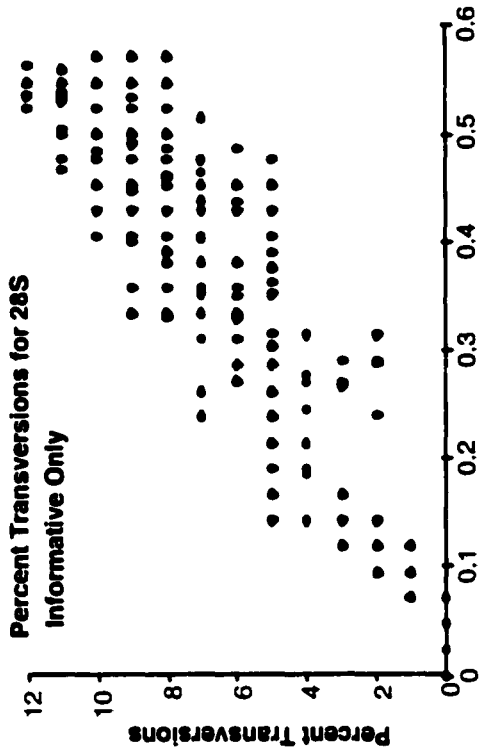


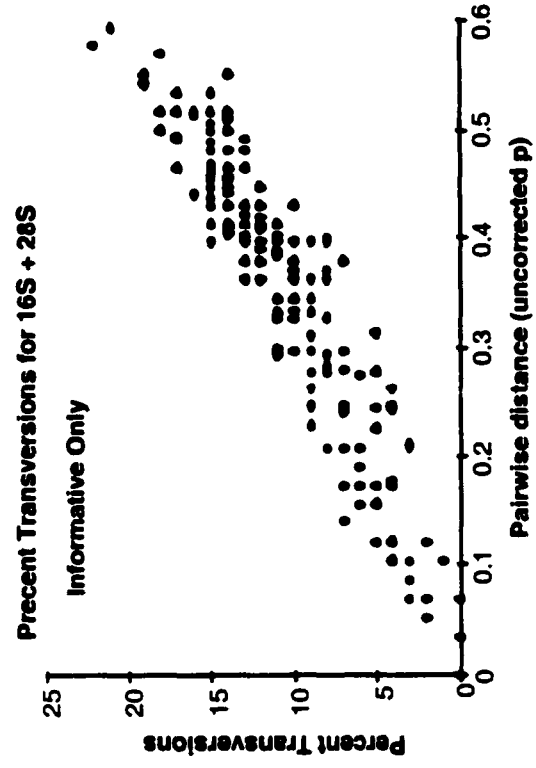


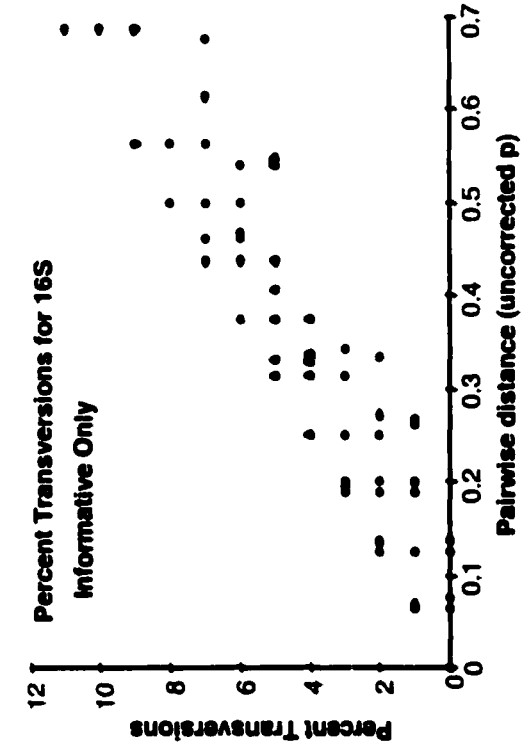


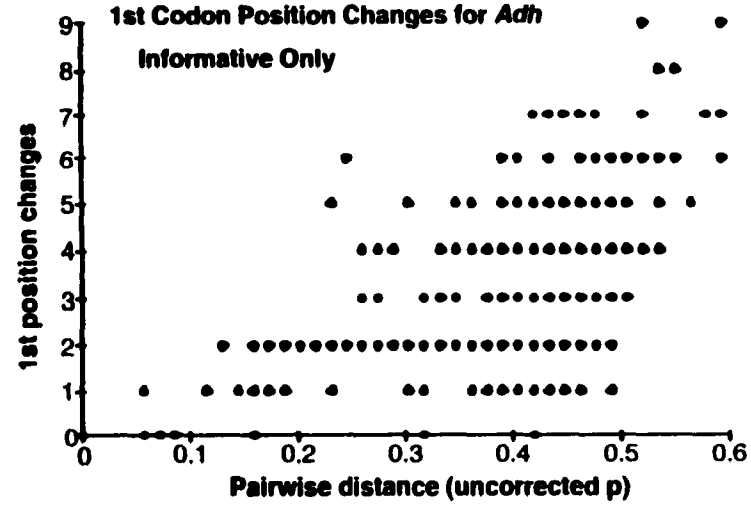


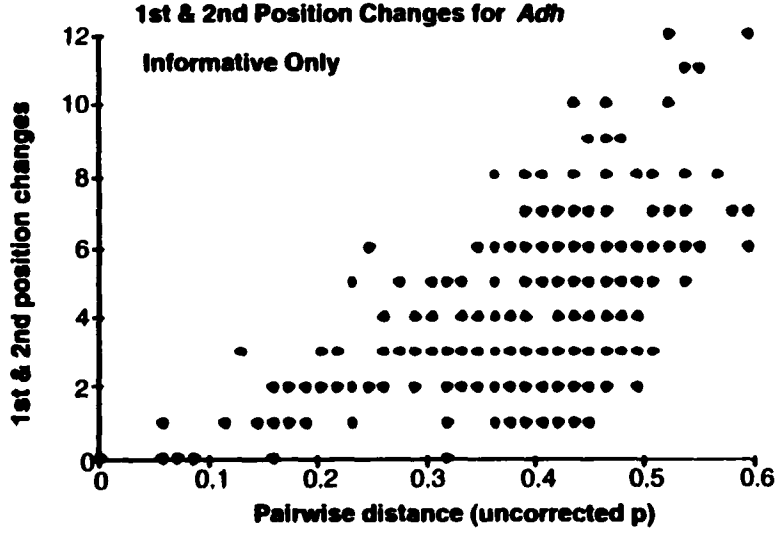


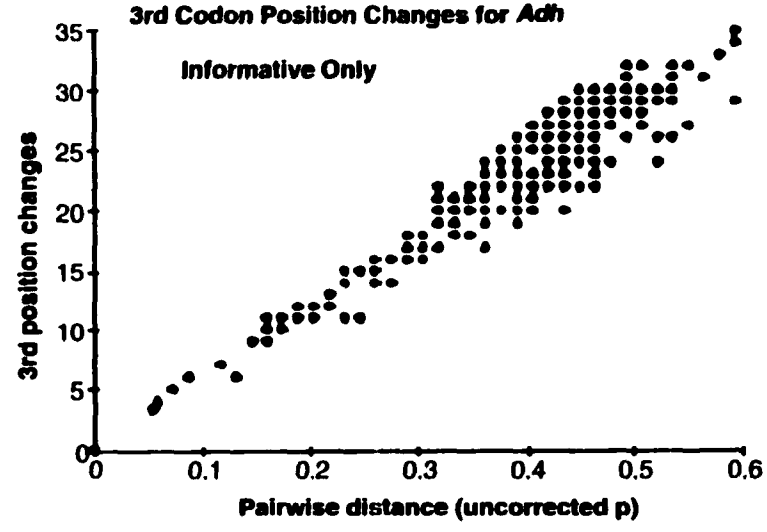


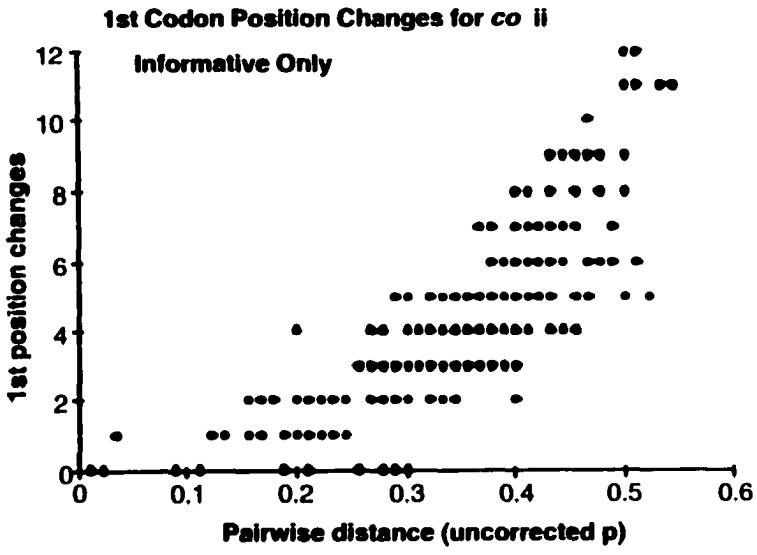


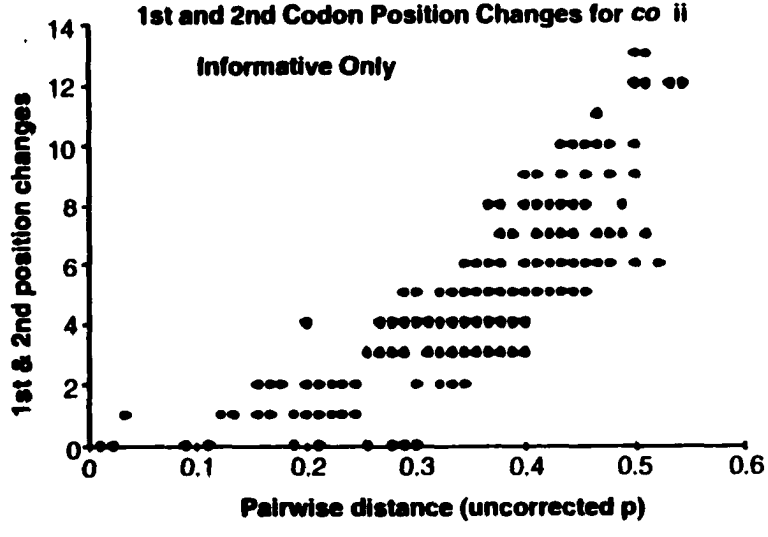


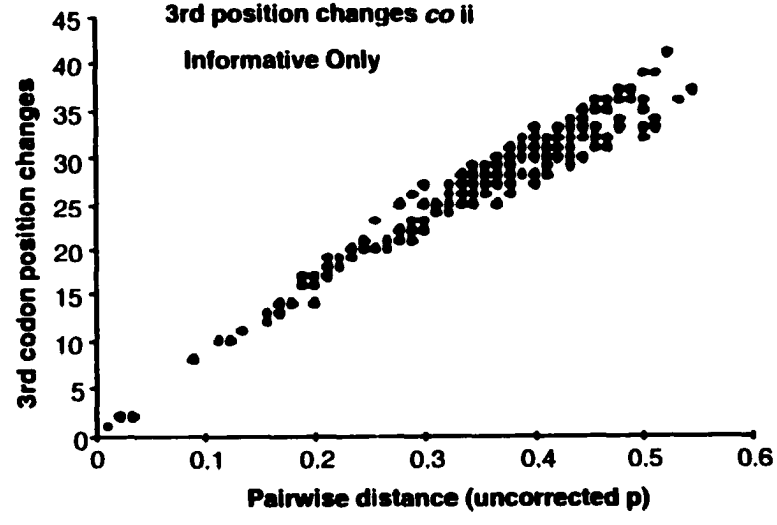


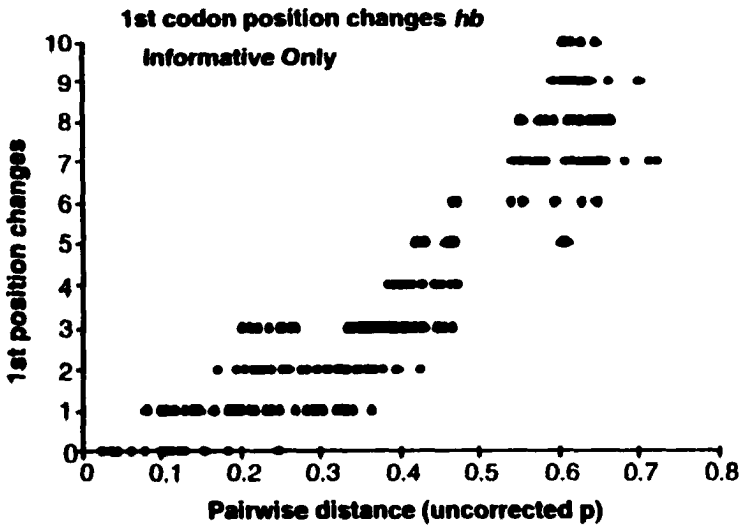


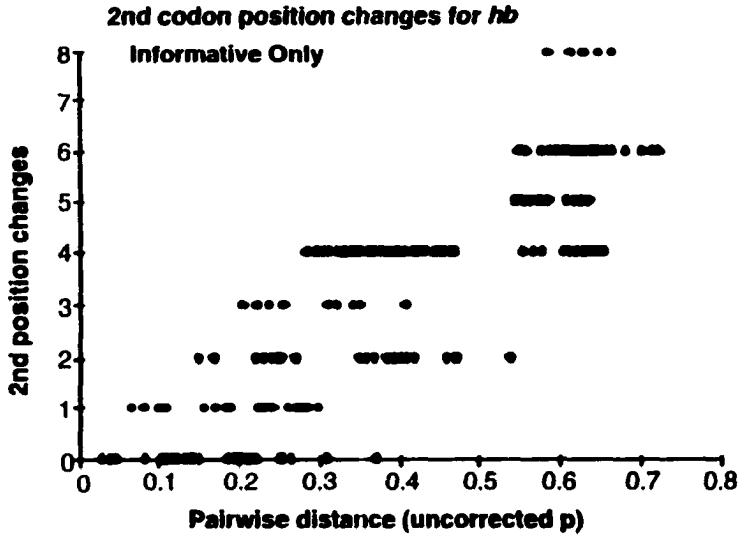


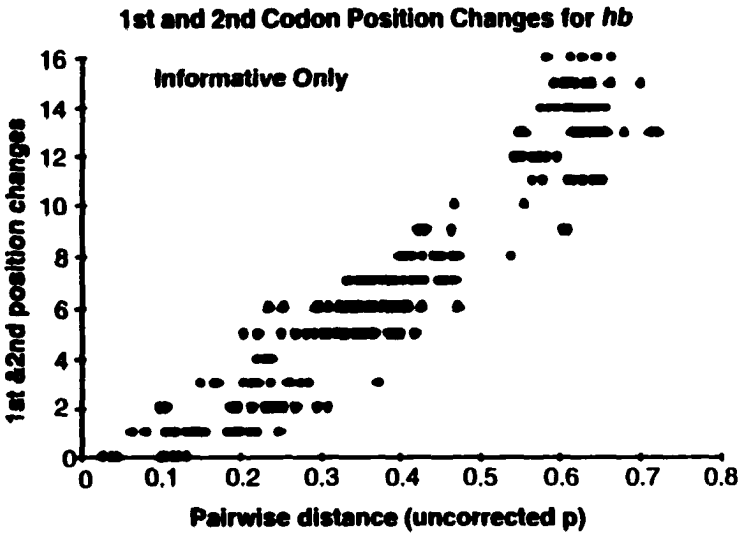


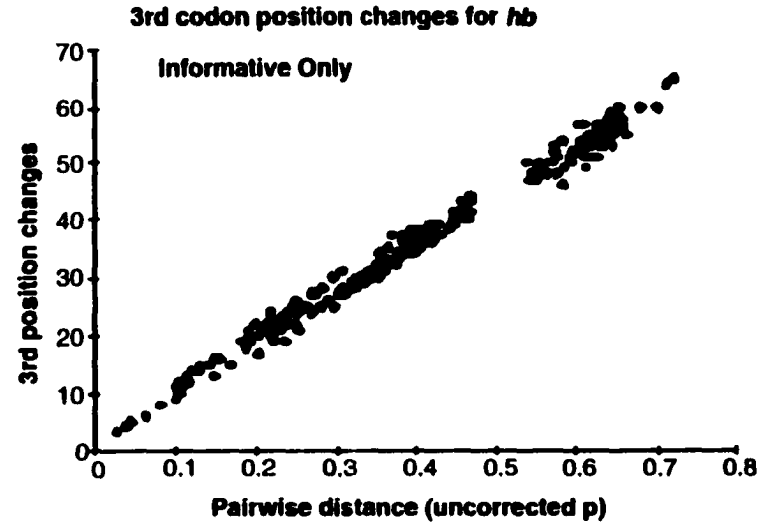


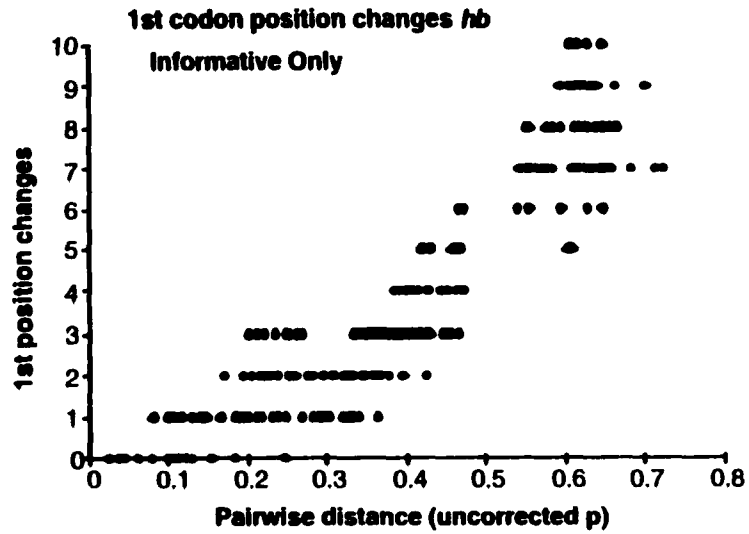


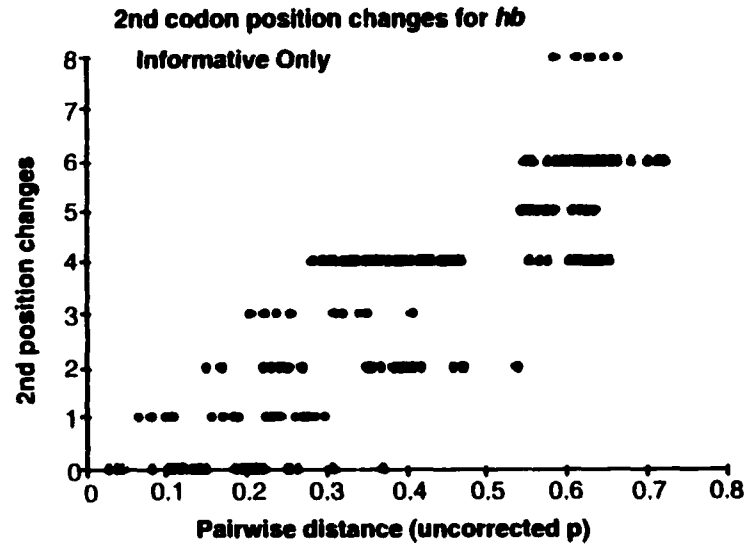


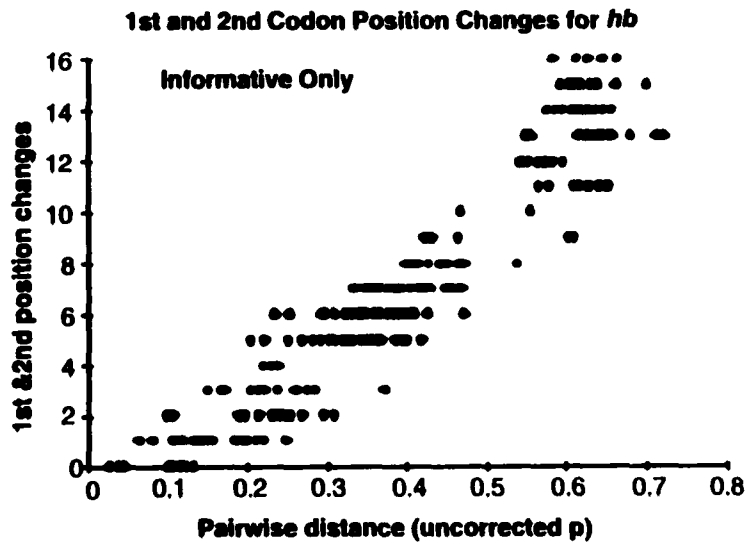


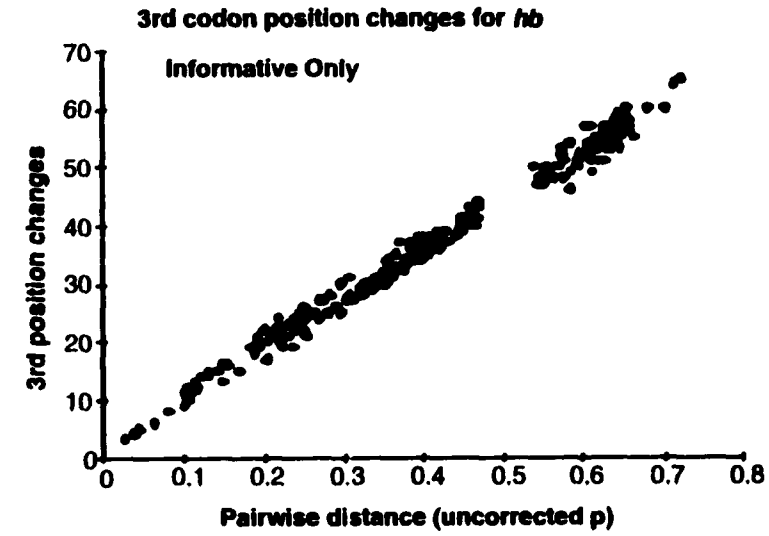


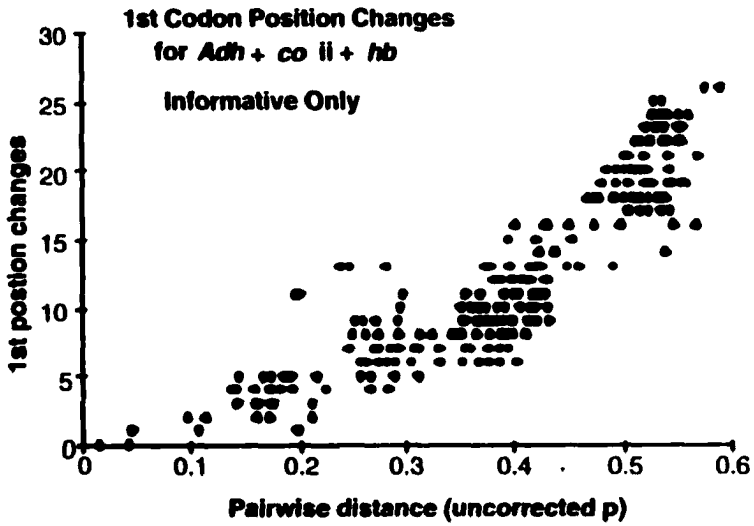


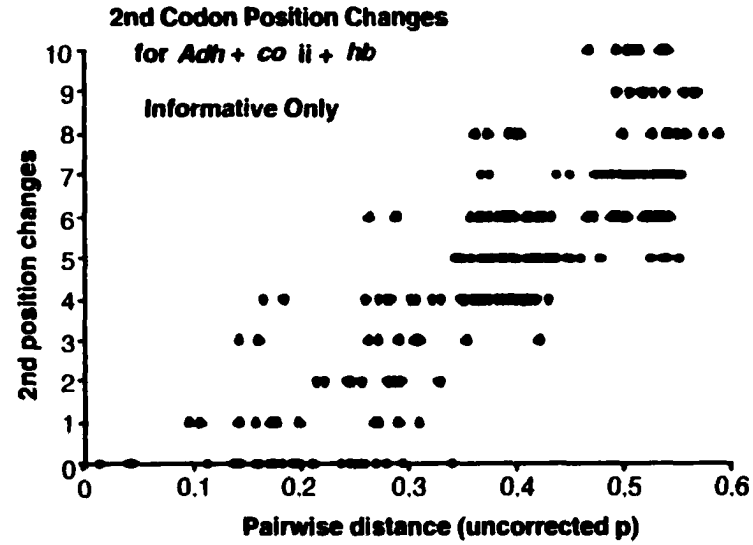


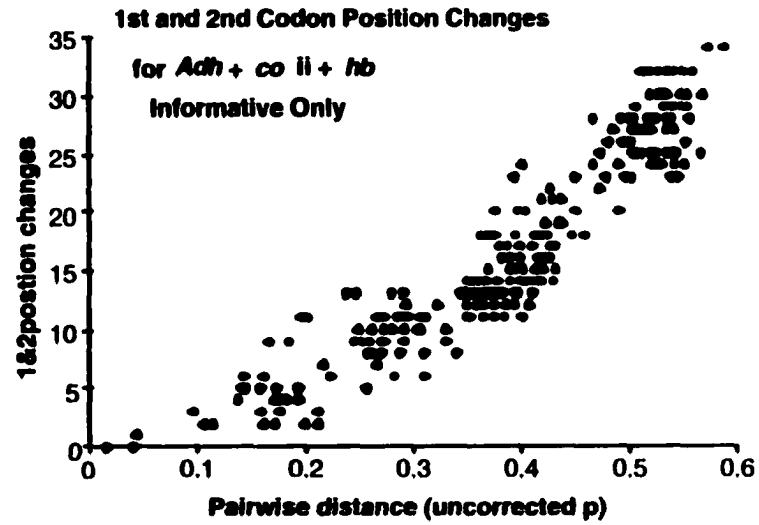


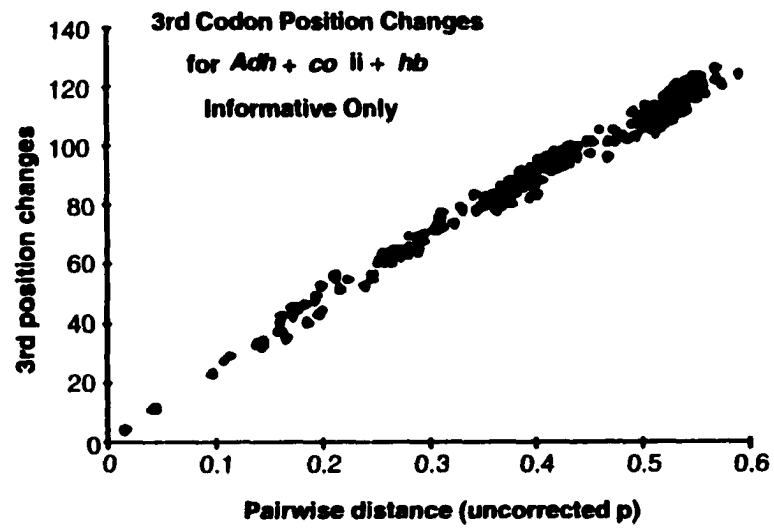






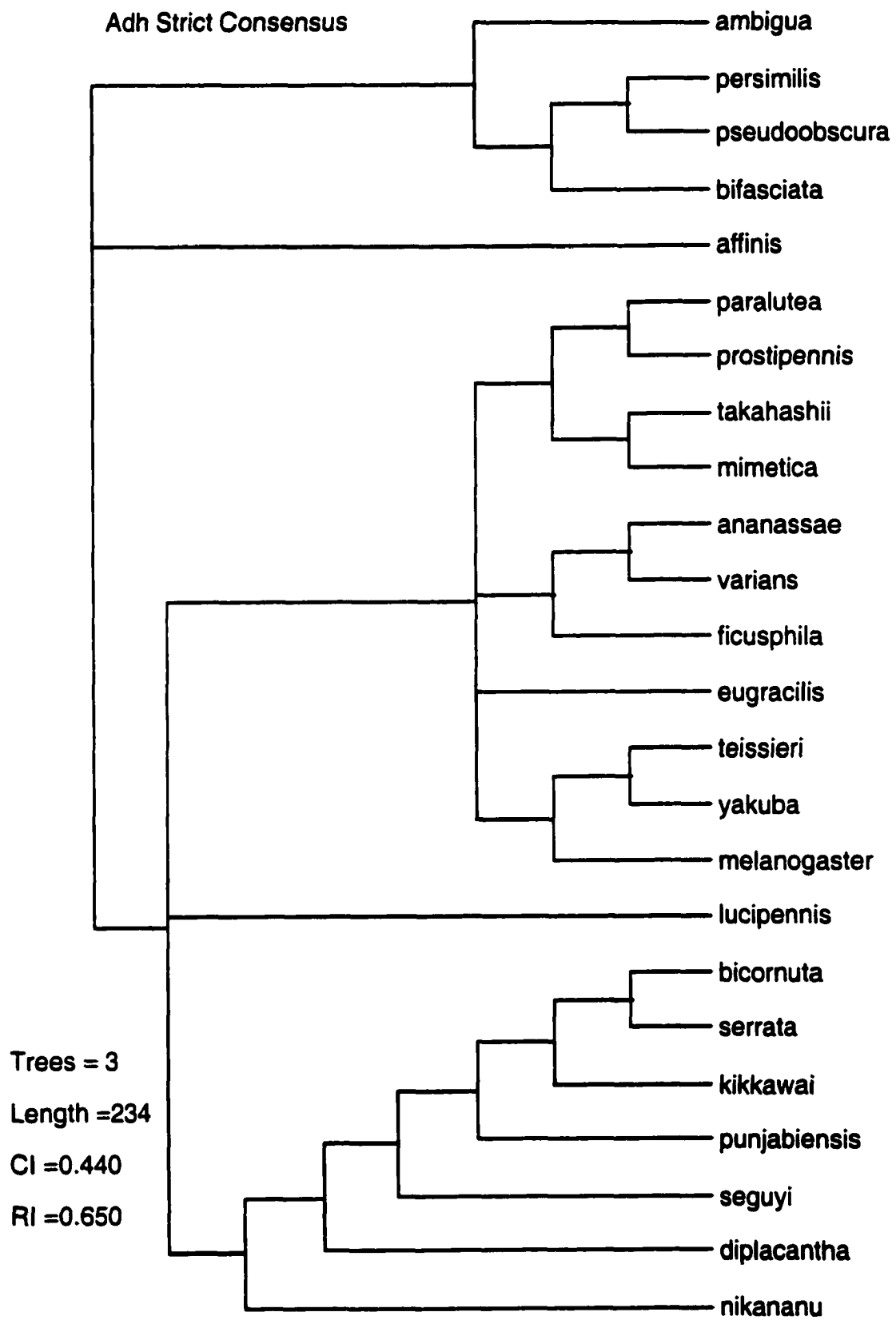


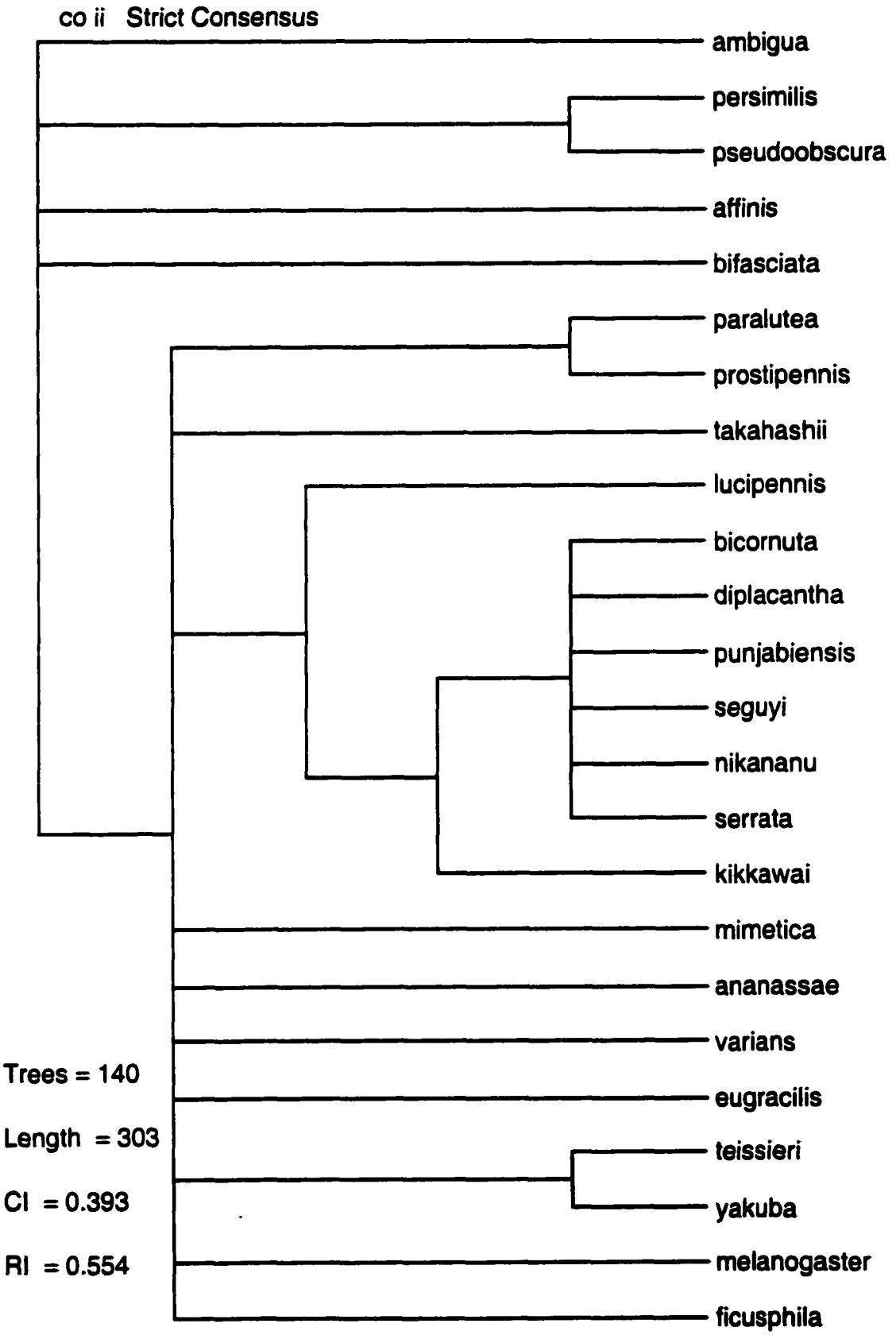


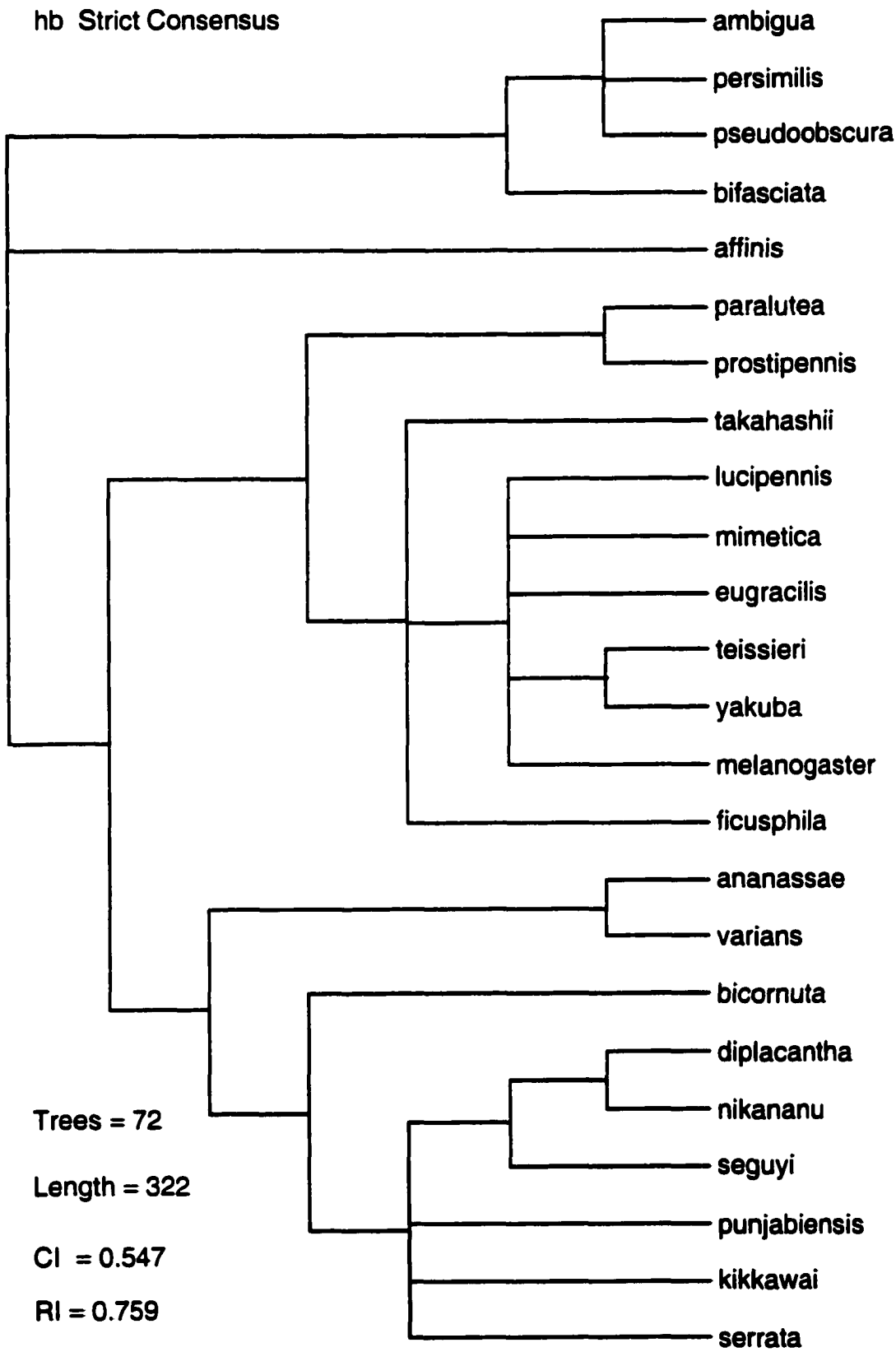


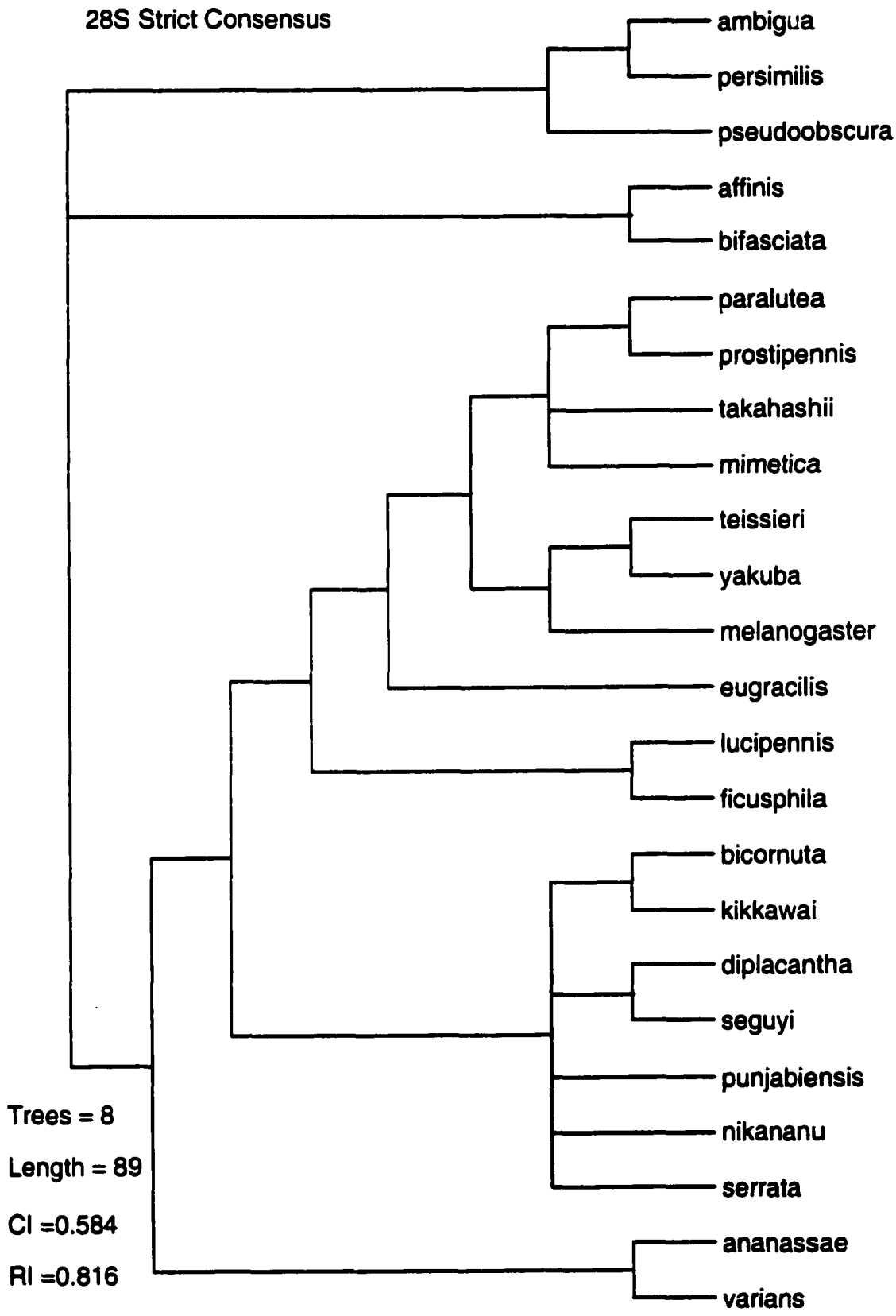
APPENDIX E: Cladograms for the Various Data Partitions (Chapter 1)

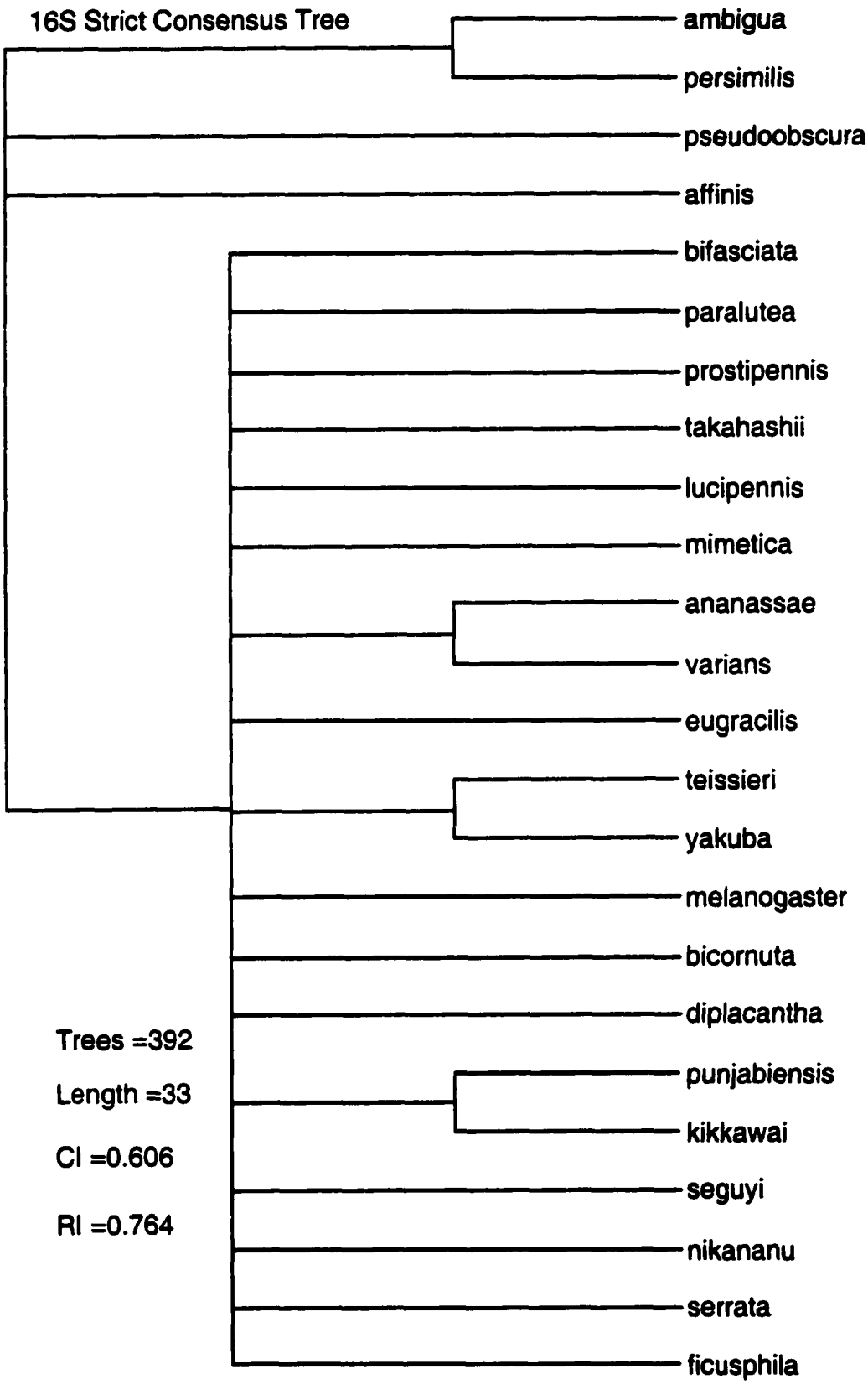
The strict consensus cladograms resulting from the phylogenetic analyses are presented here in the following order: *Adh*, *co ii*, *hb*, 28S, 16S, *Adh* + *co* ii, *Adh* + *hb*, *Adh* + 28S, *Adh* + 16S, *co* ii + *hb*, *co* ii +28S, *co* ii + 16S, *hb* + 28S, *hb* + 16S, 16S + 28S, *Adh* + *co* ii + *hb*, *Adh* + *co* ii + 28S, *Adh* + *co* ii + 16S, *Adh* + *hb* + 28S, *Adh* + *hb* + 16S, *Adh* + 16S + 28S, *co* ii + *hb* + 28S, *co* ii + *hb* + 16S, *co* ii + 16S + 28S, *hb* + 16S + 28S, *Adh* + *co* ii + *hb* + 28S, *Adh* + *co* ii + *hb* + 16S, *Adh* + *co* ii + 16S + 28S, *Adh* + *hb* + 16S + 28S and *co* ii + *hb* + 16S + 28S which had only one most parsimonous tree. Tree statistics of number of equally most parsimonious trees, tree length (steps), CI and RI are reported for each of the cladograms.

