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

UMI®

MOLECULAR SYSTEMATICS OF THE *REPLETA* SPECIES GROUP OF THE GENUS *DROSOPHILA*

by

CELESTINE M. DURANDO

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.

6/7/2000
Date

Chair of Examining Committee
Dr. Marvin Wasserman, Queens College

6/27/00
Date

Executive Officer
Dr. Richard L. Chappell

Dr. Leslie Marcus, Queens College

Dr. David Grimaldi, American Museum of Natural History

Dr. Robert DeSalle, American Museum of Natural History

Dr. Lee Ehrman, State University of New York

Supervising Committee

The City University of New York

ACKNOWLEDGMENTS

I wish to thank Marvin Wasserman for being an astute and patient advisor during the course of my studies. I will be forever grateful to have been exposed to both his wisdom and his generosity of spirit. He never failed to provide either intellectual stimulation when my enthusiasm was waning or a sympathetic ear when my spirits were lagging.

The work presented in this thesis could not have been accomplished without the tutelage of Rob DeSalle. His intense curiosity about the evolutionary process combined with his kindness and encouragement helped to make the years spent on research a truly valuable experience.

I thank the members of my committee -- Lee Ehrman, Dave Grimaldi, Max Hecht and Leslie Marcus -- for their advice and interest in this project.

Support for the research presented in this thesis was provided by NSF grant 34504734 and an American Museum of Natural History Graduate Fellowship.

I am grateful to Lee Simek and Nicholas Morales for both the friendship and the diversion that they provided, enabling me to maintain a clear perspective. I will always cherish Nicky's love and support, which created a serene place in my life and helped me find the inner strength necessary to complete this work.

TABLE OF CONTENTS

APPROVAL PAGE	ii
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	iv
LIST OF TABLES	v
LIST OF FIGURES	vii
CHAPTER 1: General Introduction	1
CHAPTER 2: Phylogenetic Analysis of the *repleta* Species Group of the Genus *Drosophila* Using Multiple Sources of Characters	14
CHAPTER 3: On Chromosomes and Crossability in Several Species Groups of *Drosophila*	44
CHAPTER 4: Relationships among and within the species of the *Drosophila mayaguana* Subcluster: A Molecular Approach	80
REFERENCES	129

LIST OF TABLES

Table 1-1. *Drosophila repleta* taxonomic list 3

Table 2-1. Taxonomic list of species used in this study 19

Table 2-2. Support for previously established species groups, subgroups, complexes and clusters 27

Table 2-3. Tree statistics for individual and combined data sets 29

Table 2-4. Partitioned Bremer support scores summed across the simultaneous analysis trees for each of the gene partitions and standardized by the minimum number of steps for each partition 31

Table 2-5. ILD values and significance values for pairwise data partition combinations 32

Table 3-1. Data sources for the four data matrices used in this study 49

Table 3-2. Tree statistics for the four analyses 51

Table 3-3. Results of partitioning Bremer support to molecular and inversion character sources 65

Table 3-4. Results obtained for t-test comparisons for direction of crosses 69

Table 3-5. Results obtained for t-test comparisons for fertile F_1 or sterile or no F_1 in four species groups of Drosophilidae 72

Table 3-6. Results obtained for t-test comparisons for sympatry and allopatry comparisons 75

Table 4-1. List of populations and strains used in this study 84

Table 4-2. Key to locale number found to the right of taxon names in Figure 4-1 91

Table 4-3. Tree statistics for individual and combined data sets 100

Table 4-4. Results of partitioning Bremer support for each of the individual data sets across the simultaneous analysis tree 101

Table 4-5. *Ace* diagnostic sites 105

Table 4-6. A+T rich diagnostic sites 107

Table 4-7. *hb* diagnostic sites 109

Table 4-8. *mam* diagnostic sites 111

Table 4-9. *vg* diagnostic sites 113

Table 4-10. Key to the species and locale designations used in the table of F_{ST} values 119

Table 4-11. F_{ST} values between populations of species in the *mayaguana* subcluster calculated for the *Ace* gene 120

Table 4-12. F_{ST} values between populations of species in the *mayaguana* subcluster calculated for the A+T rich gene 121

Table 4-13. F_{ST} values between populations of species in the *mayaguana* subcluster calculated for the *hb* gene 123

Figure 4-14. F_{ST} values between populations of species in the *mayaguana* subcluster calculated for the *mam* gene 125

Figure 4-15. F_{ST} values between populations of species in the *mayaguana* subcluster calculated for the *vg* gene 127

LIST OF FIGURES

Figure 2-1. Strict consensus of 2 most parsimonious trees from combined analysis of all the data equally weighted 25

Figure 2-2. Saturation plots of percent total sequence divergence on the X-axis versus percent partial sequence divergence for transitions, third positions, first and second positions and transversios 34

Figure 2-3. Saturation plots of percent total sequence divergence on the X-axis versus substitution ratio on the Y-axis for third positions divided by first and second positions and for transitions divided by transversions 35

Figure 2-4. Single most parsimonious tree from combined analysis including Ef1α using the pruned data set 39

Figure 2-5. Tree generated by removing third position and all transition changes from CO II partition 41

Figure 2-6. Tree generated using successive approximations weighting based on the retention index 42

Figure 3-1. Method used for counting number of nodes between species pairs 53

Figure 3-2. Cladogram showing the total evidence hypothesis for combined analyses of DNA sequences and chromosomes for the *melanogaster* species group 55

Figure 3-3. Cladogram showing the total evidence hypothesis for combined analyses of DNA sequences and chromosomes for the *virilis* species group 56

Figure 3-4. Cladogram showing the total evidence hypothesis for combined analyses of DNA sequences and chromosomes for the Hawaiian *Drosophila* species group 57

Figure 3-5. Cladogram showing the total evidence hypothesis for combined analyses of DNA sequences and chromosomes for the *repleta* species group 58

Figure 3-6. Cladograms as described in the text for the *melanogaster* species group 60

Figure 3-7. Cladograms as described in the text for the *virilis* species group 61

Figure 3-8. Cladograms as described in the text for the Hawaiian *Drosophila* species group 62

Figure 3-9. Cladograms as described in the text for the *repletar* species group 63

Figure 3-10. Plot of the consistency index of chromosomal inversion partition and the DNA sequence partition when forced to fit the parsimony tree versus the number of taxa in the data set 67

Figure 3-11. Plots showing the 95% confidence ranges for the number of nodes traversed and for patristic distances traversed in comparisons of "crossable" and "uncrossable" species pairs 71

Figure 3-12. Plots showing the 95% confidence ranges for the number of nodes traversed and the patristic distances traversed in comparisons made between sympatric and allopatric crosses 74

Figure 4-1. Strict consensus of 2364 most parsimonious trees from combined analysis of all the data equally weighted 90

Figure 4-2. Strict consensus of most parsimonious trees from analyses of individual gene partitions 94

Figure 4-3. Strict consensus of 7 most parsimonious trees from analysis of *Ace* gene partition 95

Figure 4-4. Strict consensus of 24 most parsimonious trees from analysis of A+T rich gene partition 96

Figure 4-5. Strict consensus of 354 most parsimonious trees from analysis of *hb* gene partition 97

Figure 4-6. Strict consensus of 2 most parsimonious trees from analysis of *mam* gene partition 98

Figure 4-7. Strict consensus of 1085 most parsimonious trees from analysis of *vg* gene partition 99

Figure 4-8. Strict consensus of 755 most parsimonious trees from analysis of A+T rich and *hb* gene partitions 103

Figure 4-9. Percent of the total variation attributable to among species differences, among populations with species differences and within species differences for each of the individual data partitions 128

CHAPTER 1

GENERAL INTRODUCTION

The *Drosophila repleta* species group, one of the largest species groups in the genus *Drosophila*, consists of 89 described and 2 undescribed species (Wasserman, 1992). It has been placed into the subgenus *Drosophila* along with twenty-one other species groups with which it shares the following characteristics: eggs with three or four filaments; long, fine, coiled ventral receptacles; long, spiral testes; and dorsal abdominal bands notched, narrowed or broken in the midline (Patterson and Stone, 1952). The species in this group have been placed into five subgroups -- the *D. mulleri* subgroup, the *D. hydei* subgroup, the *D. mercatorum* subgroup, the *D. repleta* subgroup, and the *D. fasciola* subgroup -- and each subgroup has been further subdivided into complexes, clusters, and subclusters. (Table 1-1). Along with the *D. aureata*, *D. canalinea*, *D. castanea* and *D. dreyfusi* species groups, the *D. repleta* group is included in the *D. repleta* section (Wasserman, 1982a,b) of the *D. virilis-repleta* radiation (Throckmorton, 1982a) and is considered one of the most successful radiations in the genus *Drosophila*. An important reason for the extensive radiation of this group was the ability of these species to inhabit the arid regions of the New World and to adapt to cactus breeding, a niche that is unfavorable to most other species of *Drosophila* (Wasserman, 1982a,b). While members of the *D. fasciola* subgroup, the *D. mercatorum* subgroup and the *D. fulvimacula* complex of the *D. repleta* subgroup inhabit wetter habitats, it is believed that these occurrences represent reinvasions of a wetter environment because in most of these cases the most primitive member of the taxon is found in a desert environment (Wasserman, 1992). The members of the *D. repleta* group can be differentiated morphologically from members of the other twenty-one species groups in the subgenus *Drosophila* by having each bristle and hair on the thorax originating from a dark spot (Patterson and Stone, 1952).