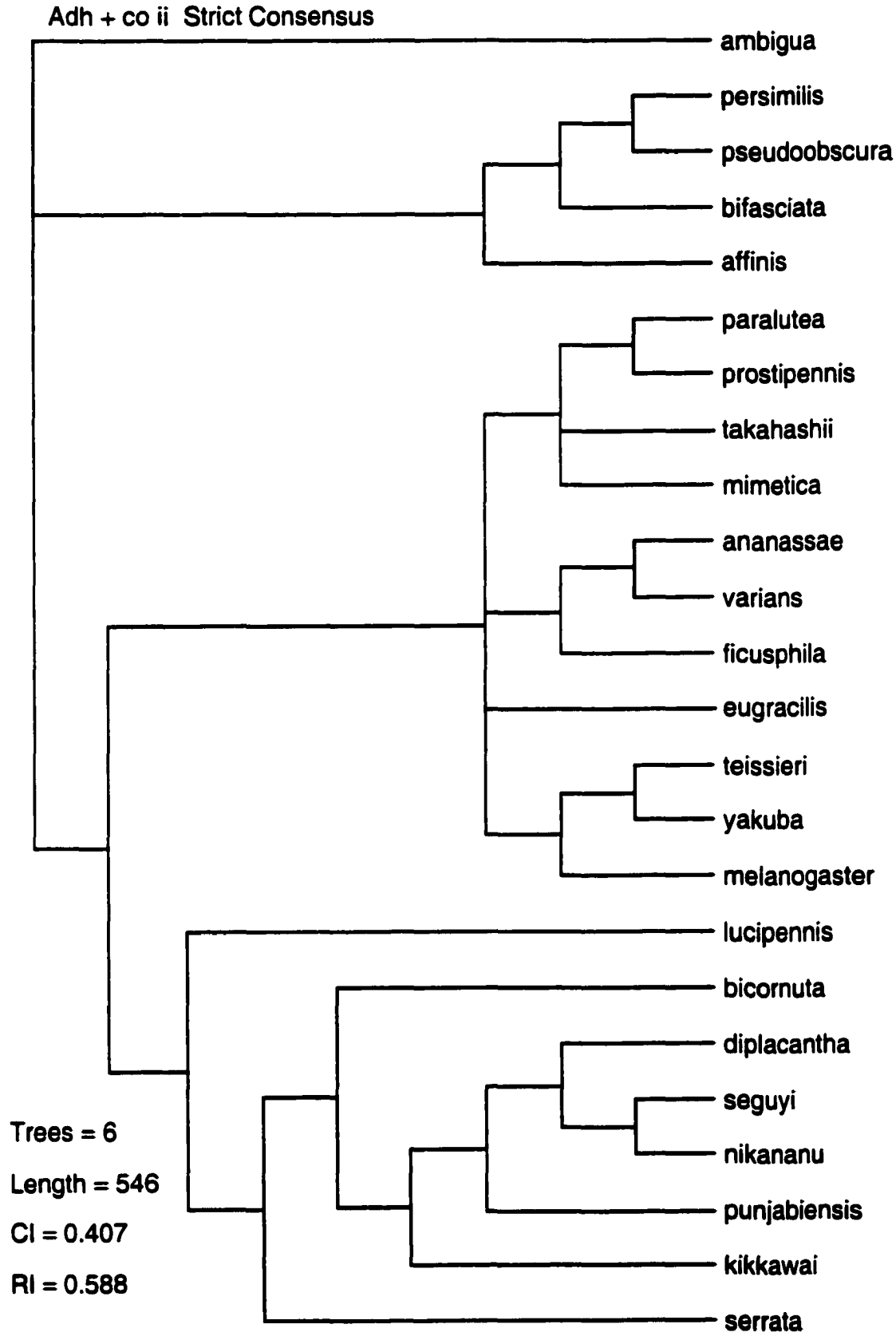


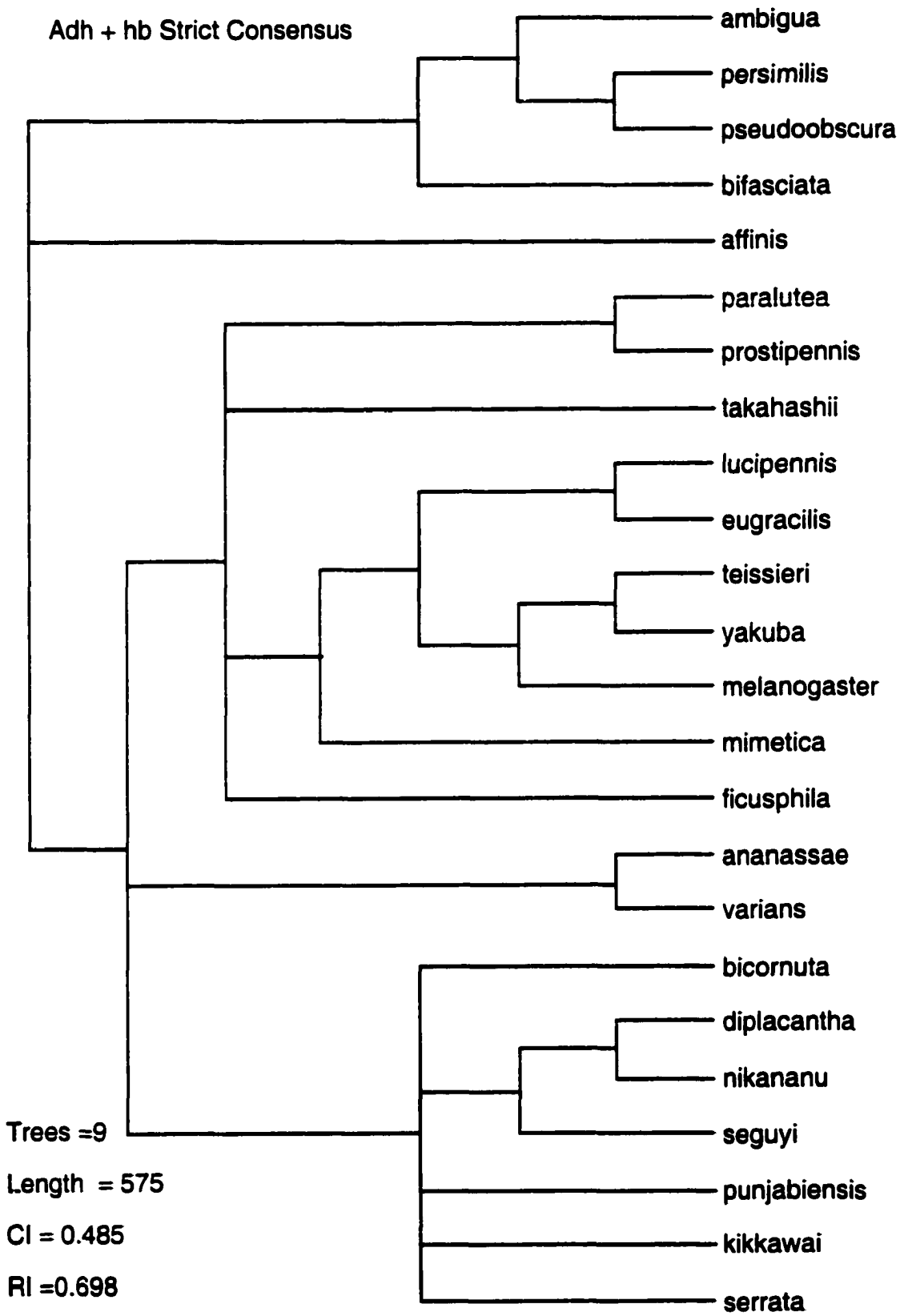


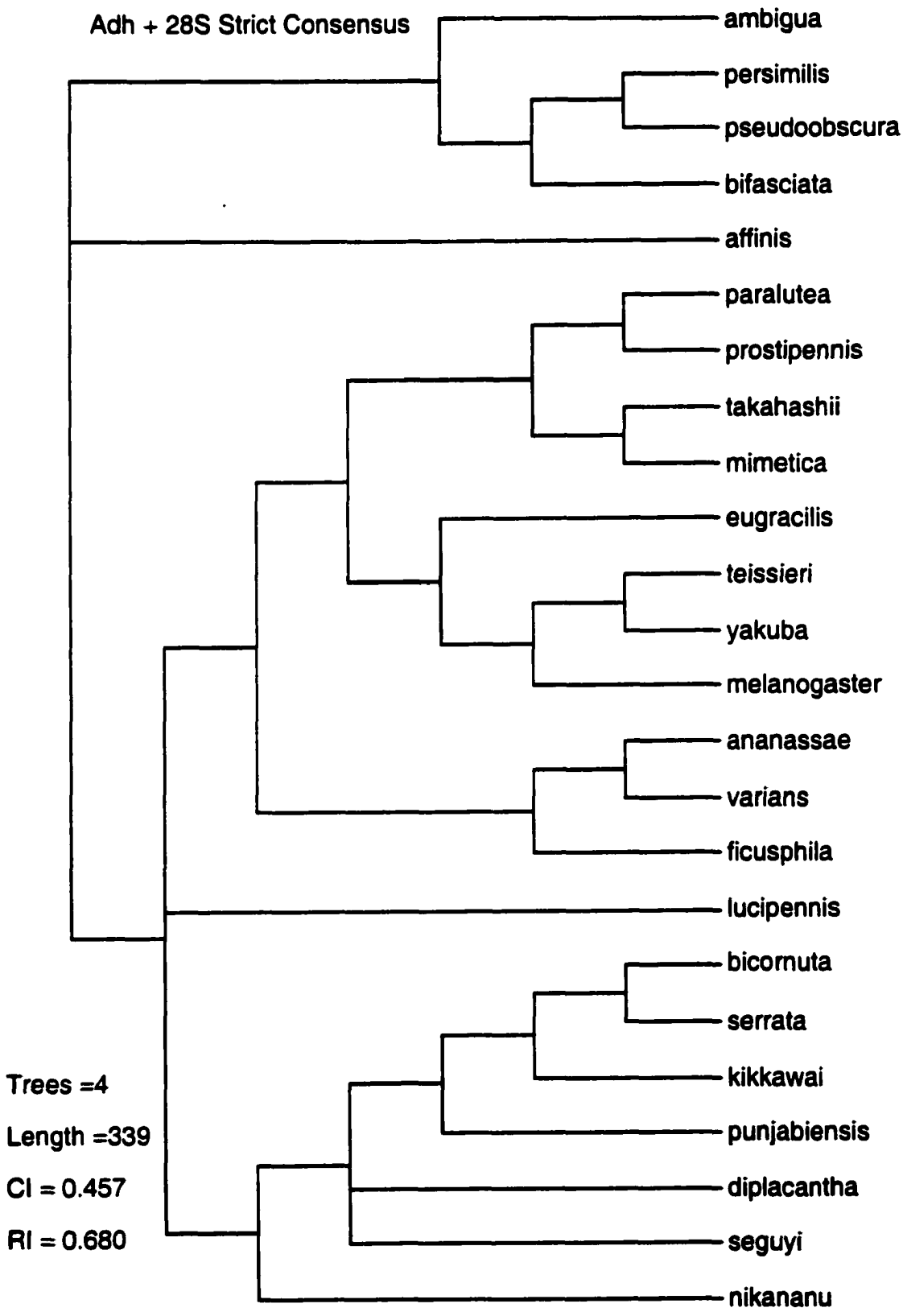


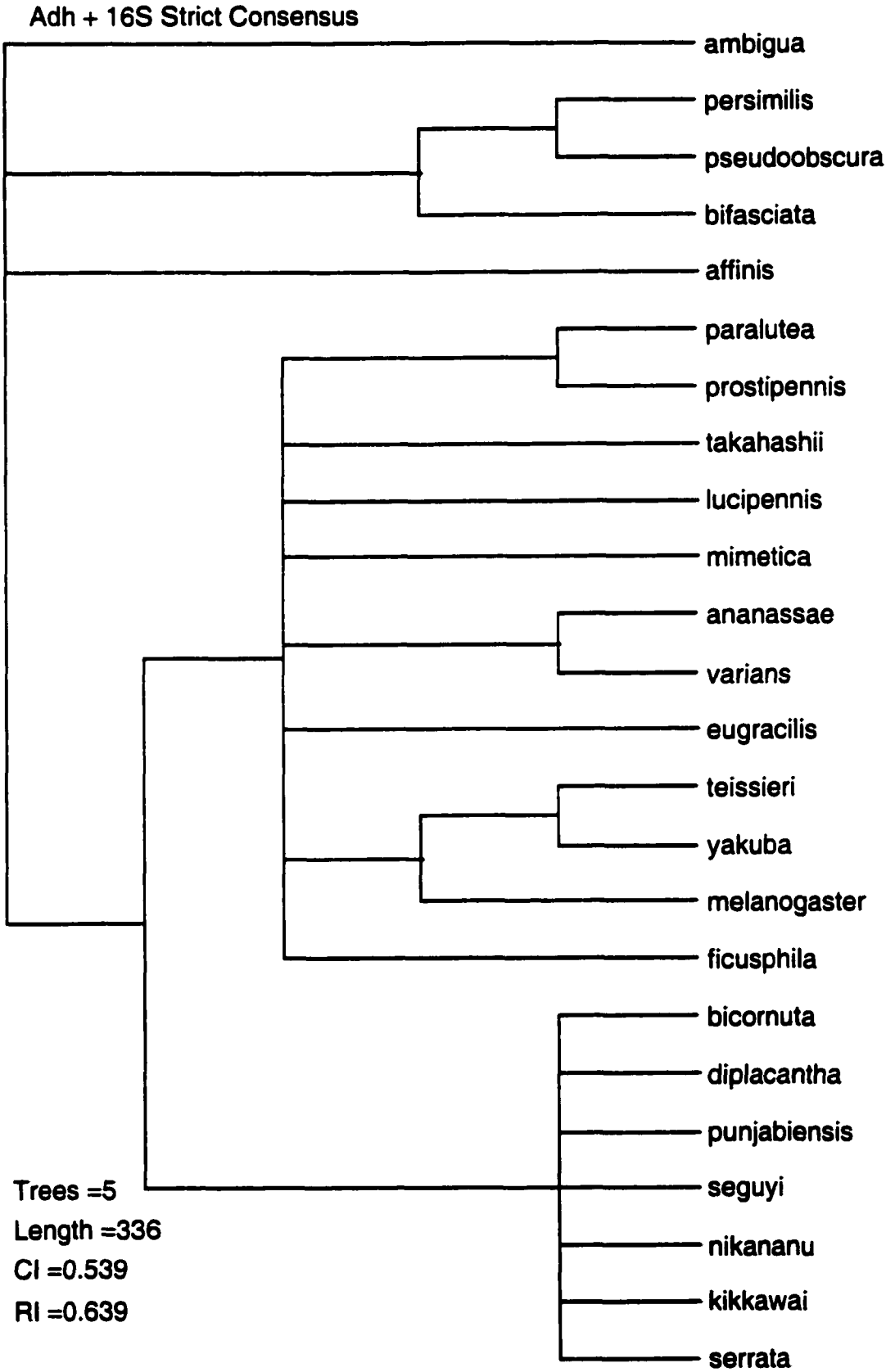


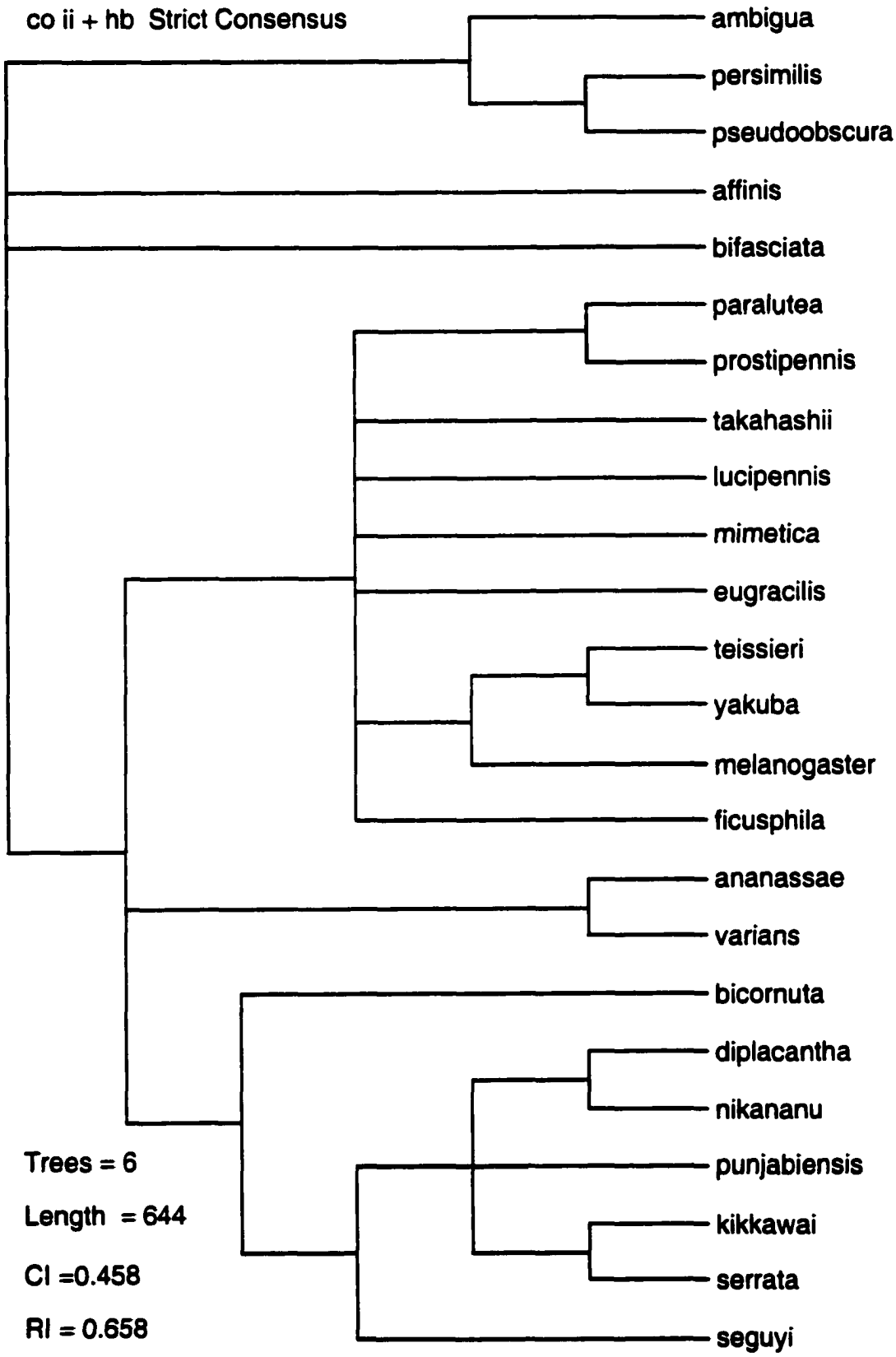


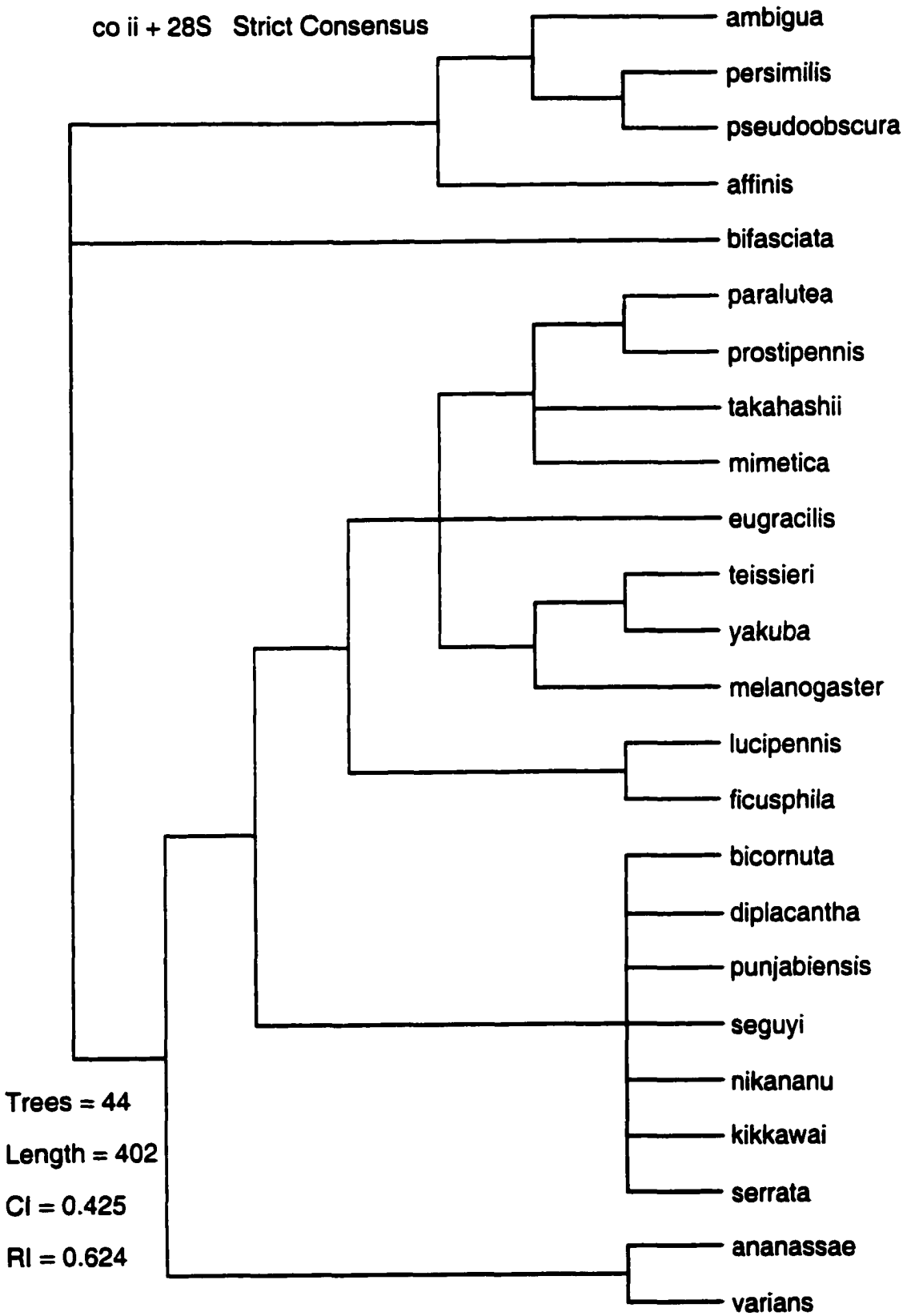


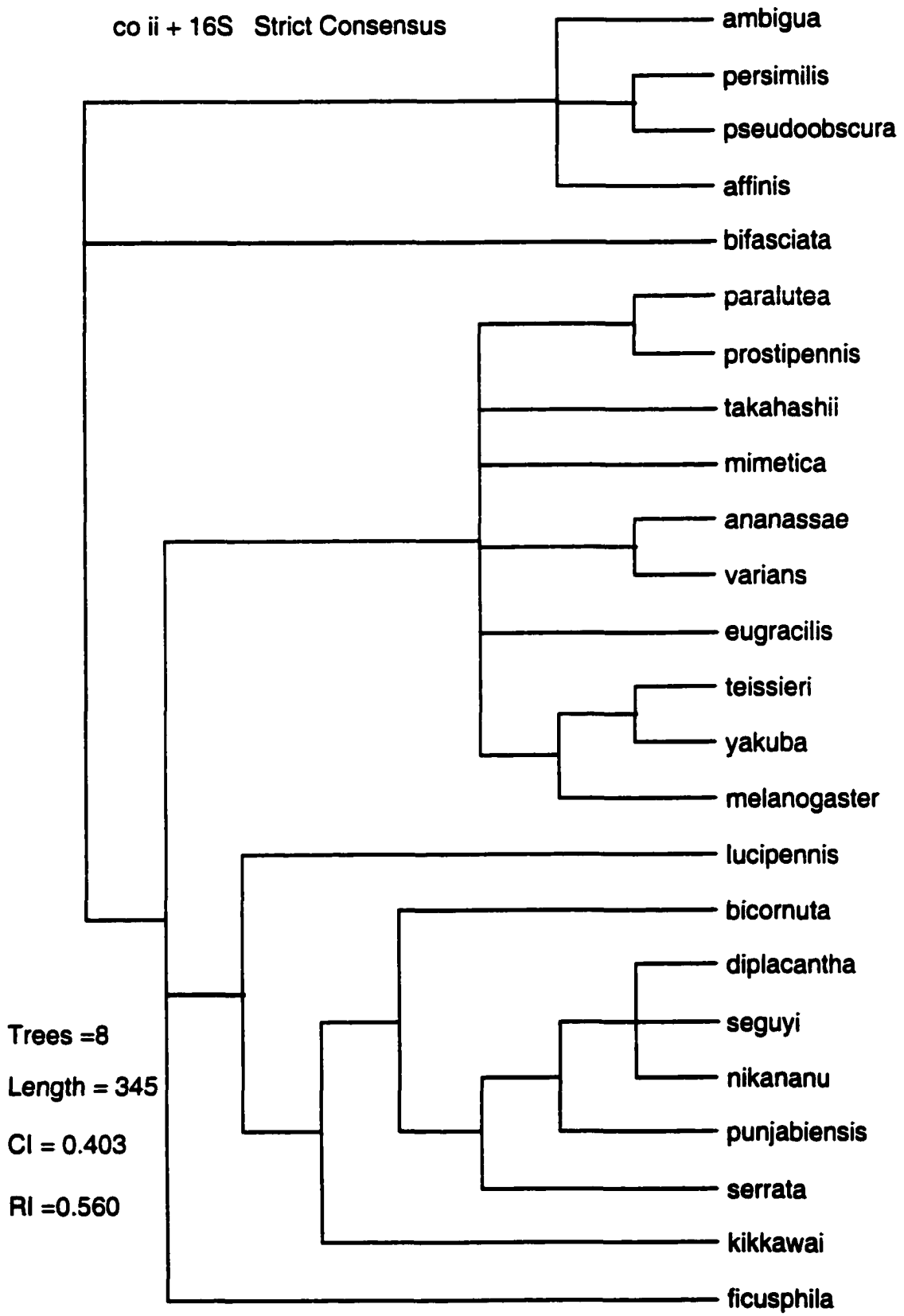


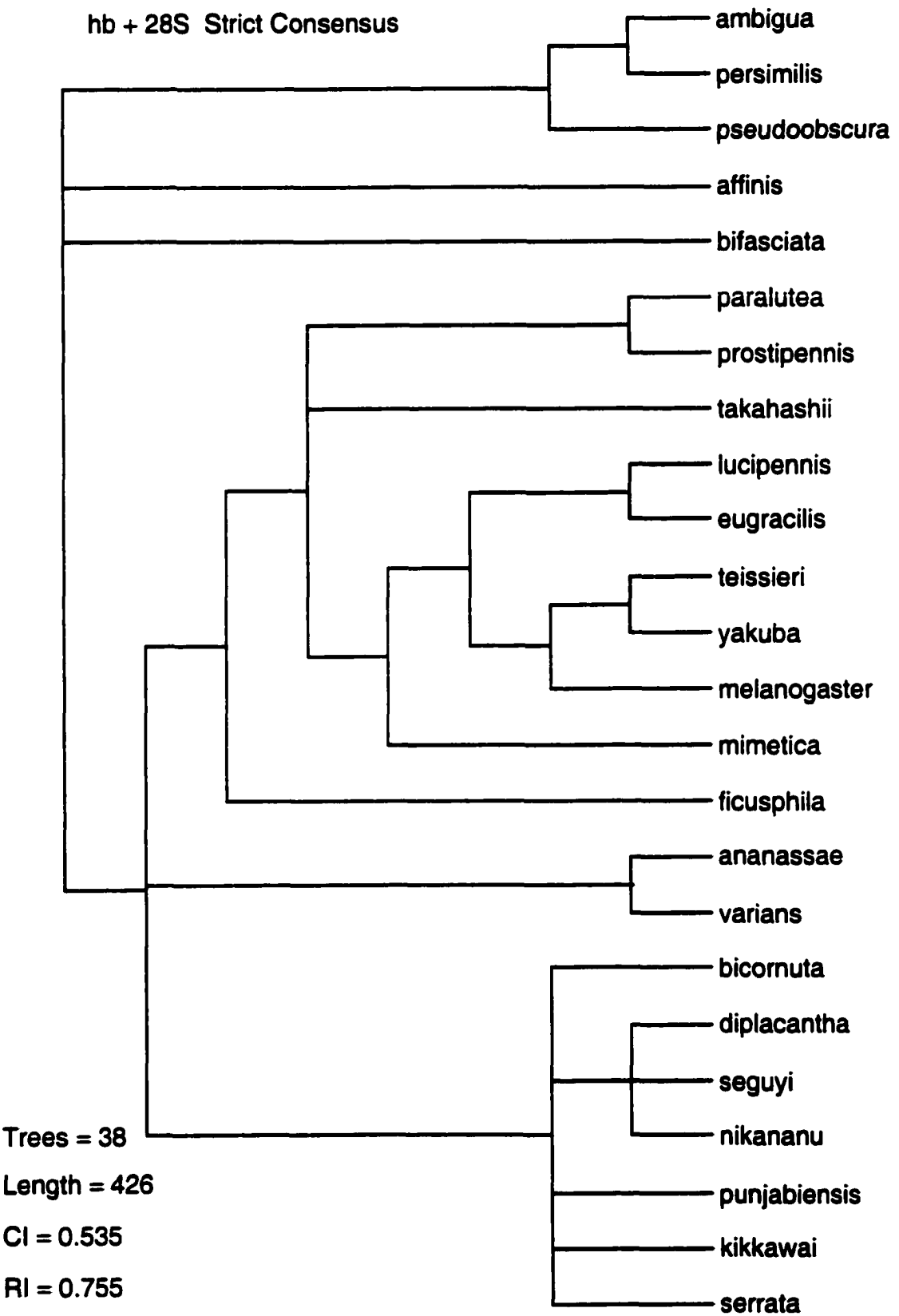


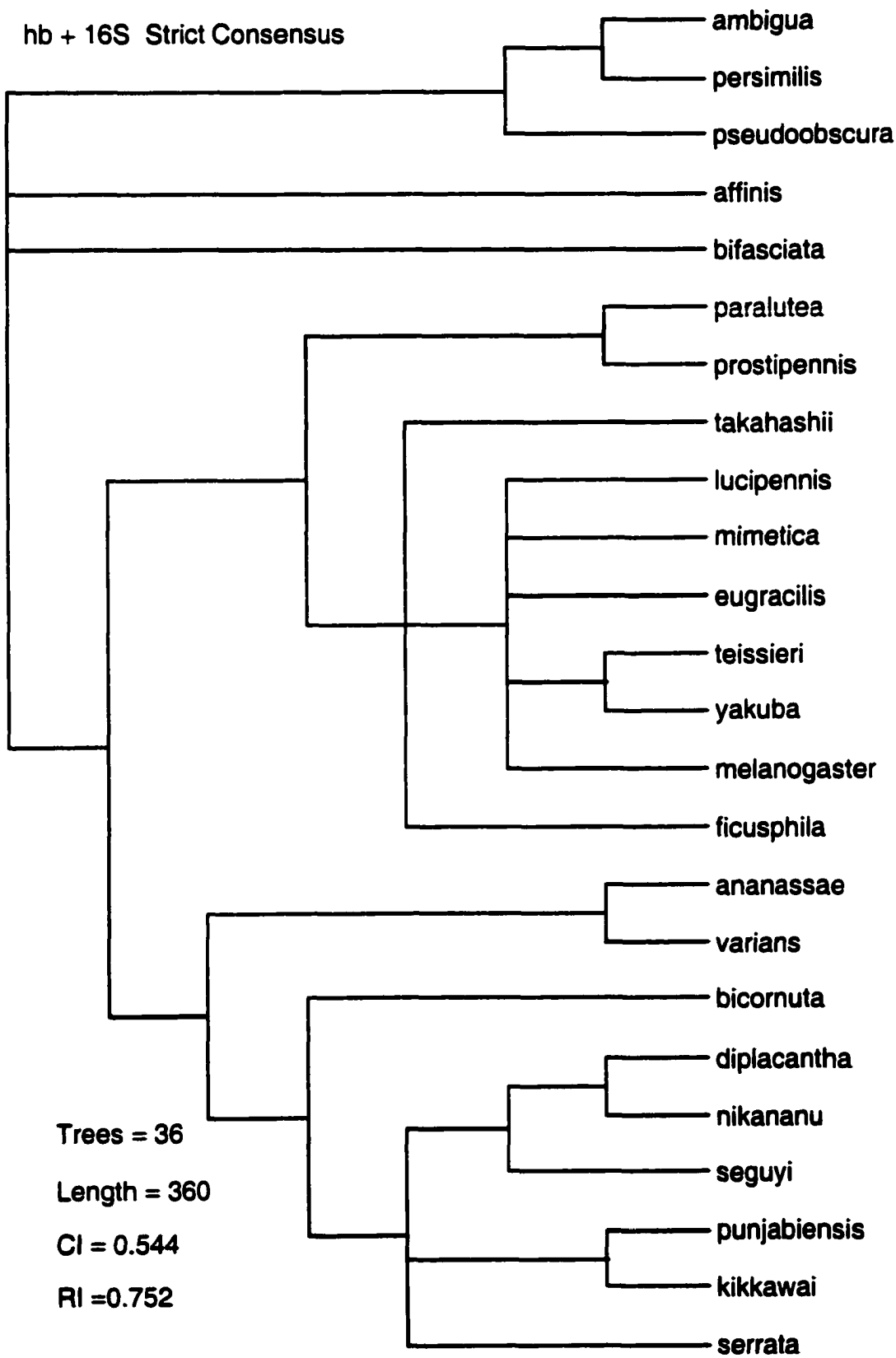


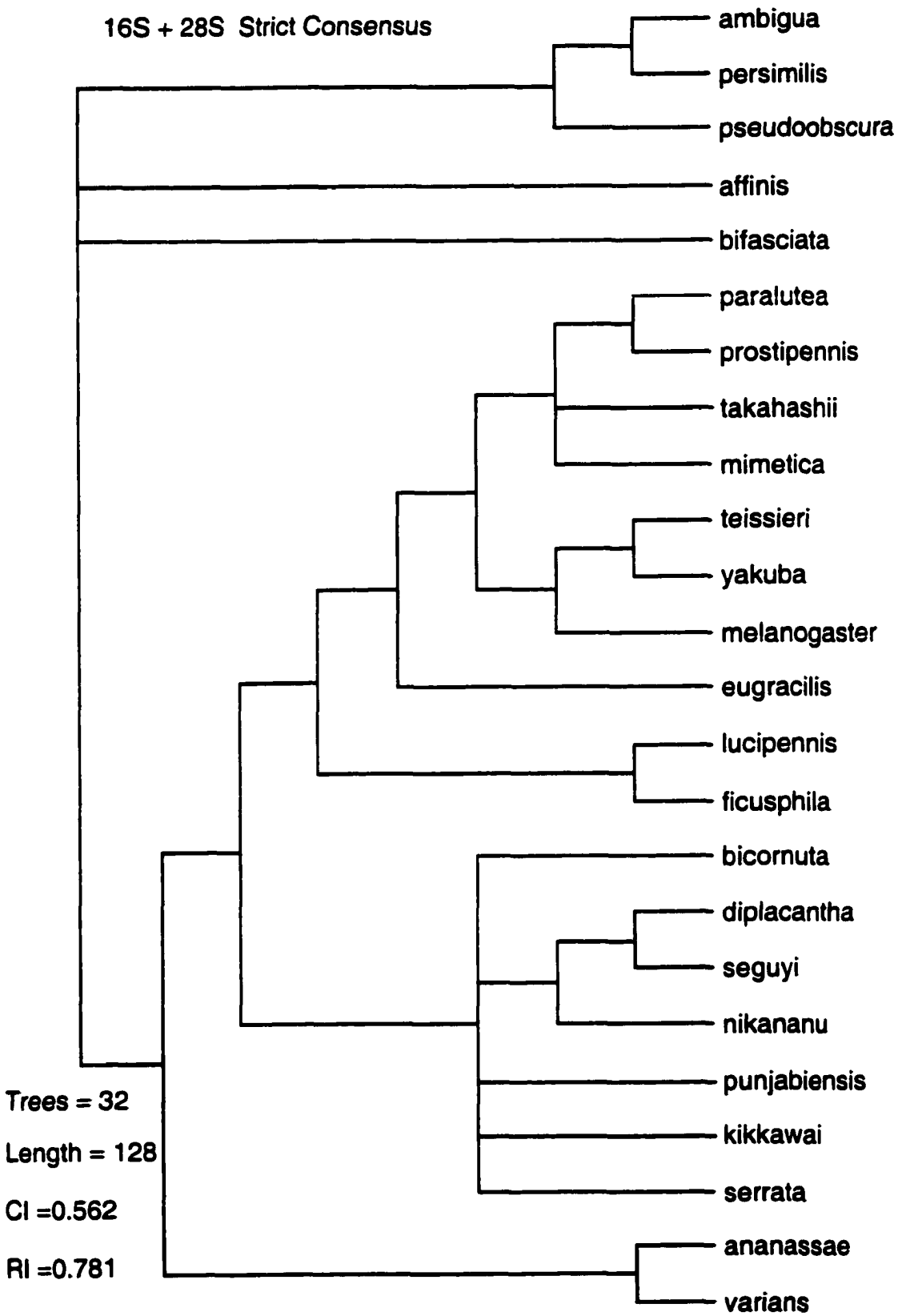


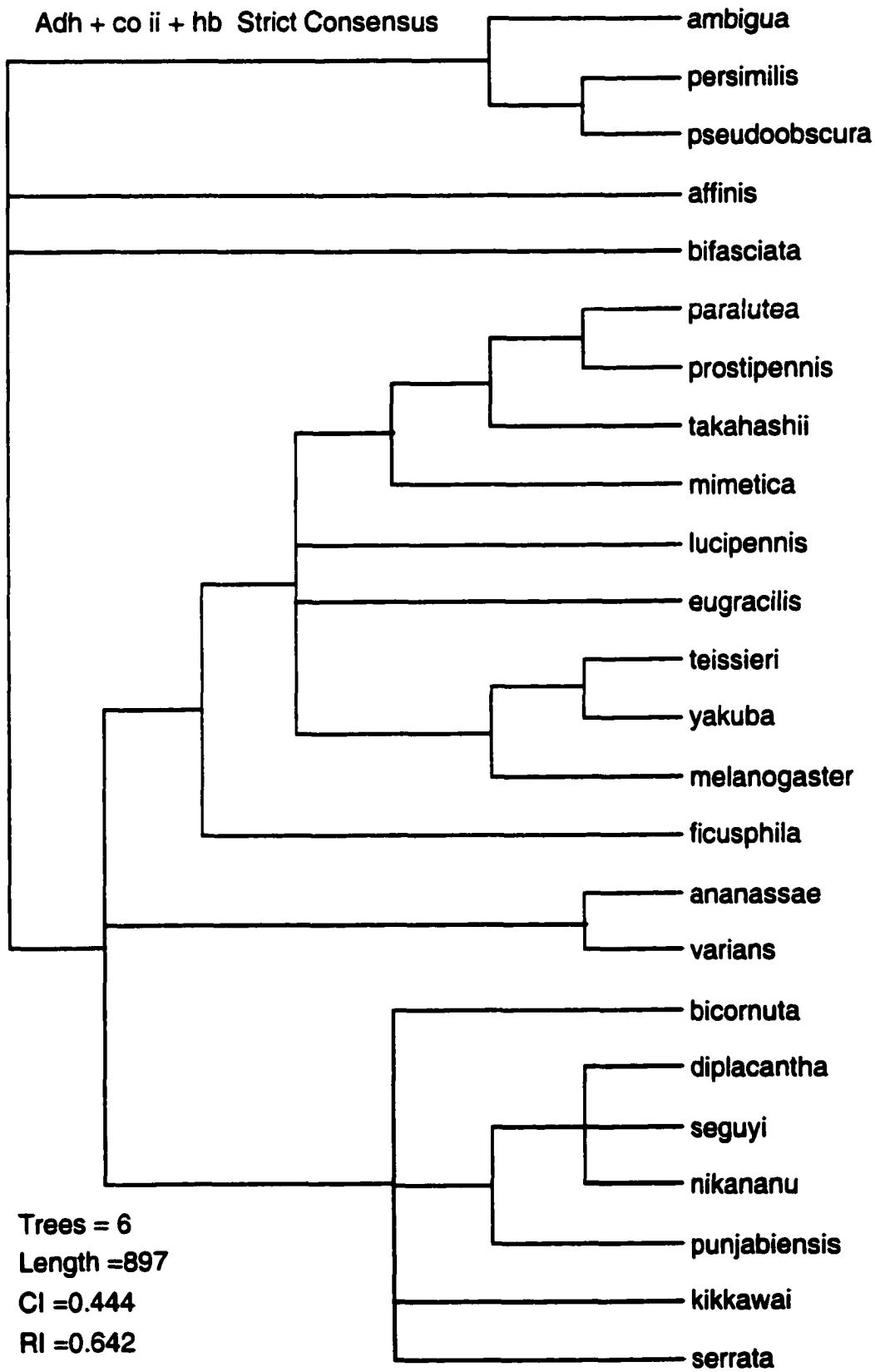


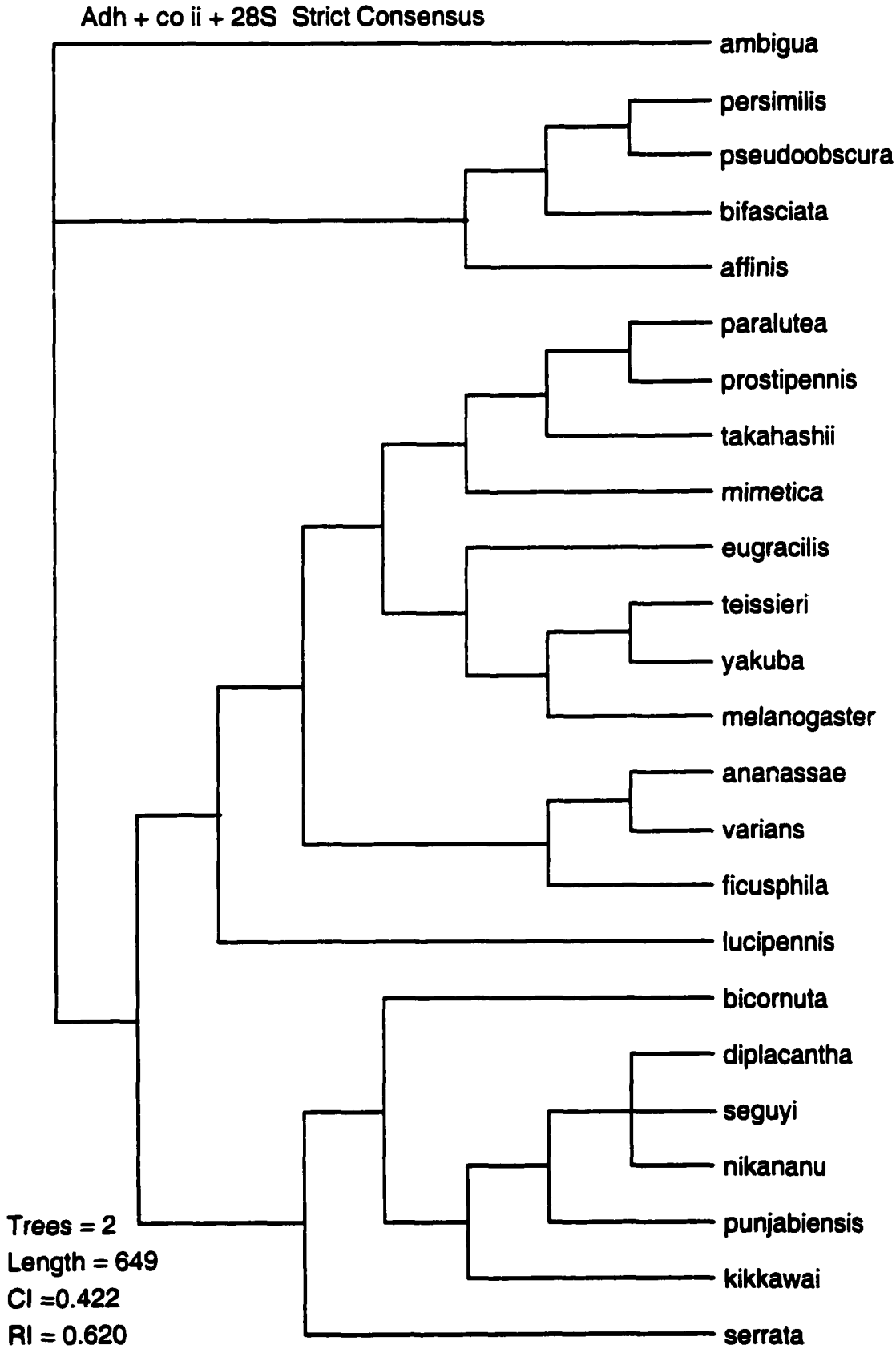


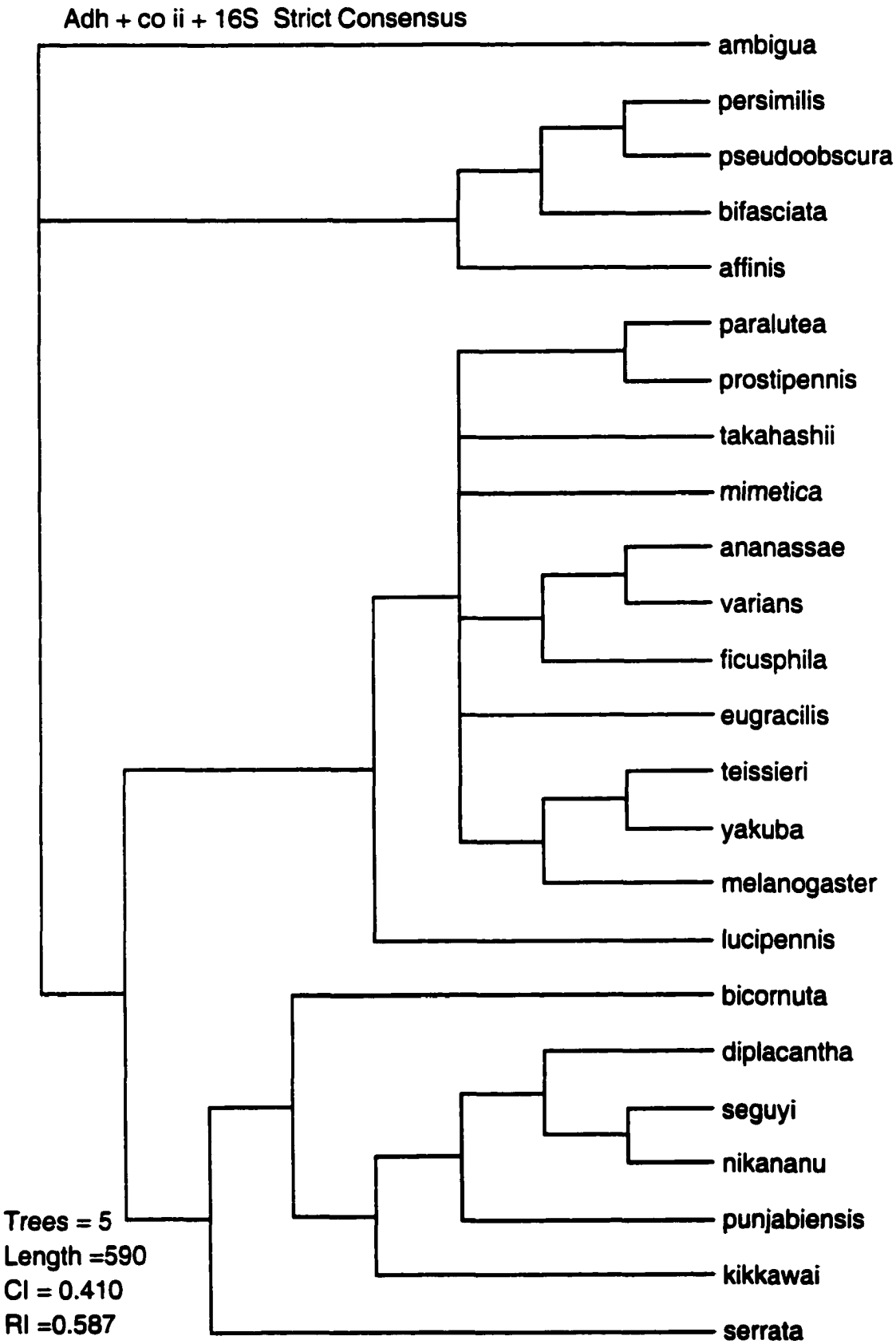


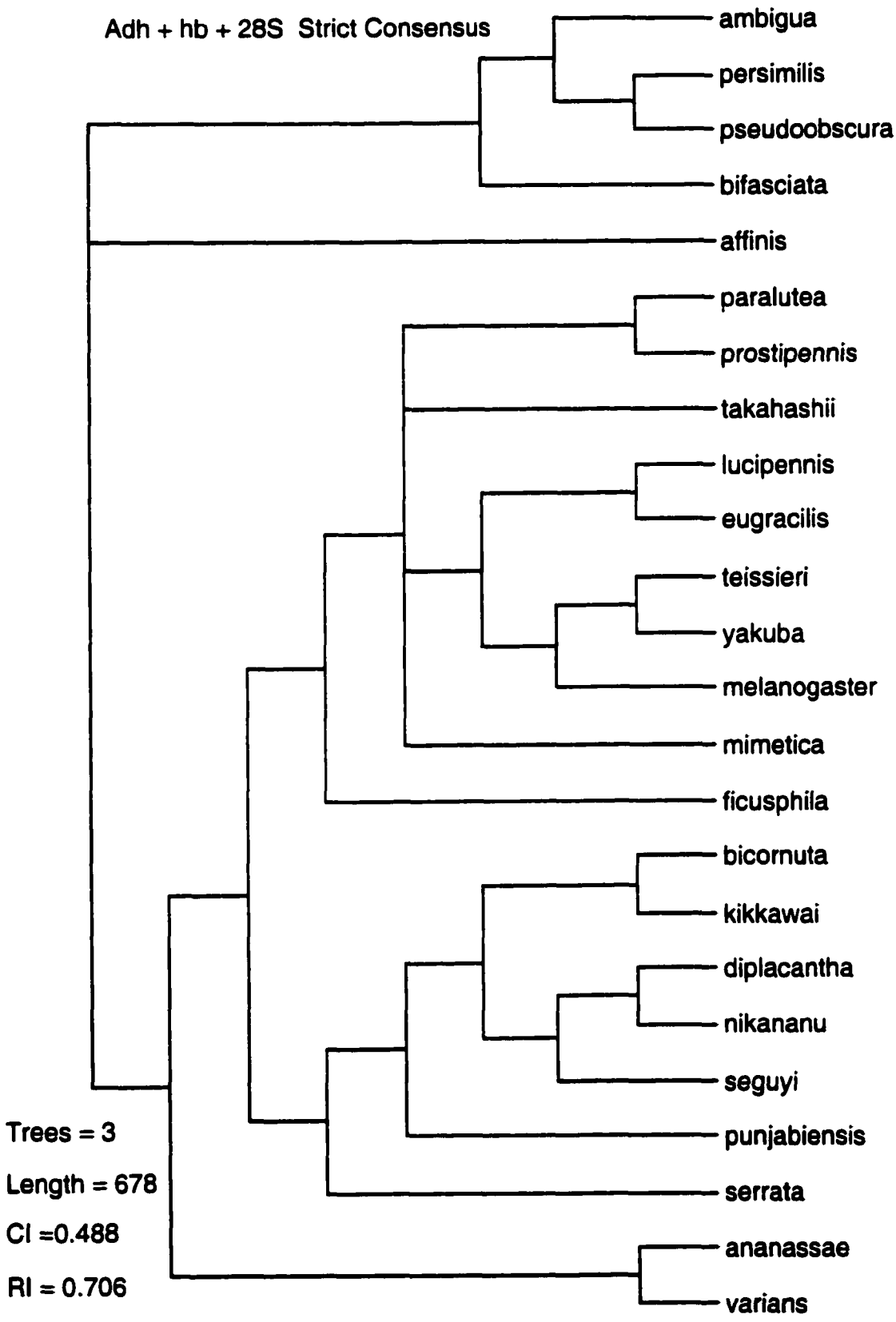


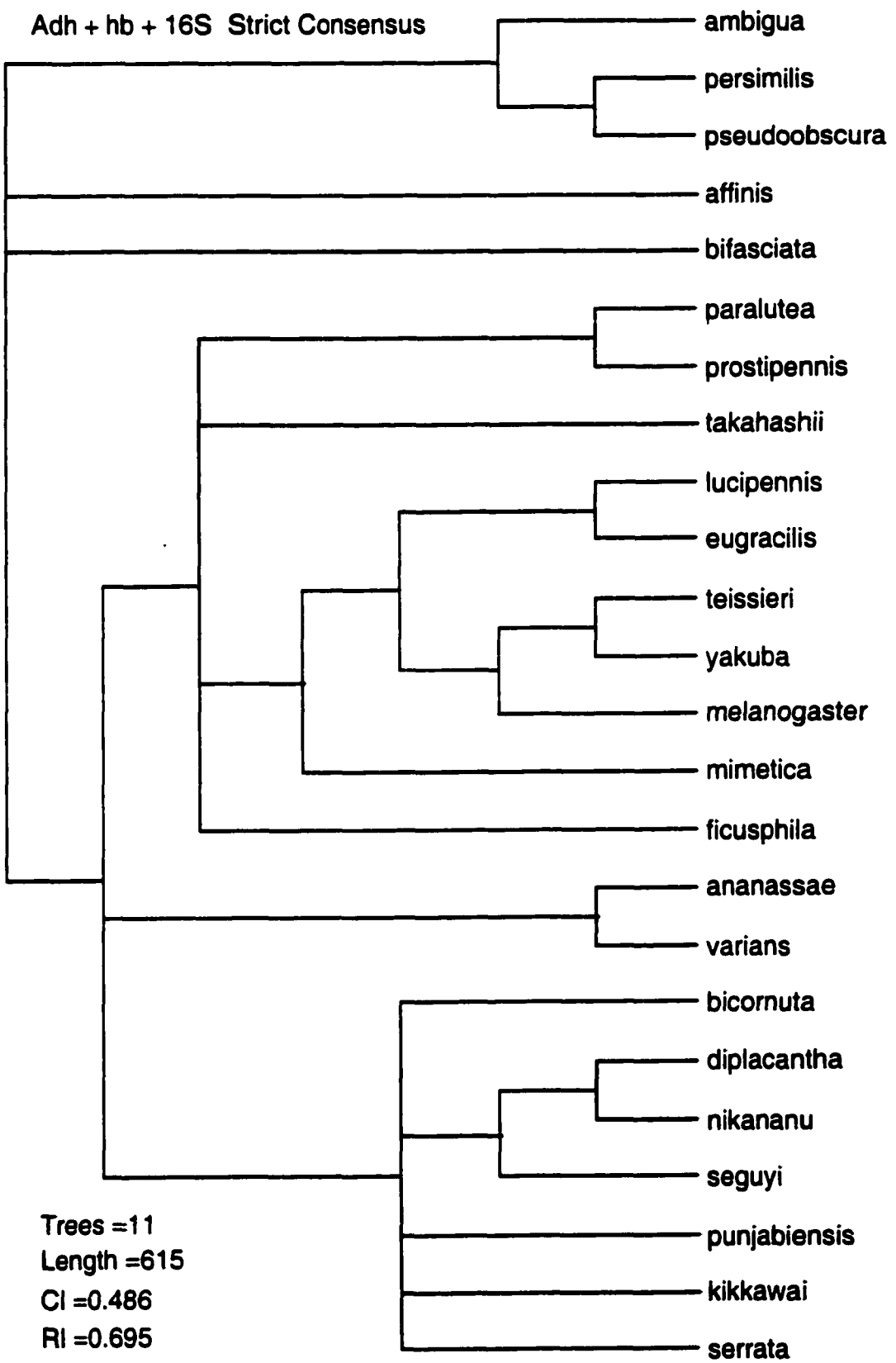


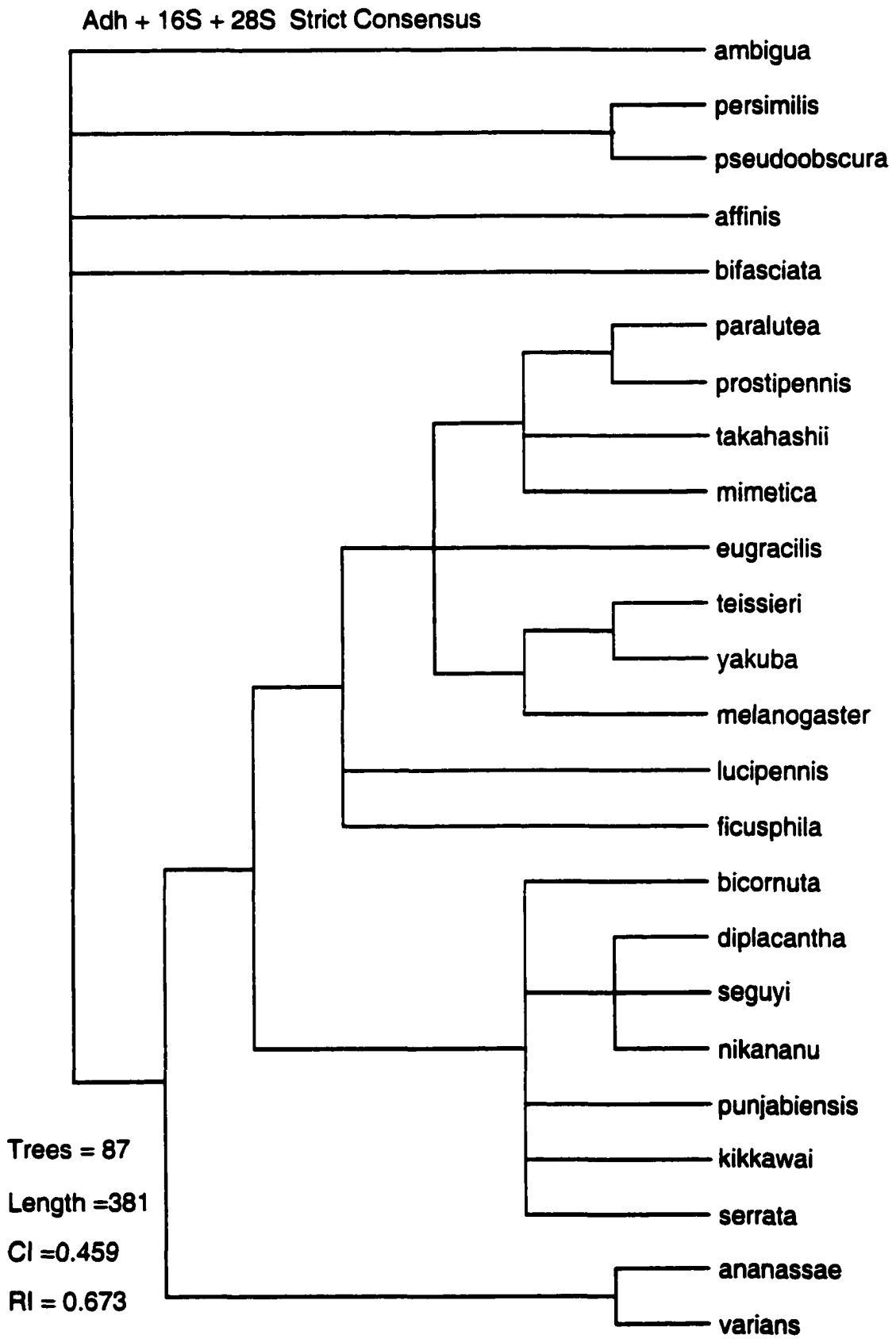


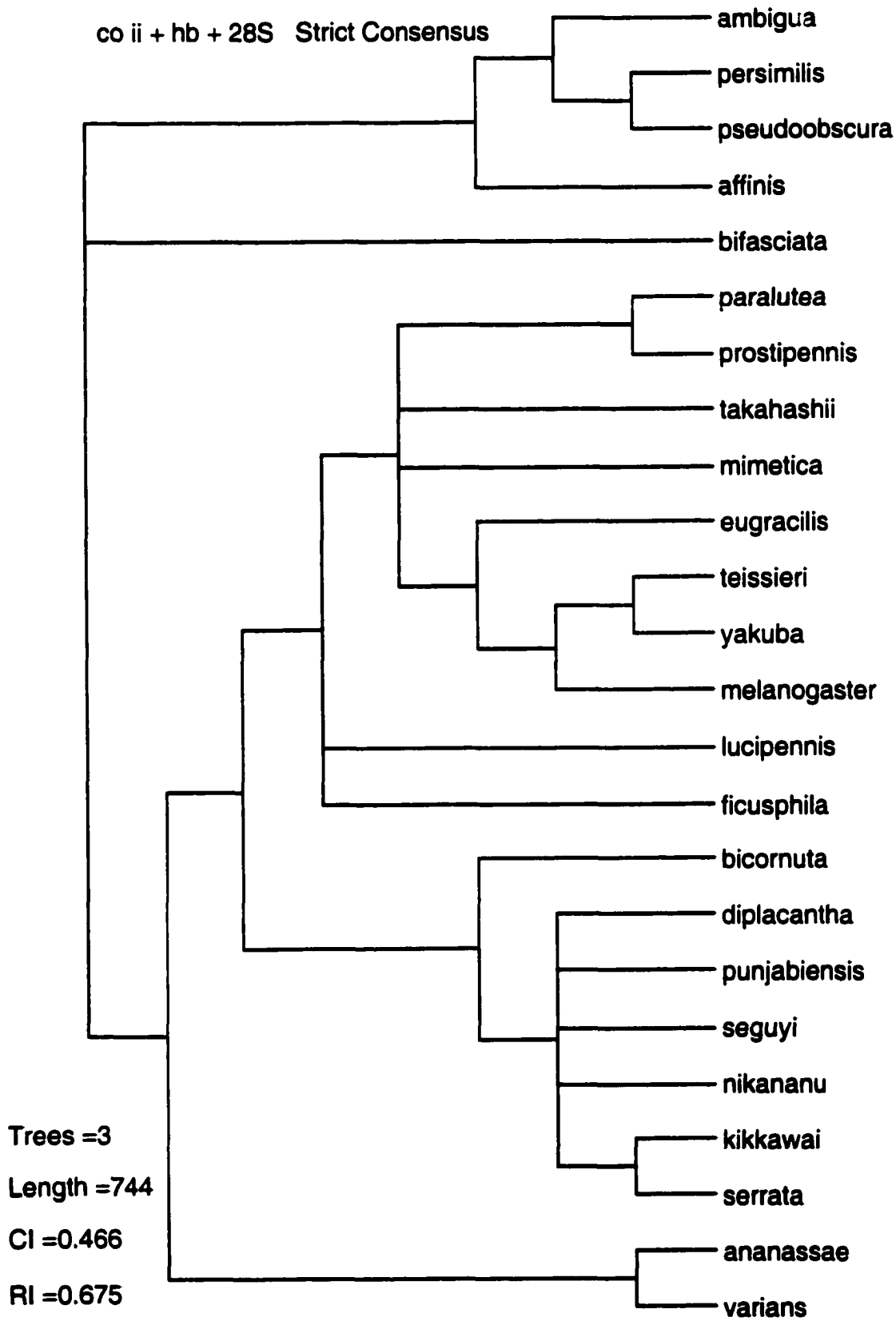


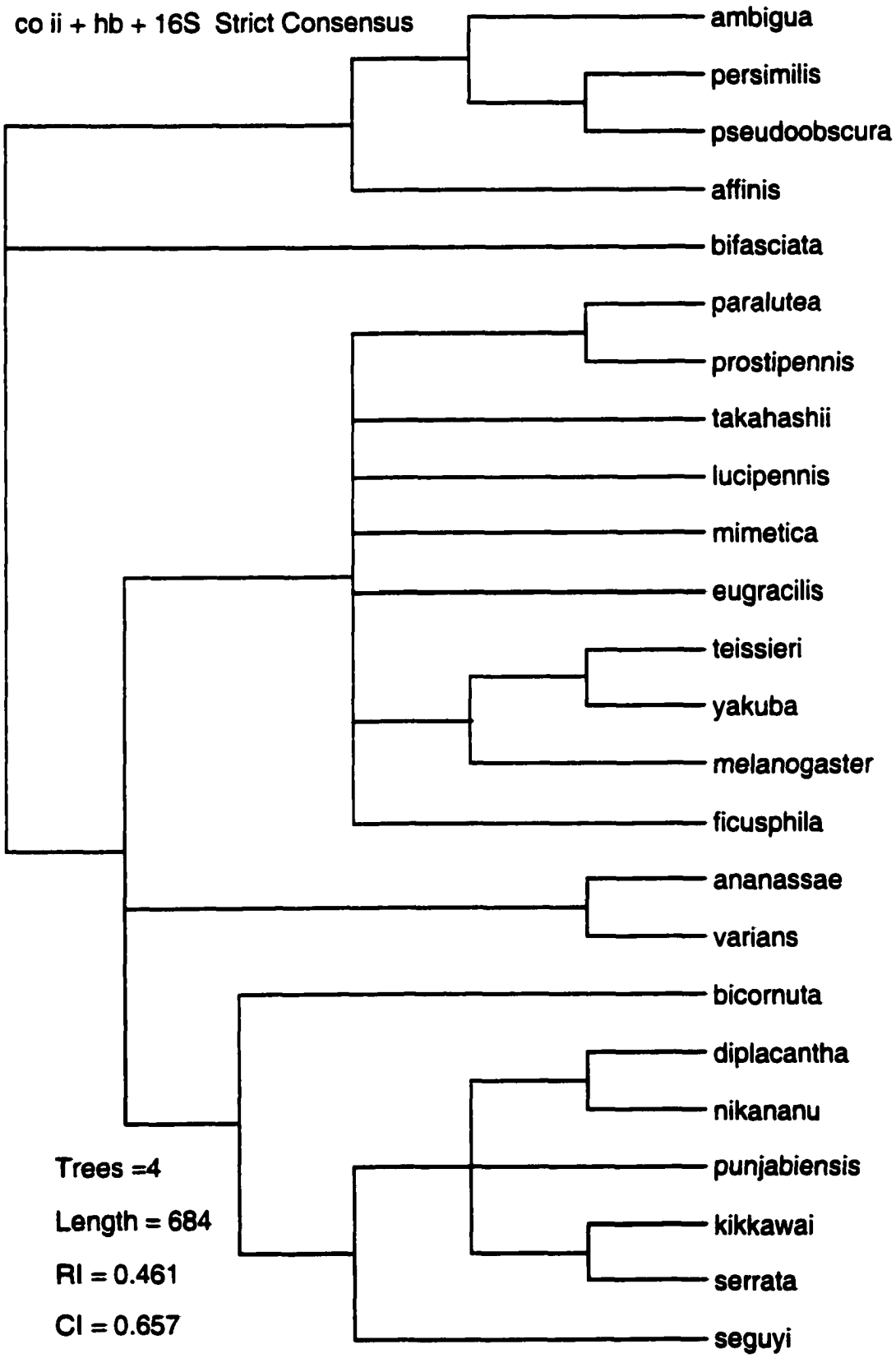


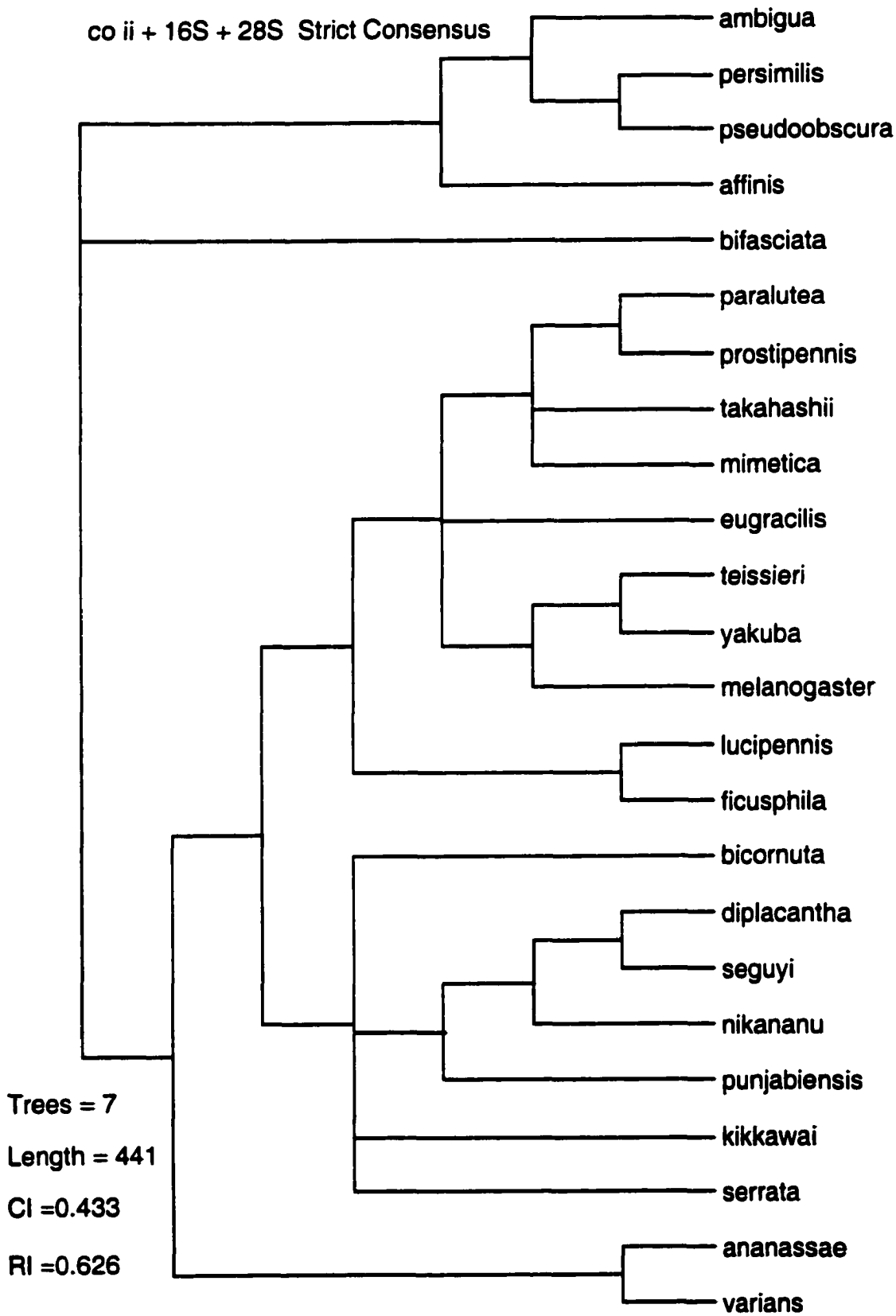