Table 1-1. *Drosophila repleta* taxonomic list.

1. *D. mulleri* subgroup
    A. *D. mulleri* complex
        1. *D. mojavensis*
            D. mojavensis baja
        2. *D. arizonae*
        3. *D. navojoa*
        4. *D. wheeleri*
        5. *D. aldrichi*
        6. *D. mulleri*
        7. *D. nigrodumosa*
        8. *D. huaylasi*
        9. *D. mayaguana*
        10. *D. straubae*
        11. *D. parisiena*
        12. *D. longicornis*
        13. *D. propachuca*
        14. *D. pachuca*
        15. *D. mainlandi*
        16. *D.* "from Sonora"
        17. *D. hexastigma*
        18. *D. spenceri*
        19. *D. desertorum*
        20. *D. ritae*

Table 1-1 (continued). *Drosophila repleta* taxonomic list.

21. *D. mathisi*
22. *D.* "mulleri complex b"

B. *D. martensis* complex

1. *D. martensis*
2. *D. starmeri*
3. *D. uniseta*
4. *D. venezolana*
5. *D. buzzatii*
6. *D. borborema*
7. *D. serido*
8. *D. koepferae*

C. *D. stalkeri* complex

1. *D. stalkeri*
2. *D. richardsoni*

D. *D. eremophila* complex

1. *D. eremophila*
2. *D. mettleri*
3. *D. micromettleri*

E. *D. meridiana* complex

1. *D. meridiana*

D. meridiana rioensis

2. *D. promeridiana*
3. *D. meridionalis*

Table 1-1 (continued). *Drosophila repleta* taxonomic list.

F. *D. anceps* complex

1. *D. anceps*
2. *D. leonis*
3. *D. nigrospiracula*

G. *D. mulleri* miscellaneous

1. *D. hamatofila*
2. *D. nigricruria*
3. *D. pegasa*
4. *D. racemova*
5. *D. subviridis*

II. *D. hydei* subgroup

A. *D. bifurca* complex

1. *D. bifurca*
2. *D. nigrohydei*
3. *D. hydeoides*
4. *D. novemaristata*
5. *D. guayllabambae*

B. *D. hydei* complex

1. *D. eohydei*
2. *D. neohydei*
3. *D. hydei*

D. hydei yucatanensis

Table 1-1 (continued). *Drosophila repleta* taxonomic list.

III. *D. mercatorum* subgroup

1. *D. carcinophila*
2. *D. paranaensis*
3. *D. mercatorum*

D. mercatorum pararepleta

IV. *D. repleta* subgroup

A. *D. peninsularis* complex

1. *D. peninsularis*

B. *D. fulvimacula* complex

1. *D. fulvimacula*

D. fulvimacula flavorepleta

2. *D. fulvimaculoides*

C. *D. repleta* complex

1. *D. repleta*
2. *D. limensis*
3. *D. melanopalpa*
4. *D. canapalpa*
5. *D. neorepleta*

D. *D. repleta* miscellaneous

1. *D. zottii*

V. *D. fasciola* subgroup

A. *D. pictilis* complex

Table 1-1 (continued). *Drosophila repleta* taxonomic list.

1. *D. pictilis*
2. *D. pictura*

B. *D. moju* complex

1. *D. paraguttata*
2. *D. moju*
3. *D. mojuoides*

C. *D. fasciola* miscellaneous

1. *D. fulvalineata*
2. *D. fasciola*
3. *D. coroica*
4. *D. ellisoni*
5. *D. fascioloides*
6. *D. linearepleta*
7. *D. onca*
8. *D. carolinae*
9. *D. hermionae*
10. *D. ivai*
11. *D. querubimae*
12. *D. rosinae*
13. *D. senei*

VI. Miscellaneous *D. repleta* species

1. *D. brevicarinata*

Table 1-1 (continued). *Drosophila repleta* taxonomic list.

2. *D. californica*
3. *D. icteroscutata*
4. *D. inca*
5. *D. marriettae*
6. *D. ramsdeni*
7. *D. vincintinae*

However, in several species these spots have fused to form dark bands, and in two species the spots have disappeared.

Current knowledge of the phylogeny of this species group is based on the morphological work of Throckmorton (1977, 1982a) and Vilela (1983) and the cytological work of Wasserman (1982; 1992 for reviews). There has also been a good deal of ecological (Barker, 1982,1990; Etges, 1990; Etges and Heed, 1987; Etges and Klassen, 1989; Fogelman and Abril, 1990), genetic (Wasserman, 1982a,b; Wasserman and Wasserman, 1992) and behavioral (Wasserman and Koepfer, 1977; Koepfer, 1987; Markow, 1982) data collected on this group. In addition, several allozyme studies (Zouros, 1973; Richardson et al., 1975; Richardson and Smouse, 1976; Richardson et al., 1977; Heed et al., 1990) and some molecular studies (Sullivan et al., 1990; Russo et al., 1995; Spicer, 1995; 1996) have contributed to our knowledge of relationships among members of subsets of this species group. However, none of the allozyme or molecular studies have attempted to address the phylogenetic relationships of the *repleta* species group as a whole.

Detailed polytene chromosome maps have been constructed for 70 of the 91 species in this group, and over 296 inversions have been mapped, of which 118 constitute fixed differences between species. Ninety-four of these 118 fixed chromosomal inversions are autapomorphic and diagnostic for various single species. Several of the chromosomal inversions that have been studied so far are variable among closely related species indicating the possible utility of inversions as phylogenetic tools at this level, yet the degree of resolution from the inversion data within species complexes and clusters is low (Wasserman, 1992). In addition, although these data have been used to infer relationships among species, they tell us

little about the direction of evolution because there is no way to determine which inversions are primitive and which are derived (Wasserman, 1992).

Cytololgical data are collected by examining the paracentric inversions found in the polytene chromosomes of the larval salivary glands. These chromosomes are characterized by specific patterns of puffs and bands. By closely inspecting these patterns one is able to determine where breakage points and rearrangements have occurred, and by comparison, to ascertain the relationships among species.

Wasserman (1954) chose the gene sequence found in the cosmopolitan species, *D. repleta*, as the standard sequence for the group. He examined each of the other species cytologically and compared their gene sequences to that of *D. repleta*. These inversions were labeled using two symbols: the first, designating the chromosome on which the inversion was found and the second, indicating the inversion. The *Basic Assumption* of this approach to analysis is that each inversion is a unique event and that any two taxa which possess the same inversion are more closely related to each other than either is to a third taxon which lacks the inversion (Wasserman, 1982a,b; 1992).

Although Wasserman (1954) had originally used the *D. repleta* gene sequence as the standard for the group, he later proposed a hypothectical ancestor possessing a gene sequence which he called Primitive I (Xabc;2ab;3b;4;5) which gave rise to the *D. repleta* group and its close relatives (Wasserman, 1982a,b). Subsequently five inversions, X-a, X-b, X-c, 2-a and 2-b were incorporated to give the sequence called Primitive III, the ancestral type of both the *repleta* and *mercatorum* subgroups. The *D. repleta* subgroup is divided into two species complexes, the *D. fulvimacula* complex and the *D. repleta* complex. These

complexes are differentiated from one another mainly by the fixation of distinct new inversions on Chromosome 2 (Wasserman, 1992). Interspecific crosses among several of the species in the *repleta* complex yielded fertile offspring in some cases, with some matings being carried to the F_2 generation (Wharton, 1942, 1944; Ward and Stone, 1952).

The *D. mercatorum* subgroup is the smallest subgroup in the *repleta* group and contains only three species. The standard gene order of the ancestor of the *mercatorum* subgroup, called Primitive V, differs from the order of Primitive III by having a fusion between Chromosomes 2 and 3, and the addition of three new inversions (Wasserman, 1982a,b; 1992).

The *D. hydei* subgroup is most closely allied to the *D. meridiana* complex of the *D. mulleri* subgroup. Both of these taxa possess the Primitive I sequence as the standard (Wasserman, 1992). However, the *hydei* subgroup species can be distinguished morphologically based on the extensive coiling in both the testes (22 to 55 coils) and the ventral receptacles (245 to 735 coils). The highest numbers reported for a non-*hydei repleta* group species are 16 coils in the testes and 116 coils in the ventral receptacle of *D. melanopalpa* (Wasserman, 1992). Crosses among members of the *hydei* subgroup yielded fertile F_1 offspring in the case of *D. nigrohydei* males crossed with *D. eohydei* females and *D. hydei* females crossed with *D. neohydei* males. In the latter case, the reciprocal cross also produced some fertile offspring (Wasserman, 1962a).

The *D. mulleri* subgroup is the largest subgroup in the *repleta* group, containing 44 described and two undescribed species. It contains all of the species which do not fall within any of the other species subgroups, but which evolved from Primitive I (Wasserman, 1982a,b; 1992). This subgroup has been divided

into six complexes (the *meridiana* complex, the *anceps* complex, the *eremophila* complex, the *stalkeri* complex, the *mulleri* complex and the *martensis* complex) plus five species which do not fit into any of the complexes.

The standard gene order of the three species that comprise the *meridiana* complex is Primitive I. Members of the *anceps* complex are grouped together based on the addition of a single new inverversion on Chromosome 3 and two new inversions on Chromosome 5. Fixation of four new inversions on Chromosome 2 defines the *eremophila* complex, while the *stalkeri* complex has added two new rearrangements on Chromosome 2 (Wasserman, 1992). The twenty-two species that encompass the *mulleri* complex are homozygous for 2g and 3a, giving them a standard gene order of Xabc; 2abg; 3ab; 4; 5 (Wasserman, 1992). An additional 61 new inversions have been found to occur in this complex, with 21inversions that are fixed interspecific differences unique to the species in which they are found. These species are remarkable in that each is homozygous or lacks some combination of four independent, assorting inversions, Xw, 2c, 2f and 3c (Wasserman, 1982a,b; 1992). Wasserman (1992) explains this phenomenon by proposing a hypothetical ancestor, which he calls Ancestor II, composed of five cytologically distinct populations or subspecies. These five subspecies gave rise to five cytologically defined species which through further changes and speciation events yielded the 22 members of the *mulleri* complex.

The *D. fasciola* subgroup (Wasserman, 1962b) has a standard gene order referred to as Primitive VII (Xabc; 2abo^{2}e^{3}l^{3};3b;4;5), and each of the species has incorporated one or more new inversions. Unlike the majority of the species in the *D. repleta* species group, the members of the *fasciola* subgroup are found in the wet, mesophylic forests of Central and South America.

The present study has been designed to explore the phylogenetic relationships among members of the *D. repleta* group using DNA sequence information. Chapter 2 incorporates the chromosomal inversion data with DNA sequence characters from several different genes in order to attempt to ascertain relationships among subgroups, among complexes within subgroups and among species in this group of *Drosophila*.

Chapter 3 integrates inversion data and DNA data with results obtained from hybridization experiments in order to determine the phylogenetic distances across which hybridization may take place. Three additional species groups in the genus *Drosophila* were examined in this context in order to ascertain if any pattern of relationship might emerge.

Several gene sequences were also examined for individuals belonging to distinct geographic populations of the species belonging to the *D. mayaguana* subcluster of the *D. mulleri* cluster for the purpose of investigating patterns of divergence and radiation.

CHAPTER 2

PHYLOGENETIC ANALYSIS OF THE *REPLETA* SPECIES GROUP OF THE GENUS *DROSOPHILA* USING MULTIPLE SOURCES OF CHARACTERS

Abstract

The species in the *repleta* group of the genus *Drosophila* have been placed into 5 subgroups -- the *mulleri*, *hydei*, *mercatorum* , *repleta* and *fasciola* subgroups. Each subgroup has been further subdivided into complexes and clusters. Extensive morphological and cytological analyses of the members of this species group have formed the foundation for the proposed relationships among the members of the *repleta* species group. Fifty-four taxa, including 46 taxa belonging to the *repleta* species group were sequenced for fragments of 4 genes - 16S ribosomal DNA (16S), Cytochrome oxidase II (CO II) and Nitrogen dehydrogenase1 (ND1) of the mitochondrial genome, and a region of the *hunchback* (*hb*) nuclear gene. We also generated a partial data set of Elongation factor 1-alpha (Ef1α) sequences for a subset of taxa. Our analysis used both DNA characters and chromosomal inversion data. The phylogenetic hypothesis we obtained supports many of the traditionally accepted clades within the *mulleri* subgroup, but the monophyly of taxonomic groups outside of this subgroup appears not to be supported. Phylogenetic analysis revealed one well supported, highly resolved clade that consists of closely related members of the *mulleri* and *buzzatii* complexes. The remaining taxa, a wide assortment of taxonomic groups, ranging from members of other species groups to members of several subgroups and members of three species complexes from the *mulleri* subgroup are found in poorly supported arrangements at the base of the tree.

Introduction

The *Drosophila repleta* species group is among the largest of all species groups in the genus *Drosophila*. Part of the *virilis-repleta* radiation (Throckmorton, 1982a), it is considered one of the most important and successful radiations in the genus *Drosophila*. For the most part the members of the *repleta* species group are found in the arid or semi-arid deserts of the New World where they live on various species of cactus (Wasserman, 1992). The species in this group have been placed into five subgroups -- the *D. mulleri* subgroup, the *D. hydei* subgroup, the *D. mercatorum* subgroup, the *D. repleta* subgroup, and the *D. fasciola* subgroup -- and each subgroup has been further subdivided into complexes, clusters, and subclusters.

Current knowledge of the phylogeny of this species group is based on the morphological work of Throckmorton (1982a) and Vilela (1983) and the cytological work of Wasserman (1982; 1992 for reviews). In addition, several allozyme studies (Zouros, 1973; Richardson et al., 1975; Richardson and Smouse, 1976; Richardson et al., 1977; Heed et al., 1990) and some molecular studies (Sullivan et al., 1990; Russo et al., 1995; Spicer, 1995; 1996) have contributed to our knowledge of relationships among members of subsets of this species group. However, none of the allozyme or molecular studies have attempted to address the phylogenetic relationships of the *repleta* species group as a whole.

Detailed polytene chromosome maps have been constructed for 70 of the 91 species in this group, and over 296 inversions have been mapped, of which 118 constitute fixed differences between species. Ninety-four of these 118 fixed chromosomal inversions are autapomorphic and diagnostic for various single species. Several of the chromosomal inversions that have been studied so far are

variable among closely related species indicating the possible utility of inversions as phylogenetic tools at this level, yet the degree of resolution from the inversion data within species complexes and clusters is low (Wasserman, 1992). In addition, although these data have been used to infer relationships among species, they tell us little about the direction of evolution because there is no way to determine which inversions are primitive and which are derived. Therefore, it is necessary to place the inversion data into a cladistic framework along with other sources of characters.

Chromosomal information (Wasserman, 1982a,b; 1992) has been an excellent base for understanding of phylogeny in this species group. However, a more detailed phylogenetic understanding of the group will most likely be obtained from DNA sequence information, which is the focus of this study. In particular, there are several questions that remain open in the phylogenetics of this species group that can be addressed by addition of molecular information. These include: 1) The phylogenetic relationships of the various subgroups to one another; 2) The monophyly of the various subgroups; 3) The phylogenetic relationships of species within species complexes, especially the *mulleri* complex; 4) The phylogenetic placement of the miscellaneous species, such as *D. pegasa* and *D. hamatofila*, into complexes.

For this study we have sequenced four genes: mt 16S rDNA, mt ND1, mt CO II (Simon et al., 1994) and the nuclear gene, *hunchback* (Treier et al., 1989). We have assessed the relative contribution of each gene partition, as well as the chromosomal inversion data, to the phylogenetic hypothesis we generated for this study. Previous studies, such as those done on Hawaiian *Drosophila* (Carson, 1972; Gillespie, 1996), the *Drosophila melanogaster* species group (Lemeunier et al.,1986), the *Drosophila virilis* species group (Throckmorton, 1982b) and the *Drosophila repleta* species group (Wasserman, 1982a,b; 1992), have indicated that

Drosophila inversion character data provide valuable information. As more data accrue for various groups of *Drosophila* , this notion of high utility of inversion data can be examined empirically.