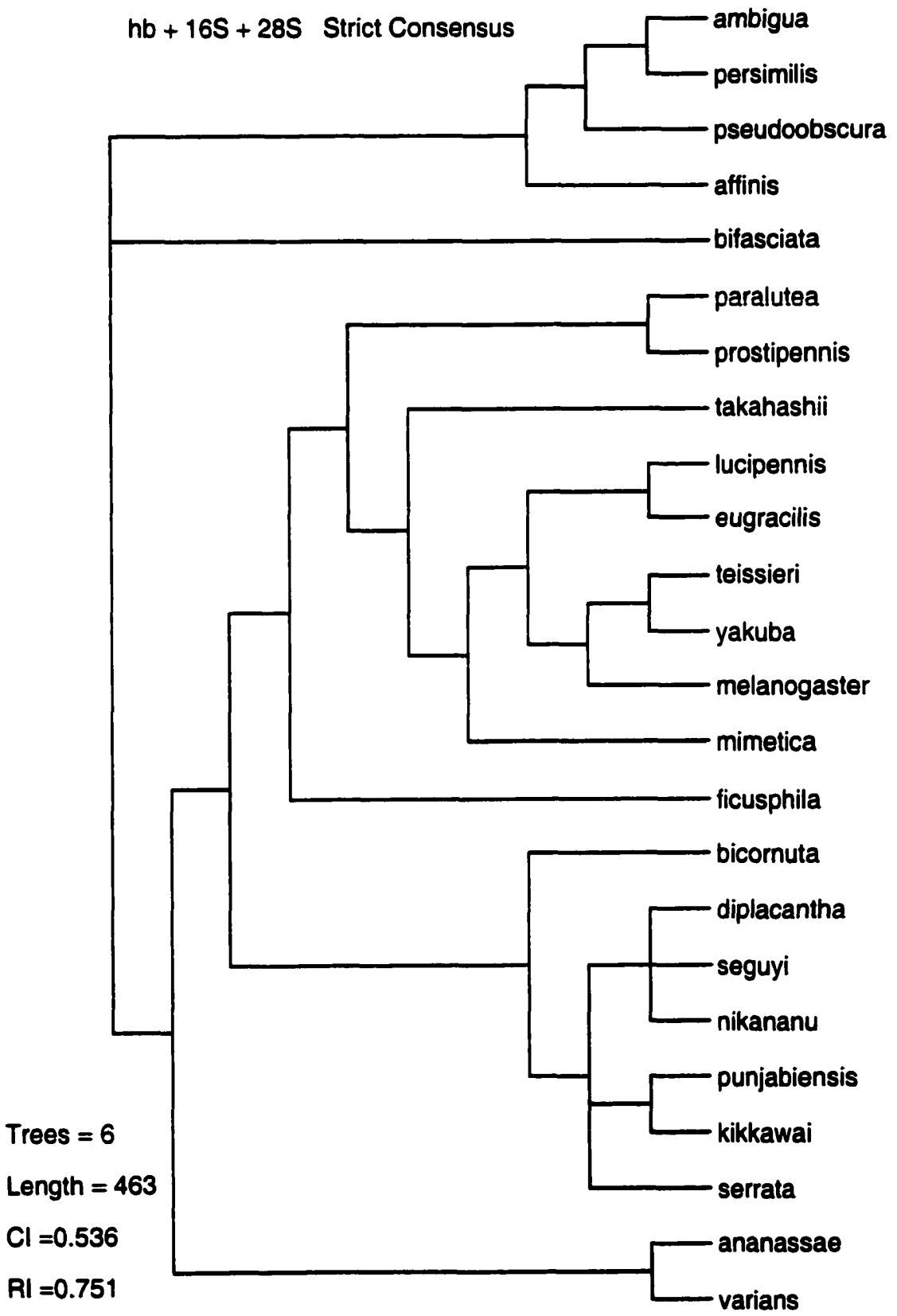


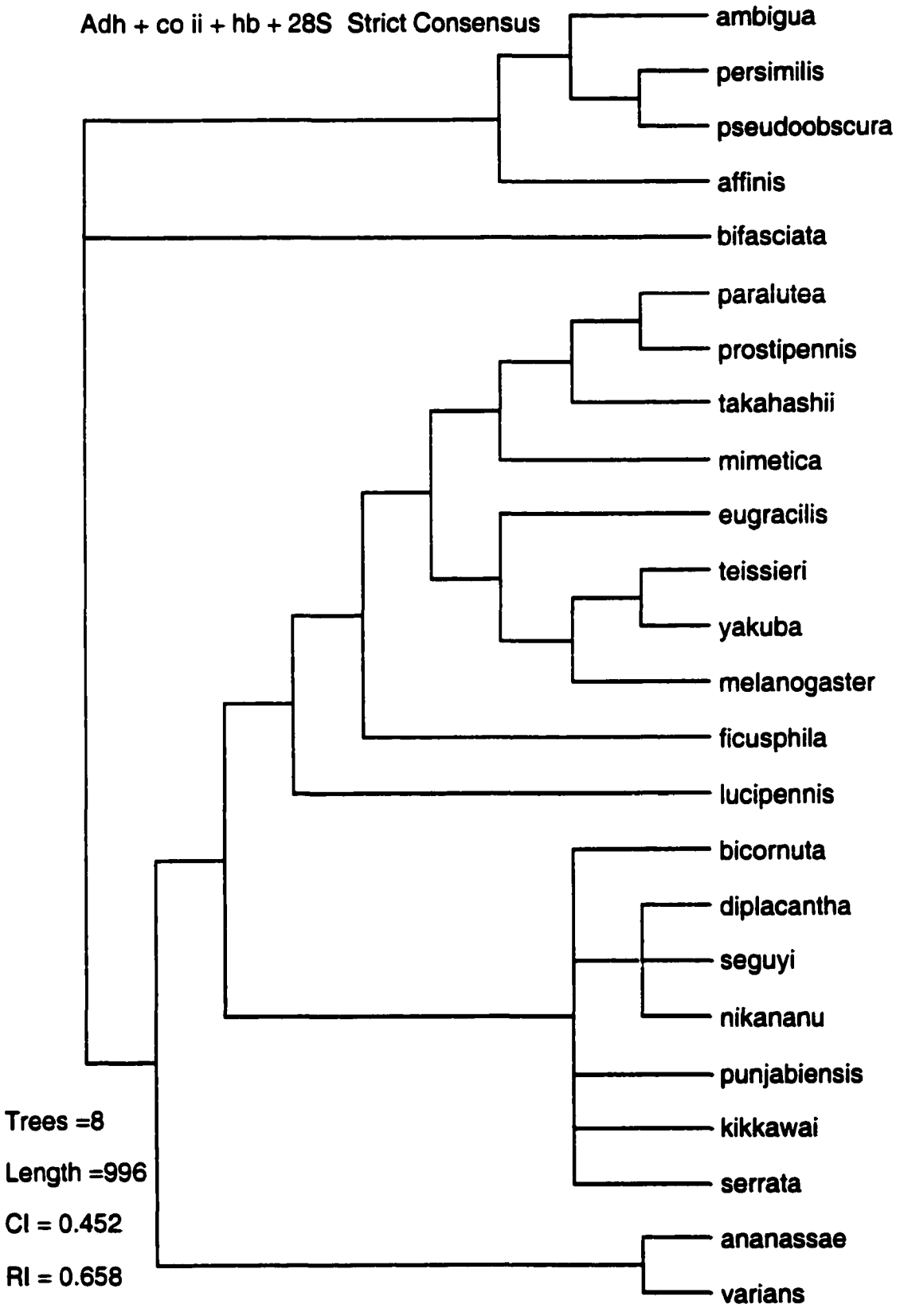


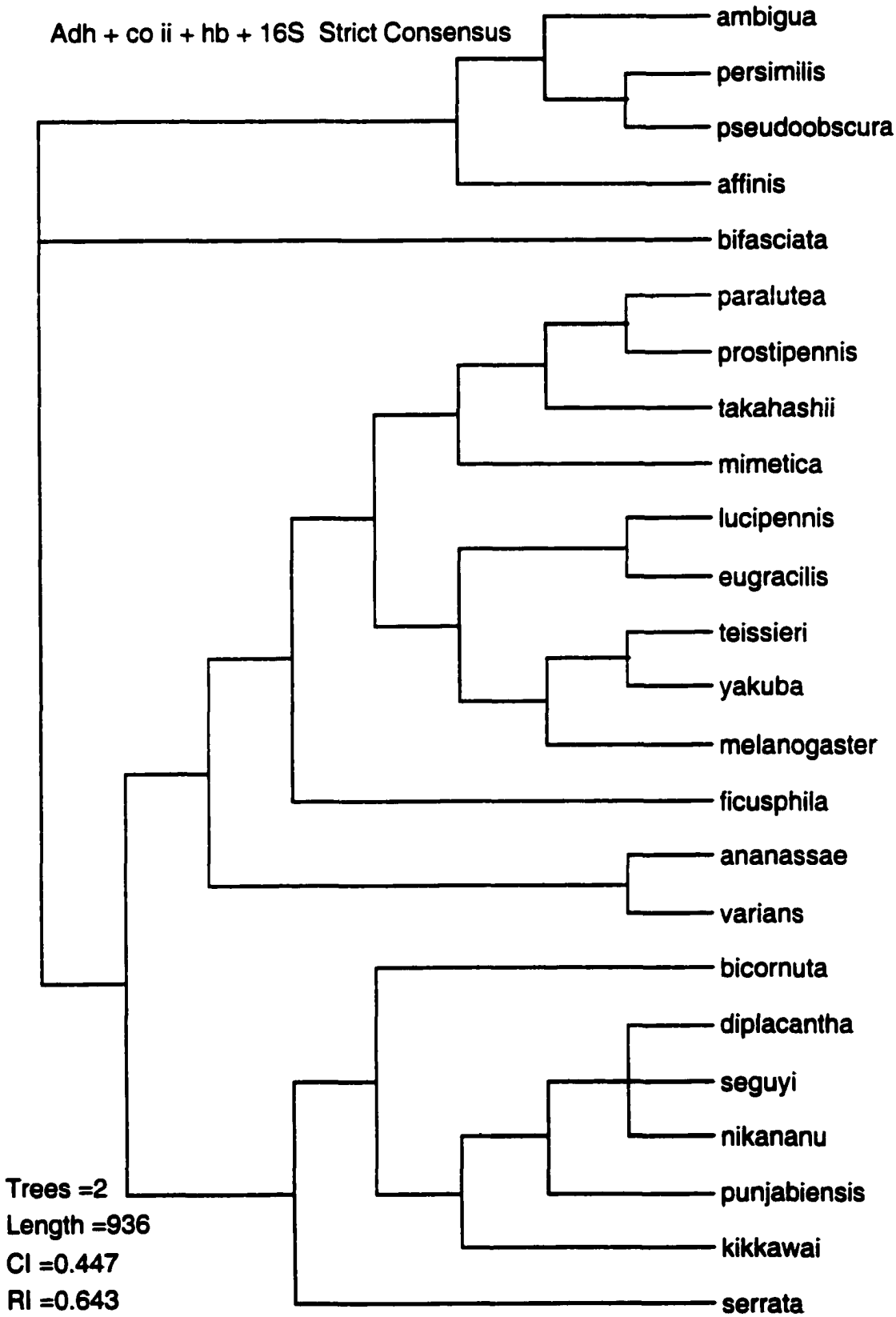


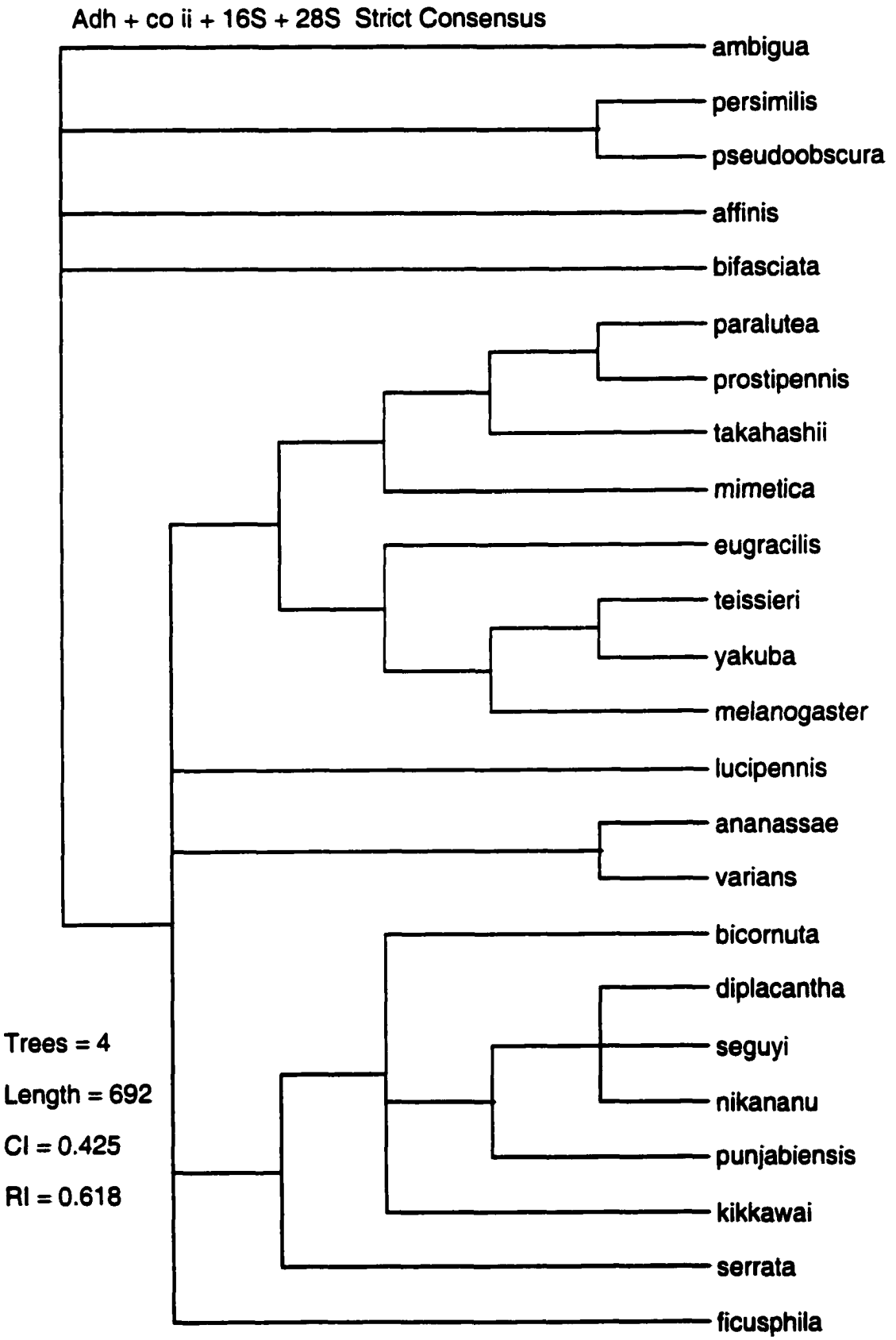


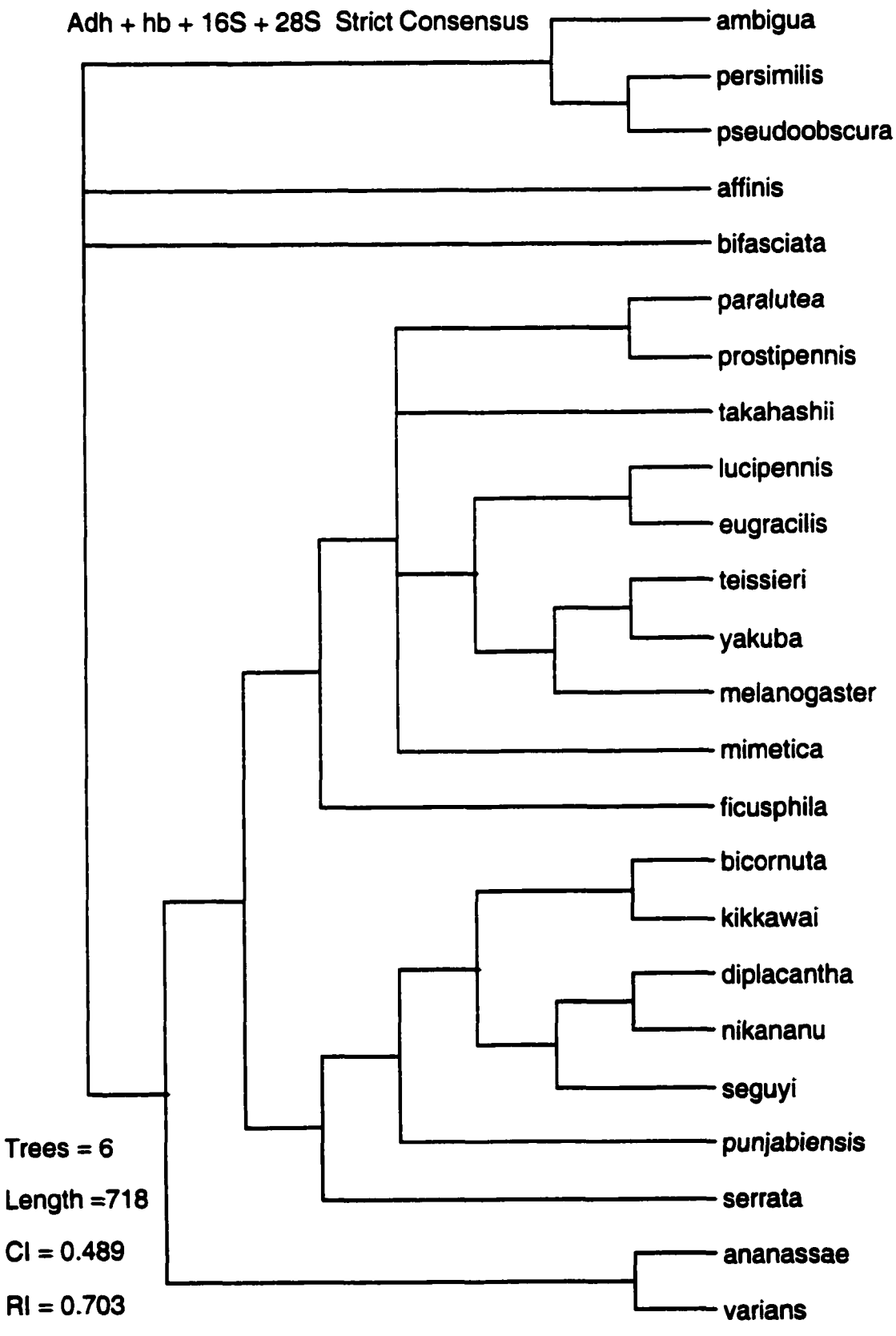


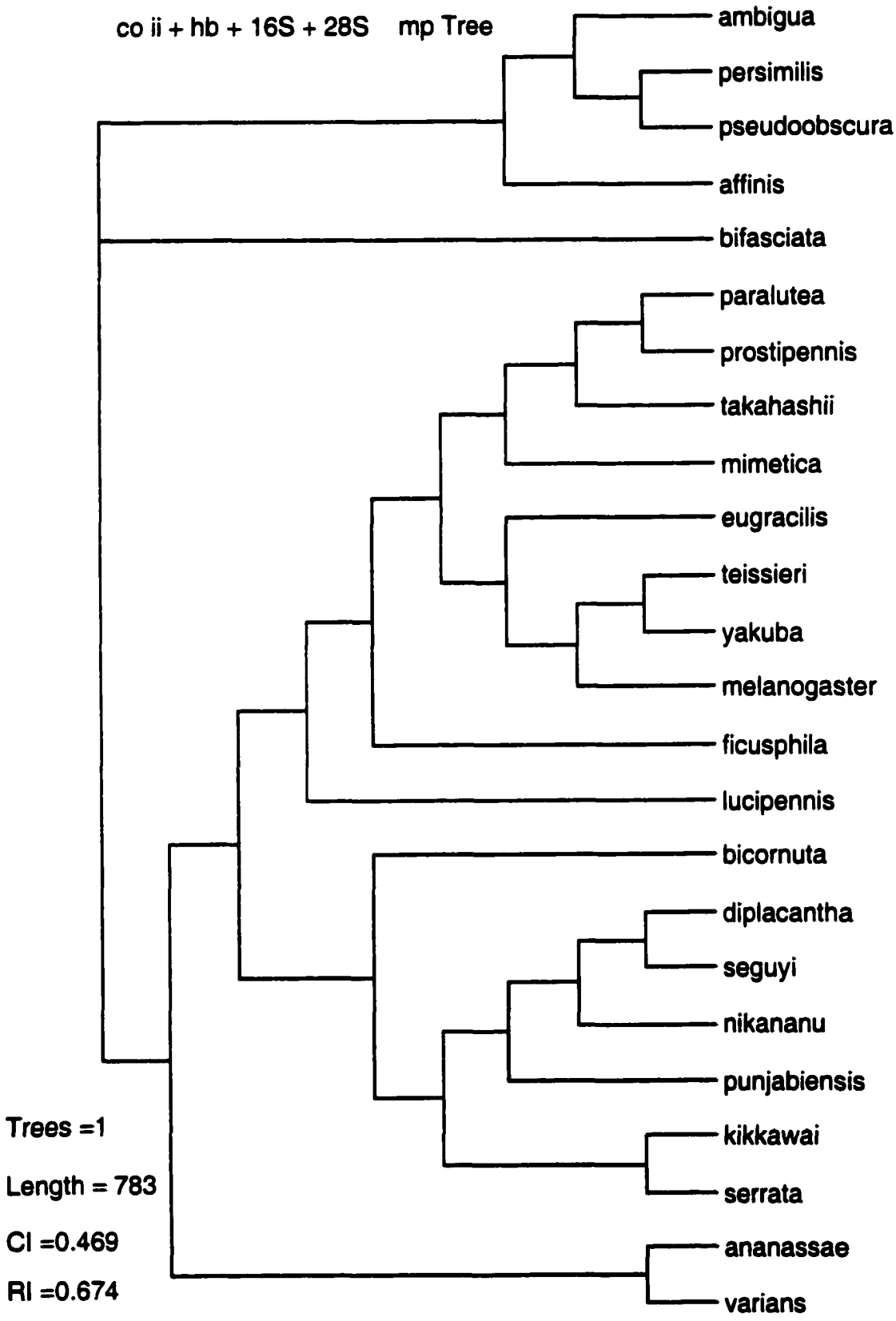












APPENDIX F : *hb* Character Assignment

By aligning the DNA or amino acid sequence, molecular systematists establish topological identity for the primary / putative homology statement (DePinna, 1991; Brower and Schawaroch, 1996). The total length of the *hb* sequence varied from 513 bp in *D. bifasciata* to 456 bp in *D. takahashii* and *D. elegans*. This caused the alignment for *hb* gene region to be more complicated in comparison to the *Adh* and *co* ii regions which had no indels (insertions or deletions).

Alignment of *hb*

It was necessary to convert *hb* nucleotide sequence to amino acid sequence for recognition of homology (i.e., topological identity sensu Brower and Schawaroch, 1996). Alignments were performed on *hb* amino acid sequences, using the Clustal method in MEGALIGN (DNASTAR, version1.02). To determine alignment ambiguous sites (Gatsey *et al.*, 1994) the cost parameters varied as follows: (1) the gap length penalty was set at a value of 10; (2) the amino acid change cost was according to the PAM250 residue weight table (Dayhoff, 1978); and (3) the gap penalty value varied from 8 to 30.

Evaluating the alignment

The resulting aligned sequence on both the amino acid and the nucleotide level is in Appendix Table 1a & b. Three stretches, amino acid positions 5-20, 106-113, and 151-166, (which correspond to bp 13-60, 316-339 and 451-498, respectively) exhibited alignment ambiguity.

The first alignment ambiguous stretch (5' end most site) shifted when the gap penalty equaled 17, thus reducing the number of gaps placed in the sequence. The gaps at amino acid site 5-7 for the *melanogaster* group taxa were eliminated (and the sequence shifted to the 5' end). The gap at amino acid site 18-20 for the *obscura* group was eliminated and the only gap maintained was at sites 17-20 for taxa sampled from the subgroups *elegans*, *eugracilis*, *ficusphila*, *melanogaster*, *suzukii* and *takahashii* (i.e., *melanogaster* + Oriental clade). Previous studies have recognized the same affinities and

distinctions among the species subgroups (Ashburner *et al.*, 1984; Lemeunier *et al.*, 1986; Ashburner, 1989; Pélandakis *et al.*, 1991; Pélandakis and Solignac, 1993).

The second ambiguous stretch corresponds to amino acid sites 106-113 (nucleotides 316-342). Here the gap present in the *melanogaster* species group taxa sampled shifted from the 3' to the 5' end of the ambiguous fragment beginning at gap penalty value of 19. Shifting this gap to the 5' side mismatched the amino acids and inflated the number of informative characters separating the *obscura* species group from the *melanogaster* species group.

The third ambiguous stretch was amino acids 151-166 (nucleotides 451-498) became ambiguous between gap penalty value of 8 and 16. For all but two of the taxa, the alignment ambiguous site would begin at amino acid positions 156-166. Due to *D. takahashii* and *D. ficusphila* the alignment ambiguous sequence began at a more 5' position (amino acid positions 151 and 155, respectively). This stretch shows good potential for characters and with more taxon sampling this site may become an excellent source of characters.

The hypervariable region predominated by Q's and H's at amino acid positions 30 - 51 oddly enough was not alignment ambiguous. The multiple repeats of Q's and H's most probably occurred by a slippage mechanism. The PAM 250 weight table (Dayhoff, 1978) which takes into account the physical and chemical features of the amino acid has a relatively low cost to switch between H and Q - a value of 2 for a range from 0-22. Putative homology statements in this region seem questionable at best. This region's alignment, however, was conserved across all the parameters tested. Therefore, this region remained in the matrix for analysis.

Determination of hb sequence for use in in phylogenic analysis

Alignment ambiguous sites were removed (Gatesy *et al.*, 1994) because topological identity (Brower and Schawaroch, 1996) could not be established. Amino acid positions 1-4, 21-105, 114-150, and 167-187 (which correspond to bp 1-12, 61-315, 340-450 and 499-

561, respectively) were invariant across all weight parameters. The remaining aligned *hb* amino acid sequence was reconverted to nucleotide sequence in an effort to maximize possible character information. The aligned *hb* nucleotide sequence now 441 bp long was inserted back as primary data in the matrix (Appendix F Table 1a & b).

Gap Coding

Gaps have traditionally been analyzed as question marks. Characters coded by a question mark can be the result of one of three conditions: the character is ambiguous, inapplicable or missing (Platnick *et al.*, 1991). Ambiguous characters seen in higher level analyses are the result of polymorphisms. In this instance, the taxon should be subdivided insuring that each character is monomorphic for the analysis (Nixon and Davis, 1991). As a result, these polymorphic taxa may be determined to be polyphyletic (Nixon and Davis, 1991). Therefore, question marks should be used only for missing or inapplicable data (Nixon and Davis, 1991). Missing data is usually encountered with fossils where the part of the specimen is actually absent or lost but has possibility of being known in the future (e.g., no soft tissues preserved, or a limb not recovered, etc.) (Nixon and Davis, 1991). Inapplicable data are characters that are not capable of being coded because the taxon never had the structure, which possesses the characters. The use of question marks in my molecular data is not because the region was not sequenced (missing data), but that the corresponding region for the other taxon does not exist and never will exist i.e., the character is inapplicable. However, there are some instances where a gap appears to convey grouping information and should be designated as a character (Wheeler, 1993; Danforth *et al.*, 1999).

Within the *hb* data some gaps seem to designate groups of taxa previously established by morphology. Therefore, the remaining gaps in the *hb* sequence were coded by one of three methods (1) all gaps 'missing," (2) all gaps as a 5th state and (3) a combination of "missing" and 5th state depending upon the alignment context. Each of these three methods of gap coding was evaluated for the influence each exerts on topology.

Combination gap coding

The third gap coding method is a combination of gaps as "missing" and 5th state. Gaps, traditionally analyzed as "missing" – that are inapplicable, may in some instances convey grouping information; therefore, these gaps should be designated as a 5th state (Wheeler, 1993). Gaps coded as a 5th state are: amino acid positions 58-59 (GA in the *obscura* group, and - - in the *melanogaster* group) and amino acid positions 138 (G in *obscura* group, and - in *melanogaster* group) (Appendix F Table 1a & b). Gaps coded as inapplicable were: (1) At amino acid positions 149-150, where the gap does not distinguish previously established taxon groups and is not cleanly demarcated since the (3') side is terminated by an ambiguous site (Appendix F Table 1a & b). And (2) at amino acid positions 98-105 where there is a Q repeat (Appendix F Table 1a & b). In this stretch, the absolute number of Q's is primarily conserved being either 3 Q's as in the *melanogaster* species group or 7 (one has 6) Q's followed by an alanine (A) in the *obscura* group. Which of the three Q's of the ingroup corresponds to the 7 or 6 Q's plus A of in the outgroup is indeterminable.

Evaluating gap coding methods

The effect of the coding methods on resulting tree topologies were compared for combined data (i.e., *co ii* + *Adh* + *hb*) and *hb* data. For both gap-coding methods (all gaps "missing" or all gaps as a 5th state) resolution decreases from the combined data to the *hb* data. In all cases the ingroup taxa divided into three major clades - the *ananassae* subgroup clade, the *montium* subgroup clade and the *melanogaster* including the Oriental subgroups clade (see also Ashburner *et al.*, 1984; Lemeunier *et al.*, 1986; Ashburner, 1989; Pélandakis *et al.*, 1991; Pélandakis and Solignac, 1993). The resulting cladogram topologies, however, varied in the relative positioning for these three clades. For gaps "missing" the combined data yielded the topology, *obscura* (*melanogaster* + oriental (*ananassae*, *montium*)), and the *hb* data yielded the topology, obscura (*ananassae* (*melanogaster* + oriental, *montium*)).

For all gaps coded as a 5th state the combined data as well as the *hb* data yielded the topology, *obscura* (*montium* (*ananassae*, *melanogaster* + oriental)).

The third gap coding method is a combination of "missing" and 5th state characters according to the matrices (Appendix F Table 2). The combined data resulted in the same topology whether coded according to matrix A or B. The *hb* data resulted in the same topology whether coded according to matrix A or B. Analysis with the combined data produced the same tree topology as when all the gaps were coded as "missing". Analysis with the *hb* data produced the same tree topology as when all the gaps were coded as "missing". In this example, tree topology was unaffected by treating gaps as either all "missing" or as a combination of "missing" and 5th state.

To insure that the number coding method had no effect, inapplicable gaps were re-coded as question marks and 5th state (character) gaps remained as dashes. Then PAUP 4.0 was run with the 'gap as 5th state' option chosen. This matrix would be equivalent to the binary coding in matrix A (Appendix F Table 2). The notation for the gap coding (question mark and dash with PAUP gap setting '5th state' versus dash and numerical with PAUP gap setting "missing") did not effect the tree topology, length, CI or RI.

Conclusion

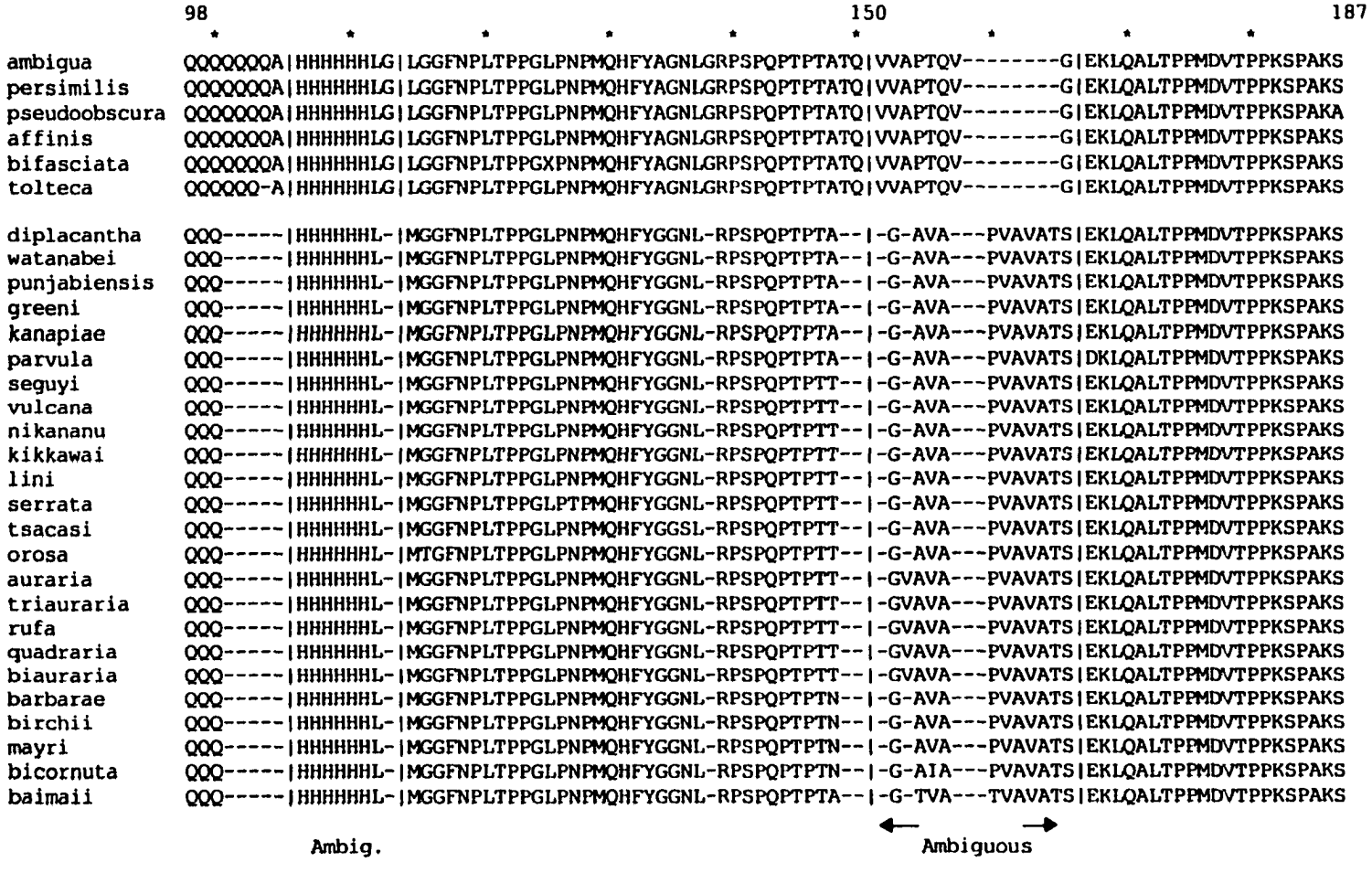
Even though these results demonstrate that the combination gap coding did not alter tree topology from the traditional coding method (gaps as "missing"), combination coding reflects the information conveyed by the gaps present in the *hb* sequence. Matrix B (Appendix F Table 2) is a summarized numerical coding of matrix A with the assumption that these gaps were single events. All phylogenetic analyses in this dissertation the gaps were combination coded (both "missing" and the 5th state). PAUP was executed with the option 'gaps as missing', and the gaps as 5th state were coded in the PAUP matrix using numerical values according to matrix B (Appendix F Table 2).

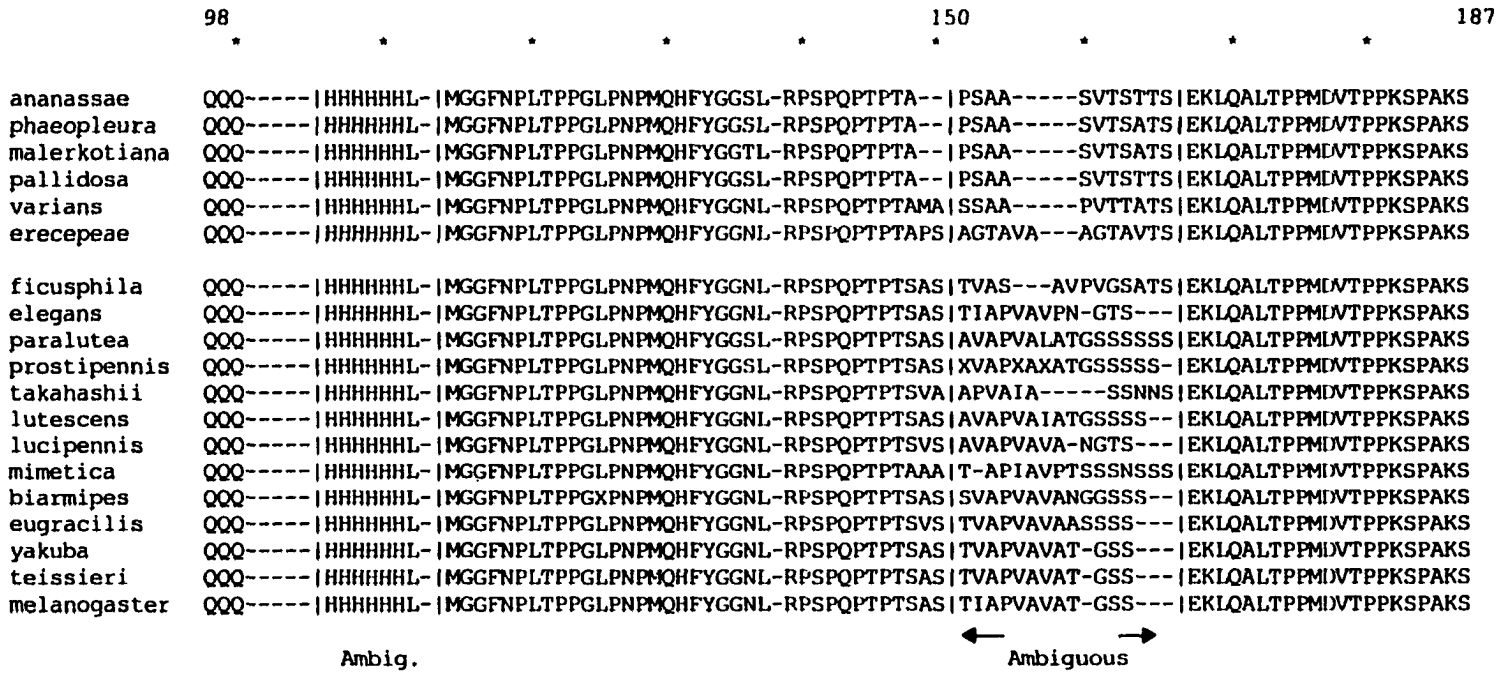
Table 1 a & b. Alignment of *hb* amino acid sequence for 49 taxa was made using MEGALIGN's (DNASTAR, version 1.02) Clustal method. Using PAM 250 residue weight table, gap length penalty of 10 and varying the gap penalty from 8 to 30, multiple alignments were run to determine alignment ambiguous sites. The exemplar alignment is the one that resulted from using a gap penalty value of 8. Table A is the amino acid sequence and Table B is the corresponding nucleotide base sequence.

hb amino acid sequence

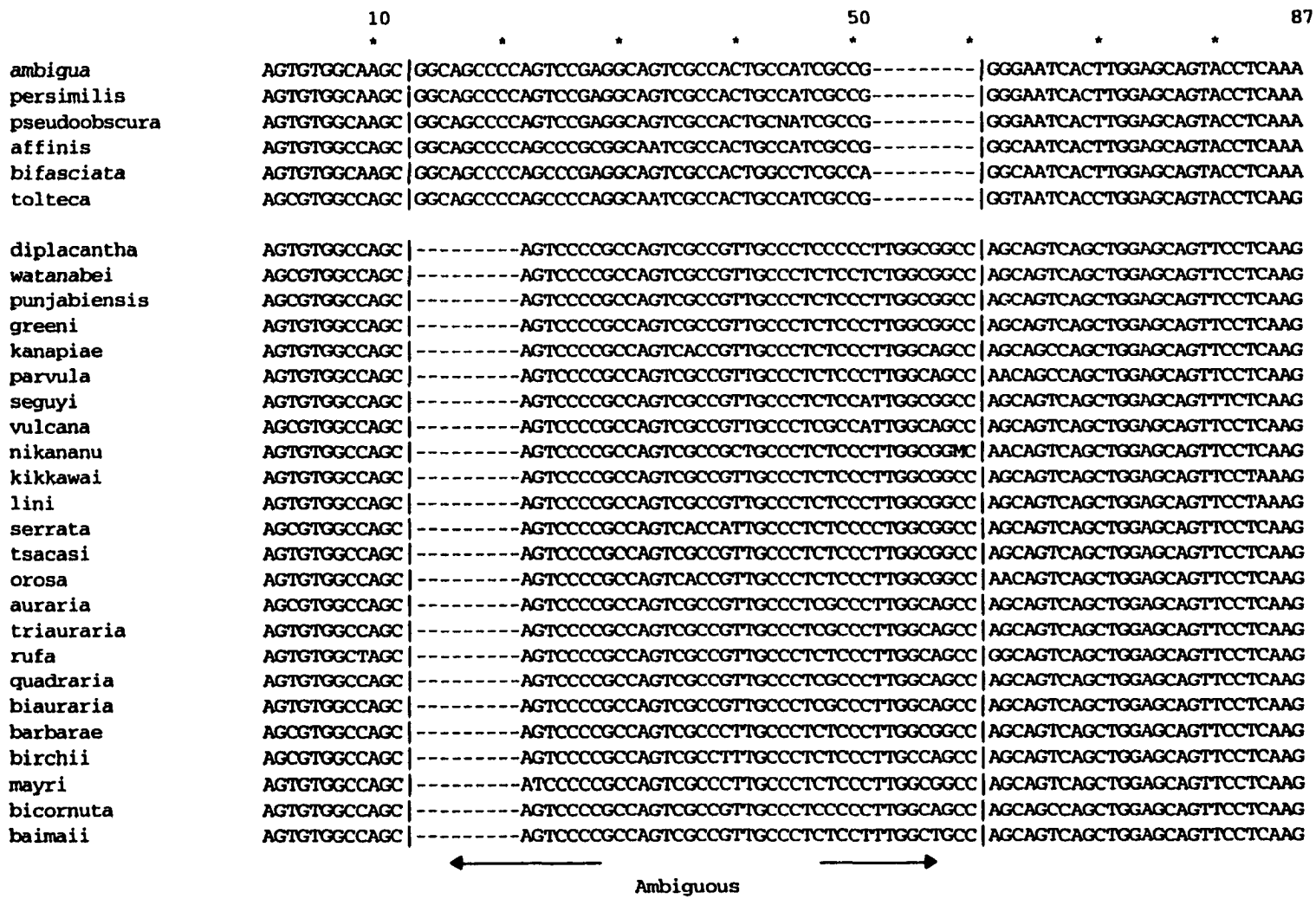
```
                          10                                                  50                                                  97
                          *         *         *         *         *         *         *         *         *         *
ambigua          SVAS|GSPSPRQSPLPSP---|GNHLEQYLKQQQQQ--HHQQQQLQ-----QQPMDTLCGAAMTPSPSQNDQNSLQHFDVTLHQQLLQQQQYQQHFQAA
persimilis       SVAS|GSPSPRQSPLPSP---|GNHLEQYLKQQQQQ--HHQQQQLQ-----QQPMDTLCGAAMTPSPSQNDQNSLQHFDVTLQQQLLQQQQYQQHFQAA
pseudoobscura    SVAS|GSPSPRQSPLXSP---|GNHLEQYLKQQQQQ--HHQQQQLQ-----QQPMDTLCGAAMTPSPSQNDQNSLQHFDVTLQQQLLQQQQYQQHFQAA
affinis          SVAS|GSPSPRQSPLPSP---|GNHLEQYLKQQQQ----HQQQQLQ-----QQPMDTMCGAAMTPSPNQNDQNSLQHFDVTLQQQLLQQQQYQQHFQAA
bifasciata       SVAS|GSPSPRQSPLASP---|GNHLEQYLKQQQQQQQHQHQQQLQ-----QQPMDTLCGAAMTPSPSQNDQNSLQHFDVTLQQQLLQQQQYQQHFQAA
tolteca          SVAS|GSPSPRQSPLPSP---|GNHLEQYLKQQQHQQQ-QQQQQLQ-----QQPMDTMCGAAMTPSPSQNDQNSLQHFDVTLQQQLLQQQQYQQHFQAA

diplacantha      SVAS|---SPRQSPLPSPLAA|SSQLEQFLKQQ-HHHQQQQQQQQ-HQSHQQQPMDTMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
watanabei        SVAS|---SPRQSPLPSPLAA|SSQLEQFLKQQ-QHHQQQQQ----HQSHQQQPMDTMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
punjabiensis     SVAS|---SPRQSPLPSPLAA|SSQLEQFLKQQ-QHHQQQQ-----HQSHQQQPMDTMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
greeni           SVAS|---SPRQSPLPSPLAA|SSQLEQFLKQQ-HHHQQQQQQQQQHQSHQQQPMDTMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
kanapiae         SVAS|---SPRQSPLPSPLAA|SSQLEQFLKQQ-HHHQQQQ-----HQTHQQQPMDTMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
parvula          SVAS|---SPRQSPLPSPLAA|NSQLEQFLKQQHHHQQQQQ-----HQTHQQQPMDTMC--TMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
seguyi           SVAS|---SPRQSPLPSPLAA|SSQLEQFLKQQQQHHHQQQ---QQHQSHQQQPMDXMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
vulcana          SVAS|---SPRQSPLPSPLAA|SSQLEQFLKQQ-HHHQQQQ---QQHQSHQQQPMDTMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
nikananu         SVAS|---SPRQSPLPSPLAX|NSQLEQFLKQQQHH----QQQQQQHQSHQQQPMDTMC--AMTPSPSQXDQNSLQHFDATLQQQFLQQQQYQQHFQAA
kikkawai         SVAS|---SPRQSPLPSPLAA|SSQLEQFLKQQ-QHHQQQQ---QHHQSHQQQPMDTMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
lini             SVAS|---SPRQSPLPSPLAA|SSQLEQFLKQQ-HHHQQQQ---QQHQSHQQQPMDTMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
serrata          SVAS|---SPRQSPLPSPLAA|SSQLEQFLKQQ-QHHQQQQQQQQQHQSHHQQPMDTMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
tsacasi          SVAS|---SPRQSPLPSPLAA|SSQLEQFLKQQ-HHHQQQQ---QQHQSHQQQPMDTMC--AMTPSPSQNDQNSLQHFDGTLQQQLLQQQQYQQHFQAA
orosa            SVAS|---SPRQSPLPSPLAA|NSQLEQFLKQQQHHHQQQQQQQQQHQSHQQQPMDTMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
auraria          SVAS|---SPRQSPLPSPLAA|SSQLEQFLKQQQHH-QQQQ---QQHQPHQQQPMDTMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
triauraria       SVAS|---SPRQSPLPSPLAA|SSQLEQFLKQQQHH-QQQQ---QQHQPHQQQPMDTMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
rufa             SVAS|---SPRQSPLPSPLAA|GSQLEQFLKQQQHH-QQQQ---QQHQSHQQQPMDTMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
quadraria        SVAS|---SPRQSPLPSPLAA|SSQLEQFLKQQQHH-QQQQ---QQHQPHQQQPMDTMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
biauraria        SVAS|---SPRQSPLPSPLAA|SSQLEQFLKQQQHHHQQQQ---QQHQPHQQQPMDTMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
barbarae         SVAS|---SPRQSPLPSPLAA|SSQLEQFLKQQ-QHHQQQ------HQSHQQQPMDTMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
birchii          SVAS|---SPRQSPLPSPLPA|SSQLEQFLKQQQHHHHQQQQ----HQSQQQQPMDTMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
mayri            SVAS|---IPRQSPLPSPLAA|SSQLEQFLKQQHHH---QQHQEQQHQSHQQQPMDTMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
bicornuta        SVAS|---SPRQSPLPSPLAA|SSQLEQFLKQQ-HHQQQQQQQQQQHQSHQQQLMDTMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
baimaii          SVAS|---SPRQSPLPSPLAA|SSQLEQFLKQQQHHQQQQQHQ---HPSHQQQPMDTMC--AMTPSPSQNDQNSLQHFDATLQQQLLQQQQYQQHFQAA
```