Materials and Methods:

Flies and DNA sequences: Table 2-1 lists all flies used in this study. We chose eight outgroup taxa to root our phylogenetic trees - *D. melanogaster* , a member of the subgenus *Sophophora* ; *D. cyrtoloma* and *D. longipedis*, members of the Hawaiian *Drosophila* group; *D. virilis*; *D. pavani* and *D. gaucha* , members of the *D. mesophragmatica* group; *D. canalinea* of the *D. canalinea* species group; and *D. camargoi*, a member of the *D. dreyfusi* group. *D. melanogaster* was chosen as a distant outgroup as it is a member of another subgenus within the genus *Drosophila* , and the Hawaiian *Drosophila* were chosen as a closer outgroup as they are one of the potential sister groups to the *virilis-repleta* radiation (Remsen and DeSalle, 1998; Kwiatowski and Ayala, 1999). We also chose to examine the placement of *D. virilis* in relation to the *repleta* species group flies as well as the relationship of the species in the *D. mesophragmatica* species group and the *D. canalinea* and *D. dreyfusi* species groups because these species have long been considered the closest sister taxa to the *D. repleta* group. DNA was isolated from single and multiple flies using the methods outlined in Vogler et al. (1993). The polymerase chain reaction (PCR) was used to amplify fragments for sequencing. Primers used are described in Baker and DeSalle (1997) and Baker et al. (1998). PCR products were cleaned using Geneclean kits (BIO 101) and sequenced directly using either manual or automated sequencing methods. Manual sequencing was accomplished using Sequenase (US Biochemicals) and S^{35} labelling. Autoradiograms were used

Table 1. Taxonomic list of *Drosophila* species used in this study. Taxa marked with an * are those belonging to the subset of taxa used to perform the analysis with Ef1α. Numbers refer to Bowling Green Stock Center numbers. Localities refer to locations where MW, WBH and WJE collected specimens. MW, WBH refer to specimens archived in the labs or Wasserman or Heed.

Species Group	Subgroup	Complex	Cluster	Species	Source
D. repleta	*D. mulleri*	*D. mulleri*	*D. mojavensis*	*D. mojavensis**	Vallecito CA
				D. arizonae	Tomatlan, Jalisco, Mex.
				D. navojoa	Las Bocas, Sonora, Mex.
			D. mulleri	*D. wheeleri**	Catalina Island CA
				D. aldrichi	Zapotitlan, Puebla, Mex.
				D. mulleri	Big Pine Key FLA
				D. nigrodumosa	Merida, Venezuela
				D. huaylasi	WBH
				D. mayaguana	Grand Inagua Island
				*D. straubae**	Navassa Island
				D. parisiena	Fond Parisien, Haiti
			D. longicornis	*D. longicornis*	Zapotitlan, Puebla, Mex.
				D. propachuca	MW
				*D. pachuca**	MW
				D. mainlandi	Catalina Island CA
				D. hexastigma	Zapotitlan, Puebla, Mex.

			D. spenceri	Guayamas, Sonora, Mex.
		D. ritae	*D. desertorum*	MW
			D. ritae	1471.2
	D. buzzatii	*D. martensis*	*D. martensis**	MW
			D. starmeri	La Palmares, Venezuela
			D. uniseta	MW
			D. venezolana	MW
		D. buzzatii	*D. buzzatii*	1291.1
			D. borborema	MW
			*D. serido**	1431.2
			D. koepferae	Vipos, Argentina
		D. stalkeri	*D. stalkeri*	Big Pine Key FLA
			D. richardsoni	Spanish Point, Monserrat
	D. eremophila		*D. eremophila**	Guayalejo, Tamaulipas, Mex.
			D. mettleri	MW
			*D. micromettleri**	Skyline Drive, Cuba
	D. meridiana		*D. meridiana*	1342.0
			*D. meridionalis**	MW
	D. anceps		*D. anceps**	MW
			*D. leonis**	1395.0
			*D. nigrospiracula**	1503.0
	Miscellaneous		*D. hamatofila*	MW
			*D. pegasa**	Oaxaca, Oaxaca, Mex.
D. hydei	*D. bifurca*		*D. nigrohydei*	MW
	D. hydei		*D. hydei*	Tequila, Jalisco, Mex.
D. mercatorum			*D. paranaensis**	MW
			*D. mercatorum**	MW
D. repleta	*D. fulvimacula*		*D. fulvimacula**	MW

	D. fasciola	*D. repleta* Miscellaneous
D. canalinea		
D. dreyfusi		
D. mesophragmatica		
D. virilis		
Hawaiian Drosophila		
D. melanogaster		

*D. neorepleta**	1611.2
D. ellisoni	MW
*D. canalinea**	1221.1
D. camargoi	1221.2
*D. gaucha**	1231.0
D. pavani	1241.0
*D. virilis**	See Baker and DeSalle (1997)
D. cyrtoloma	See Baker and DeSalle (1997)
D. longipedis	See Baker and DeSalle (1997)
*D. melanogaster**	See Baker and DeSalle (1997)

to visualize the sequences, and all sequences were generated in both directions. Automated sequencing was accomplished using the ABI dye terminator system and sequenced on an ABI 373 machine. Inspection and correction of automated sequences were accomplished using the SEQUENCHER software (Gene Codes Corp., 1995).

Data matrices and phylogenetic analysis: All sequences were compiled into NEXUS files after alignment of individual gene partitions. We explored the alignment space using the methods outlined in Gatesy et al. (1994) in which alignments are obtained for several gap:change costs and various multiple alignments are examined for regions of "ambiguity". We removed ("culled") regions in our alignments that exhibited ambiguity as defined by Gatesy et al. (1994). CO II alignments were trivial as no indels occurred in this gene region for all of the taxa in our matrix. The mt 16S rDNA sequences required alignment, and approximately 60-70 bases were removed in the culling process. *Hunchback* (*hb*) alignments were accomplished by first translating the DNA sequences into amino acid sequences and performing alignments on the amino acid sequences using ClustalX (Gibson et al., 1994). Areas of alignment ambiguity were removed as described above and the corresponding DNA sequences were substituted for the amino acids. This process was necessary due to the existence of large stretches of poly-glycine repeats in the *hb* sequences that produced several regions of alignment ambiguity. Chromosomal inversion data were also coded as presence or absence of particular inversions and included in the data matrix. Phylogenetic analysis was accomplished using PAUP 4.01b (Swofford, 1999). We analyzed each of the four genes and the inversions separately and in combination in order to explore the interaction of the various gene regions in phylogenetic analysis. In addition, we

examined the congruence of the various gene partitions and inversions using the incongruence length difference (ILD; Farris et al., 1994; 1995) and the associated statistical test for congruence implemented in PAUP 4.01b. Bootstrap values (Felsenstein, 1985) were generated using PAUP 4.01b. Bremer supports (Bremer, 1988; 1994) were calculated using the AUTODECAY program (Eriksson, 1997). Partitioned Bremer supports for the various character partitions were calculated using the methods outlined in Baker and DeSalle (1997) and Baker et al. (1998).

Higher level versus lower level analyses: Initially our phylogenetic analysis was accomplished using a broad sampling of species in the *repleta* species group as well as the outgroup and sister taxa mentioned above. Our initial analysis using the four gene regions listed above resulted in a lack of resolution at the base of the tree where relationships between the species subgroups would be most evident. To examine relationships of the species subgroups we generated sequences for a subset of taxa for Ef1α (Table 2-1). We sequenced Ef1α for only a subset because at this level we were interested in the overall relationships of the species subgroups, and there is a relatively low level of variability in Ef1α (Cho et al., 1995). We analyzed the Ef1α data in a combined analysis by pruning our data matrix to only those taxa that had complete sequences for all gene partitions. Alignments for Ef1α were trivial as no indels were apparent in these sequences for the taxa we examined. All phylogenetic analyses using this higher level matrix were accomplished as described above.

Results and Discussion

A phylogenetic hypothesis for relationships of species in the repleta species group

Two parsimony trees were obtained by analysis of the 46 ingroup taxa and eight outgroup taxa (two Hawaiian *Drosophila* species and *D. melanogaster*). There was a total of 501 characters that were phylogenetically informative in the combined character matrix. The strict consensus of the two most parsimonious trees (CI = 0.31, RI = 0.52 and a total length of 2491 steps) is shown in Figure 2-1. This consensus tree contains one major clade showing a high degree of resolution and a second set of taxa at the base of the tree that shows a low degree of resolution. The clade that is highly resolved contains the *mulleri* and *buzzatii* complexes as well as a single "miscellaneous" species, *D. hamatofila* , all from the *mulleri* subgroup. The less resolved group of taxa contains a mixture of all five species subgroups, including three *mulleri* subgroup complexes and species from three of the species groups allied to the *repleta* group (*mesophragmatica, canalinea* and *dreyfusi* species groups). The *meridiana* complex, a member of the *mulleri* subgroup, is shown as the most basal clade in the consensus tree. While this unresolved group of flies is shown as a monophyletic group in Figure 2-1, support for this hypothesis is low. In addition, within this group only relationships at the tips of the tree are well resolved, with relationships among the species subgroups and complexes depicted as a polytomy.