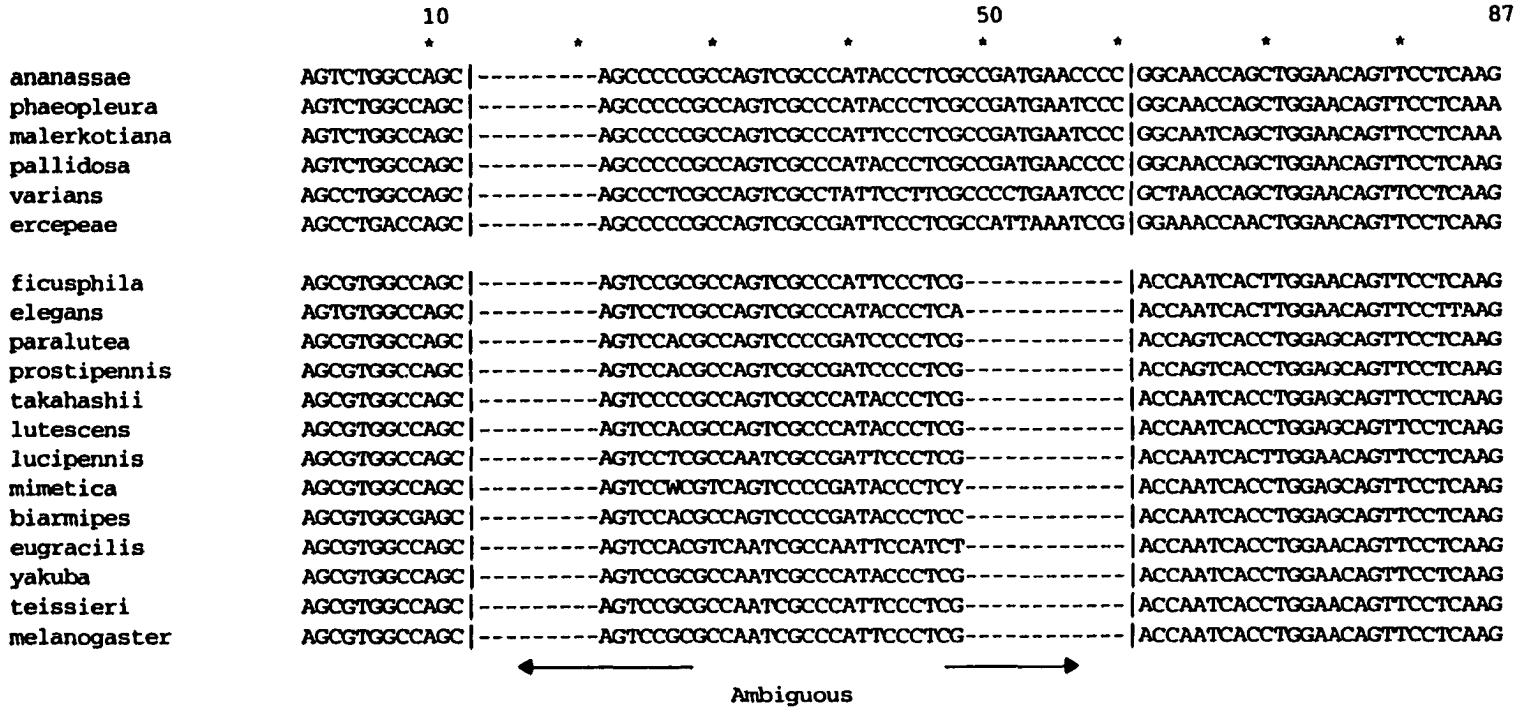
← Ambiguous →

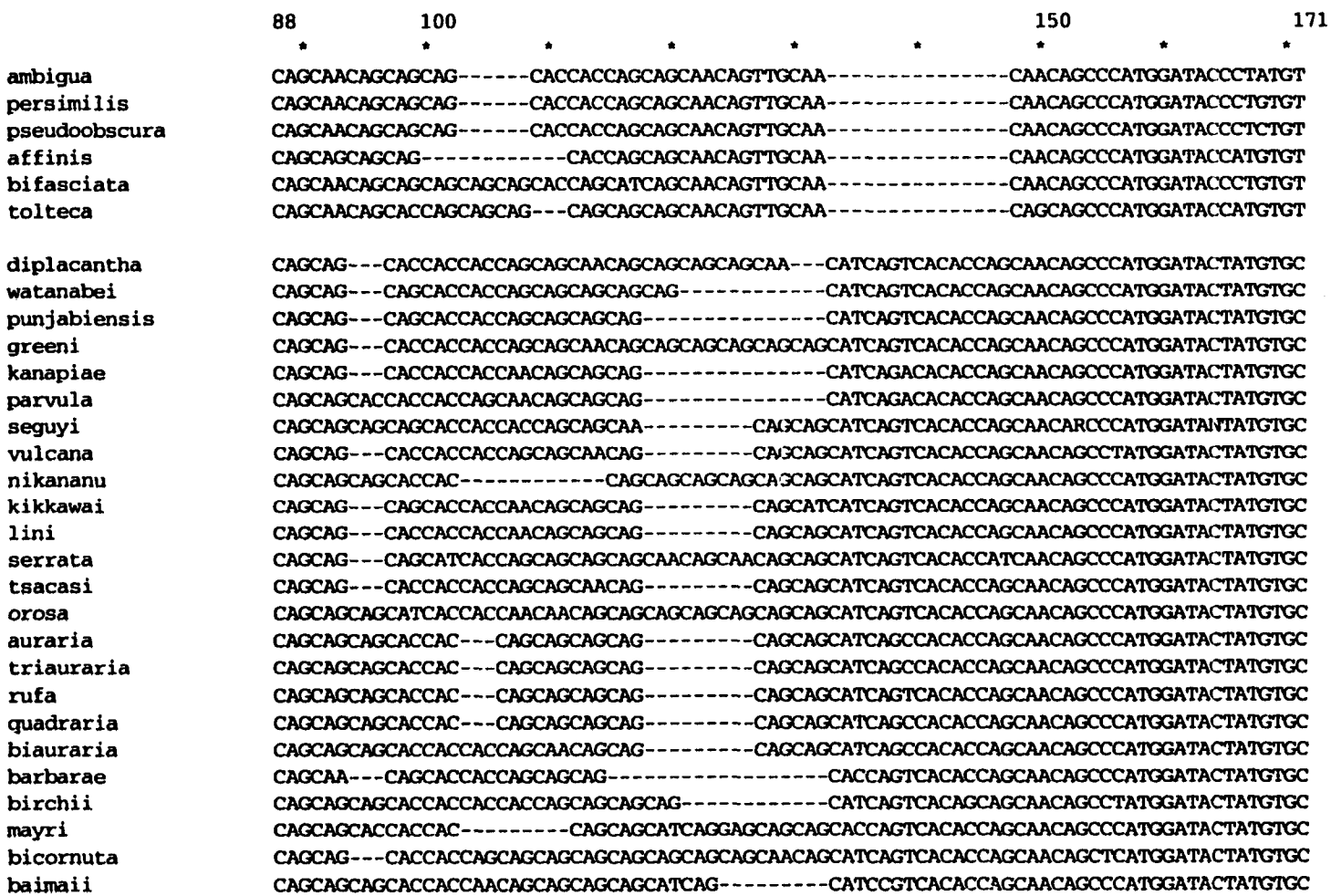


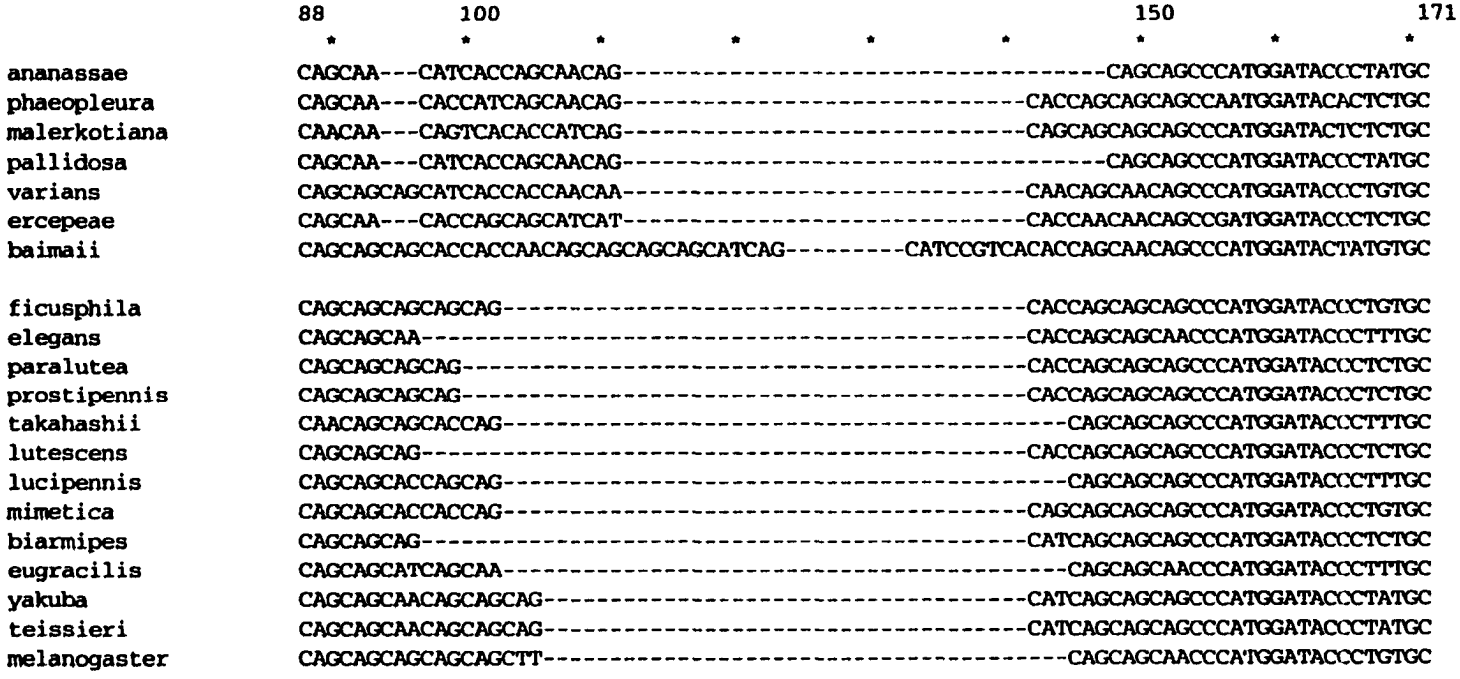


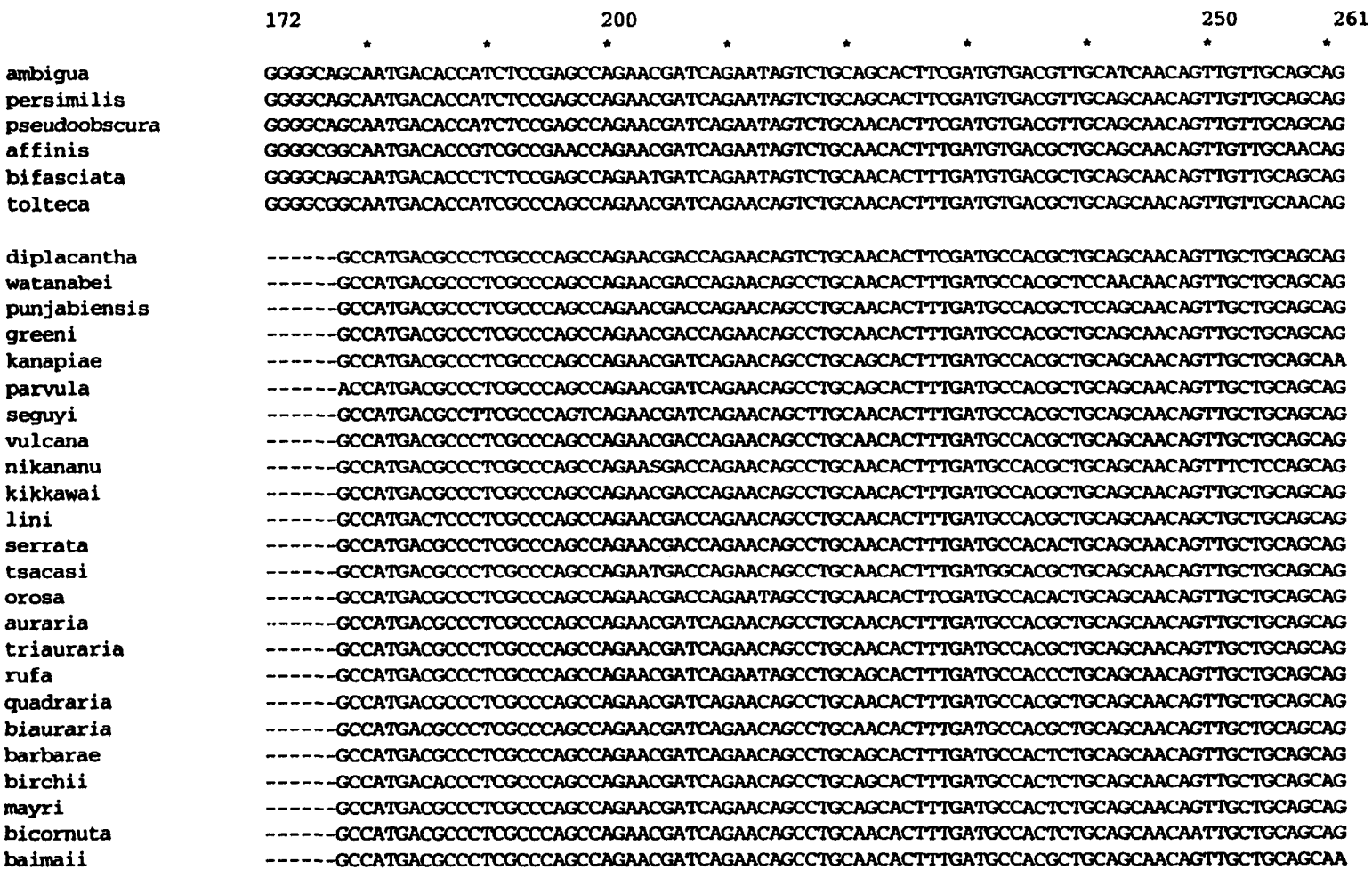
hb nucleotide sequence

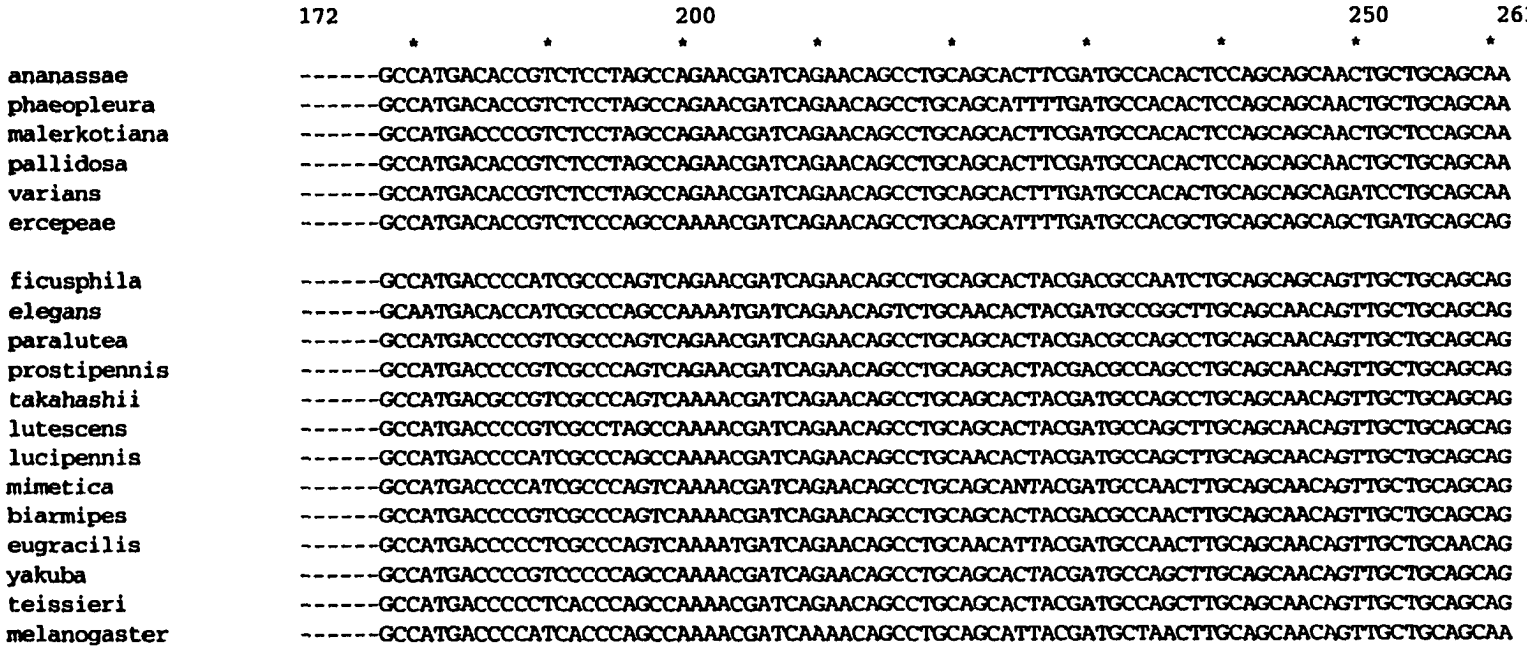


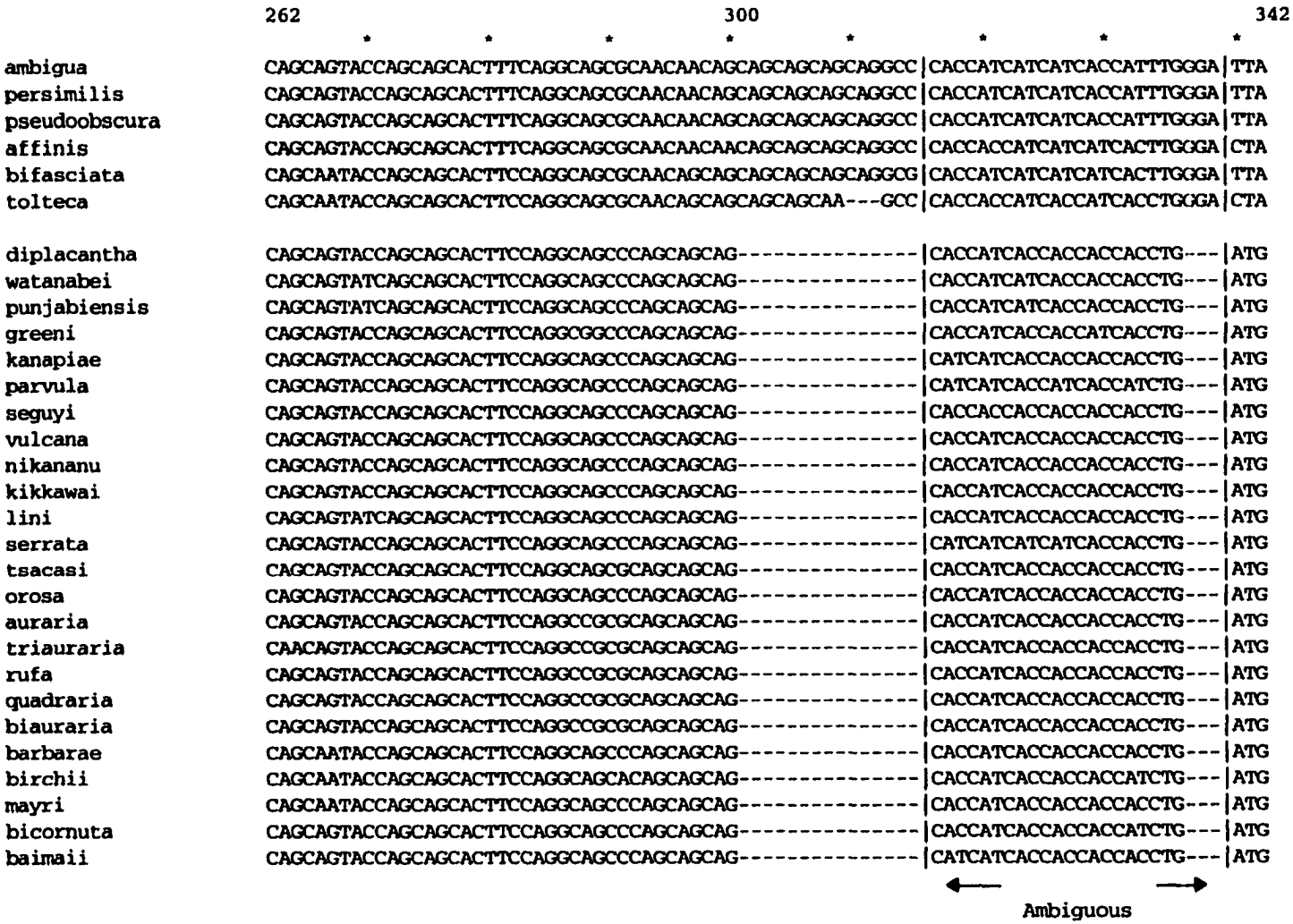


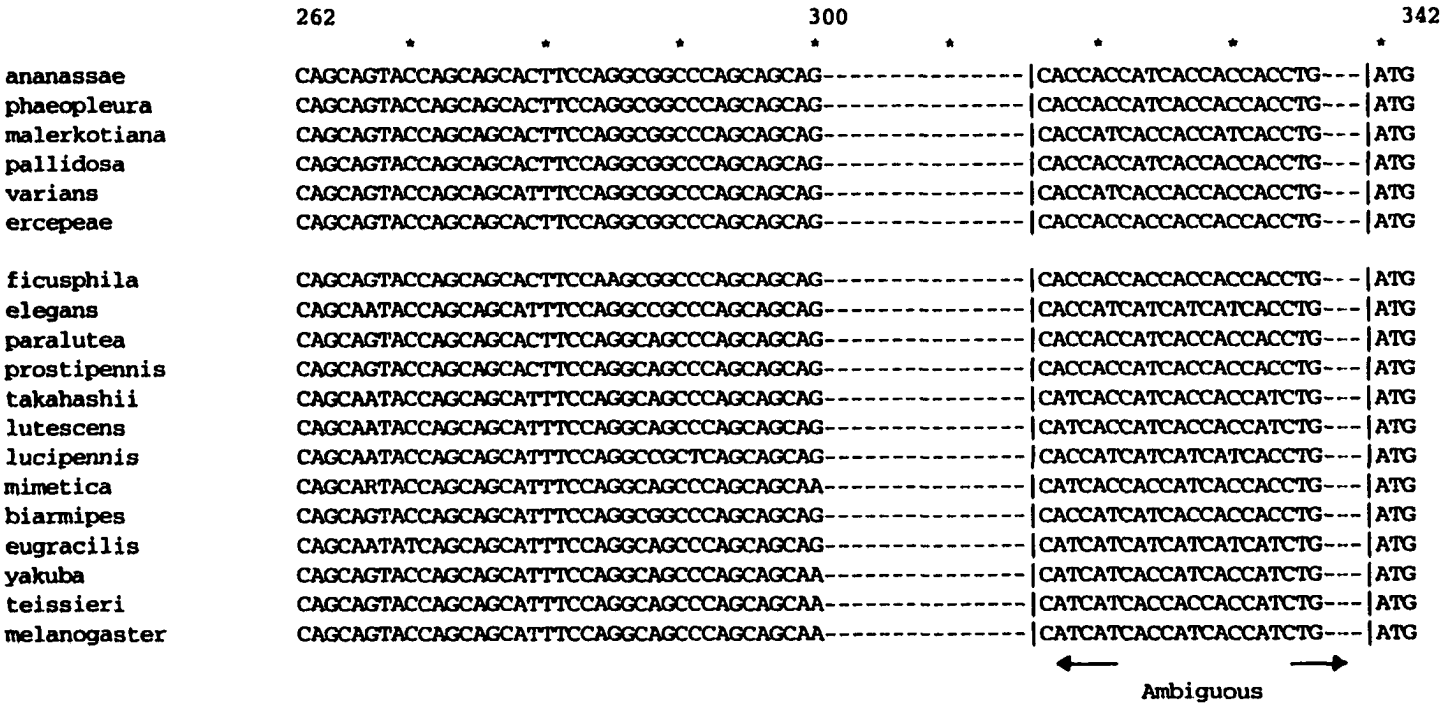


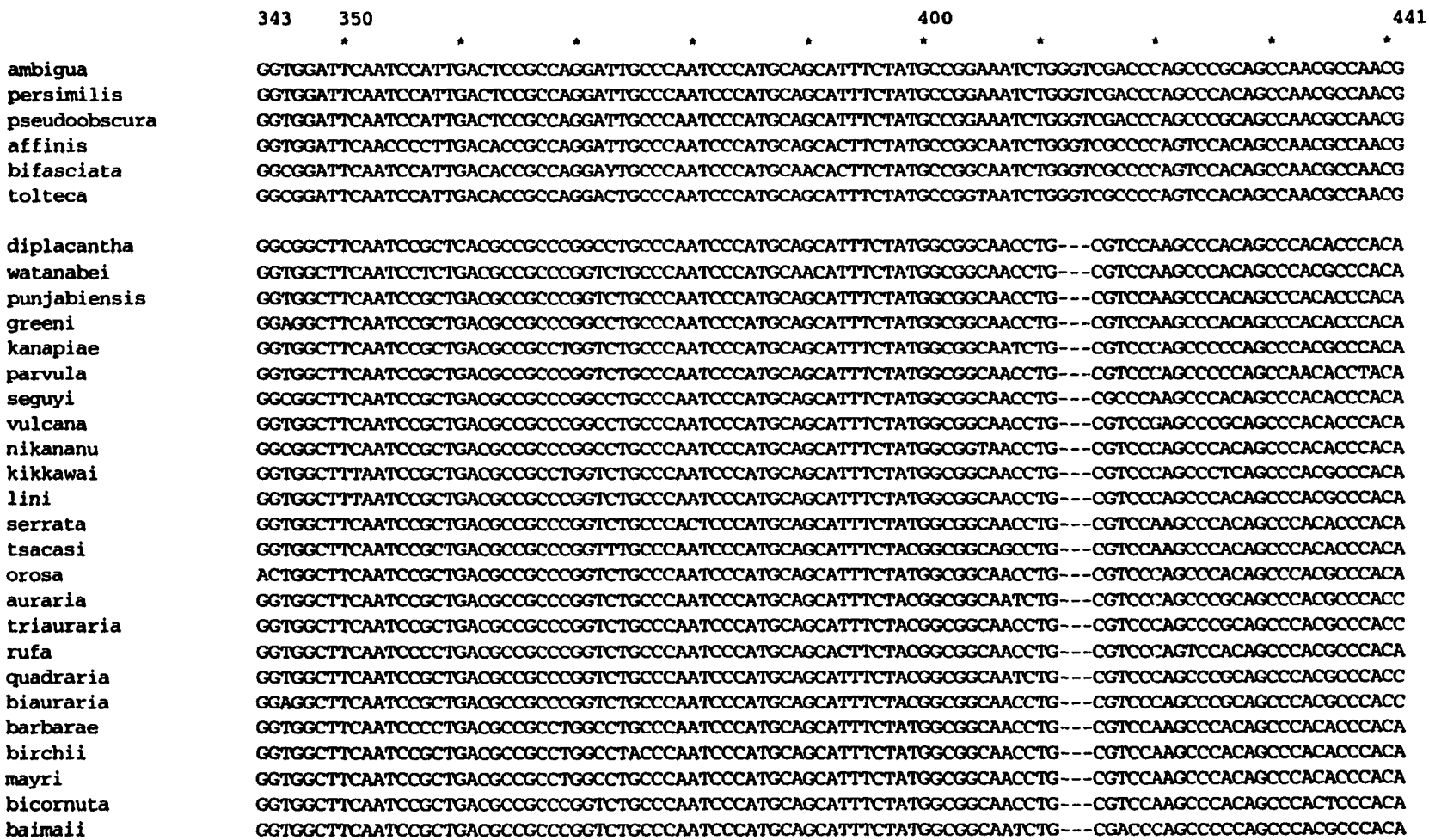


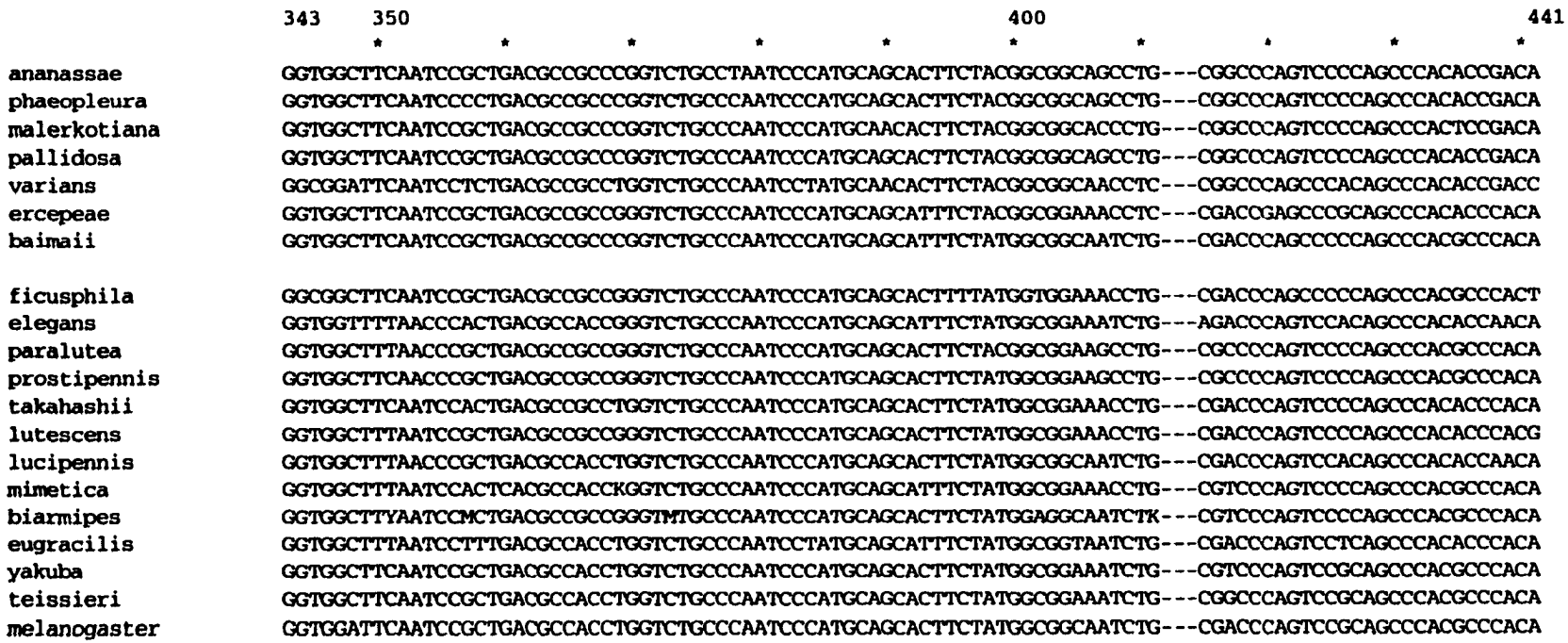


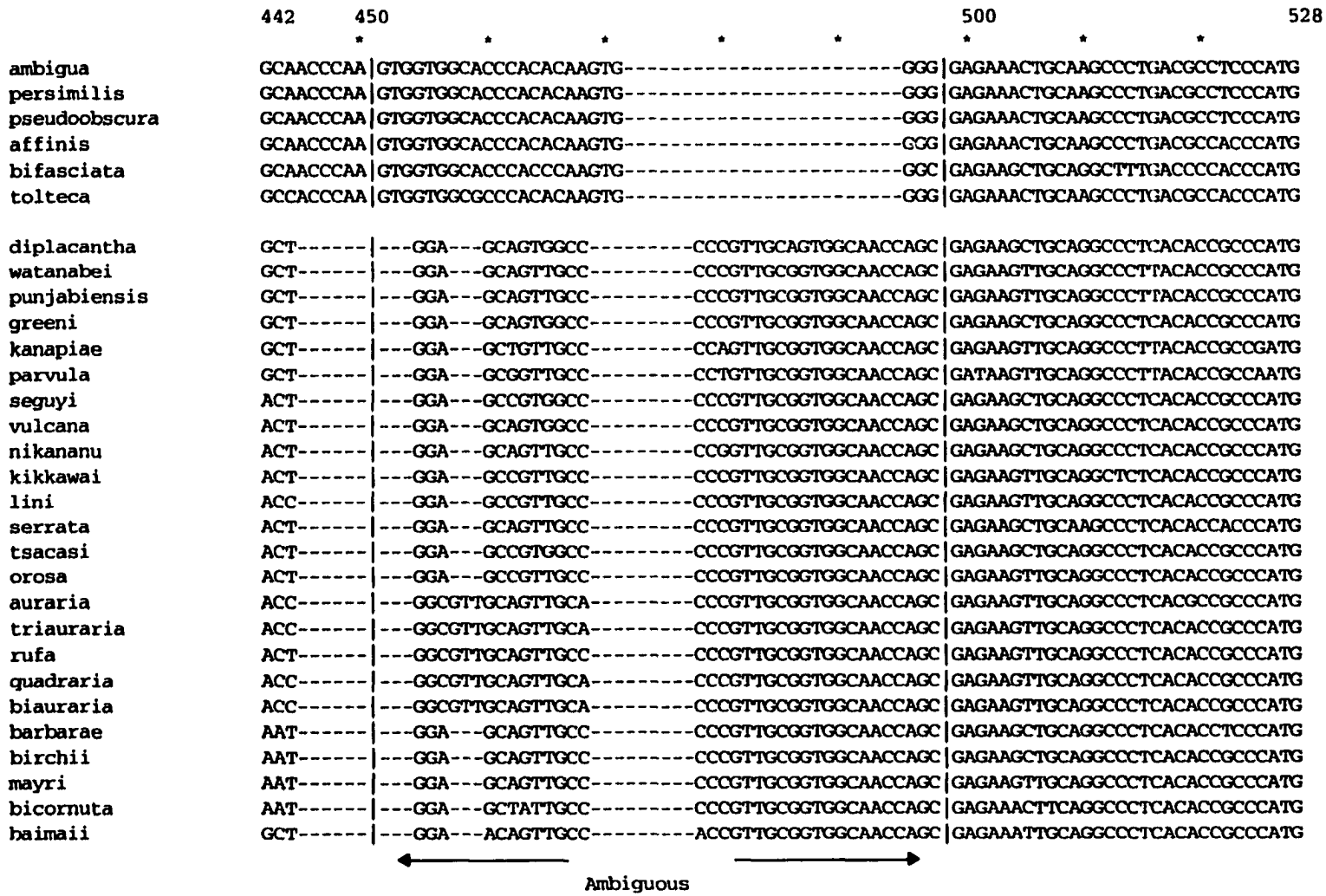


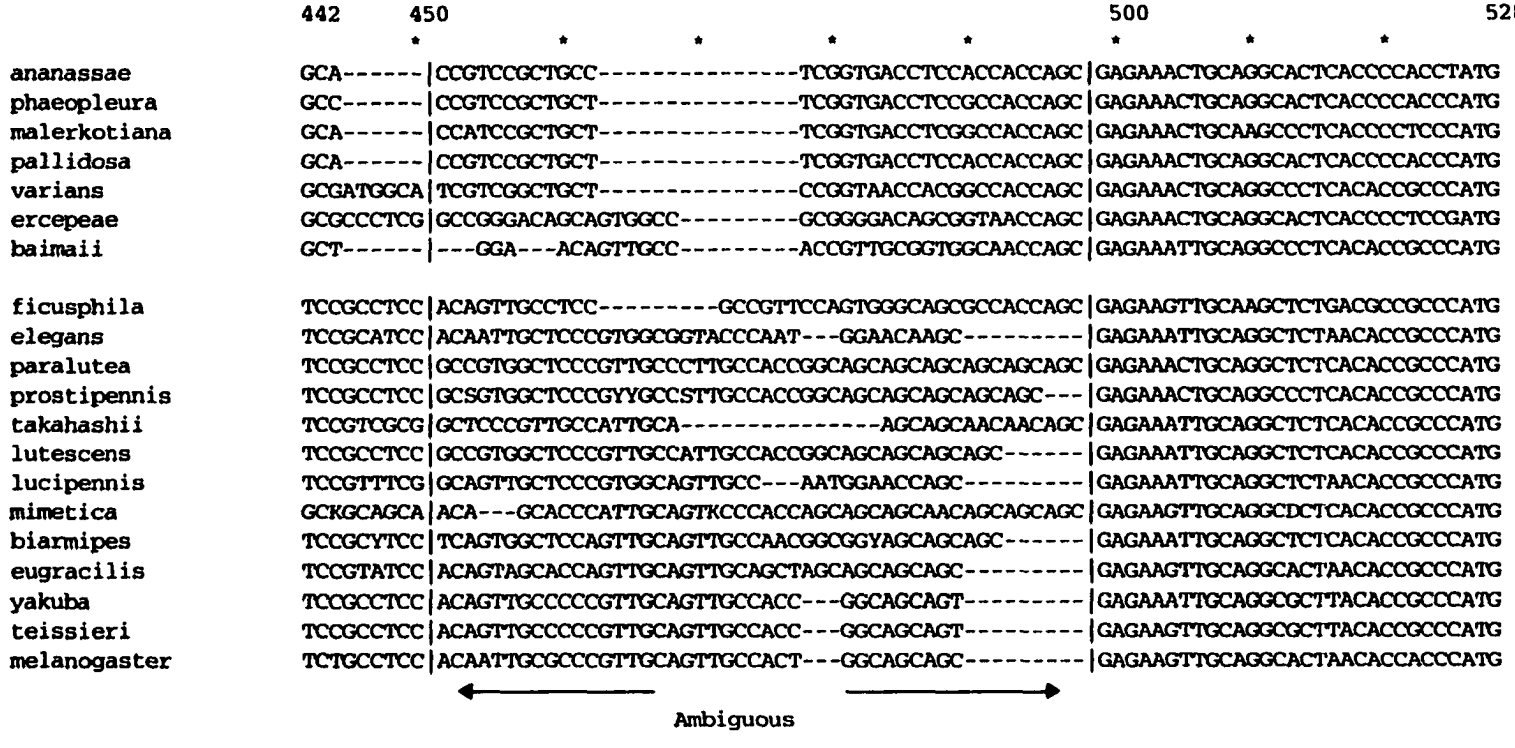












```
                    529                 550         561
                     *         *         *         *
ambigua             GATGTGACGCCACCAAAGTCACCAGCAAAATCC
persimilis          GATGTGACGCCACCAAAGTCACCAGCGAAATCC
pseudoobscura       GATGTGACGCCACCAAAGTCACCAGCGAAAGCC
affinis             GATGTGACTCCACCAAAGTCACCAGCCAAGTCC
bifasciata          GATGTGACGCCACCAAAGTCACCAGCGAAATCC
tolteca             GACGTGACGCCCCCCAAGTCGCCGGCCAAGTCC

diplacantha         GATGTGACGCCGCCCAAGTCGCCGGCCAAGTCC
watanabei           GATGTGACGCCGCCCAAGTCGCCGGCCAAGTCC
punjabiensis        GATGTGACGCCGCCCAAGTCGCCGGCCAAGTCC
greeni              GATGTGACGCCGCCCAAGTCGCCGGCCAAGTCC
kanapiae            GATGTGACGCCGCCCAAGTCGCCGGCCAAGTCC
parvula             GATGTGACGCCGCCCAAGTCGCCGGCCAAGTCC
seguyi              GACGTGACGCCGCCCAAGTCGCCGGCCAAGTCC
vulcana             GATGTGACGCCGCCCAAGTCGCCGGCCAAGTCC
nikananu            GATGTGACGCCGCCCAAGTCGCCGGCCAAGTCC
kikkawai            GATGTGACGCCGCCCAAGTCGCCGGCCAAGTCC
lini                GATGTGACGCCGCCCAAGTCGCCGGCCAAGTCC
serrata             GATGTTACGCCGCCCAAGTCGCCGGCCAAGTCC
tsacasi             GATGTGACGCCGCCCAAGTCGCCGGCCAAGTCC
orosa               GATGTGACGCCGCCCAAGTCGCCGGCCAAGTCC
auraria             GATGTGACGCCGCCAAAGTCGCCGGCCAAGTCC
triauraria          GATGTGACGCCGCCAAAGTCGCCGGCCAAGTCC
rufa                GATGTGACGCCGCCCAAGTCGCCGGCCAAGTCC
quadraria           GATGTGACGCCGCCAAAGTCGCCGGCCAAGTCC
biauraria           GATGTGACGCCGCCAAAGTCGCCGGCCAAGTCC
barbarae            GATGTGACGCCGCCCAAGTCGCCGGCCAAGTCC
birchii             GATGTGACGCCGCCCAAGTCGCCGGCCAAGTCC
mayri               GATGTGACGCCGCCCAAGTCGCCGGCCAAGTCC
bicornuta           GATGTGACGCCGCCCAAGTCGCCGGCCAAGTCC
baimaii             GATGTGACGCCGCCCAAGTCGCCGGCCAAGTCT
```

```
                    529                 550         561
                     *         *         *         *
ananassae           GACGTGACACCGCCCAAGTCGCCCGCCAAGTCC
phaeopleura         GACGTGACGCCGCCCAAGTCGCCTGCCAAGTCC
malerkotiana        GACGTGACGCCGCCAAAGTCGCCCGCCAAGTCT
pallidosa           GACGTGACACCGCCCAAGTCGCCCGCCAAGTCC
varians             GACGTGACGCCGCCCAAGTCGCCCGCCAAGTCC
ercepeae            GACGTGACGCCGCCAAAGTCGCCCGCCAAGTCC
baimaii             GATGTGACGCCGCCCAAGTCGCCGGCCAAGTCT

ficusphila          GATGTCACGCCGCCCAAGTCGCCGGCCAAGTCC
elegans             GATGTCACACCACCGAAATCGCCGGCCAAGTCG
paralutea           GATGTCACACCGCCCAAGTCGCCGGCCAAGTCG
prostipennis        GATGTCACGCCGCCCAAGTCGCCGGCCAAGTCG
takahashii          GATGTCACGCCGCCCAAGTCGCCGGCCAAGTCG
lutescens           GATGTCACACCGCCCAAGTCGCCGGCCAAGTCG
lucipennis          GATGTCACACCGCCCAAATCGCCGGCCAAGTCG
mimetica            GATGTCACACCGCCCAAGTCGCCGGCCAAGTCC
biarmipes           GATGTCACACCGCCCAAGTCGCCGGCCAARTCC
eugracilis          GATGTCACACCACCCAAGTCGCCGGCCAAGTCG
yakuba              GATGTCACACCGCCCAAGTCGCCGGCCAAGTCT
teissieri           GATGTCACACCGCCCAAGTCGCCGGCCAAGTCG
melanogaster        GATGTCACACCGCCTAAGTCGCCGGCCAAGTCG
```

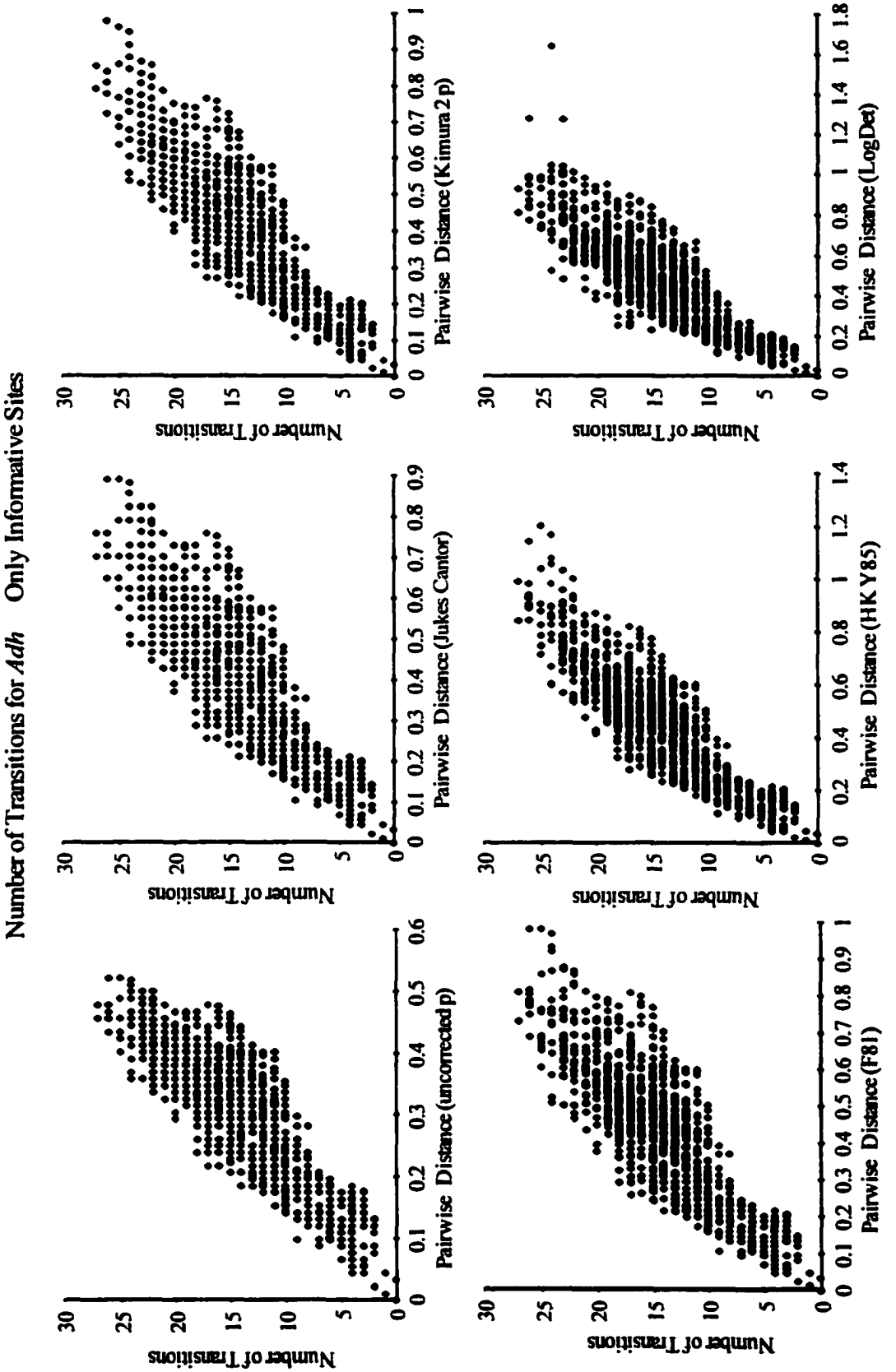
Table 2. Coding of gaps as characters. Gaps at amino acid positions 58-59 and 138 were recoded as characters then reinserted as the primary data in the matrix. Each of these amino acid sequences are present in the six outgroup taxa and are lost for the taxa of the melanogaster group. Thus this deletion supports the hypothesis of ingroup monophyly. In this instance, the presence of a gap is an informative character. There is no variation exhibited in the size of the gap; therefore, the deletion producing the gap may have occurred only once. These data were condensed into a single unordered binary or multistate character.

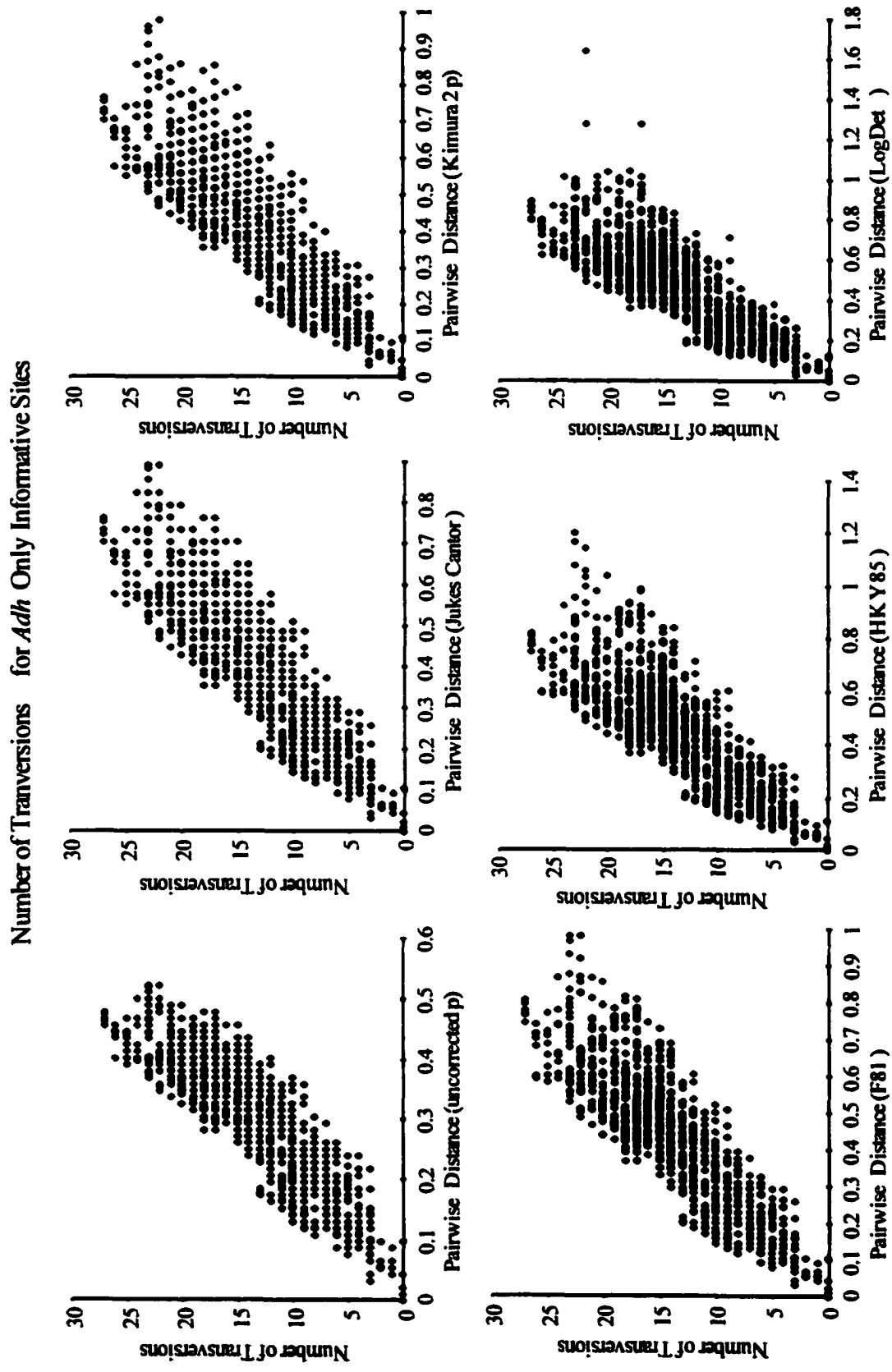
Species	Sequence	Binary Coding Matrix	
		A	B
ambigua	GGGGCA	000000	0
persimilis	GGGGCA	000000	0
pseudoobscura	GGGGCA	000000	0
affinis	GGGGCG	000001	1
bifasciata	GGGGCA	000000	0
tolteca	GGGGCG	000001	1
paralutea	------	111112	2
prostipennis	------	111112	2
takahashii	------	111112	2
lutescens	------	111112	2
lucipennis	------	111112	2
mimetica	------	111112	2
biarmipes	------	111112	2
ananassae	------	111112	2
varians	------	111112	2
phaeopleura	------	111112	2
greeni	------	111112	2
malerkotiana	------	111112	2
pallidosa	------	111112	2
eugracilis	------	111112	2
teissieri	------	111112	2
yakuba	------	111112	2
melanogaster	------	111112	2
bicornuta	------	111112	2
diplacantha	------	111112	2
watanabei	------	111112	2
punjabiensis	------	111112	2
seguyi	------	111112	2
vulcana	------	111112	2
nikananu	------	111112	2
auraria	------	111112	2
barbarae	------	111112	2
birchii	------	111112	2
kikkawai	------	111112	2
lini	------	111112	2
quadraria	------	111112	2
serrata	------	111112	2
triauraria	------	111112	2
tsacasi	------	111112	2
baimaii	------	111112	2
biauraria	------	111112	2
kanapiae	------	111112	2
mayri	------	111112	2
orosa	------	111112	2
parvula	------	111112	2
rufa	------	111112	2
ercepeae	------	111112	2
elegans	------	111112	2
ficusphila	------	111112	2

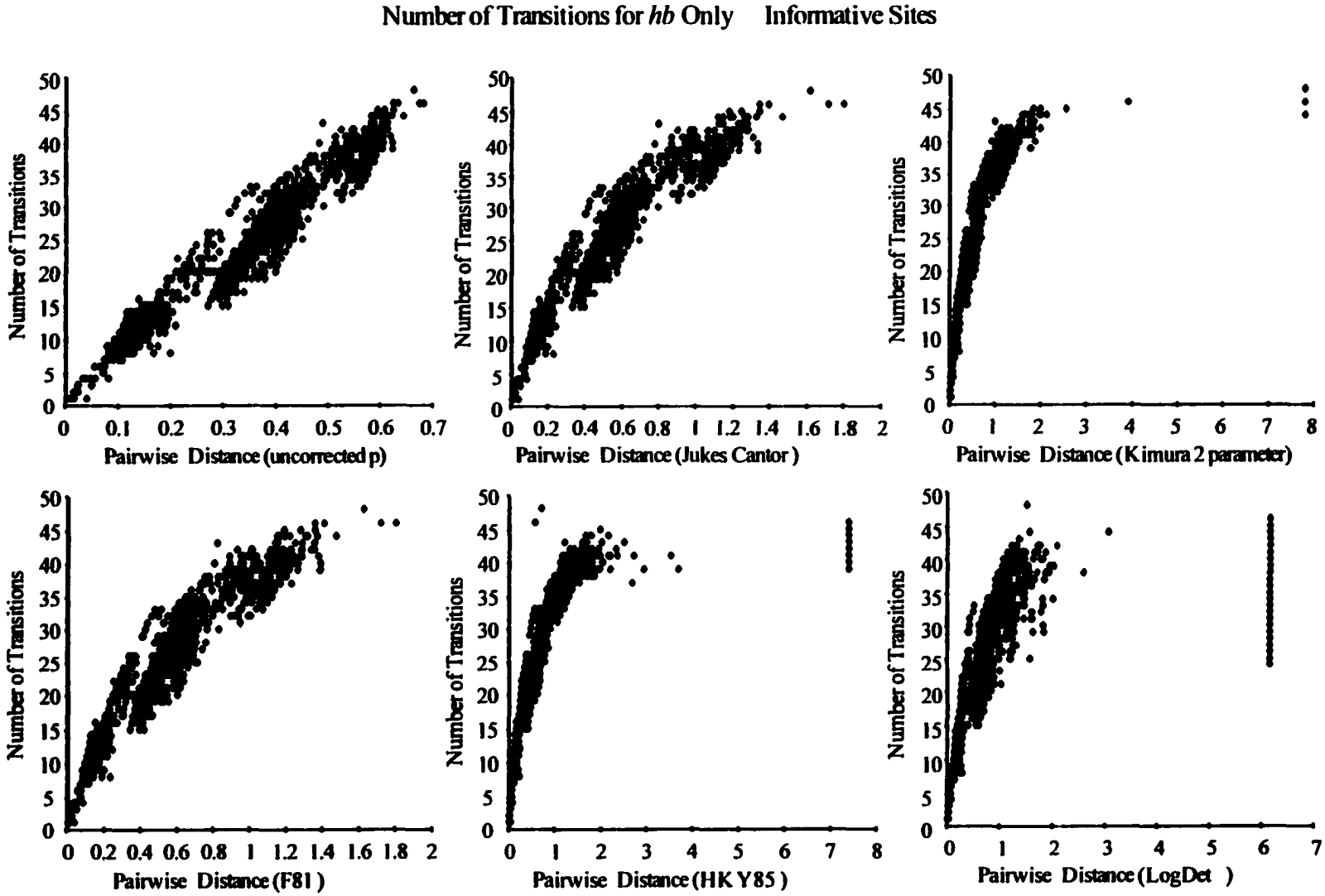
Species	Sequence	Binary Coding Matrix	
		A	B
ambigua	GGT	000	0
persimilis	GGT	000	0
pseudoobscura	GGT	000	0
affinis	GGT	000	0
bifasciata	GGT	000	0
tolteca	GGT	000	0
paralutea	---	111	1
prostipennis	---	111	1
takahashii	---	111	1
lutescens	---	111	1
lucipennis	---	111	1
mimetica	---	111	1
biarmipes	---	111	1
ananassae	---	111	1
varians	---	111	1
phaeopleura	---	111	1
greeni	---	111	1
malerkotiana	---	111	1
pallidosa	---	111	1
eugracilis	---	111	1
teissieri	---	111	1
yakuba	---	111	1
melanogaster	---	111	1
bicornuta	---	111	1
diplacantha	---	111	1
watanabei	---	111	1
punjabiensis	---	111	1
seguyi	---	111	1
vulcana	---	111	1
nikananu	---	111	1
auraria	---	111	1
barbarae	---	111	1
birchii	---	111	1
kikkawai	---	111	1
lini	---	111	1
quadraria	---	111	1
serrata	---	111	1
triauraria	---	111	1
tsacasi	---	111	1
baimaii	---	111	1
biauraria	---	111	1
kanapiae	---	111	1
mayri	---	111	1
orosa	---	111	1
parvula	---	111	1
rufa	---	111	1
ercepeae	---	111	1
elegans	---	111	1
ficusphila	---	111	1

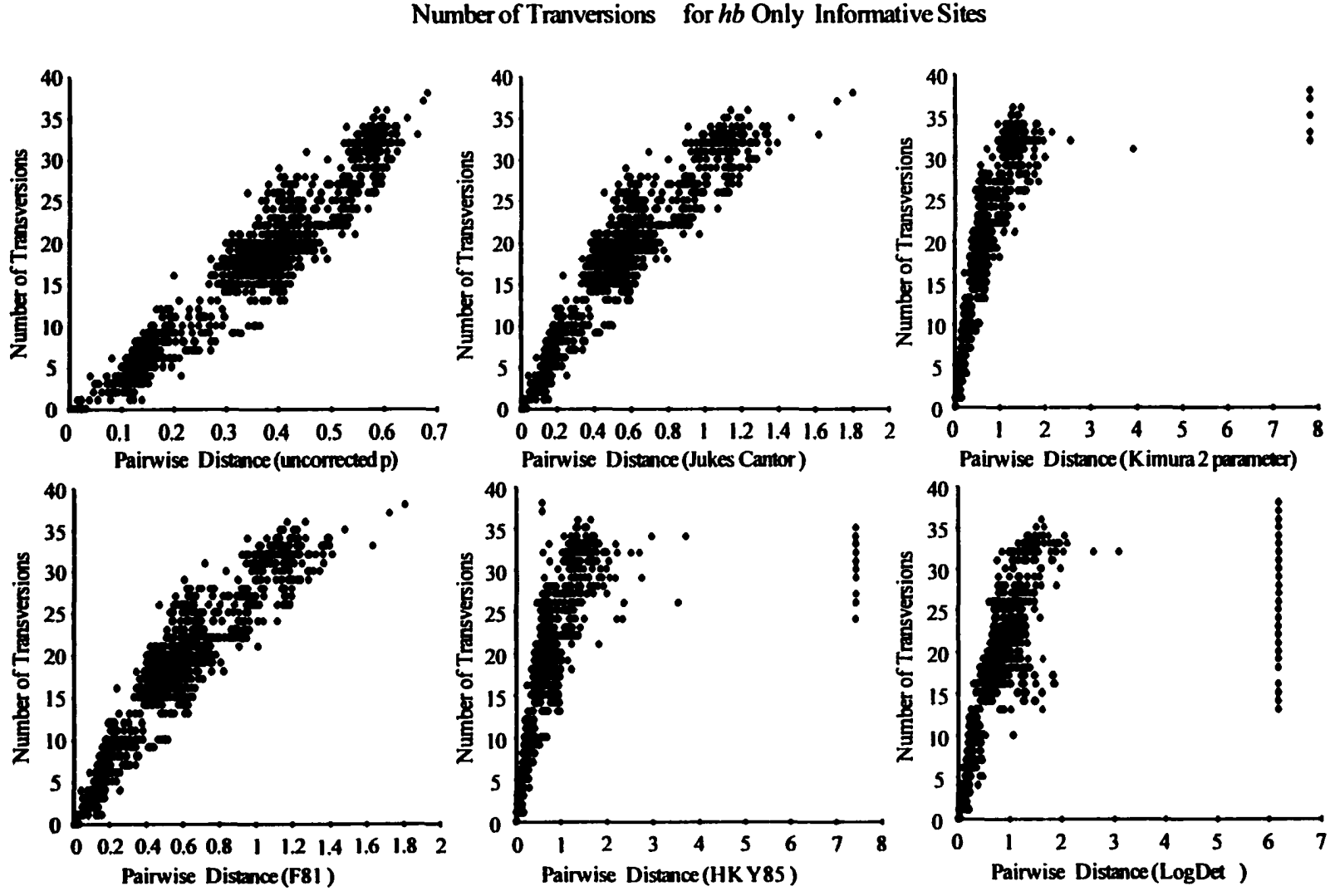
APPENDIX G: Saturation Plots (Chapter2)

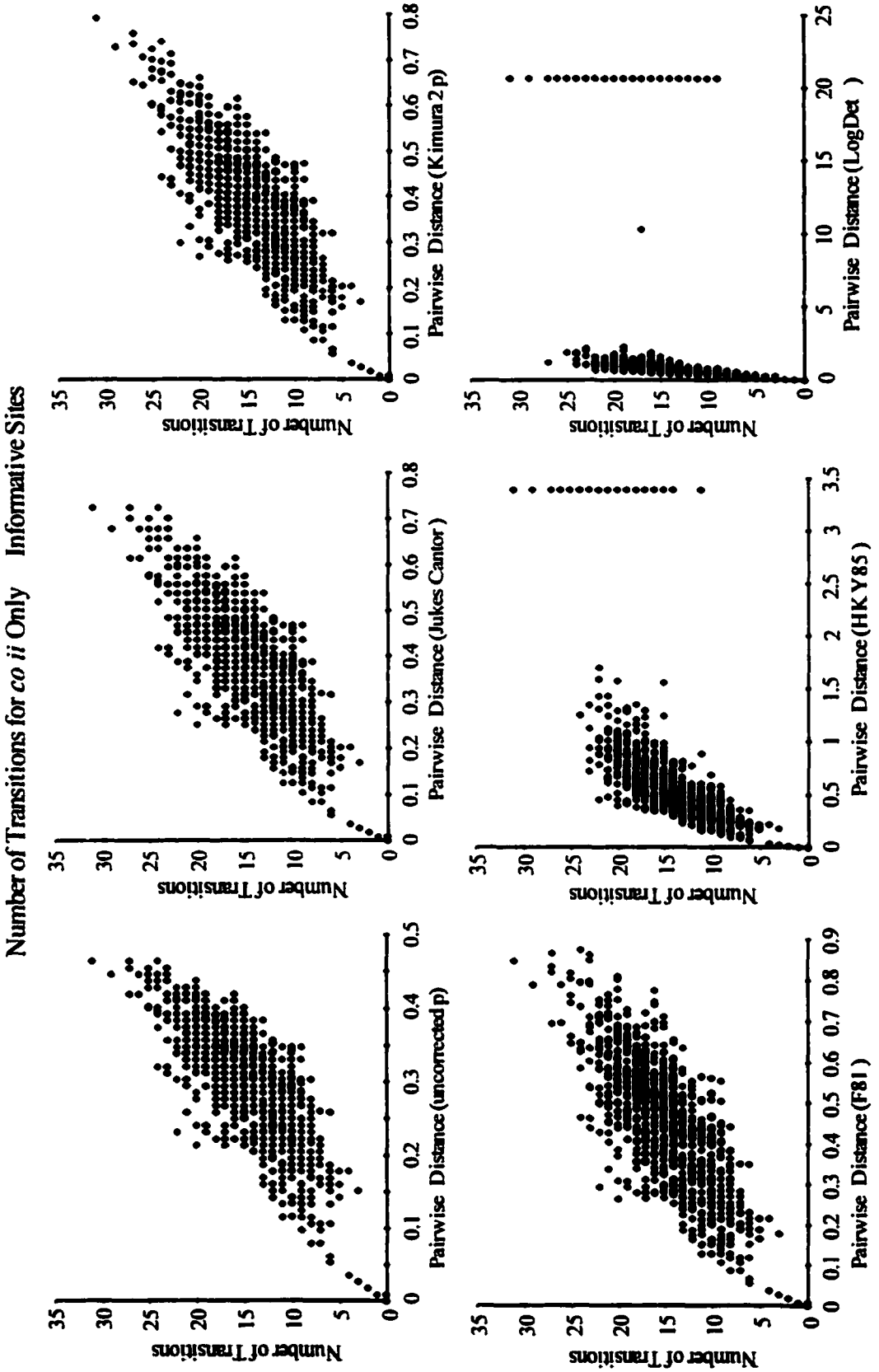
Saturation plots were made for nucleotide composition of transitions, transversions, and codon positions as listed below.

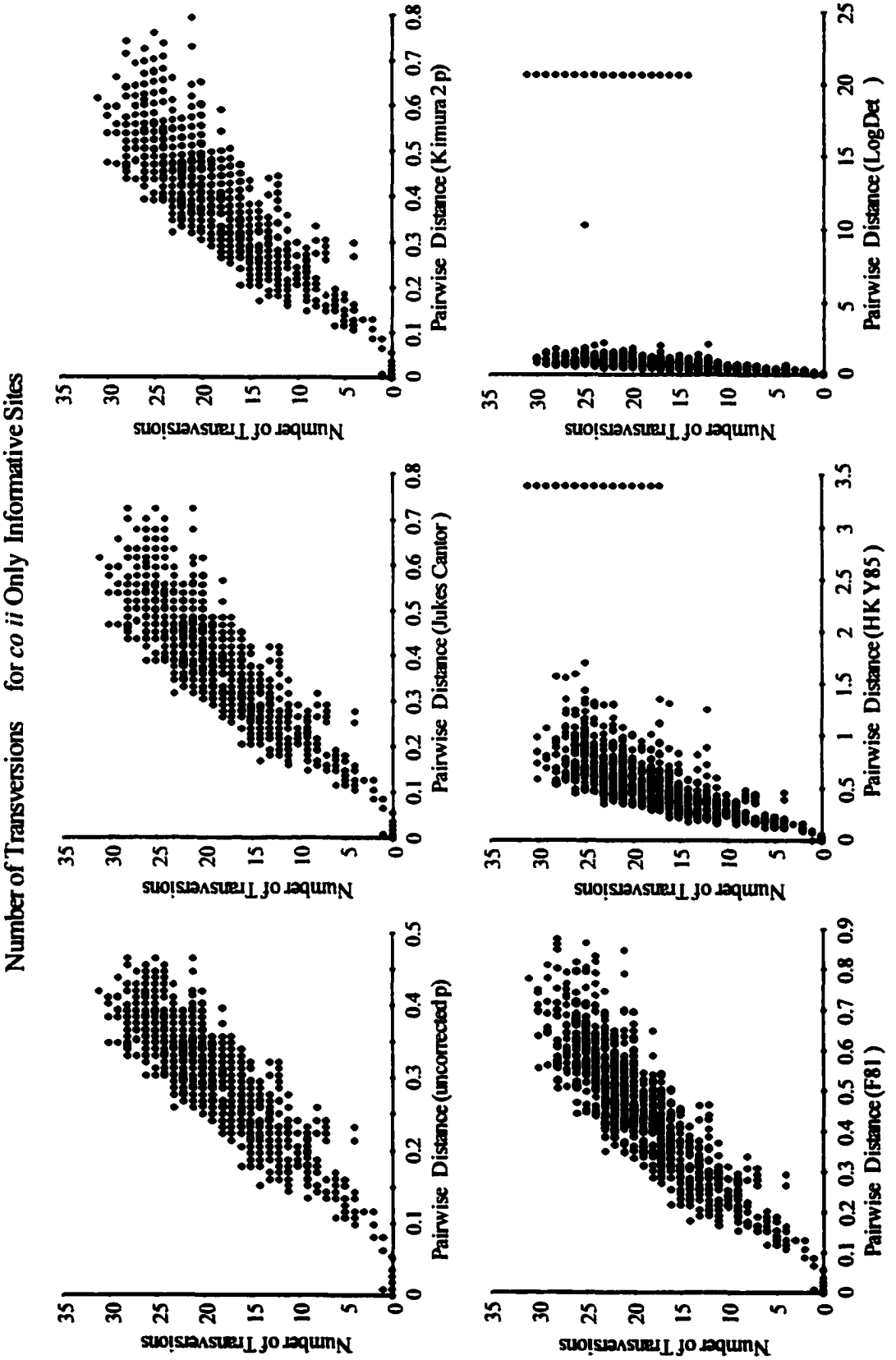


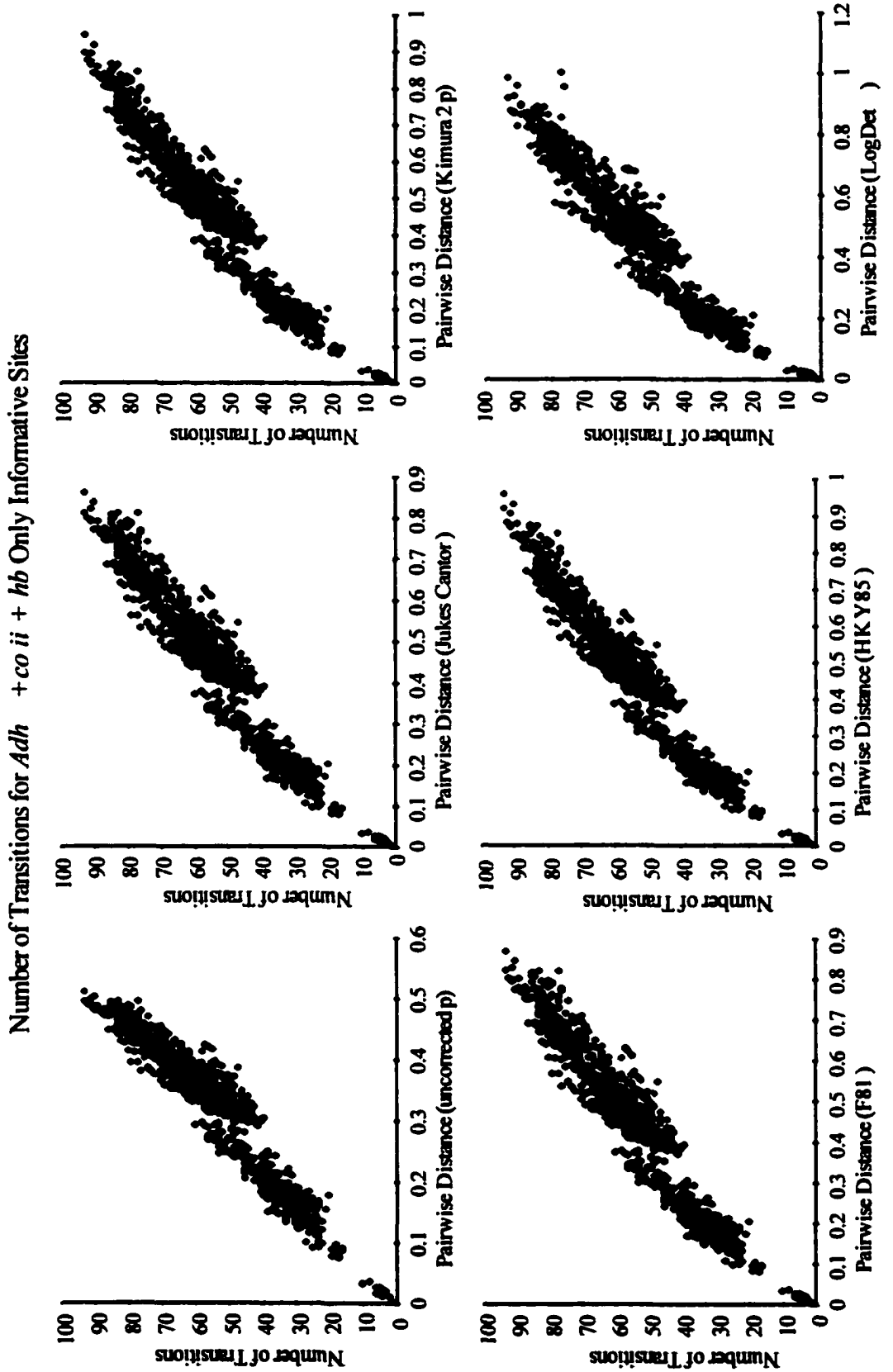


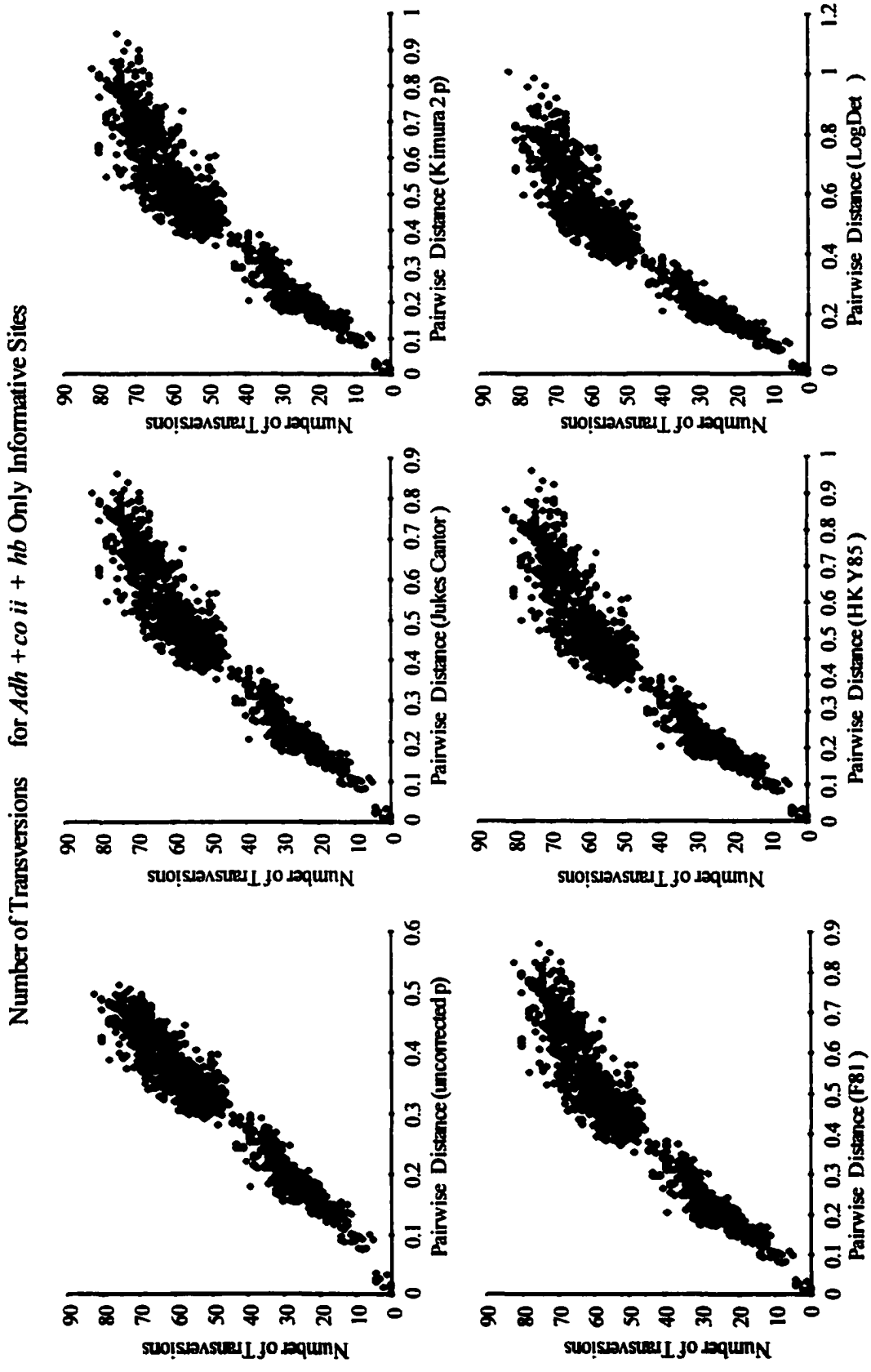


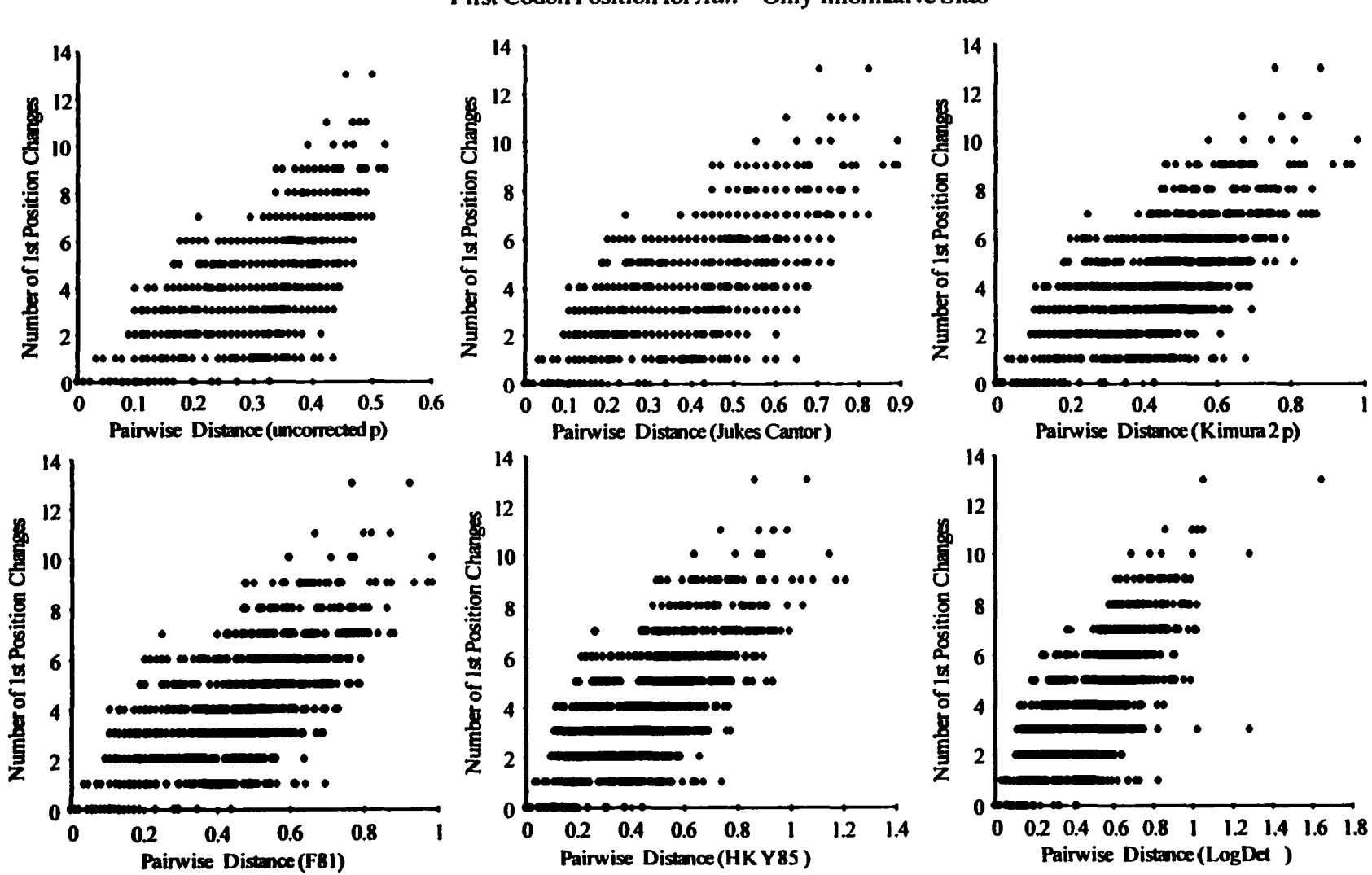


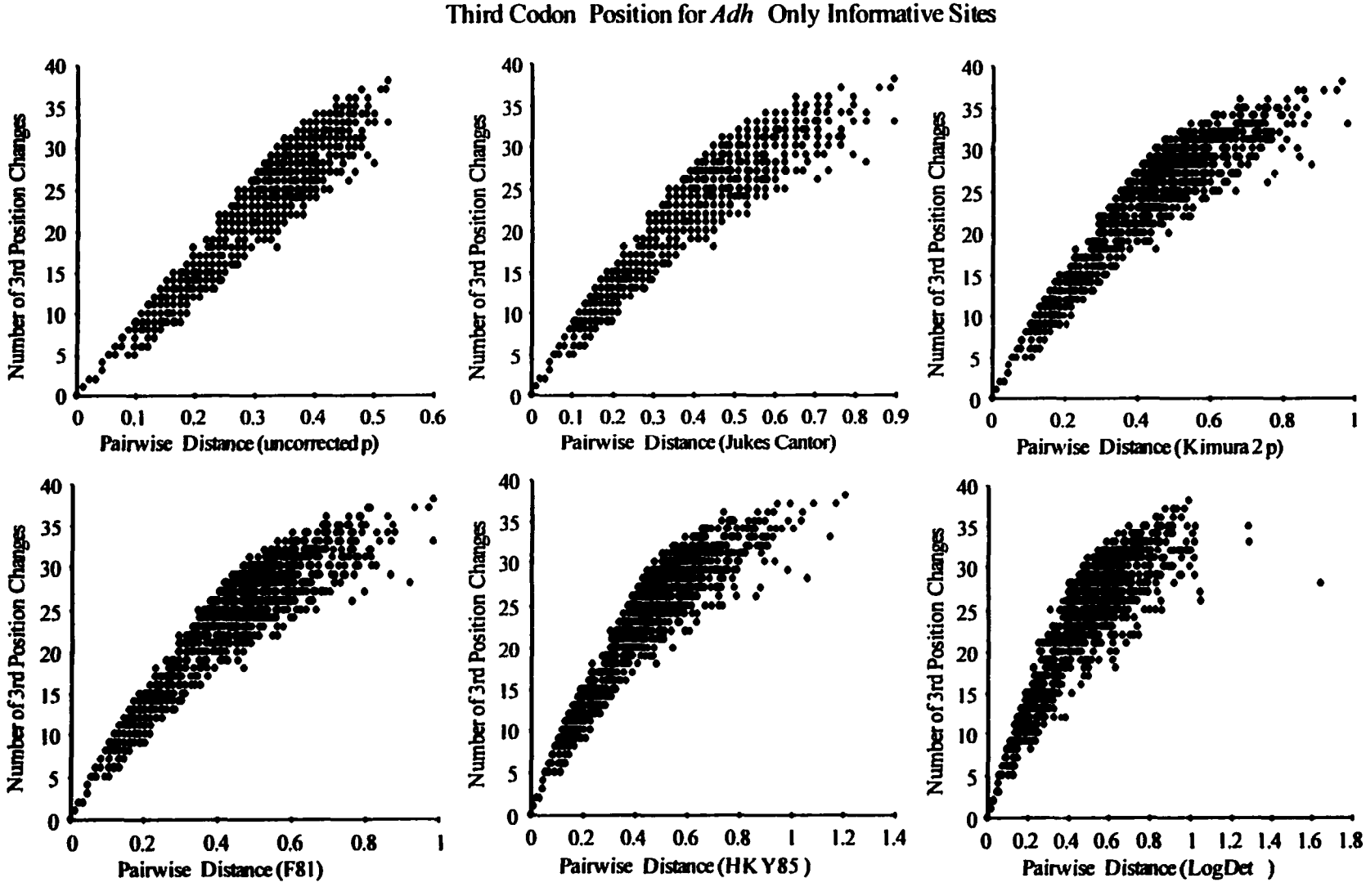


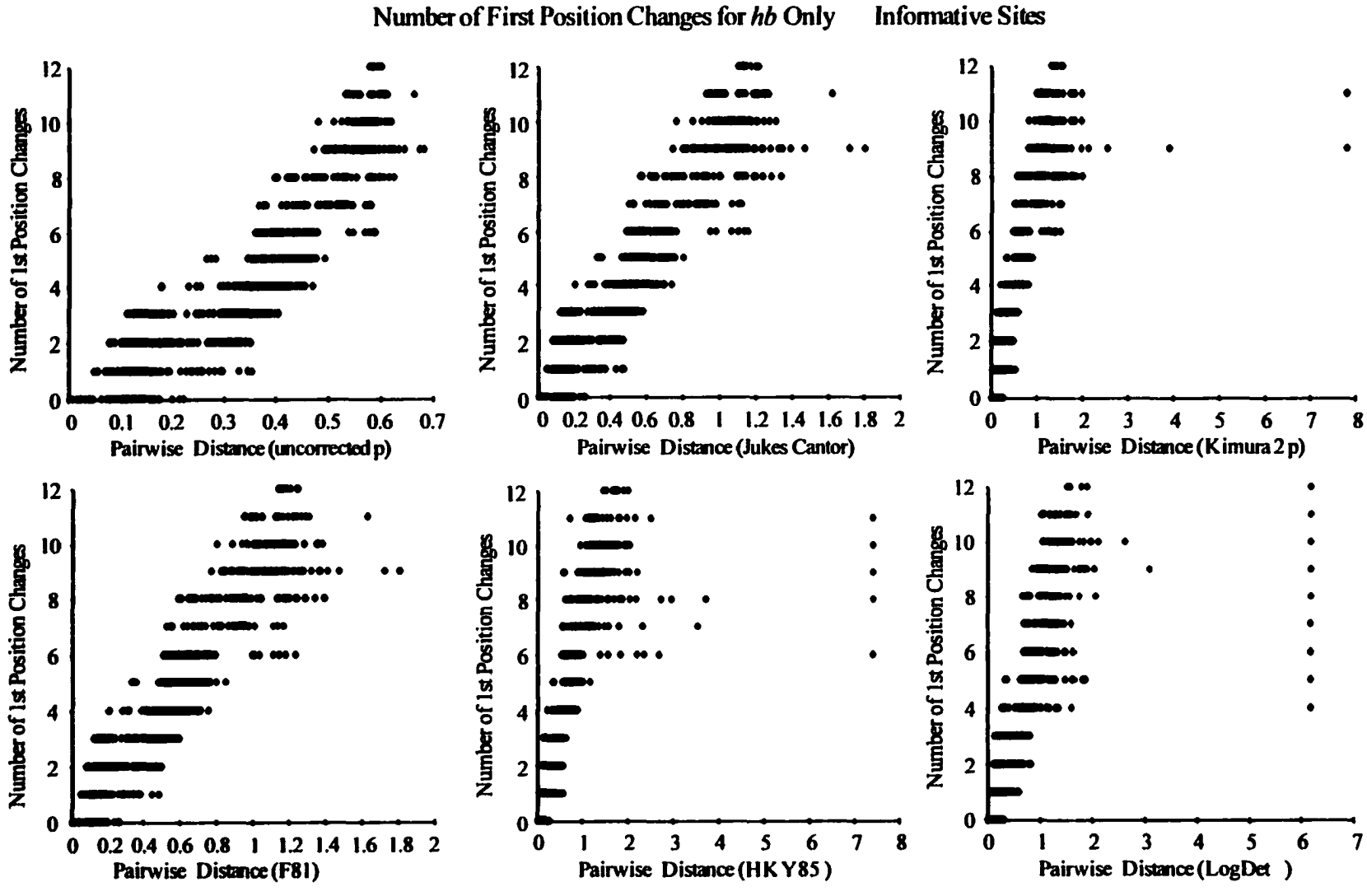


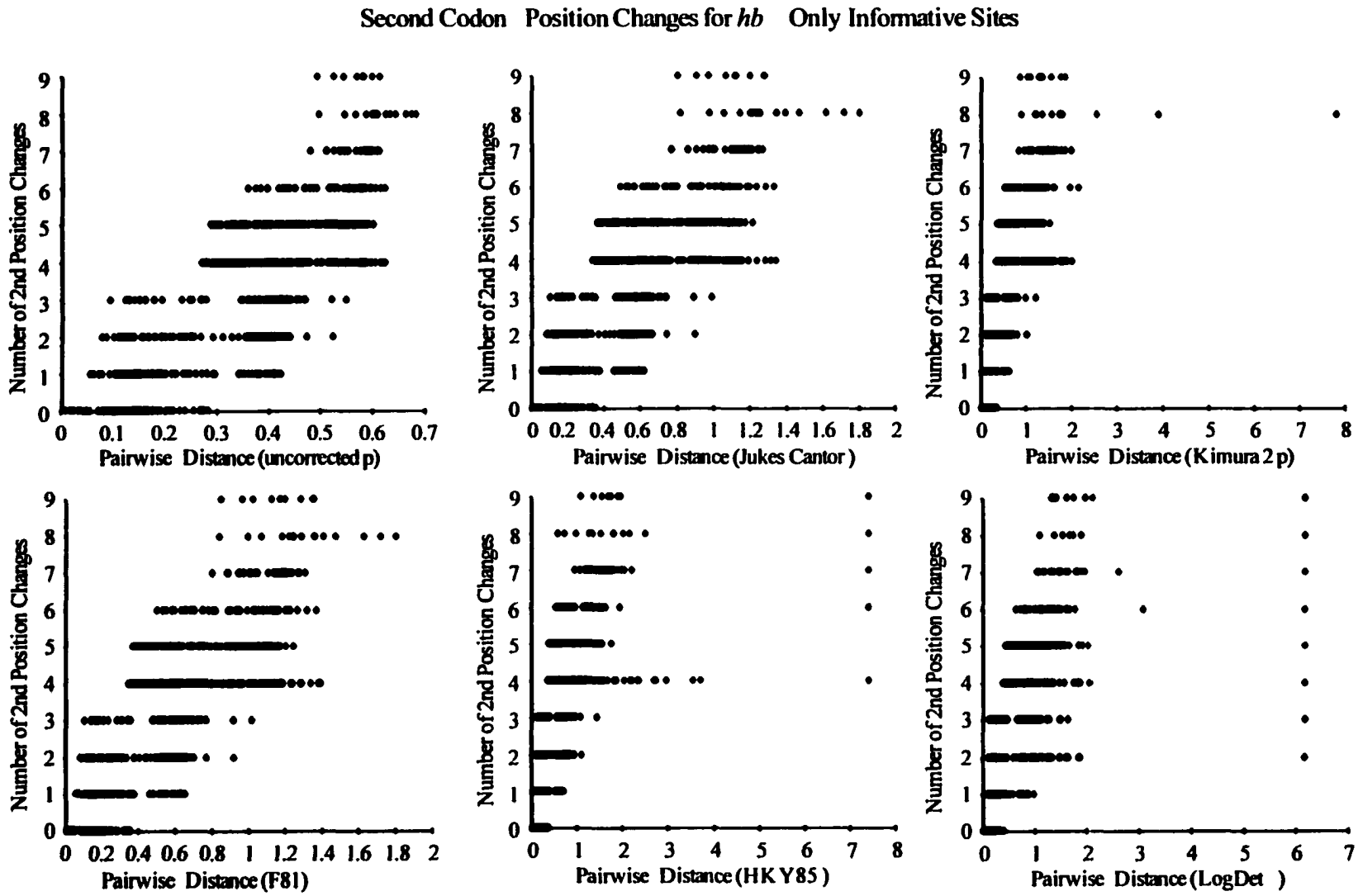


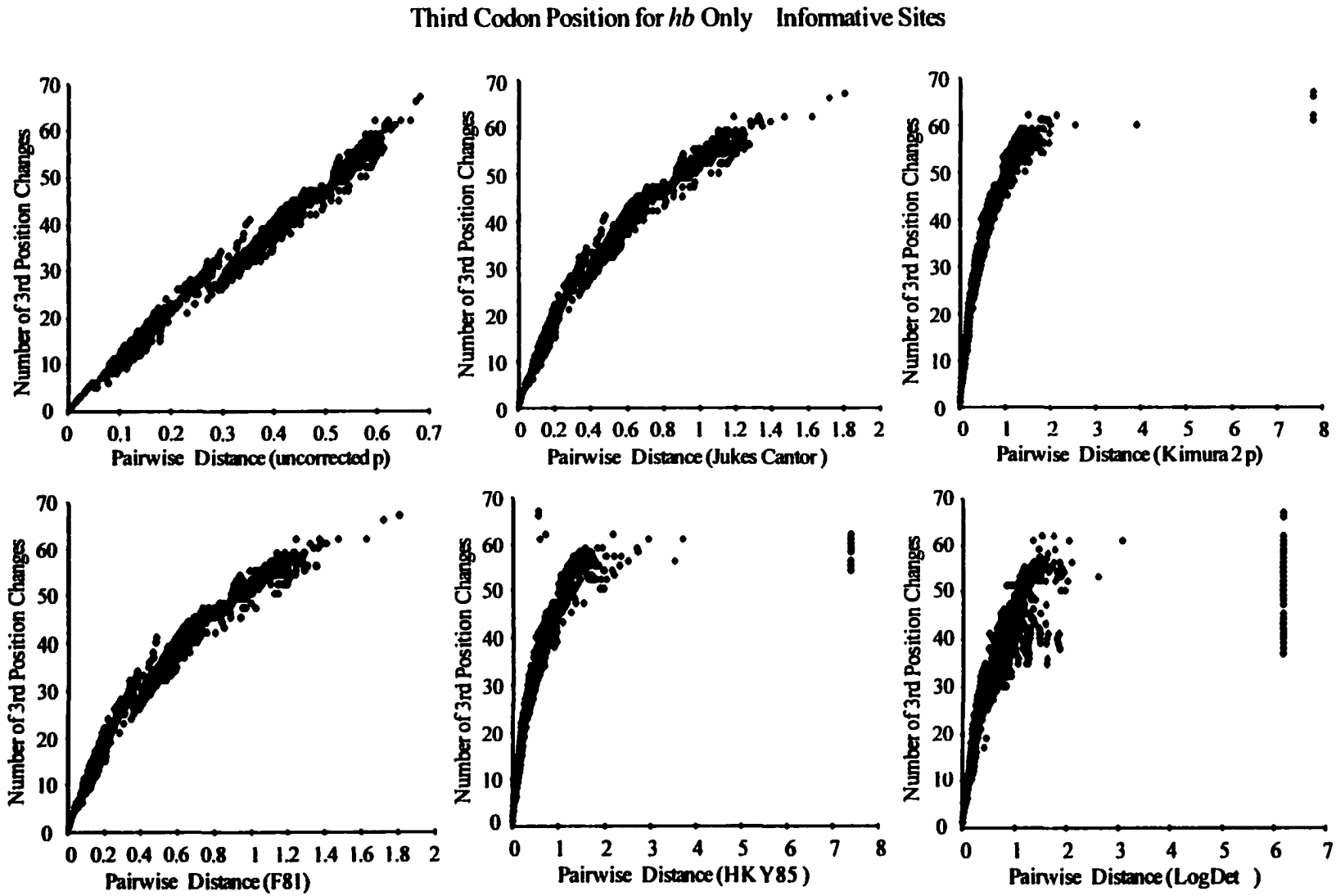


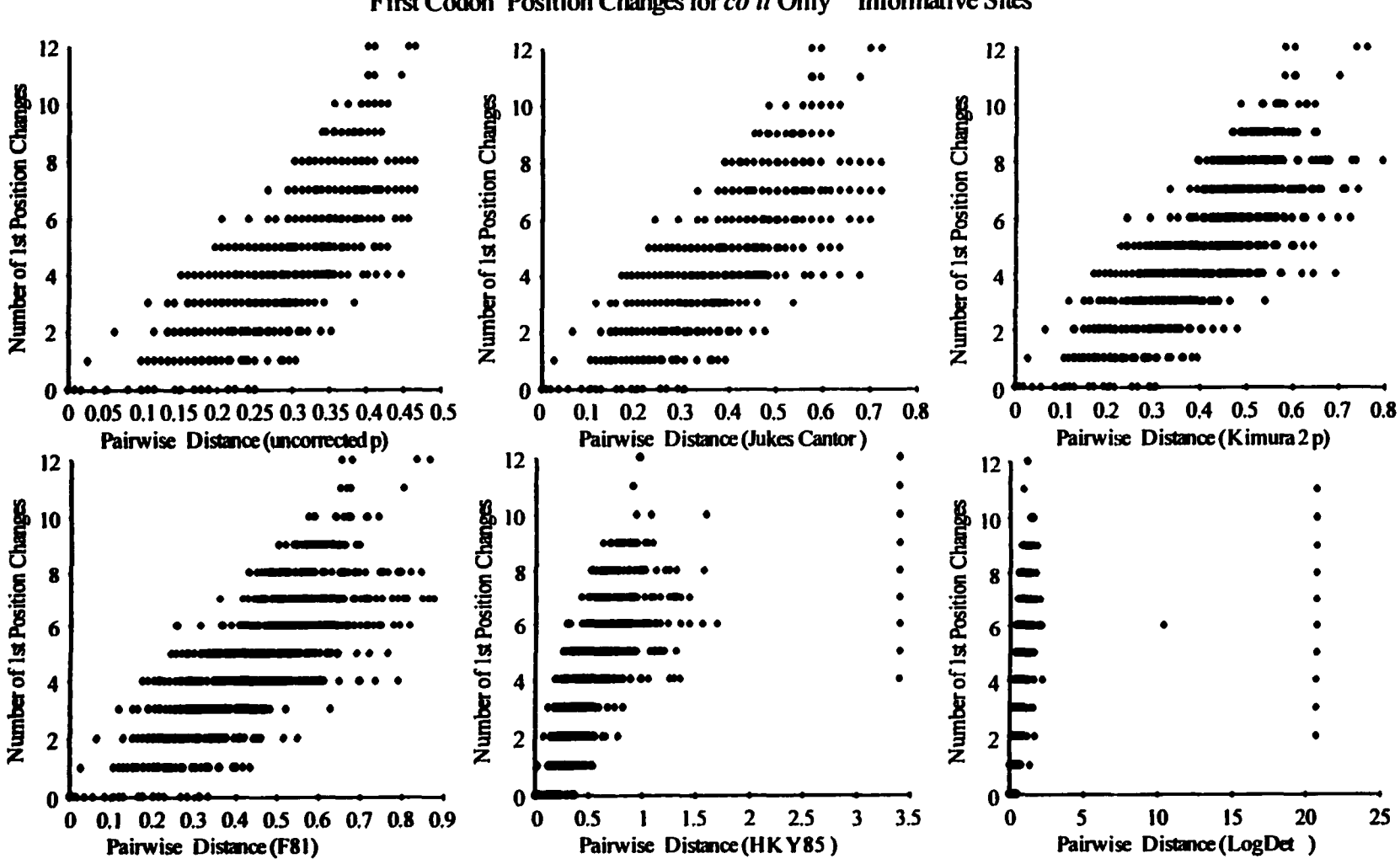


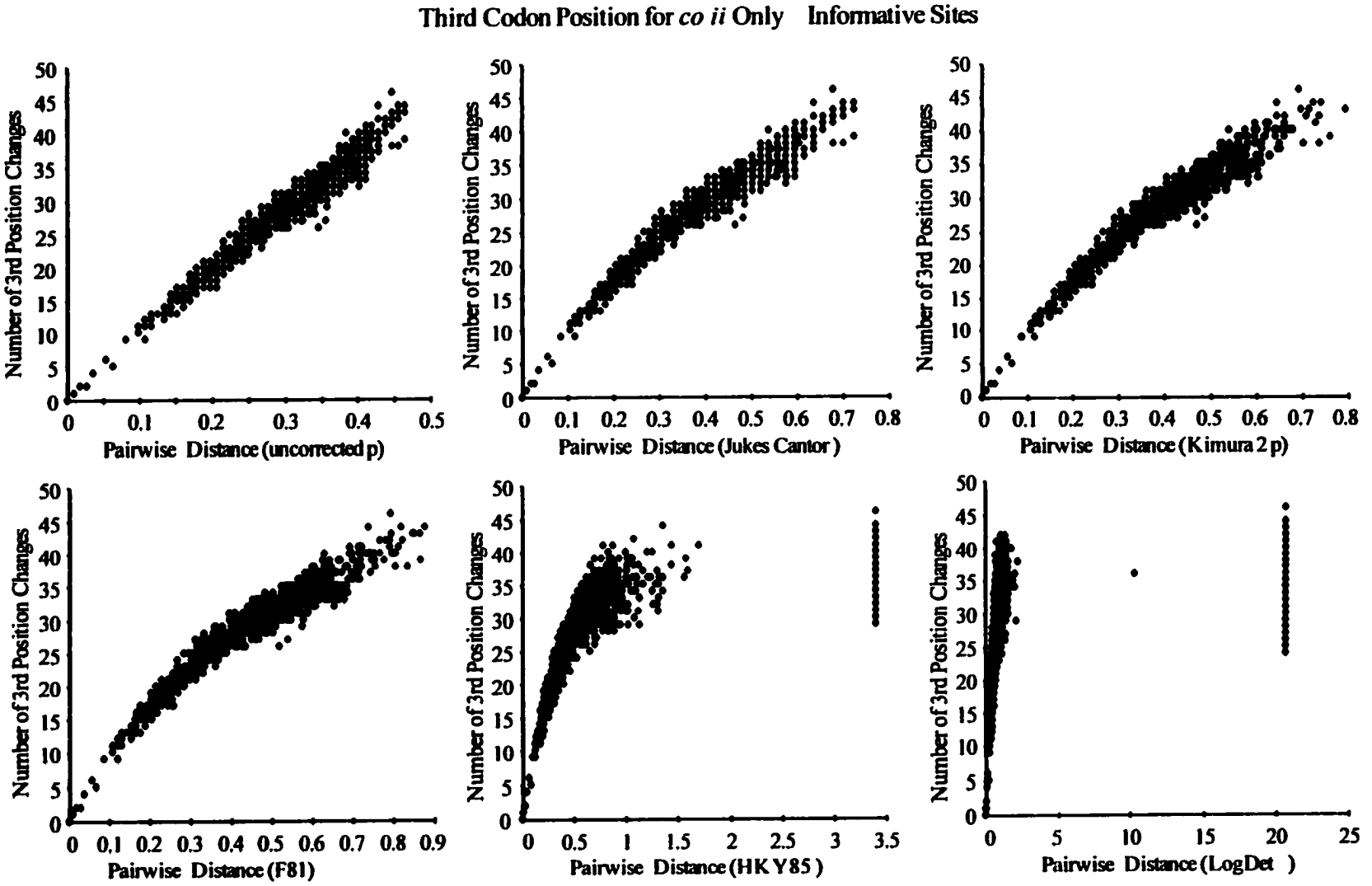


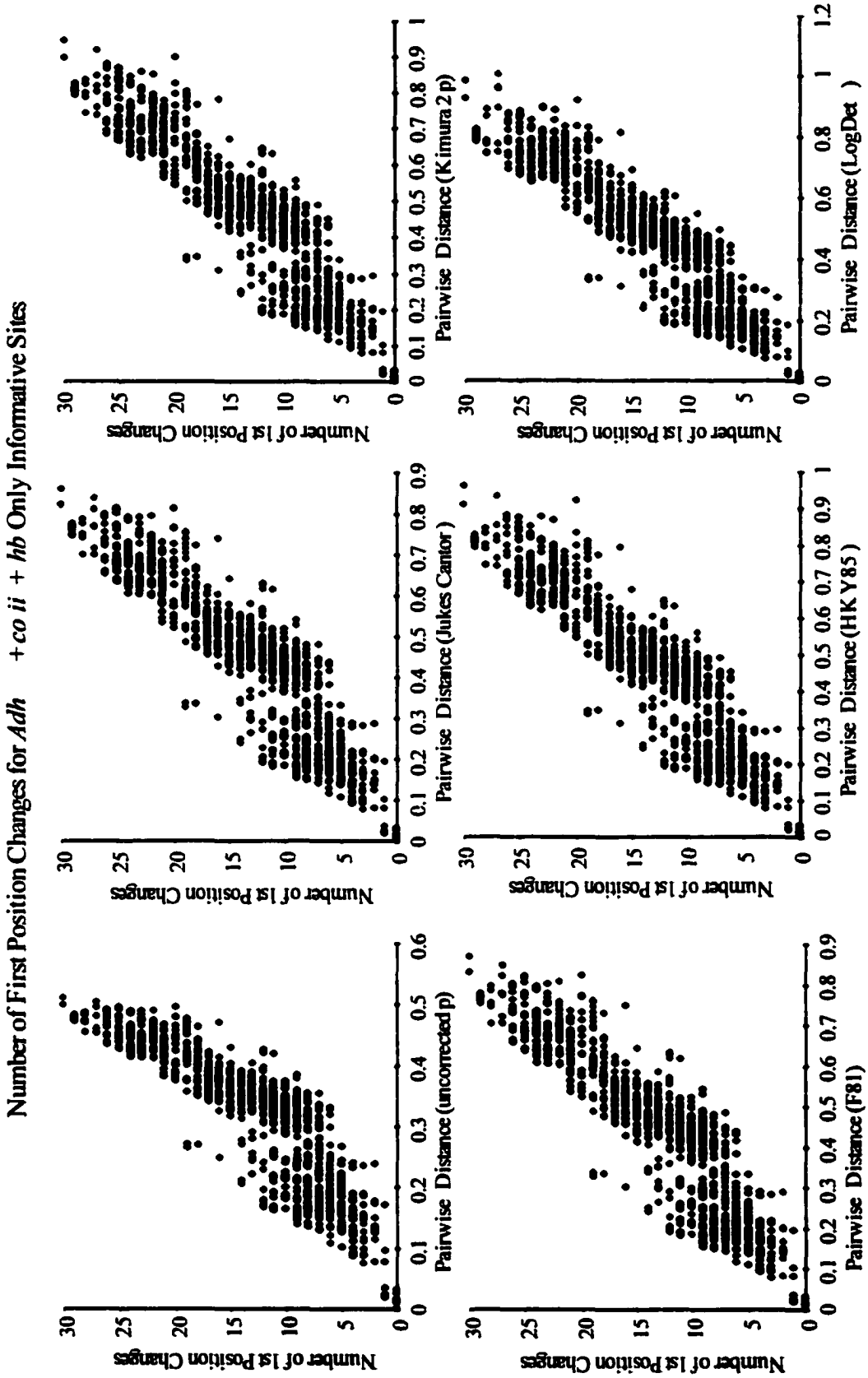


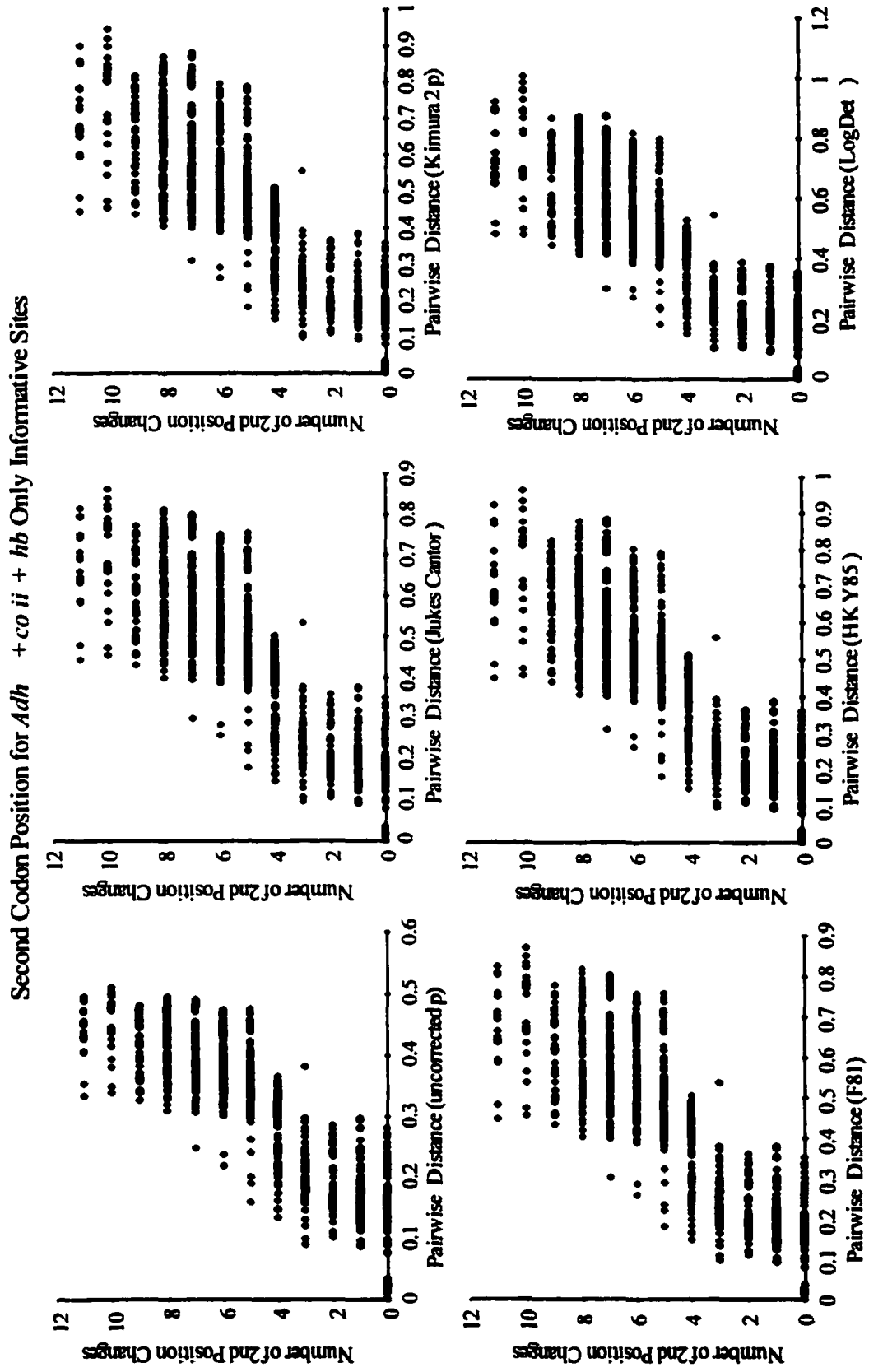


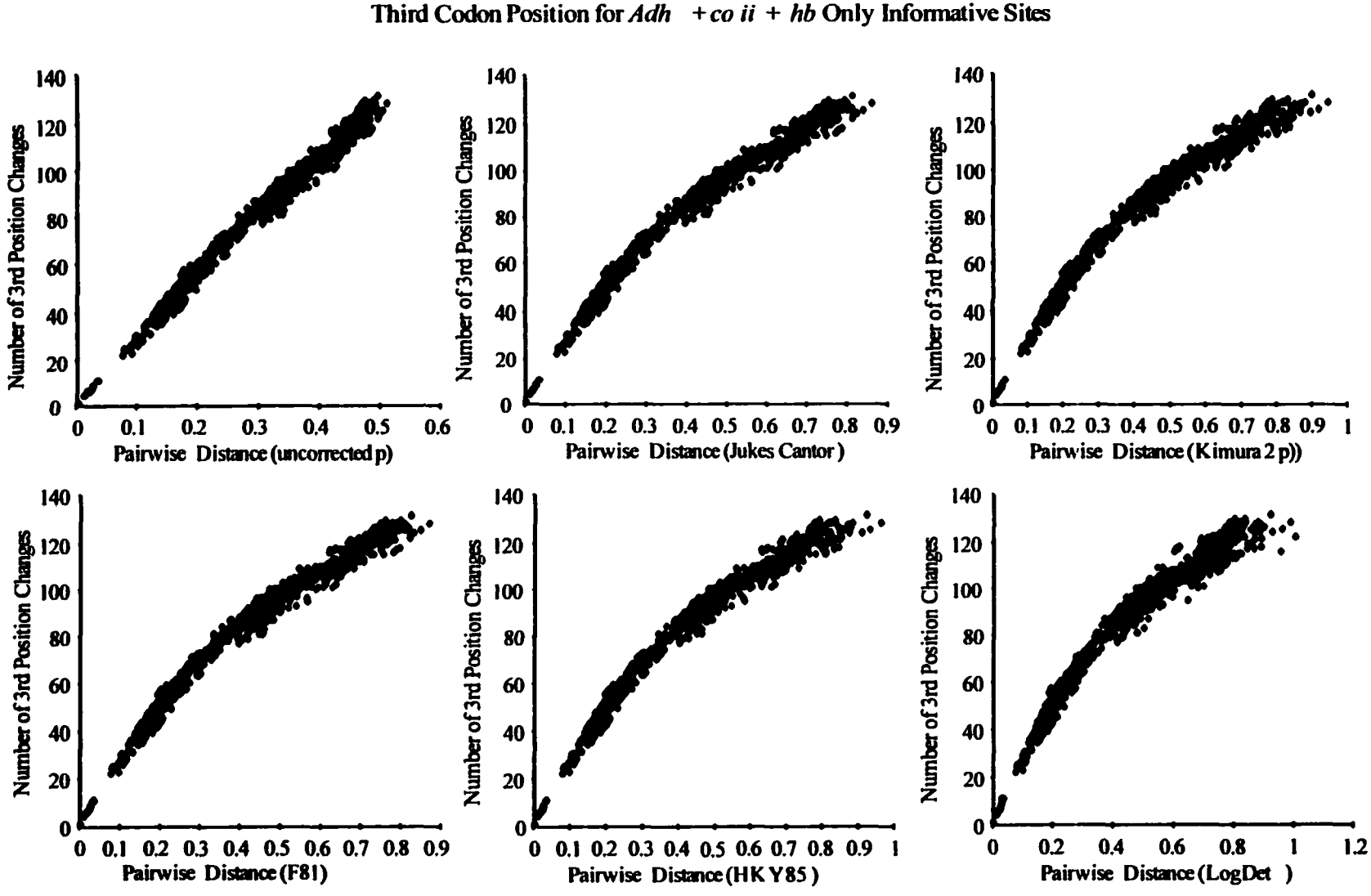






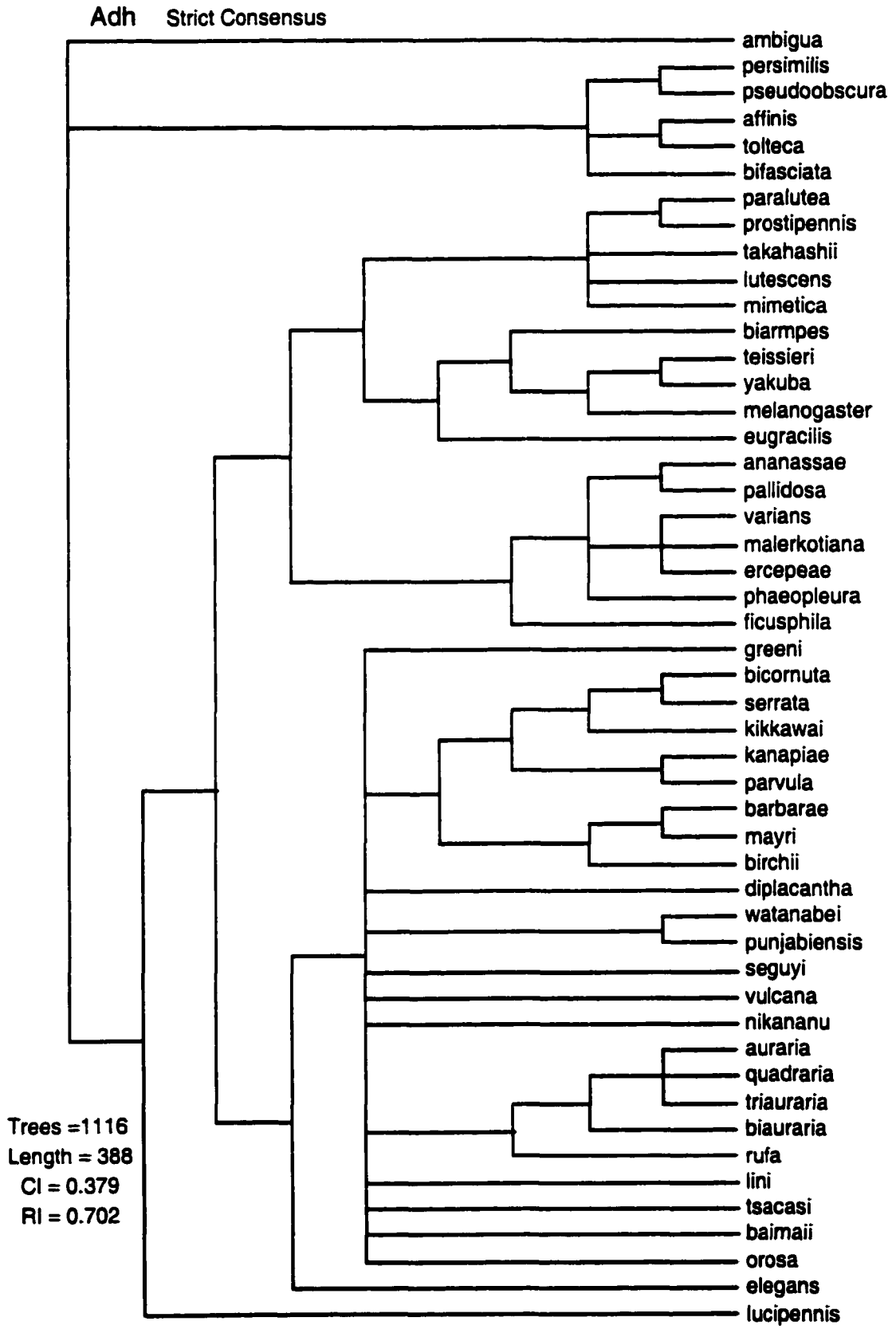


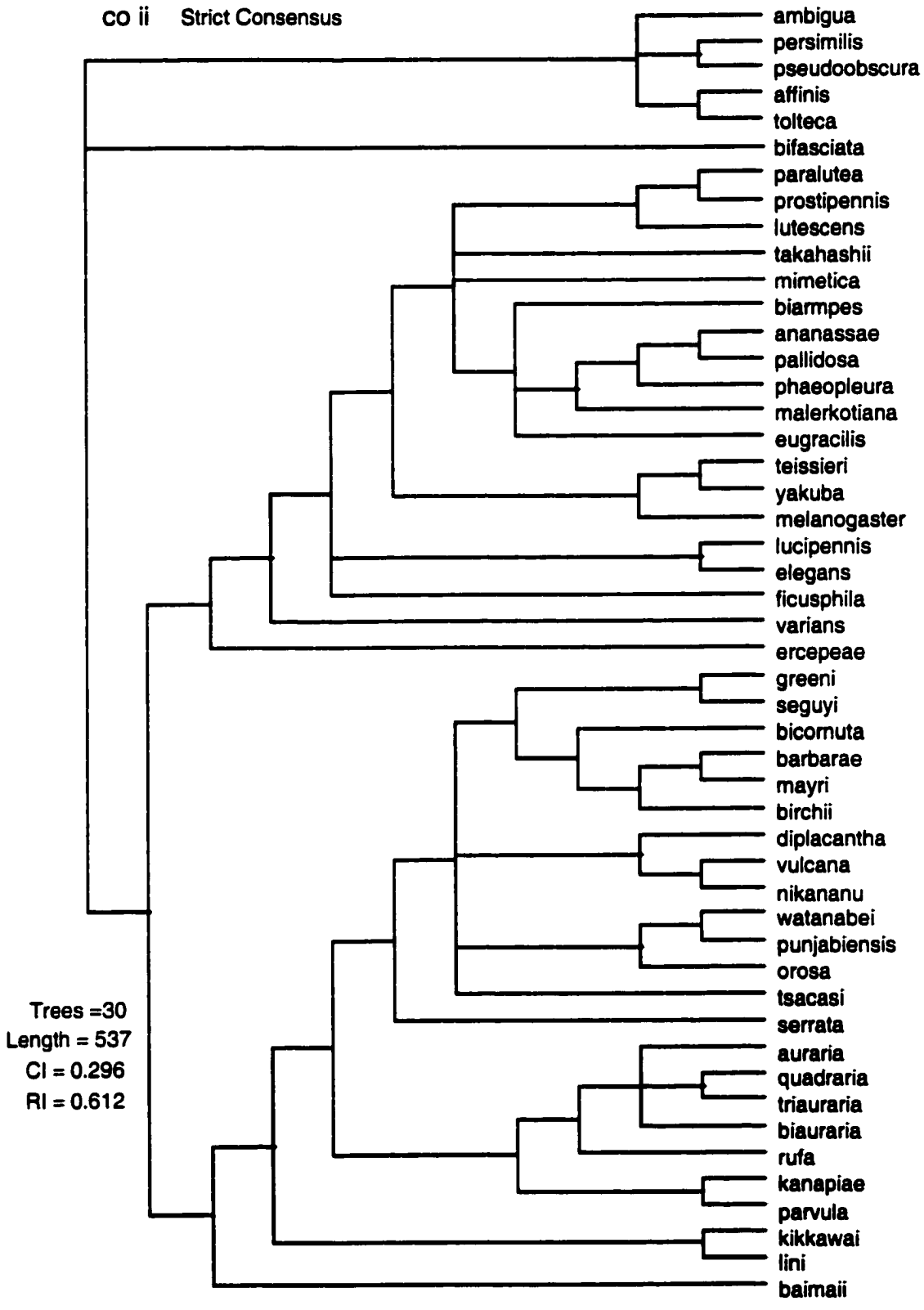


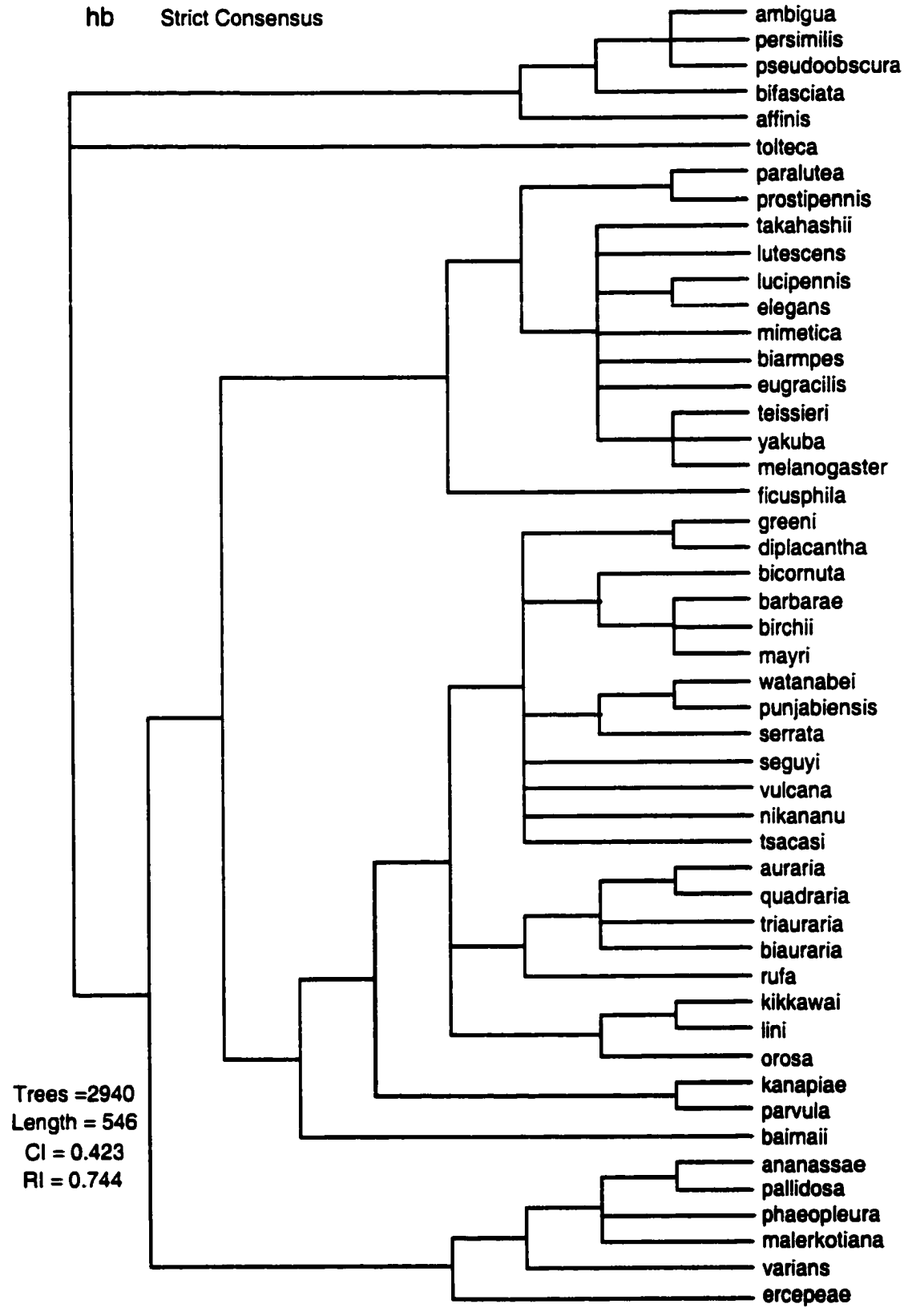


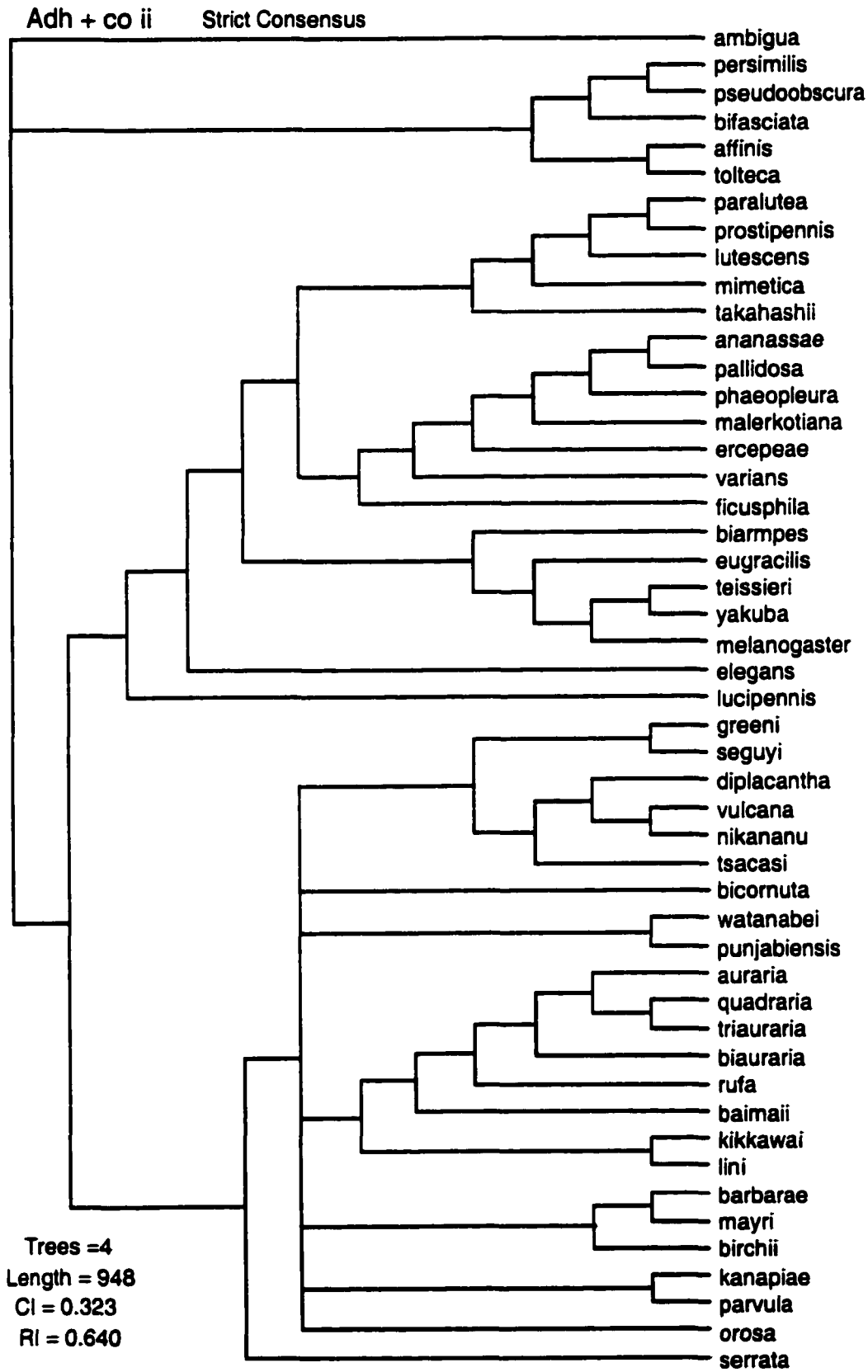
APPENDIX H: Cladograms for the Various Data Partitions (Chapter 2)

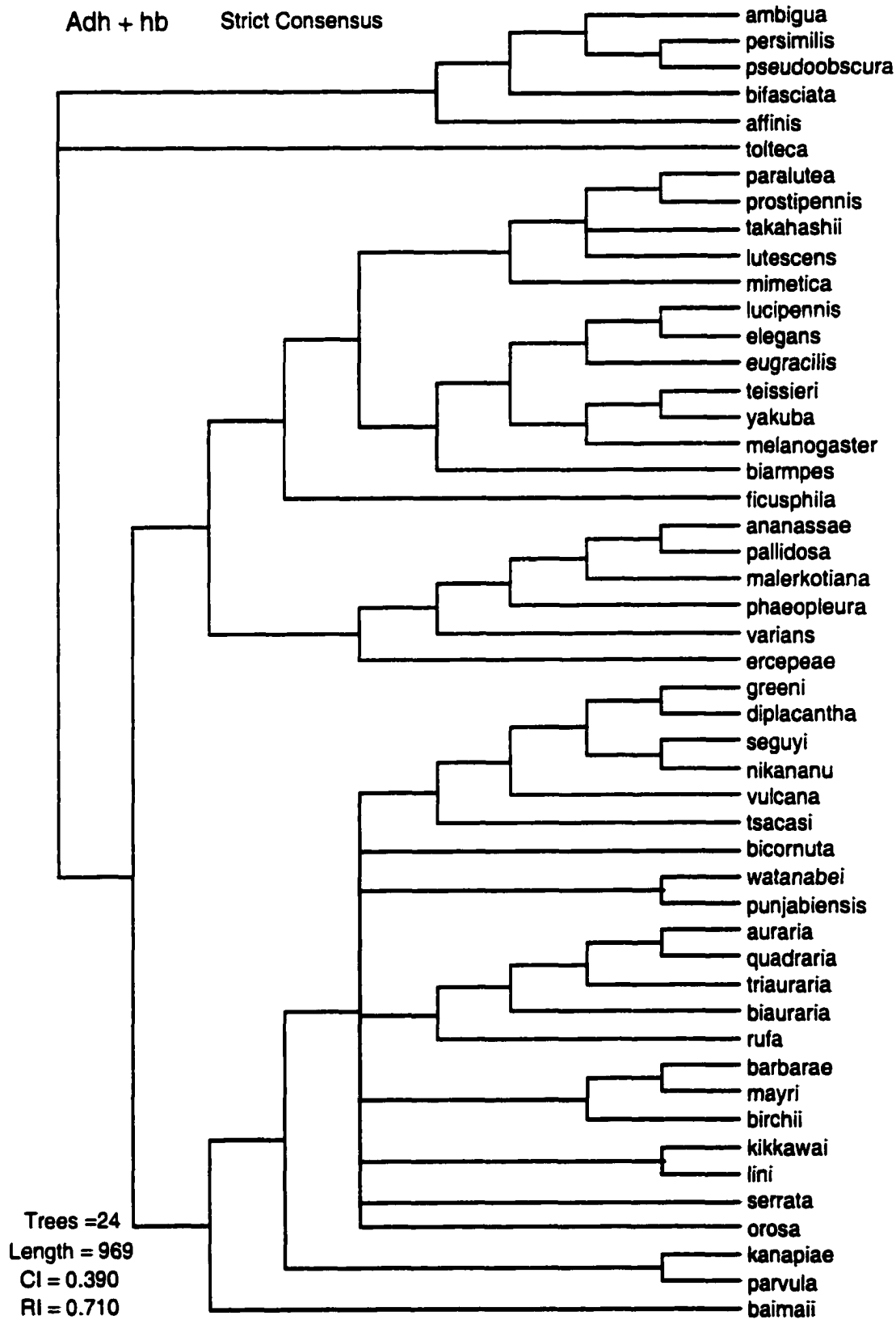
The strict consensus cladograms resulting from the phylogenetic analyses are presented here in the following order: *Adh*, *co ii*, *hb*, *Adh* + *co ii*, *Adh* + *hb*, *hb* + *co ii*. Since *Adh* + *co ii* resulted in only four most parsimonious trees they are presented here. Successive weighting / successive approximations analyses were performed and the resulting tree is presented here (all are a strict consensus except for the *Adh* + *hb* successive weighting tree).