The high degree of resolution in the *mulleri* and *buzzatii* complexes can be contrasted with the lack of significant resolution in the rest of the taxa in the analysis. Of the five subgroups (*hydei, repleta, fasciola, mulleri* and *mercatorum*) only the *mercatorum* and *hydei* subgroups are seen as monophyletic in the consensus tree in Figure 2-1. The species subgroup for which we have the best

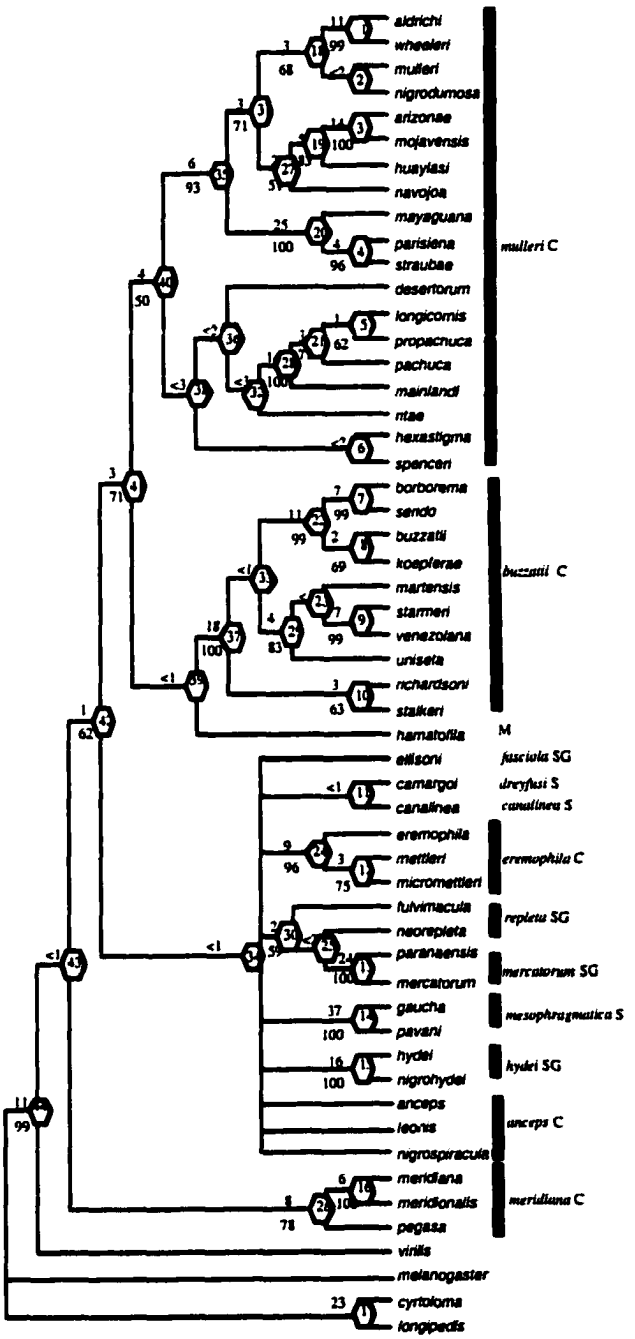


Figure 2-1. Strict consensus of 2 most parsimonious trees from combined analysis of all the data equally weighted. Numbers in hexagons are node designations. Bremer support values are provided above each node and bootstrap values below each node. SG = subgroup, C = complex, S = species group.

sampling, the *mulleri* subgroup, is polyphyletic and the representatives of the *repleta* subgroup are paraphyletic with respect to the *mercatorum* subgroup flies. Of the five species complexes in the *mulleri* subgroup, four are clearly monophyletic with moderate to high character support (See Table 2-2). The fifth complex is embedded in the unresolved clade described above. The fact that the ten species comprising the *buzzatii* complex are recovered as a monophyletic group supports the conclusions of the cytological re-analysis of these species performed by Ruiz and Wasserman (1993). Our consensus tree shows that the *buzzatii* complex is sister to the *mulleri* complex andthat the *buzzatii* complex (BS = 18 and BP = 100%) is monophyletic. Within the *mulleri* complex there are four so-called clusters (*mojavensis*, *ritae*, *longicornis* and *mulleri*). The *mulleri* cluster and the *mojavensis* cluster are intermixed in Figure 2-1 as four of the *mulleri* cluster species (*aldrichi*, *wheeleri*, *mulleri* and *nigrodumosa*) are observed as sister to the *mojavensis* cluster (*mojavensis*, *navojoa* and *arizonae*) with the remaining *mulleri* cluster species (*parisiena*, *mayaguana* and *straubae*) as sister to these. Support for these relationships is relatively strong as the node defining the break between the *mulleri* cluster flies (node 31) has BS = 3 and BP = 71%. It is interesting to note that *D. huaylasi*, which had been placed into the *mulleri* cluster based on the fact that it is homosequential with the other the species in that cluster, is recovered as a member of the *mojavensis* cluster in Figure 2-1. The males in the *mojavensis* cluster possess a characteristic penis, and when we went back to the original description of *D. huaylasi* (Fontevila et al., 1990), we found that it, too, possessed the *mojavensis* cluster penis. The *longicornis* and *ritae* clusters are also intermixed, but the relationships of the flies in these two clusters is not robust as BS and BP values are relatively low at the base of this clade.

Table 2-2. Support for previously established species groups, subgroups, complexes and clusters. See also Figure 2-1. BS = Bremer support, BP = bootstrap, MONO = number of extra steps required to make a non-monophyletic group monophyletic.

Previously established group	Figure 1 (BS, BP)	Figure 4a (BP)	Figure 4b (BP)	MONO
GROUP				
repleta	N	Y (< 50)	N	2
mesophragmatica	Y (37, 100)	Y (100)	Y (88)	
SUBGROUP				
mulleri	N	N	N	7
hydei	Y (16, 100)	Y (100)	Y (100)	
repleta	N	N	N	1
mercatorum	Y (24, 100)	Y (100)	Y (100)	
COMPLEX				
mulleri	Y (4, 50)	Y (78)	Y (73)	
buzzatii	Y (18,100)	Y (100)	Y (100)	
meridiana	Y (8, 78)	Y (100)	Y (100)	
anceps	N	Y (63)	N	3
eremophila	Y (9, 96)	Y (100)	Y (100)	
CLUSTER				
mojavensis	N	Y (56)	N	5
ritae	N	N	N	6
longicornis	N	N	N	8
mulleri	N	N	N	5
martensis	Y (4, 83)	Y (95)	Y (90)	
buzzatii	Y (11,99)	Y (92)	Y (100)	
stalkeri	Y (3, 63)	Y (59)	Y (78)	

We also estimated the number of steps that need to be added to the parsimony tree in order to make the groups in Figure 2-1 monophyletic. These trees indicate that although a group is not monophyletic, it could easily become monophyletic with the addition of more data, as evidenced by the low number of extra steps required to make it monophyletic (*repleta* group, *repleta* subgroup, *anceps* complex). On the other hand, the lack of monophyly of several clusters (*mojavensis*, *ritae*, *longicornis* and *mulleri*) and the *mulleri* subgroup is more strongly supported by our data, as a larger number of extra steps are necessary to make each of them monophyletic (Table 2-2).

In general, resolution at the base of the tree is extremely poor. For instance, there are 19 nodes in the tree (Figure 2-1) with bootstrap values greater than 80%. Of these 19 strongly supported nodes only 2 unite species from different clusters, and both of these occur within the *mulleri* subgroup.

Separate analyses and conflicting signal among molecular partitions

In order to assess the relative contribution of the various gene regions to the simultaneous analysis (SA) hypothesis we analyzed each gene separately and calculated both partitioned Bremer support and incongruence length differences. The general pattern which emerges from these analyses indicates significant disagreement between, on the one hand, COII and ND1 and, on the other hand, *hb*. All of the separate analyses are characterized by low consistency (Table 2-3) and a low degree of resolution. As with the SA trees, the individual gene trees exhibit strongly supported relationships only among closely related taxa. All of the genes show substantial topological disagreement with each other and with the SA hypothesis (Figure 2-1). Only 4 nodes on the SA tree (numbers 3, 13, 14, 16 in

Table 2-3. Tree statistics for individual and combined data partitions. Tot. no. chars. = total number of characters in the data partition, No. PI = number of phylogenetically informative characters in the data partition, No. trees = number of most parsimonious trees obtained in the analysis, Steps = length of most parsimonious trees, CI = consistency index, RI = retention index.

	Tot. no. Chars.	No. PI	No. trees	Steps	CI	RI
Inv	119	40	4	80	0.89	0.93
CO II	442	154	72	972	0.25	0.47
ND1	129	40	60	217	0.29	0.60
16S	521	71	550	257	0.37	0.58
hb	527	196	1710	848	0.40	0.61
Total	1738	501	2	2491	0.31	0.52

Figure 2-1) appear in each of the separate analyses. The *hb* tree shows the greatest topological similarity to the SA tree, with 17 nodes in common, followed by COII (13), ND1 (10) and 16S (9). The separate analyses of each of the three mitochondrial genes are particularly divergent from the SA topology in several areas. For instance, within the COII strict consensus, *D. virilis* has a highly derived placement within the ingroup, *D. uniseta* is strongly separated from the other *buzzatii* complex species and *D. hydei* and *D. micromettleri* are sister taxa. In the ND1 consensus tree, *D. virilis* again has a highly derived placement within the ingroup, and members of the *anceps* and *eremophila* complexes are united with the most distantly related outgroup taxa. The 16S consensus tree places the outgroup taxa at several derived positions, and representatives from the *hydei* subgroup, *repleta* subgroup and *anceps* complex are widely separated.

Because of the poor resolution in the separate analyses it is difficult to establish to what degree individual genes support relationships that emerge in the combined analysis. Partitioned Bremer support provides one means for assessing the contribution of different genes to the SA topology. For this data set, the genes differ significantly in the extent to which they contribute to the total Bremer support of the SA hypothesis. Relative to its size, 16S provides over twice the support of any of the other genes, whereas ND1 provides virtually no support (Table 2-4). Differences in support provided by the various genes may result either from differences in internal homoplasy within each gene or conflicting signals among the genes. An incongruence length test for all the genes combined indicates significant disagreement (standardized ILD = 0.047, p = 0.01). Table 2-5 shows the results of ILD tests for all pairwise gene comparisons and for each gene against the rest of the data combined. The pattern of ILD scores from Table 2-5 suggests that the major

Table 2-4. Partitioned Bremer support scores summed across the simultaneous analysis trees (Figure 2-1) for each of the gene partitions and standardized by the minimum number of steps for each partition.

Gene	Summed PBS	Min Steps	Summed PBS/Min Steps
COII	86.05	243	.35
16S	79.39	96	.82
ND1	6.08	63	.10
hb	111.75	336	.33

Table 2-5. ILD (incongruence length difference) values and significance of value for pairwise data partition combinations. The upper half of the matrix provides the significance values (NS = not significant) and the lower half of the matrix provides the ILDs standardized by the length of the most parsimonious tree(s).

	CO II	16S	ND1	hb	Rest
CO II	-----	NS	NS	0.01	0.01
16S	0.038	-----	NS	NS	NS
ND1	0.035	0.067	-----	0.02	NS
hb	0.036	0.028	0.034	-----	0.01
Rest	0.031	0.013	0.014	0.020	-----

area of conflict is between the mitochondial protein coding genes (COII-ND1) and *hb*. The *hb* gene is significantly incongruent with both mitochondrial genes, which are not different from each other. A breakdown of the partitioned Bremer support values at each node also indicates substantial conflict between these partitions. Of the 42 resolved nodes on the SA consensus tree 25 have conflicting PBS values (i.e., negative values for one partition, positive values for the other) for COII-ND1 and *hb*.

Given the conflict between COII-ND1 and *hb*, it is important to identify whether this disagreement is isolated to specific characters or taxa. Several studies have demonstrated differences in rates of evolutionary change among different classes of molecular data (Brown et al., 1982; DeSalle et al., 1987; Helm-Bychowski and Cracraft, 1993; Knight and Mindell, 1993), and this result is often used as a criterion for downweighting these characters in phylogenetic analysis (Martin, 1995; Yoder et al., 1996; Murphy and Collier, 1997; Bloomer and Crowe, 1998, Danforth and Ji, 1998; Martin and Bermingham, 1998; Simons and Mayden, 1998). Saturation curves have generally been used as a means for assessing relative rates of change for different types of character data although the utility of these plots has recently been questioned (Zang, 1998; Allard et al., 1999; Baker et al., submitted). A comparison of pairwise sequence distances for *hb* and COII-ND1 (Figure 2-2) shows a marked contrast in the pattern of divergences between the two partitions. Divergences for *hb* are linear with respect to the total amount of sequence change whereas COII-ND1 divergences clearly asymptote at a point approximately corresponding to ingroup-outgroup comparisons. Figure 2-3 breaks down the saturation in COII-ND1 with respect to codon position and transitions/tranversions. Figure 2-3a suggests that thirds position sites are changing at slightly faster rates

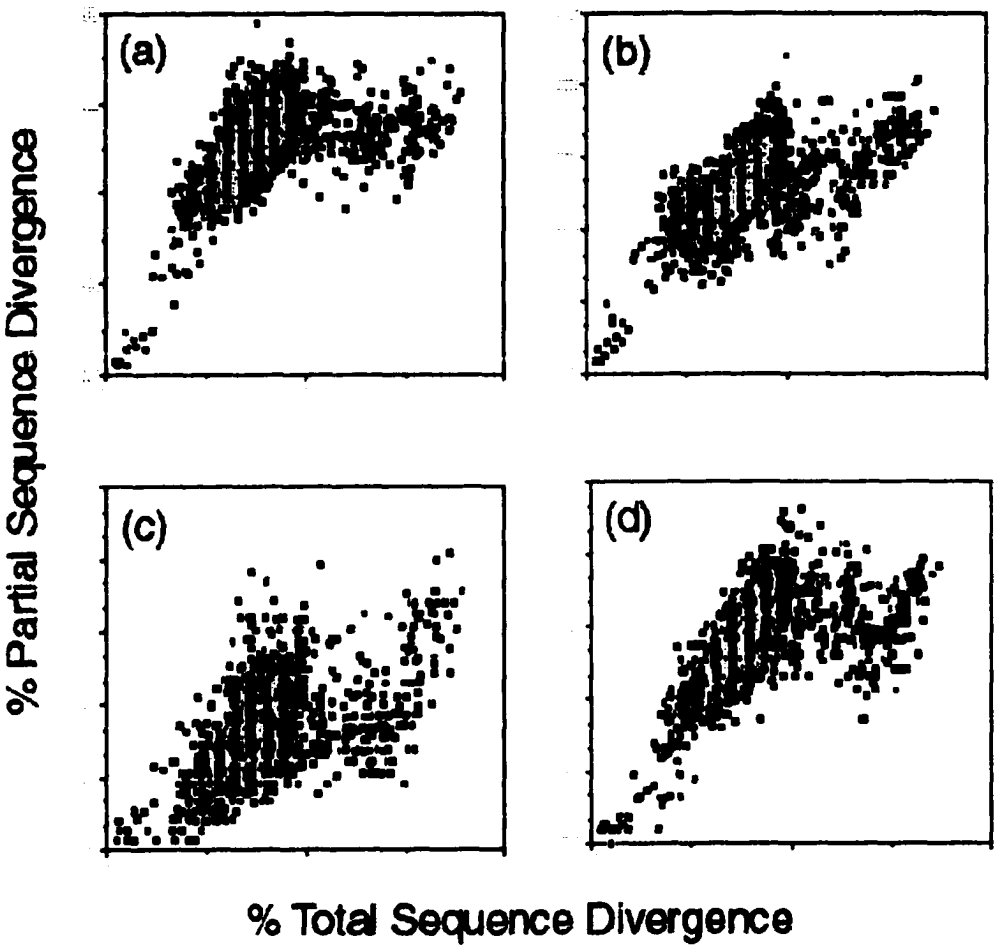


Figure 2-2. Saturation plots of Percent total sequence divergence on the X-axis versus Percent partial sequence divergence for third positions (a), transistions (b), first and second positions (c) and transversions (d). Spread sheets were created in Microsoft Excel with the appropriate variables from PAUP 4.01b (Swofford, 1998) and graphed in Excel.

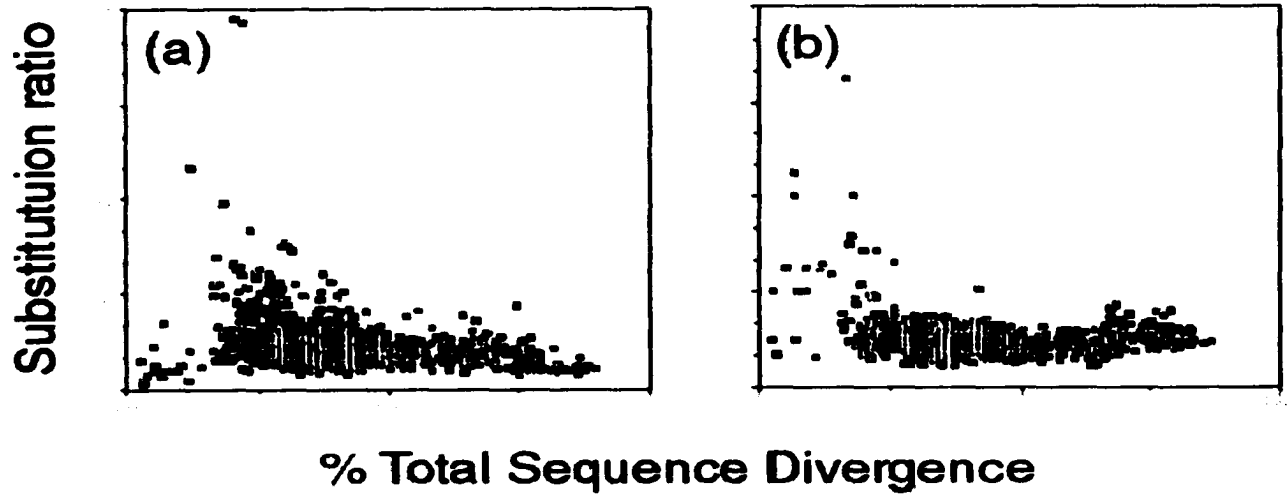


Figure 2-3. Saturation plots of Percent total sequence divergence on the X-axis versus Substitution ratio on the Y-axis for third posistions divided by first and second position (graph on left) and for transitions divided by transversions (graph on right). Spread sheets were created in Microsoft Excel with the appropriate variables from PAUP 4.01b (Swofford, 1998) and graphed in Excel.

than first and second position sites because the most closely related taxa have the highest ratio of third position distances to first and second position distances. A similar pattern exists for transitions relative to tranversions (Figure 2-3b). This type of pattern of sequence change is often used as a justification for downweighting the effects of third position sites and transistions (Martin, 1995; Martin and Bermingham, 1998; Simons and Mayden, 1998).