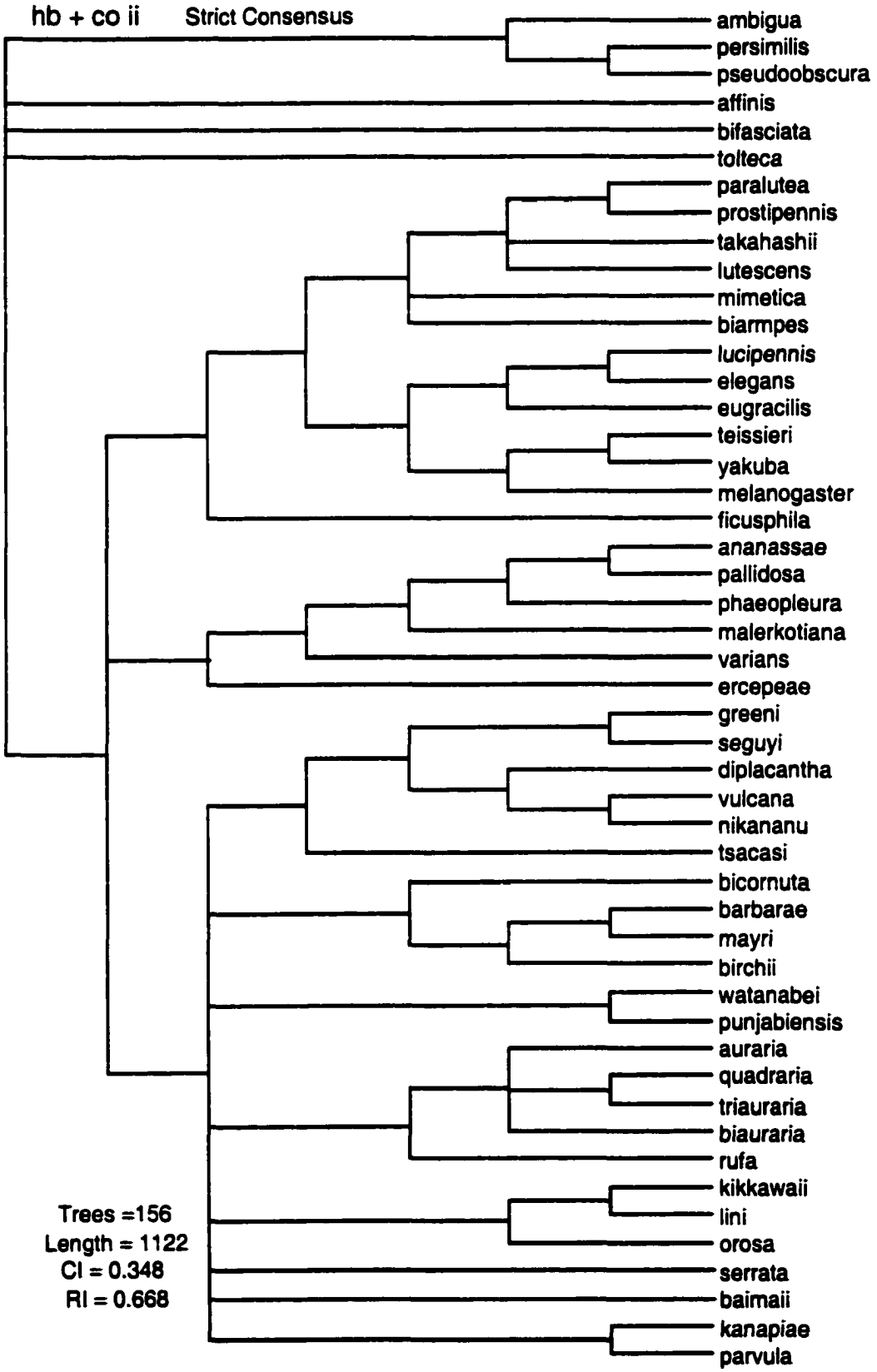


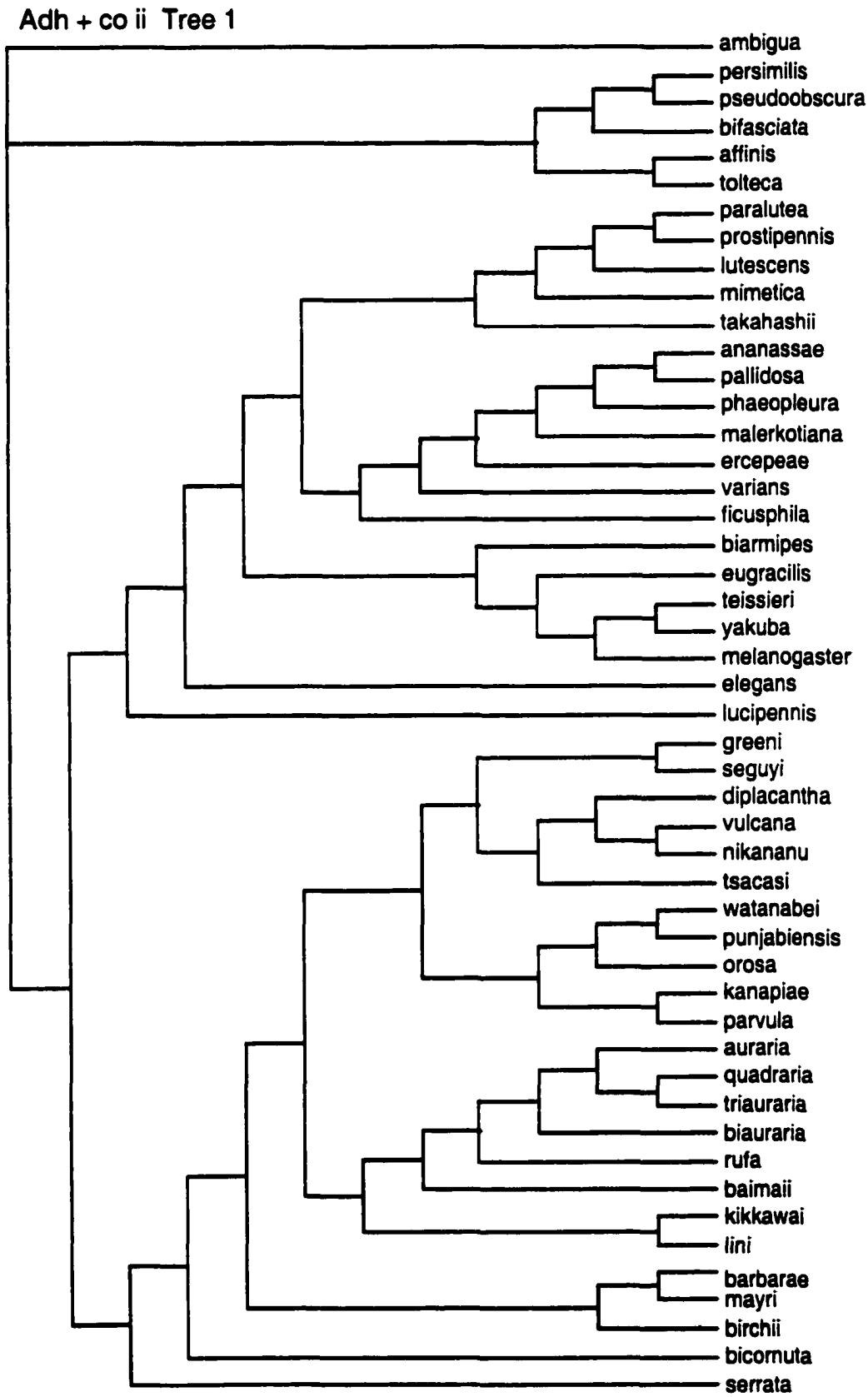


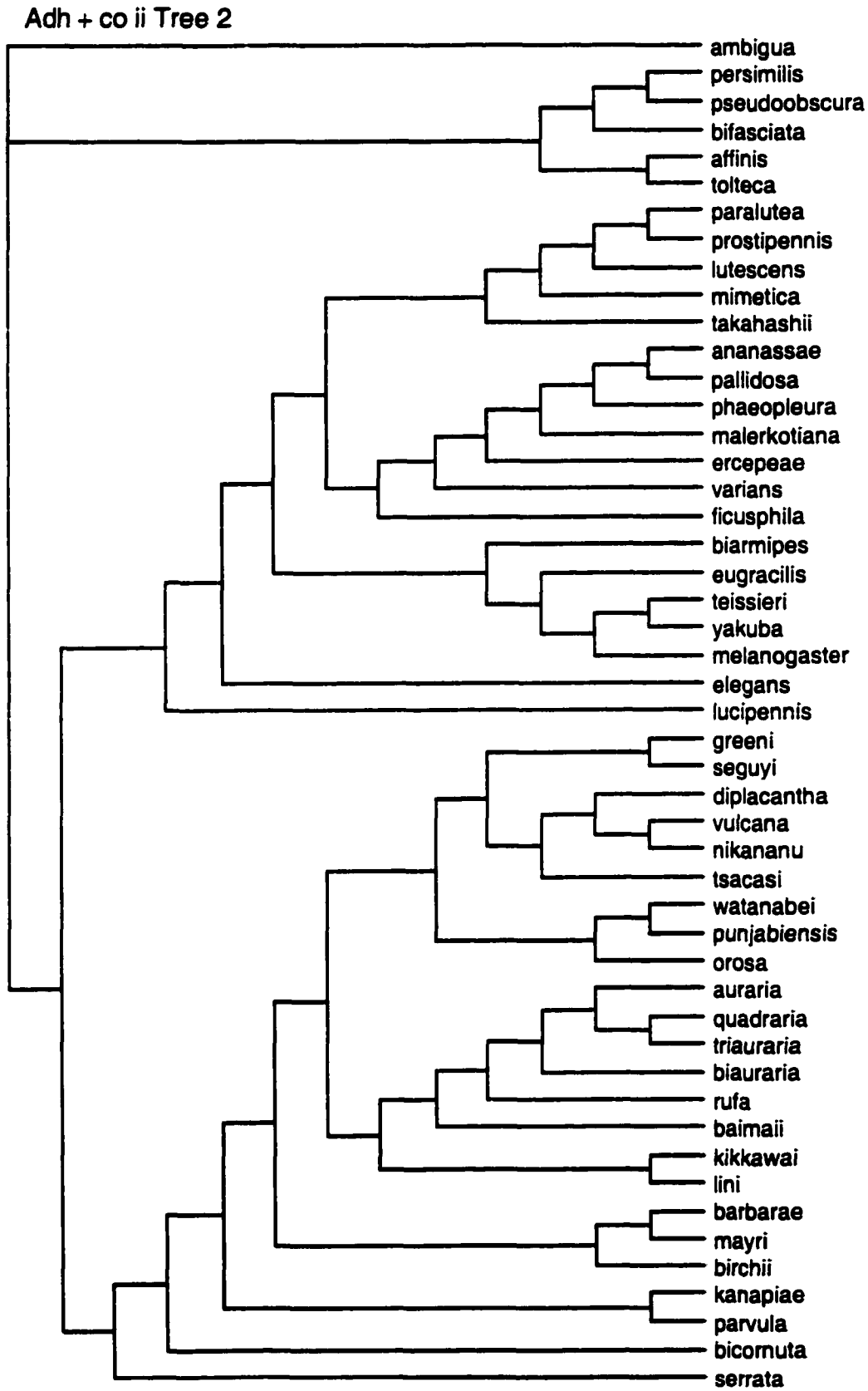


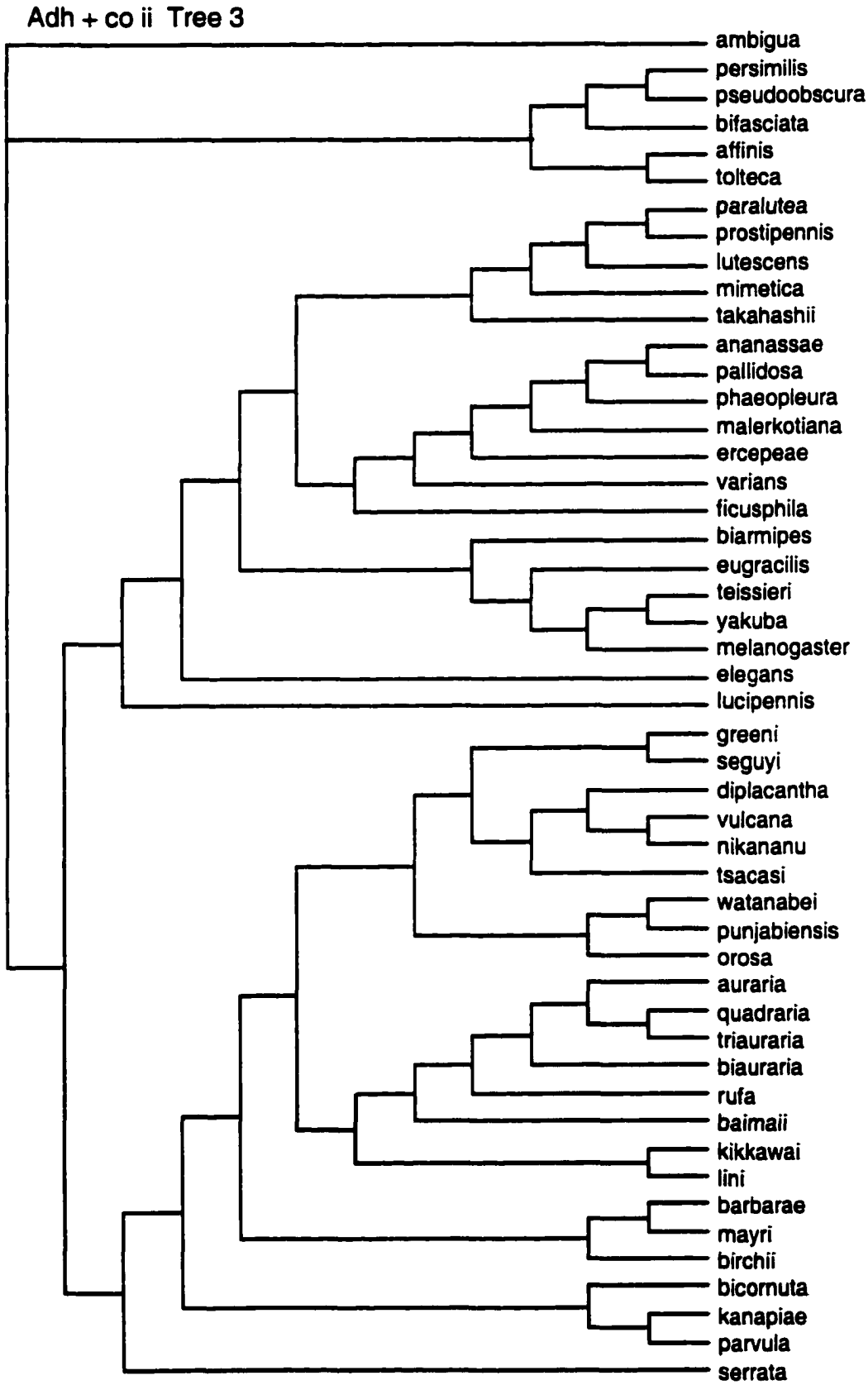


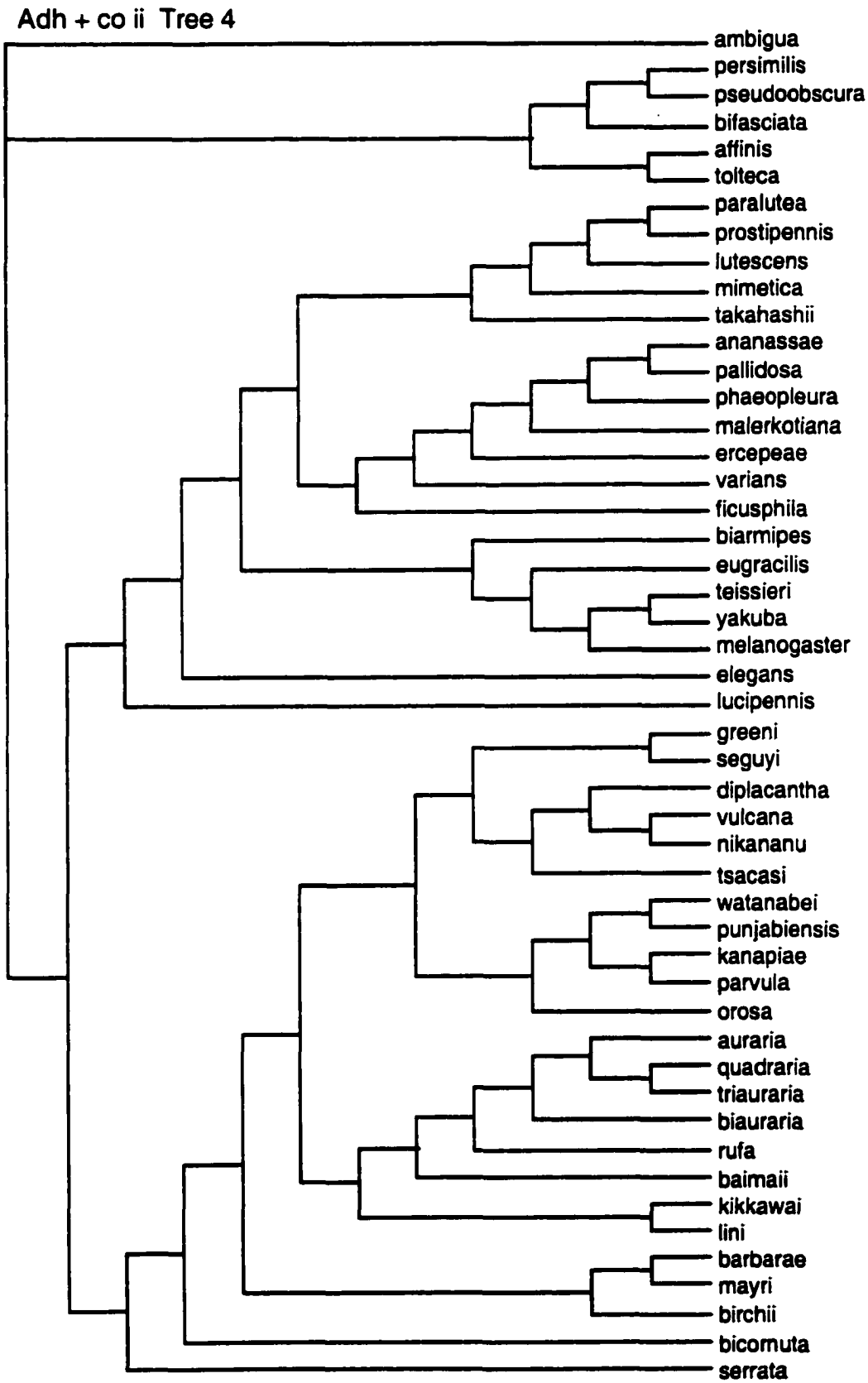


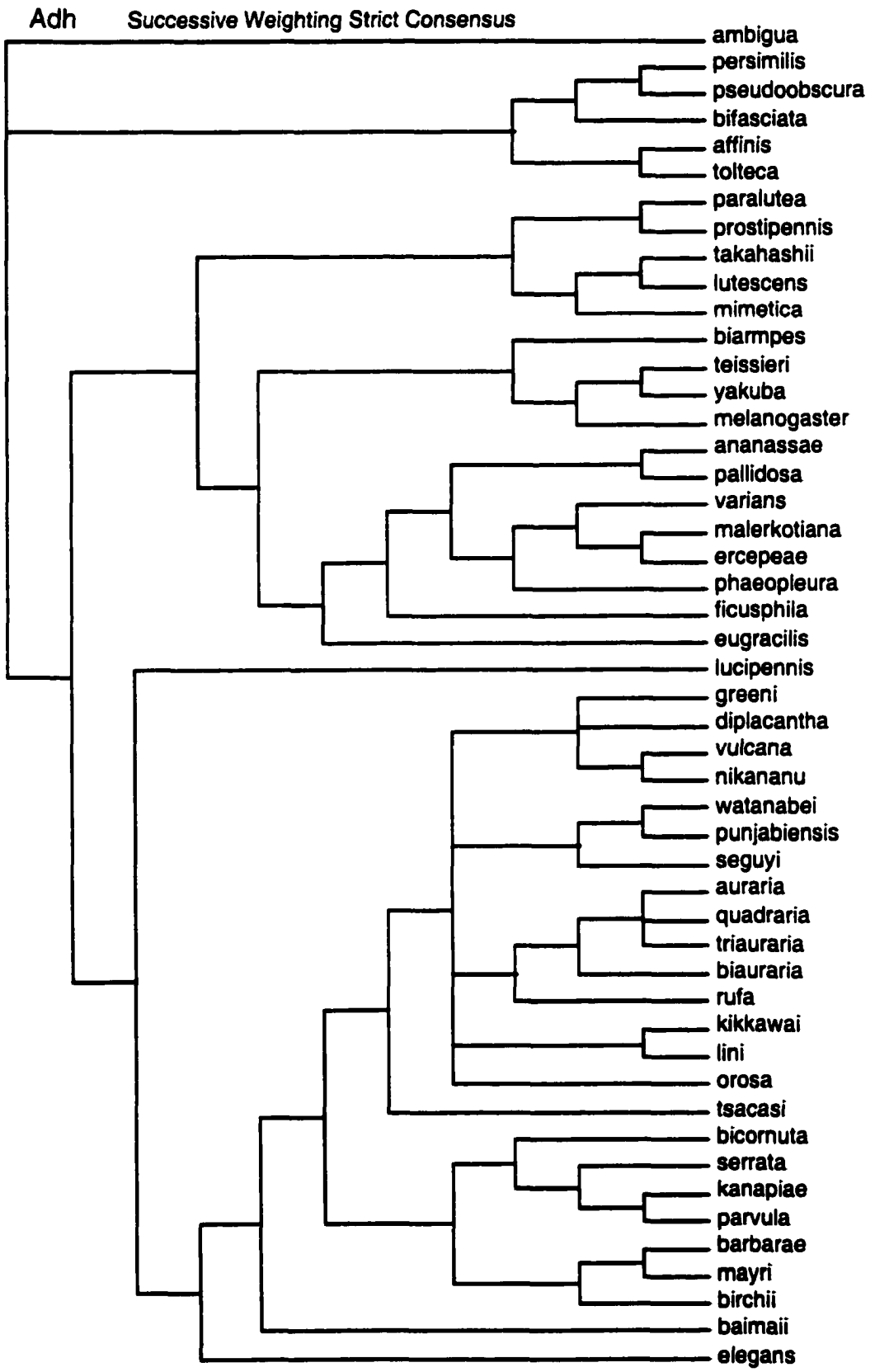


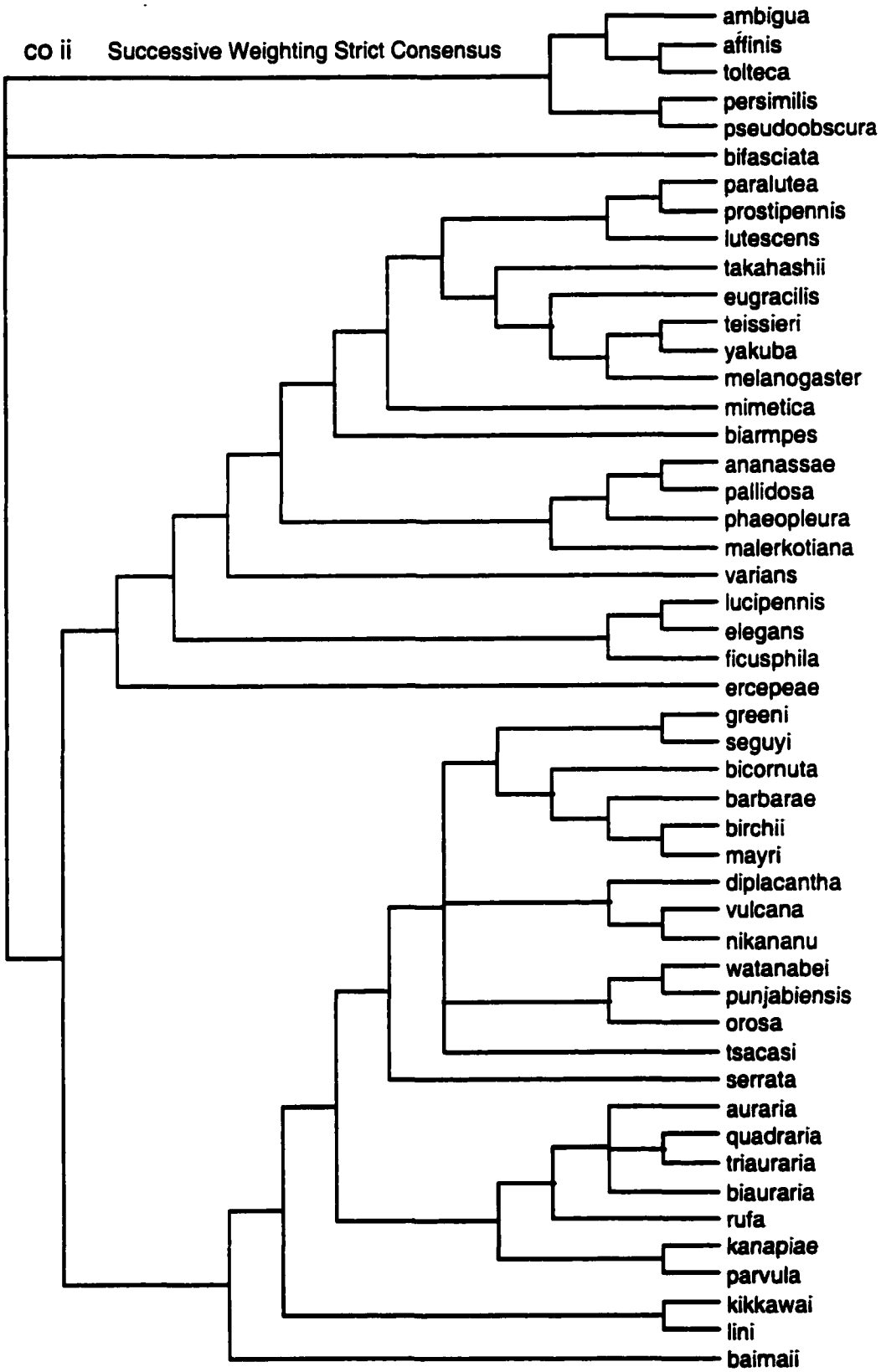




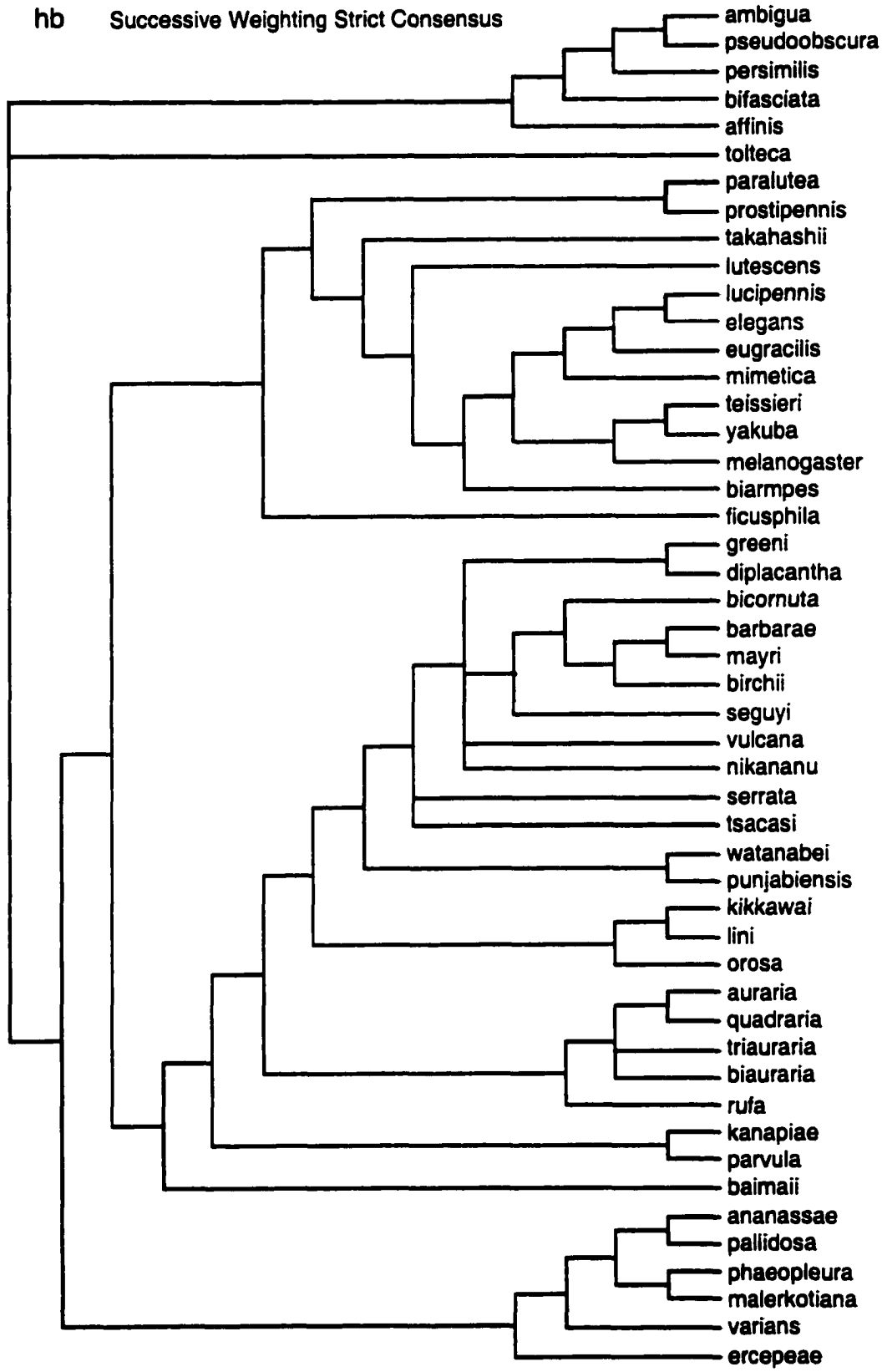


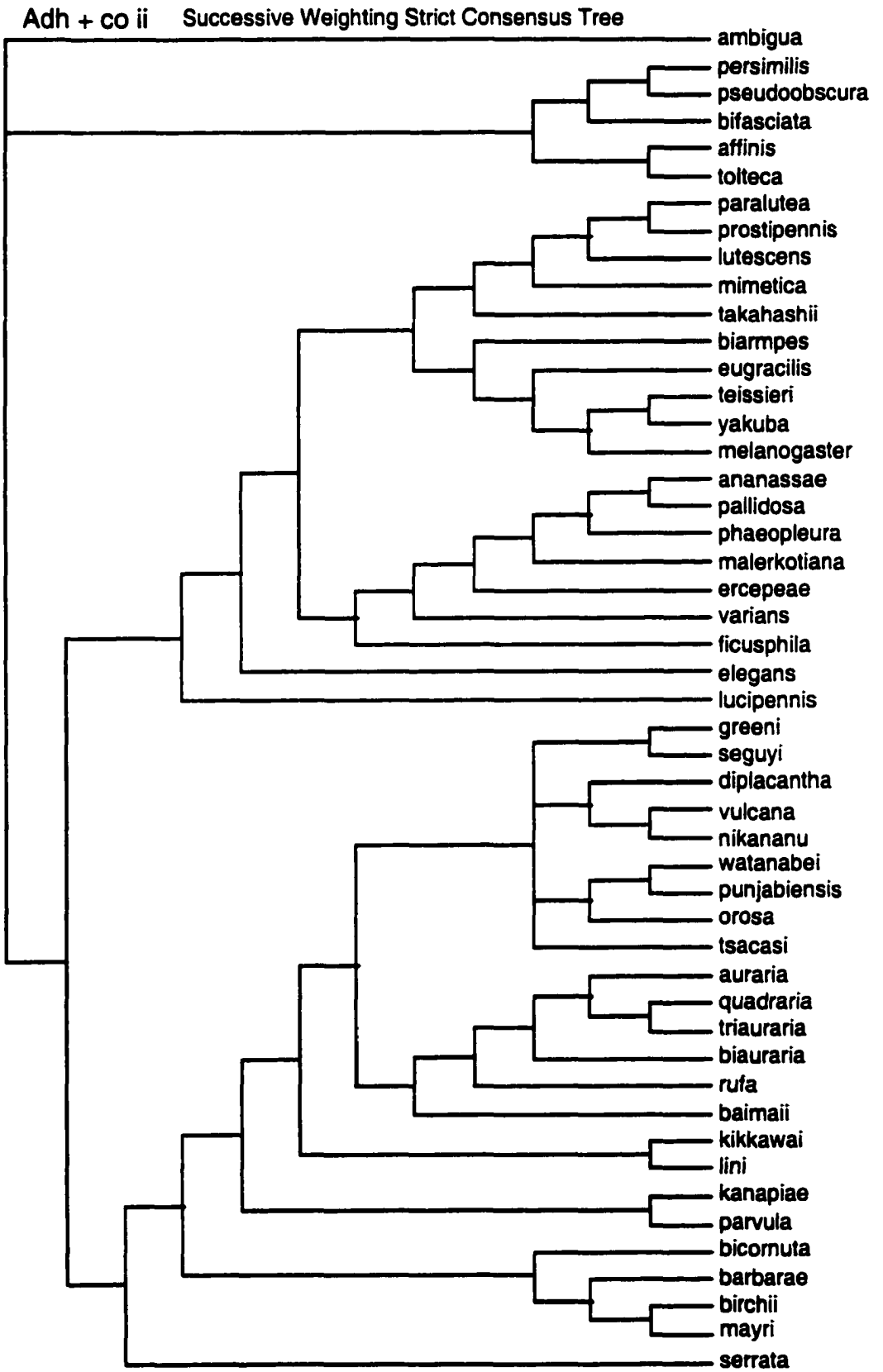


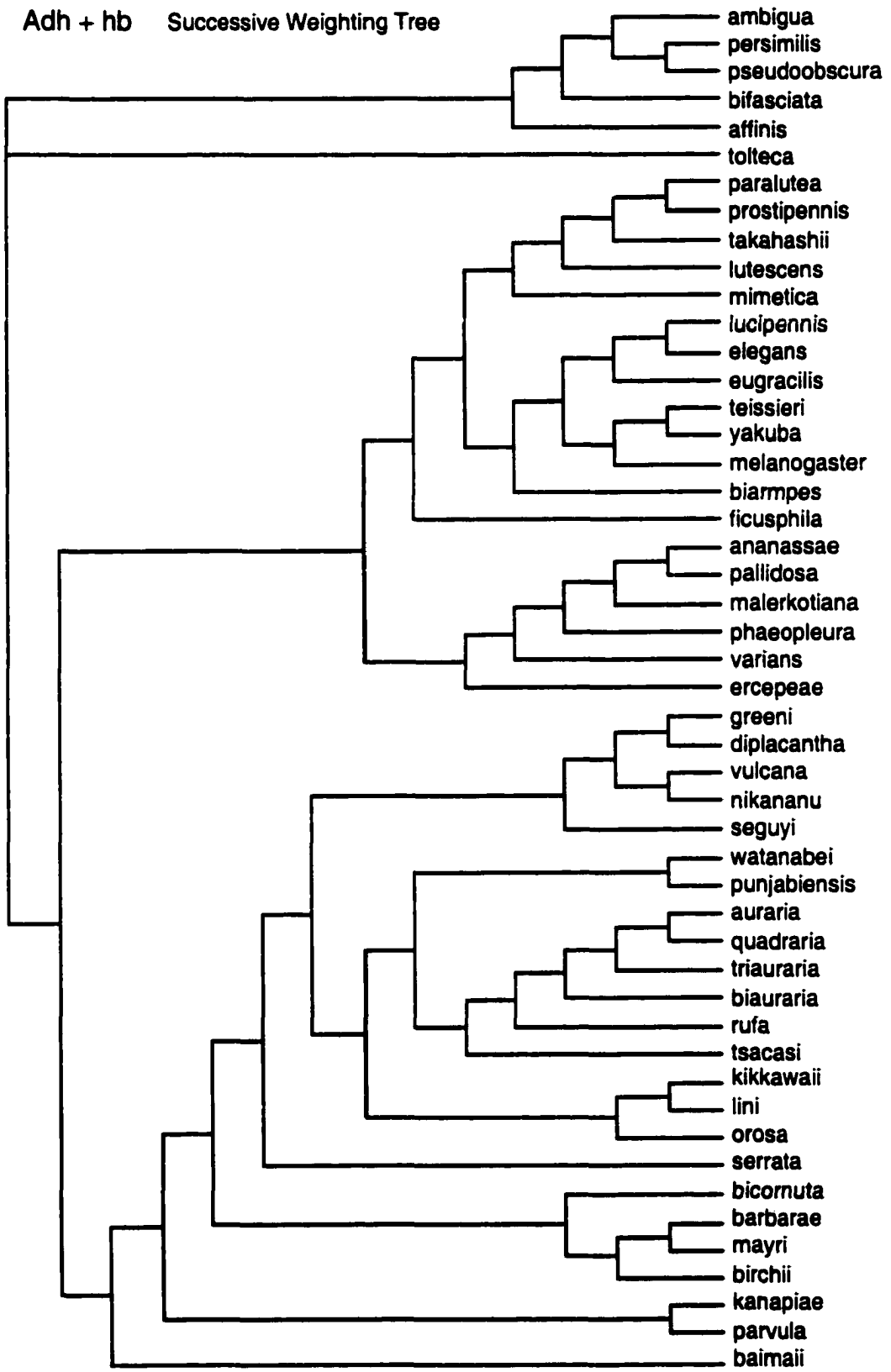


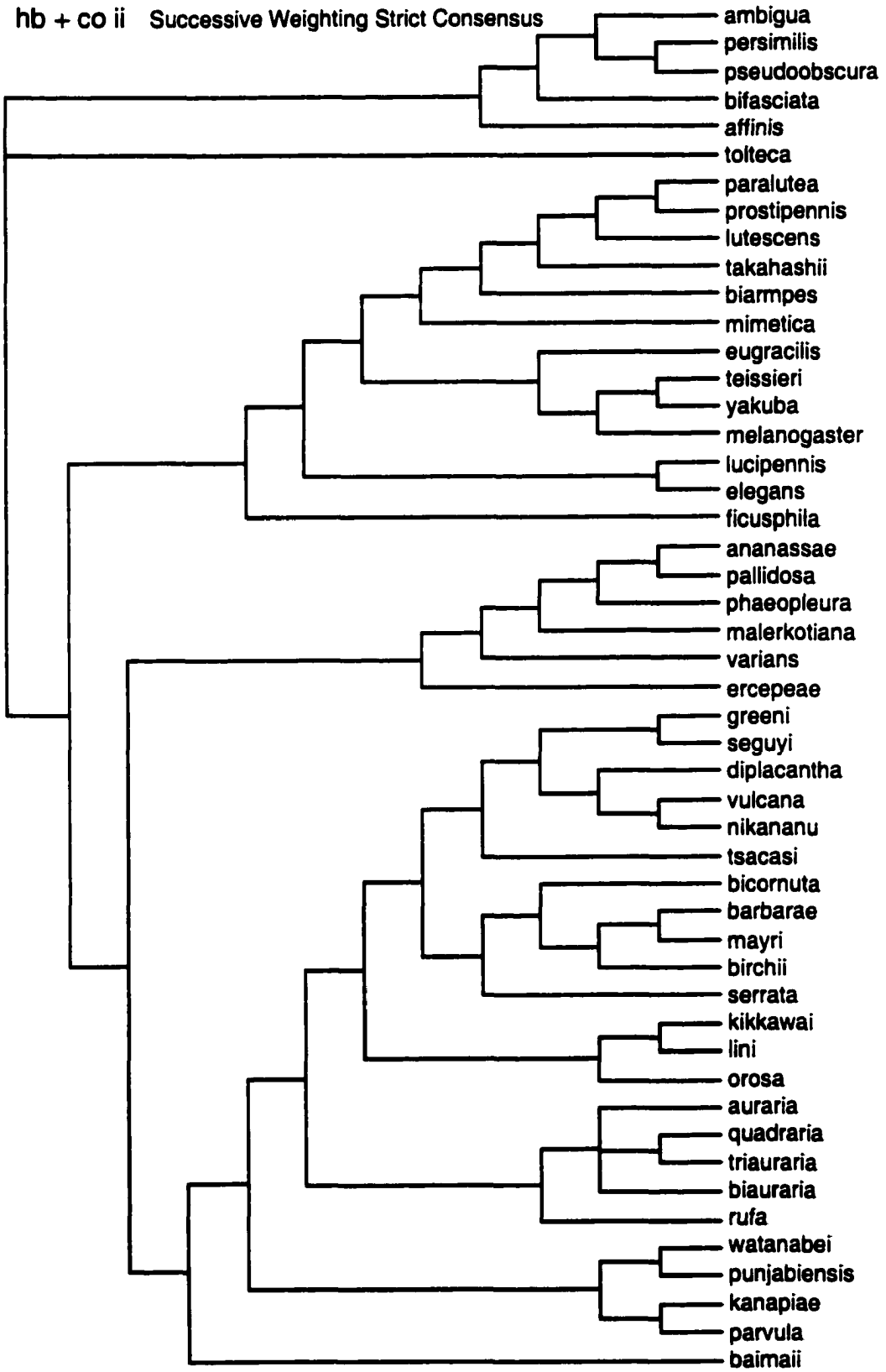


hb Successive Weighting Strict Consensus









LITERATURE CITED

Ashburner, M. 1989. *Drosophila*: A laboratory handbook. New York: Cold Spring Harbor Laboratory Press, xiii + 1331 pp.

Ashburner, M., Bodmer, M. and Lemeunier, F. 1984. On the evolutionary relationships of *Drosophila melanogaster*. Developmental Genetics. 4:295-312.

Avise, J. C. 1994. Molecular markers, natural history and evolution. Chapman and Hall, Inc. New York.

Ayala, F. 1965a. *Drosophila dominicana*, a new sibling species of the *serrata* group. Pacific Insects. 7: 620-622.

Ayala, F. 1965b. Sibling species and the *Drosophila serrata* group. Evolution. 19: 538-545.

Baimai, V. 1970a. Additional evidence on sexual isolation within *Drosophila birchii*. Evolution. 24: 149-155

Baimai, V. 1970b. Chromosomal polymorphism in *Drosophila birchii*. J. Hered. 61: 22-34.

Baimai, V. 1980. Metaphase karyotypes of certain species of the *Drosophila montium* subgroup. Jpn. J. Genet. 55: 165-175.

Baker, R. H., X. Yu and R. DeSalle. 1998. Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees. Mol. Phylog. Evol. 9(3): 427-436.

Baker, R.H., and DeSalle, R. 1997. Multiple scores of molecular characters and the phyolgeny of Hawaiian drosophilids. Syst. Biol. 46: 654-673.

Barrett, M., M.J. Donoghue and E. Sober. 1991. Against consensus. Syst. Zool. 40(4): 486-493.

Barrio, E., A. Latorre, A. Moya. 1994. Phylogeny of the *Drosophila obscura* species group deduced from mitochondrial DNA sequences. J. Mol. Evol. 39: 478-488.

Beckenbach, A.T., Y. W. Wei and H. Liu. 1993. Relationships in the *Drosophila obscura* species group, inferred from mitochondriai cytochrome oxidase ii sequences. Mol. Biol. Evol. 10: 619-634.

Bock, I. R. 1976. Drosphilidae of Australia. I. *Drosophila* (Insecta: Diptera). Aust. J. Zoolo. Suppl. Ser. 40: 1-105.

Bock, I.R. 1980. Current status of the *Drosophila melanogaster* species-group (Diptera). Syst. Entomol. 5:341-356.

Bock, I.R. and Wheeler, M.R. 1972. The *Drosophila melanogaster* species group. Univ. Texas Publ. 7213:1-102.

Borror, D. J., C.A. Triplehorn and N.F. Johnson. 1989. An intorduction to the study of insects, Sixth edition. Fort Worth, Texas: Saunders College Publishing, xiv + 875 pp.

Bremer, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evolution. 42: 795-803.

Bremer, K. 1994. Branch support and tree stability. Cladistics. 10: 295-304.

Brower, A.V.Z. and V. Schawaroch. 1996. Three steps of homology assessment. Cladistics. 12: 265-272.

Bull, J.J., J. Huelsenbeck, C. Cunningham, D. Swofford, and P. Waddell. 1993. Partitioning and combining data in phylogenetic analysis. Syst. Biol. 42: 384-397.

Caccone, A., E.N. Moriyama, J.M. Gleason, L. Nigro and J.R. Powell. 1996. A molecular phylogeny for the *Drosophila melanogaster* subgroup and the problem of polymorphism [*sic*] data. Mol. Biol. Evol. 13(9): 1224-1232.

Caccone, A., G.D. Amato, and J.R. Powell. 1988. Rates and patterns of scnDNA and mtDNA divergence within the *Drosophila melanogaster* subgroup. Genetics. 118: 671-683.

Caccone, A., Gleason, J.M., and Powell, J.R. 1992. Complementary DNA-DNA hybridization in *Drosophila*. J. Mol. Evol. 34: 130-140.

Cariou, M. L., D. Lachaise, L. Tsacas, J. Sourdis, C. Krimbas and M. Ashburner. 1988. New African species in the *Drosophila obscura* species group: genetic variation, differentiation and evolution. Heredity. 61: 73-84.

Carpenter, J.M. 1988. Choosing among multiple equally parsimonious cladograms. Cladistics. 4(3): 291-296.

Chassagnard, M.-T. 1991. Une nouvelle espèce afrotropicale du sous-groupe *Drosophila montium* groupe *melanogaster* (Diptera, Drosophilidae). Rev. fr. Ent. (N.S.) 13: 119-122.

Chassagnard, M.-T., L. Tsacas and D. Lachaise. 1997. Drosophilidae (Diptera) of Malawi. Ann. Natal Mus. 38: 61-131.

Clary, D. O. and D. R. Wolstenholme. 1985. The mitochondrial DNA molecular of *Drosophila yakuba*: Nucleotide sequence, gene organization, and genetic code. J. Mol. Evol. 22: 252-271.

Clayton, F.E. and M.R. Wheeler.1975. A catalog of *Drosophila* metaphase chromosome configurations. In: "Handbook of genetics" R.C. King, ed. Vol. 3 pp. 471-512. Plenum: New York.

Coyne, J.A. and M. Kreitman. 1986. Evolutionary genetics of two sibling species *Drosophila simulans* and *D. sechellia*. Evolution. 40: 673-691.

Danforth, B.N., H. Sauquert and L. Packer. 1999. Phylogeny of the bee genus *Halictus* (Hymenoptera: Halictidae) based on parsimony and likelihood analyses of nuclear EF-1α sequence data. Mol. Phylog. Evol. 13(3): 605-618.

Daniels, S. B., A. Chovnick, and I. Boussy. 1990. Distribution of *hobo* transposable elements in the genus *Drosophila*. Mol. Biol. Evol. 7(6): 589-606.

Davis, J.I. and K. C. Nixon, 1992. Populations, genetic variation, and the delimitation of phylogenetic species. Syst. Biol. 41(4): 421-435.

Dayhoff, M.O. 1978. Atlas of protein sequence and structure, vol. 5, suppl. 3. Natl. Biomed. Res. Found.: Silver Spring, Maryland.

De Pinna, M.C.C. 1991. Concepts and tests of homology in the cladistic paradigm. Cladistics. 7: 367- 391.

De, A. and J. P. Gupta. 1996. Records of drosophilid species from west Bengal with description of one new and two previously unrecorded species from India (Insecta: Diptera: Drosophilidae). Senckenbergiana biologica 76: 129-133.

DeSalle, R. 1992. The origin and possible time of divergence of the Hawaiian Drosophilidae: Evidence from DNA sequences. Mol. Biol. Evol. 9: 905-916.

DeSalle, R. and D. A. Grimaldi. 1991. Morphological and molecular systematics of the Drosophilidae. Annu. Rev. Ecol. Syst. 22:447-475.

DeSalle, R., A. K. Williams and M. George. 1993. Isolation and characterization of animal mitochondrial DNA. *In:* E. A. Zimmer, T. J. White, R. L. Cann and A. C. Wilson (ed.s), Methods in enzymology, Vol. 224: 176-204. London: Academic Press.

DeSalle, R., T. Freedman, E.M. Prager and A.C. Wilson. 1987. Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila*. J. Mol. Evol. 26: 157-164.

DeSalle, R., T. Freedman, E.M. Prager and A.C. Wilson. 1987. Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila*. J. Mol. Evol. 26: 157-164.

DeSalle, R.and E. Carew. 1992. Phyletic phenocopy and the role of developmental genes in morphological evolution in the Drosophilidae. J. Evol. Biol. 5(3): 363-374.

Dobzhansky, T. and C. Epling. 1944. Taxonomy, geographic distribution, and ecology of *Drosophila pseudoobscura* and its relatives. Carnegie Inst. Wash. Publ. 554: 1-46.

Dobzhansky, T. and J. R. Powell. 1975. *Drosophila pseudoobscura* and its American relatives *Drosophila persimilis* and *Drosophila miranda*. *In:* R.C. King (ed.), Handbook of genetics, Vol. 3: 537-587. New York: Plenum Press.

Dobzhansky, T. and W.B. Mather. 1961. The evolutionary status of *Drosophila serrata*. Evolution. 15: 461-467.

Drosopoulou, E. and Z.G. Scouras. 1995. The ß-tubulin gene family evolution in the *Drosophila montium* subgroup of the *melanogaster* species group. J. Mol. Evol. 41: 293-298.

Drosopoulou, E. and Z.G. Scouras. 1995. The ß-tubulin gene family evolution in the *Drosophila montium* subgroup of the *melanogaster* species group. J. Mol. Evol. 41: 293-298.

Drosopoulou, E., I. Konstantopoulou and Z. G. Scouras. 1996. The heat shock genes in the *Drosophila montium* subgroup: chromosomal localization and evolutionary implications. Chromosoma. 105: 104-110.

Duda, O. 1924. Beitrag zur Systematik der Drosophiliden under besonderer Berücksichtigung der Palärktischen und Orientalischen Arten (Dipteren). Arch. Naturforsch., Abt. A 90: 172-259.

Eriksson, T. 1997 AutoDecay version 2.9.8 (Hypercard stacks distributed by the author). Botaniska institutionen. Stockholm University, Stockholm.

Farris, J. S. 1979. The information content of the phylogenetic system. Syst. Zool. 28: 483-519.

Farris, J. [S.] 1969. A successive approximations approach to character weighting. Syst. Zool. 18: 374-385.

Farris, J.S. 1989. The retention index and the rescaled consistency index. Cladistics. 5: 417-419.

Farris, J.S., Kallersjo, M., Kluge, A.G., and Bult, C. 1994. Testing significance of congruence. Cladistics. 10:315-320.

Farris, J.S., Kallersjo, M., Kluge, A.G., and Bult, C. 1995. Constructing a significance test for incongruence. Syst. Biol. 44:570-572.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. J. Mol. Evol. 17: 368-376.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution. 39: 783-791.

Gatsey, J., R. DeSalle and W.C. Wheeler. 1994 Alignment-ambiguous nucleotide sites and the exclusion of data. Mol. Phylog. Evol. 2: 152-157.

Geer, B.W., M. L. Langevin and S.W. McKechnie. 1985. Dietary ethanol and lipid synthesis in *Drosophila melanogaster*. Biochem. Genet. 23(7/8): 607-622.

Gleason, J.M., A. Caccone, E.N. Moriyama, K.P.White and J.R. Powell. 1997. Mitochondrial DNA phylogenies for the *Drosophila obscura* group. Evolution. 51: 433-439.

Goloboff, P. A. 1991. Homoplasy and the choice among cladograms. Cladistics. 7: 215-232.

Goloboff. P.A. 1998. NONA. phylogentic analysis program distribution from author and J. Carpenter Amer. Mus. Natl. Hist.

Graybeal, A. 1993. The phylogenetic utility of cytochrom *b*: lessons from bufonid frogs. Mol. Phylog. Evol. 2: 256-269.

Graybeal, A. 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? Syst. Biol. 47: 9-17.

Grimaldi, D. A. 1991. Cladistics and the classification of the Drosophilidae. *In:* Proceedings of the second international congress of Dipterology. Weismann, L.I. Orzßgh and A. C. Pont (ed.s). SPB Academic Publishing: The Hague, the Netherlands p. 85-97 + one figure.

Grimaldi, D.A. 1990. A phylogenetic, revised classification of genera in the Drosophilidae (Diptera). Bull. Am. Mus. Nat. Hist. 197:1-139.

Gupta, J. P. and A. K. Sundaran. 1990. Further record of two new and one known species of *Drosophila* (Diptera: Insecta) from Karnataka, India. Proc. Zool. Soc. (Calcutta). 43: 31-35.

Gupta, J. P. and K. K. Gupta. 1992. *Drosophila watanabei*, a new species of the *montium* subgroup of the *melanogaster* species group of the subgenus *Sophophora* (Diptera: Drosophilidae). Oriental Insects. 26: 201-205.

Hasegawa, M., H. Kishino and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochodrial DNA. J. Mol. Evol. 21: 160-174.

Heed, W. B. and R. L. Mangan. 1986. Community ecology of the Sonoran desert *Drosophila*. Volume 3e. M.Ashburner, H.L.Carson, J.N. Thompson (ed.s). Academic Press, Inc.: London. p. 311-345.

Hennig, W. 1966. Phylogenetic Systematics. University of Illinios Press: Urbana.. xiii +263pp.

Hey, J. and R. M. Kliman. 1993. Population genetics and phylogenetics of DNA sequence variation at multiple loci within the *Drosophila melanogaster* species complex. Mol. Biol. Evol. 10: 804-822.

Hillis, D. M. 1987. Molecular versus morphological approaches to systematics. Annu. Rev. Ecol. Syst. 18: 23-42.

Hillis, D.M. 1998.Taxonomic sampling, phylogenetic accuracy and investigator bias. Syst. Biol. 47: 3-8.

Hsu, T.C. 1949. The external genital apparatus of male Drosophilidae in relation to systematics. Univ. Texas Publ. 4920, 80-142.

Huelsenbeck, J.P. D.L. Swofford, C.W. Cunningham, J.J. Bull and P. Waddell. 1994. Is character weighting a panacea for the problem of data heterogeneity in phylogenetic analysis? Syst. Biol. 43: 288-291.

Jukes, T.H. and C. R. Cantor. 1969. Evolution of protein molecules. *In*: H.N. Munro (ed.), Mammalian Protein Metabolism: 21-132. New York: Academic Press.

Kalanti-Makri, M. C., L. H. Margaritis, J. Sourdis. 1985. Phylogenetic relationships within the *montium* species subgroup of the genus *Drosophila* (*Sophophora*) based on electrophoretic data of the major chorion proteins. Annls. Soc. Ent. Fr. (N. S.) 21(4): 357-366.

Kim, B. K. and T. Okada. 1988. A new species of the *Drosophila montium* species subgroup (Diptera, Drosophilidae) from Amami-oshima Island. Proc. Jpn. Soc. Syst. Zool. No. 38: 57-61.

Kim, B. K., T. K. Watanabe and O. Kitagawa. 1989. Evolutionary genetics of the *Drosophila montium* subgroup. I. Reproductive isolations and the phylogeny. Jpn. J. Genet. 64: 177-190.

Kim, B.K., T. Aotsuka and O. Kitagawa. 1993. Evolutionary genetics of the *Drosophila montium* subgroup. II. Mitochondrial DNA variation. Zool. Sci. 10: 991-996.

Kim, N.W. and T.J. Lee. 1991. Genetic relationships among the eight species of the *Drosophila melanogaster* species group by allozyme analysis. Korean J. Genet. 13(4): 297-309.

Kim, N.W., T.J. Lee, E.S. Song. 1992 Evolutionary genetic study on the eight species of the *Drosophila melanogaster* group from Korea: Reproductive isolation and protein analysis. Korean J. Zool. 35(2): 21-218.

Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16:11-120.

Kimura, M. T. 1987. Habitat differentiation and speciation in the *Drosophila auraria* species-complex (Diptera, Drosophilidae). Kontyú. 55(3): 429-436.

Kimura, M.T., T. Ohtsu, T. Yoshida, T. Awasaki and F.-J. Lin. 1994 Climatic adaptations and distributions in the *Drosophila takahashii* species subgroup (Diptera: Drosophilidae). J. Nat. Hist. 28: 401-409.

Kluge, A. G. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). Syst. Zool. 38(1):7-25.

Kluge, A.G. and A. J. Wolf. 1993. Cladisitcs: What is in a word? Cladistics. 9: 183-199.

Kohler, R. E.. 1994. Lords of the Fly: *Drosophila* genetics and the experimental life. Chicago: The University of Chicago Press.

Konstantopoulou, I., E. Drosopoulou and Z.G. Scouras. 1997. Variations in the heat-inducted protein pattern of several *Drosophila montium* subgroup species (Diptera: Drosophilidae). Genome. 40: 132-137.

Lachaise, D., M.-L. Cariou, J. R. David, F. Lemeunier, L. Tsacas and M. Ashburner. 1988. Historical Biogeography of the *Drosophila melanogaster* species subgroup. *In*: M. K. Hecht, B. Wallace, and G. T. Prance (ed.s), Evolutionary biology, Vol. 22: 159-225. New York: Plenum Press.

Lakovaara, S. and A. Saura. 1982. Evolution and speciation in the *Drosophila obscura* group. *In:* M. Ashburner, H. L. Carson and J. N. Thompson, Jr. (ed.s), The genetics and biology of *Drosophila*, Vol. 3b: 1-59. London: Academic Press.

Lecointre, G. H. Philippe, H.L. Van Le, and H. Le Guyader. 1993. Species sampling has a major impact on phylogentic inference. Mol. Phylog. Evol. 2: 205-224.

Lee, T. J., K. J. Hong and N.W. Kim. 1994. Genetic relationships and protein variations during development within the *Drosophila melanogaster* species group. II. Analysis of soluble protein by 2DE. Korean J. Zool. 37: 249-254.

Lee, T.J., K.J. Hong and E.S. Song. 1993. Genetic relationships and protein variations during development within the *Drosophila melanogaster* species group.1.Analysis of soluble protein by SDS-PAGE. Korean J. Genet. 15: 269-276.

Lemeumier, F., J.R. David, L. Tsacas, and M. Ashburner. 1986. The *melanogaster* species group. *In*: M.Ashburner, H.L.Carson, J.N. Thompson (ed.s), The genetics and biology of the *Drosophila*. Volume 3e: 147-256. London: Academic Press.