A more detailed analysis of the incongruence patterns, however, suggests that the disagreement between *hb* and COII-ND1 is not simply caused by the faster evolving characters. This is supported by several results. First, within COII and ND1, the phylogenetic behavior from the faster evolving characters does not differ substantially from that of the slower evolving characters. Relative to the rest of the data, first and second positions sites and transversions for COII-ND1 show a pattern of saturation similar to third position sites and transitions (Figure 2-2). There is also no significant conflict, within COII and ND1, between the first and second positions sites and third position sites (ILD= 31, p = .94). Second, the removal of third position characters or transitions does not eliminate the conflict with *hb*. COII-ND1 partitions including only first and second positions sites (ILD = 38, p = .01) or transversions (ILD = 56, p = .01) are both significantly incongruent with *hb*. Third, the disagreement between COII-ND1 and *hb* is not isolated to the most divergent taxa. The pairwise distance plots for COII-ND1 (Figure 2-2) asymptote approximately at the point of ingroup-outgroup comparisons. Saturation curves for the ingroup taxa alone exhibit a linear pattern of change. Therefore, if saturated characters are the primary cause of conflict between COII-ND1 and *hb*, we would expect agreement between the partitions for analyses including only ingroup taxa. To examine this possibility, we conducted ILD tests for both the repleta group species and the *mulleri* complex species (node 40 in Figure 2-1). For

both of these subtaxa analyses, there is still significant disagreement between COII-ND1 and *hb* (*repleta*: ILD = 39, p = .01 ; *mulleri*: ILD = 26 , p = .01). Overall, the saturated characters in COII-ND1 provide only a limited explanation for the disagreement with *hb*, and there is little evidence that removing these characters improves phylogenetic estimation.

Inversions and congruence

We found that the inversion data conflict with the simultaneous analysis hypothesis at two nodes (nodes 19 and 26) while the combined molecular partitions conflict with the simultaneous analysis hypothesis at three nodes (nodes 29, 40 and 41). The total Bremer support from inversions at all nodes in the simultaneous analysis tree is 29 and for the molecular partition is 270. If these values are standardized by dividing the total Bremer support by the minimum steps for each partition (Baker et al., 1998), both the inversion partition and the molecular partition contribute 0.37 Bremer support units each per phylogenetically informative character. In addition, the consistency indices of the inversion data, whether analyzed alone or forced onto the SA hypothesis, are three times higher than the consistency indices for the molecular data (CI(for inversions) = 0.89, CI(for molecules) = 0.30 for data analyzed alone; CI(for inversions) = 0.89, CI(for molecules) = 0.29 for data forced on the SA tree). Because a consistency index of one is an indication of no homoplasy in a data set, the CI(for inversions) = 0.89 appears to suggest some conflict with the *Basic Assumption* of Wasserman (1992) that chromosomal inversions are unique events. However, we may conclude that the inversion data set contains a greater amount of information than the molecular data set as it possesses a far lesser degree of homoplasy than the molecular

partition. Finally, we note that there is no significant incongruence between the molecular and inversion partitions (ILD = 0.011; NS).

A phylogenetic hypothesis for the relationships of species subgroups in the repleta species group using Ef1α

Due to the lack of resolution at the base of our tree and the lack of inference on the species subgroup relationships, we constructed a second data matrix by adding Ef1α (Cho et al., 1995) sequences. Because of the slow rate of evolution of Ef1α sequences, we decided to sequence only a subset of taxa from our original sampling to see if the Ef1α sequences could add resolution to the cladogram. We used representatives from four of the five species subgroups. Because Figure 2-1 suggests that the *mulleri* subgroup is not monophyletic, we included representatives of all five species complexes in this subgroup. Table 2-1 shows the taxa for which we obtained Ef1α sequences. Phylogenetic analysis of this pruned data set resulted in a single parsimony tree (CI = 0.42; RI = 0.37; steps = 1568) shown in Figure 2-4. This cladogram demonstrates the utility of the Ef1α sequences in resolving relationships in this problematic clade as all nodes are resolved and have relatively high Bremer support, although bootstrap values were low.

The cladogram shows the *mulleri* subgroup splitting into two major clades as observed in Figure 2-1. The *mulleri* and *buzzatii* complexes are sisters in one of the monophyletic groups and the *meridiana*, *anceps* and *eremophila* complexes

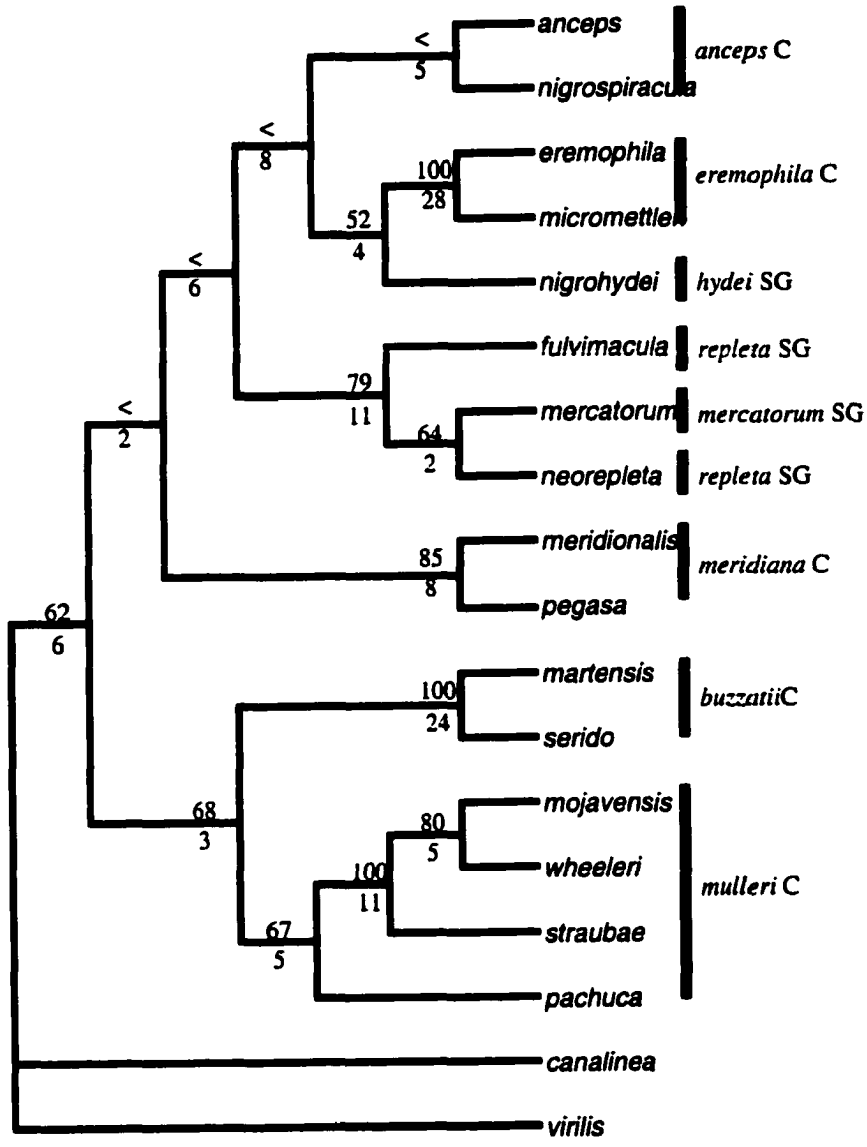


Figure 2-4. Single most parsimonious tree from combined analysis including Ef1a using the pruned data set. Bootstrap values are provided above each node and Bremer support values below each node. SG = subgroup, C = complex.

reside in the second monophyletic group. The *hydei* subgroup representative is observed as the sister taxon to the *eremophila* complex of the *mulleri* subgroup (BS = 4 and BP= 57%). The *mercatorum* and *repleta* subgroups are observed as members of a well supported clade (BS = 11 and BP= 79%) that is sister to the *hydei, eremophila, anceps* clade (BS = 6 and BP= 50%). The *meridiana* complex is the most basal representative of this clade. While Ef1α adds to the resolution of the relationships of taxa in this problematic clade, other data will probably be required to resolve the relationships of these species subgroups.