Lemeunier, F., S. Aulard, M. Arienti, J.-M. Jallon, M.-L. Cariou, and L. Tsacas. 1997. The *ercepeae* complex: new cases of insular speciation within the *Drosophila ananassae* species subgroup (*melanogaster* group) and descriptions of two new species (Diptera: Drosophilidae). Ann. Entomol. Soc. Amer. 90: 28-42.

Liu, H. and A.T. Beckenbach. 1992. Evolution of the mitochondrial cytochrome oxidase II gene among ten orders of insects. Mol. Phylog. Evol. 1: 41-52.

Lockhart, P.J., M. A. Steel, M.D. Hendy and D. Penny. 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Mol. Biol. Evol. 11: 605-612.

Lutzoni, F. M. 1997. Phylogeny of lichen- and non-lichen-forming omphalinoid mushrooms and the utility of testing for compatability among multiple data sets. Syst. Biol. 46: 373-406.

Mavragani-Tsipidou, P., A. Zambetaki, K. Kleanthous, E. Pangou and Z. G. Scouras. 1994. Cytotaxonomic differentiation of the Afrotropical *Drosophila montium* subgroup: *D. diplacantha* and *D. seguyi*. The major role of reverse tandem duplications. Genome. 37: 935-944.

McEvey, S. F., J. R. David and L. Tsacas. 1987. The *Drosophila ananassae* complex with description of a new species from French Polynesia (Diptera: Drosophilidae). Ann. Soc. Entomol. (N.S.) 23(4): 377-385.

McEvey, S.F., S. Aulard and O. Ralisoa-Randrianasolo. 1989. An Australian drosophilid (Diptera) on *Eucalyptus* and *Eugena* (Myrtacea) flowers in Madagascar. J. Aust. Ent. Soc. 28(1): 53-54.

Meier, R., P. Kores, and S. Darwin. 1991. Homoplasy slope ratio: A better measurement of observed homoplasy in cladistic analyses. Syst. Zool. 40(1): 74-88.

Mickevich, M. and Farris, S. 1981. The implications of congruence in Menidia. Syst. Zool. 30:351-370.

Miller, J.S., A.V.Z. Brower, and R. DeSalle. 1997. Phylogeny of the neotropical moth tribe Josiini (Notodontidae: Dioptinae): Evidence from DNA sequences and morphology. Biol. J. Linn. Soc. 60: 297-316.

Muniyappa, N., G. S. Reddy and N. B. Krishnamurthy, 1981. Two new species of *Drosophila* from India (Diptera, Drosophilidae). Oriental Insects. 15:215-220.

Murphy, R. W., J. W. Sites, Jr., B. G. Buth and C. H. Haufler. 1996. Proteins: Isozyme electrophoresis. *In*: D. M. Hillis, C. Moritz and B. K. Mable (ed.s), Molecular Systematics. 2nd Edition: 51-120. Sunderland, MA.:Sinauer Associates.

Nelson, G. J. and N. I. Platnick. 1981. Systematics and biogeography: Cladistics and vicariance. New York: Columbia University Press.

Nigro, L., M. Solignac and P.M. Sharp. 1991. Mitochondrial DNA sequence divergence in the *melanogaster* and Oriental species subgroups of *Drosophila*. J. Mol. Evol. 33: 156-162.

Nikolaidis, N. and Z. G. Scouras. 1996. The *Drosophila montium* subgroup species. Phylogenetic relationships based on mitochondrial DNA analysis. Genome. 39: 874-883.

Nixon, K. C. and J. I. Davis, 1991. Polymorphic taxa, missing values and cladistic analysis. Cladistics. 7: 233-241.

Ohnishi, S., and T. K. Watanabe. 1984. Systematics of the *Drosophila montium* species subgroup: A biochemical approach. Zool. Sci. 1: 801-807.

Ohnishi, S., K.-W. Kim and T. K. Watanabe. 1983. Biochemical phylogeny of the *Drosophila montium* species subgroup. Jpn. J. Genet. 58: 141-151.

Ohnishi, S., M. Kawanishi and T. K. Watanabe. 1983. Biochemical phylogenies of *Drosophila*: protein differences detected by two-dimensional electrophoresis. Genetica. 61: 55-63.

Ohno, S. 1970. Evolution by gene duplication, Springer-Verlag, N.Y.

Okada, T. 1954. Comparative morphology of the drosophilid flies. I. Phallic organs of the *melanogaster* group. Kontyú. 22:36-46.

Okada, T. 1984. New or little known species of *Drosophila* (*Lordiphosa*) with taximetrical analyses (Diptera, Drosophilidae). Kontyú. 52:565-575.

Okada, T. 1988. Family Drosophilidae (Diptera) from the Lund University Ceylon Expedition in 1962 and Borneo Collections in 1978-1979. Entomologica Scandinavica Supplement. No. 30: 111-151.

Okada, T. 1989. A proposal for establishing tribes for the family Drosophilidae with keys to tribes and genera (Diptera). Zool. Sci. 6: 391-399.

Okada, T. and H. L. Carson. 1982. Drosophilidae associated with flowers in Papua New Guinea. 4. Araceae, Compositae, Convolvulaceae, Rubiaceae, Leguminosae, Malvaceae. Kontyú. 50(4): 511-526.

Okada. T. 1964. Drosophilidae (Diptera) of Southeast Asia collected by the Thai-Japanese Biological Expedition 1961-62. *In:* T. Kira, and T. Umesao (ed.s), Nature and life in southeast Asia. Vol. 3: 439-466. Kyoto, Japan: Fauna and Flora Research Society.

Olmstead, R. G. and J.A. Sweere. 1994. Combining data in phylogenetic systematics: An empirical approach using three molecular data sets in the Solenaceae. Syst. Biol. 43: 467-481.

Pardali, E. E. Feggou, E. Drosopoulou, I. Konstantopoulou, Z. G. Scouras and P. Mavragani-Tsipidou. 1996. The Afrotropical *Drosophila montium* subgroup: Balbiani ring 1, polytene chromosomes, and heat shock response of *Drosophila vulcana*. Genome. 39:588-597.

Parkash, R., Jyoutsna and Vandna. 1994. Allozyme phylogeny of five species of *takahashii* species subgroup of *Drosophila*. Korean J. Genetics. 16(3): 187-196.

Patterson, C. 1982. Morphological characters and homology. *In*: K. A. Joysey and A. E. Friday (eds), Problems of phylogenetic reconstruction. 21-74. London: Academic Press.

Patterson, J.T. and W. S. Stone. 1952. Evolution in the genus *Drosophila*. New York: The MacMillian Co., 610 pp.

Pélandakis, M. and M. Solignac. 1993. Molecular phylogeny of *Drosophila* based on ribosomal RNA sequences. J. Mol. Evol. 37: 525-543.

Pélandakis, M., D.G. Higgins and M. Solignac. 1991. Molecular phylogeny of the subgenus *Sophophora* of *Drosophila* derived from the large subunit of ribosomal RNA sequences. Genetica. 84: 87-94.

Pissios, P. and Z. G. Scouras. 1993. Mitochondrial DNA evolution in the *montium*-species subgroup of *Drosophila*. Mol. Biol. Evol. 10(2): 375-382.

Platnick, N.I., C.E. Griswold and J. A. Coddington. 1991. On missing entries in cladistic analysis. Cladistics. 7: 337-343.

Poe, S. 1998. Sensitivity of phylogeny estimation to taxonomic sampling. Syst. Biol. 47: 18-31.

Powell, J.R. and R. DeSalle. 1995. *Drosophila* molecular phylogenies and their uses. Evol. Biol. 28: 87-138.

Rafael, V. 1984. Relations interspecifiques dans le nouveau complexe Africain de *Drosophila bakoue* du groupe *melanogaster*, sous-groupe *montium* (Diptera, Drosophilidae). Bull. Soc. Zool. Fr. 109: 179-189.

Remsen, J., and R. DeSalle. 1998. Character congruence of multiple data partitions and the origin of the Hawaiian Drosophilidae. Mol. Phylog. Evol. 9: 225-235.

Saitou, N. and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4: 406-425.

Sanderson, M.J. 1995. Objections to bootstraping phylogenies: A critique. Syst. Biol. 44(3) 299-320.

Sanderson, M.J. and M.J. Donoghue, 1989. Patterns of variation in levels of homoplasy. Evolution. 43(8): 1781-1795.

Schawaroch, V. 2000. This dissertation and related papers in prep.

Scouras, Z. G. 1995. The *Drosophila montium* subgroup species: Recent cytogenetic, molecular, development and evolutionary studies. Bios (Macedonia, Greece) 3: 125-158.

Shaffer, H. B., P. Meylan, and M. L. McKnight. 1997. Tests of turtle phylogeny: molecular, morphological, and paleontological approaches. Syst. Biol. 46: 235-268.

Shyamala, B. V., and H. A. Ranganath. 1990. Biochemical phylogeny of seven Indian species of the *montium* subgroup of *Drosophila*. Genetica. 81: 71-75.

Shyamala, B.V., and H. A. Ranganath. 1989. Metaphase karyotype differentiation in eight species of the *montium* subgroup of *Drosophila*. Genetica. 79: 191-196.

Siebert, D.J. 1992. Tree statistics; trees and 'confidence'; consensus trees; alternatives to parsimony; character weighting; character conflict and its resolution. *In*: P.L. Forey, C.J. Humphries, I.J. Kitching, R. W. Scotland, D.J. Siebert, and D. M. Williams (ed.s), Cladistics: A practical course in systematics: 72-88. Oxford, England: The Systematc Association and Oxford University Press.

Simmons, G. M. 1992. Horizontal transfer of the *hobo* transposable elements within the *Drosophila melanogaster* species complex: Evidence from DNA sequencing. Mol. Biol. Evol. 9(6): 1050-1060.

Singh, B. K. and J. P. Gupta. 1977[1979]. Two new and two unrecorded species of the genus *Drosophila* Fallen (Diptera: Drosophilidae) from Shillong, Meghalaya, India. Proc. Zool. Soc. (Calcutta). 30: 31-38. Yes 2 years. On back cover it says "Issued December 1979".

Singh, B. K. and S. Dash. 1998. Drosophilidae of Kumaun Region, India with the description of four new species (Insecta: Diptera). Proc. Zool. Soc. (Calcutta). 51: 45-56.

Sneath, P. H. A. and R. R. Sokal. 1973. Numerical Taxonomy. San Francisco: W. H. Freeman.

Sturtevant, A. H. 1942. The classification of the genus *Drosophila* with the description of nine new species. Univ. Texas Publ. 4213: 5-51.

Sturtevant, A.H. 1965. A history of genetics. New York: Harper and Row, 165 pp.

Swofford, D. L., G. J. Olsen, P. J. Waddell and D. M. Hillis. 1996. Phylogenetic inference. *In*: D. M. Hillis, C. Moritz and B. K. Mable (ed.s), Molecular Systematics. 2nd Edition: 407-514. Sunderland, MA.: Sinauer Associates.

Swofford, D.L. 2000 PAUP*: Phylogentic analysis using parsimony (and other methods), version 4.0 Sinauer.

Takamori, H. and T. Okada. 1983. *Drosophila tsukubaensis*, a new species of the *obscura* group of the genus *Drosophila* (Diptera, Drosophilidae) form Japan. Kontyú. 51: 265-268.

Tautz, D. Lehmann, R., Schnurch, H., Schuh, R., Seifert, E., Kienlin, A., Jones, K. and Jackle, H. 1987. Finger protein of novel structure encoded by hunchback, a second member of the gap class of *Drosophila* segmentation genes. Nature. 327 (4 June 1987): 383-389.

Thomas, R.H. and J.A. Hunt. 1993. Phylogenetic relationships in *Drosophila*: A conflict between molecular and morphological data. Mol. Biol. Evol. 10: 362-374.

Throckmorton, L.H. 1962. X. The problem of phylogeny in the genus *Drosophila*. Univ. Texas Publ. 6205:207-373.

Throckmorton, L.H. 1965. Similarity versus relationship in *Drosophila*. Syst. Zool. 14: 221-236.

Throckmorton, L.H. 1975. The phylogeny, ecology, and geography of *Drosophila*. *In*: R.C. King (ed.), Handbook of genetics, vol. 3: 421-469. New York: Plenum.

Toda, M.J. 1991. Drosophilidae (Diptera) in Myanmar (Burma) VII. The *Drosophila melanogaster* species-group, excepting the *D. montium* species-subgroup. Oriental Insects. 25:69-94.

Triantaphyllidis, C.D. J. N. Panourgias and Z. G. Scouras. 1978. Isozyme variation and phylogenetic relationships among six species of the *montium* subgroup of the *Drosophila melanogaster* species group. Genetica. 48(3): 223-227.

Tsacas, L. 1975. *Drosophila davidi* n. sp. du groupe *melanogaster* du congo (Diptera , Drosophilidae). Annls. Univ. Brazzaville 11(C): 127-130.

Tsacas, L. 1979. Contribution des données africaines à la compréhension de la biogéographie et de l'évolution du sous-genre *Drosophila* (*Sophophora*) Sturtevant (Diptera, Drosophilidae). C.R. Soc. Biogéogr. 48:29-51.

Tsacas, L. 1980. Les groupes d'espèces du sous-genre *Sophophora* Sturtevant (Diptera, Drosophilidae, *Drosophila*) et le rôle du functionnement des génitalia mâles dans la définition des taxons supraspécifiques. Bull. Soc. Zool. Fr. 105: 529-543.

Tsacas, L. 1981. Quatre nouvelles espèces africaines du groupe *Drosophila melanogaster* (Diptera, Drosophilidae). Rev. Fr. Entomol. (N. S.) 3: 97-104.

Tsacas, L. 1984. Nouvelles données sur la biogéographie et l'évolution du groupe *Drosophila melanogaster* en Afrique. Description de six nouvelles espèces. (Diptera, Drosophildae). Annls. Soc. Ent. Fr. (N.S.) 20(4): 419-438.

Tsacas, L. and D. Lachaise. 1974. Quatre nouvelles especes de la cote-d'Ivoire du genre *Drosophila*, groupe *melanogaster*, et discussion de l'origine du sous-groupe *melanogaster* (Diptera: Drosophilidae). Ann. Univ. Abidjan, serie E Ecologie 7(1): 193-211.

Tsacas, L. and J. David. 1978. Systematics and biogeography of the *Drosophila kikkawai*-complex, with descriptions of new species (Diptera, Drosophilidae). Annls. Soc. Ent. Fr. (N.S.). 13(1977): 675-693. [Note: Date of publication February 28, 1978.]

Tsacas, L. and M.-T. Chassagnard. 1988. Dipteres Drosophilidae de Nouvelle-Caledonie. 1. *Drosophila*: sous genres *Drosophila* et *Sophophora*. Memoires du Museum National d'Histoire Naturelle Serie A Zoologie. 142: 143-154.

Tsacas, L. and M.-T. Chassagnard. 1991[1992]. Le complexe *Drosophila nikananu*: description d'une nouvelle espèce africaine et analyse de quelques caracteres morphologiques du groupe *melanogaster* (Diptera, Drosophilidae). Nouv. Revue Ent. (N.S.). 8: 385-398.

Tsakas, S. C., and L. Tsacas. 1984. A phenetic tree of eighteen species of the *melanogaster* group of *Drosophila* using allozyme data as compared with classifications based on other criteria. Genetica. 64: 139-144.

Vilela, C. R. and G. Bächli. 1991. Case 2767. *Drosophila hydei* Sturtevant, 1921 (Insecta, Diptera): proposed conservation of the specific name. Bull. Zool.Nomenclature. 48:22-224.

Wantanabe, T.K., M. Matsuda, S. Ohnishi, and F. Hihara. 1982. Notes on the systematics of *Drosophila jambulina*. Jpn. J. Genet. 57: 561-567.

Wasserman, M. 1982. Evolution of the *repleta* group. *In:* The genetics and biology of *Drosophila*, Vol. 3b, Michael Ashburner, Hampton L. Carson and James N. Thompson, Jr. (ed.s) Academic Press Inc. (London) Ltd. p. 61-139.

Watanabe, H. and D. Sperlich. 1997. *Drosophila hubeiensis* (Diptera, Drosophilidae), a new species of the *Drosophila obscura* species-group from the mainland of China. Jpn. J. Entomol. 65: 621-633.

Watanabe, H., R. Terasawa and F.-J. Lin. 1996. A new species of the *Drosophila obscura* species-group (Diptera, Drosophilidae) from China. Jpn. J. Entomol. 64: 489-495.

Watanabe,T. K. and M. Kawanishi. 1979. Mating preference and the direction of evolution in *Drosophila*. Science. 205: 906-907.

Watanabe,T. K. and M. Kawanishi. 1981. Asymetrical mating success and the phylogeny of *Drosophila*. Zool. Mag. (Tokyo) 90: 217-224.

Wheeler, Marshall R. 1981. The Drosophilidae: a taxonomic overview. *In*: M. Ashburner, H. L. Carson and J. N. Thompson, Jr. (ed.s), The genetics and biology of *Drosophila* Vol. 3a: 1-97. London: Academic Press.

Wheeler, Marshall R. 1986. Additions to the catalog of the world's Drosophilidae. *In*: M. Ashburner, H. L. Carson and J. N. Thompson, Jr. (ed.s), The genetics and biology of *Drosophila* Vol. 3e: 395-409. London: Academic Press.

Wheeler, W. D. Gladstein. MALIGN version 2.7. American Museum of Natural History. Contact wheeler@amnh.org

Wheeler, W.C. 1993. The triangle inequality and character analysis. Mol. Biol. Evol. 10(3): 707-712.

Whiting, M. F., J. C. Carpenter, Q. D. Wheeler and W. C. Wheeler. 1997. The Strepsiptera problem: Phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. Syst. Biol. 46: 1-68.

Wiley, E. O. 1981. Phylogenetics. The Principles and Practice of Phylogenetic Systematics. New York: John Wiley & Sons.

Wiley, E. O., D. Siegel-Causey, D. R. Brooks and V. A. Funk, 1991. The Compleat Cladist. A primer of phylogenetic procedures. Special Publication No. 19. Lawrence, Kansas: The University of Kansas Museum of Natural History.