Conclusions

The present study suggests that the *repleta* species group is not monophyletic. However, this inference may be the result of poor resolution at the base of the phylogenetic tree we obtained (Figure 2-1). Other existing taxonomic groups, such as subgroups, complexes and clusters (Table 2-1) in the *repleta* lineage, were also examined for monophyly. In general these other taxonomic groupings are upheld by our analyses (Table 2-2). A notable exception, however, is the monophyly of the *mulleri* species subgroup, which is paraphyletic at best and most likely polyphyletic (See Figures 2-1, 2-5 and 2-6). However, this could be explained by the fact that Wasserman used the *mulleri* subgroup as a "waste paper basket" to include several complexes and clusters which he could not place elsewhere. At the level of relationships of species subgroups our results indicate close affinity of the *repleta* and *mercatorum* subgroups. A well resolved clade comprised of the *mulleri* and *buzzatii* species complexes and a poorly resolved mixed taxonomic group comprised of a mixture of *mulleri* subgroup species complexes and all other species subgroups, as well as flies from other species

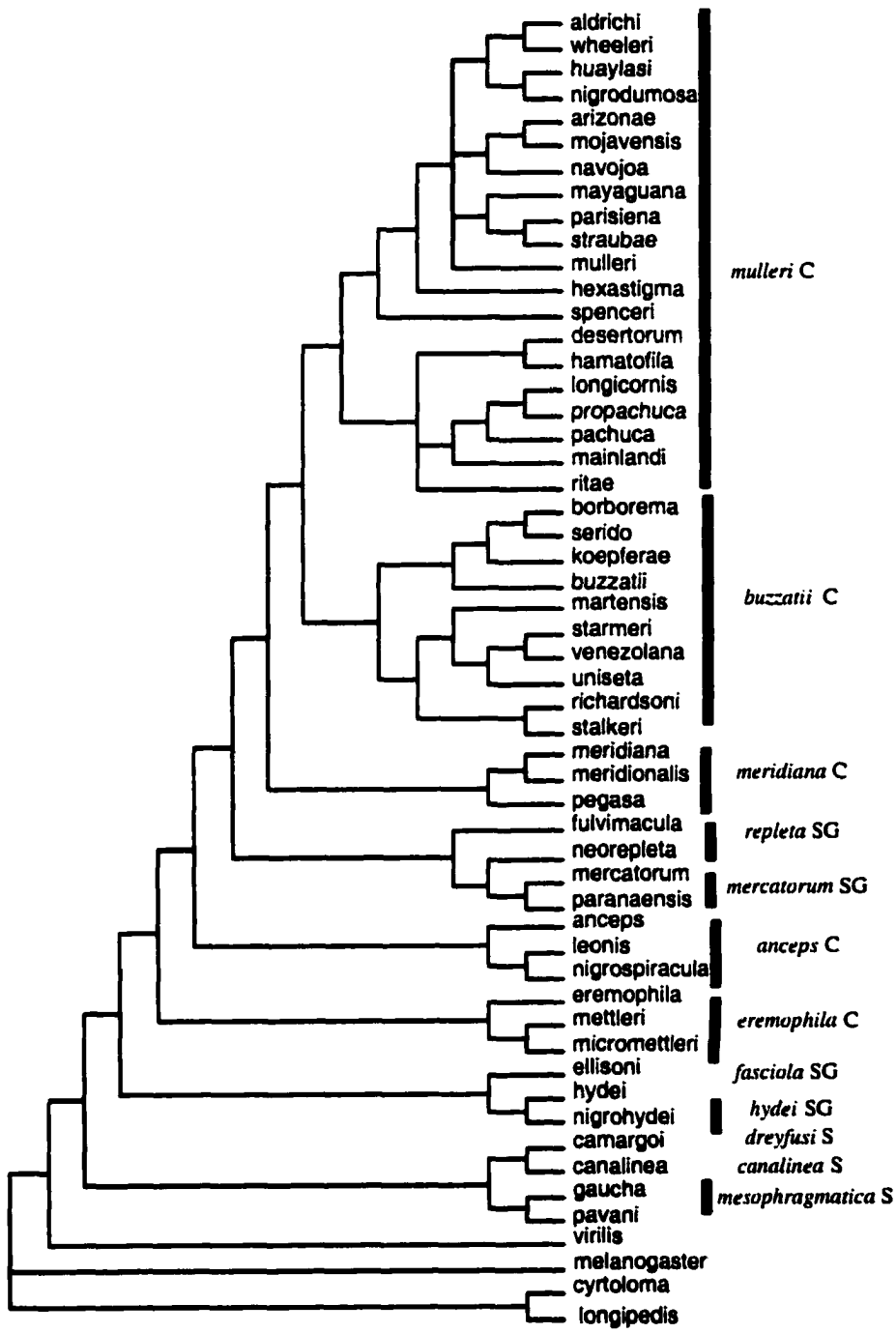


Figure 2-5. Tree generated by removing third position and all transition changes from the CO II partition.

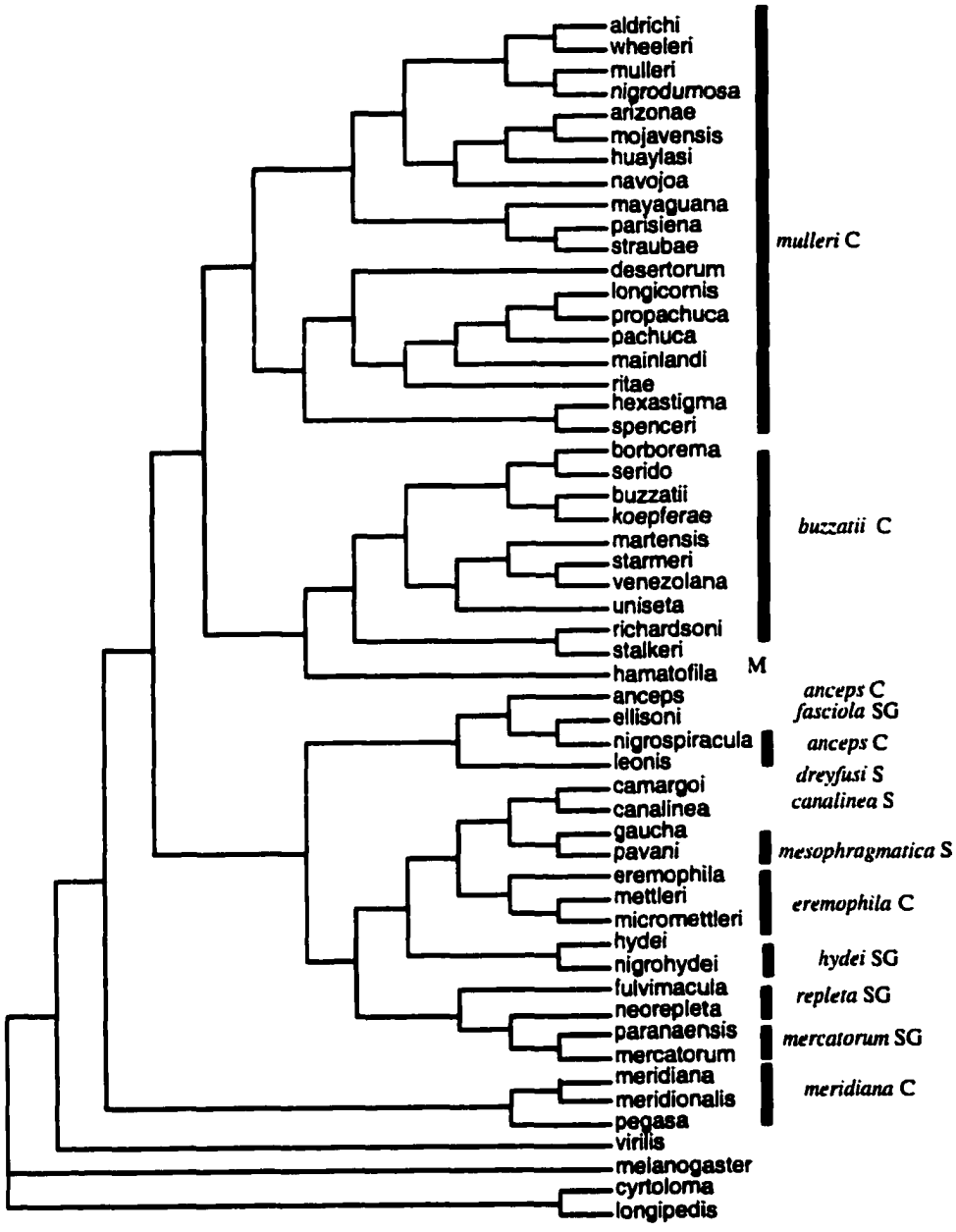


Figure 2-6. Tree generated using successive approximations weighting based on the retention index.

groups (*dreyfusi*, *canalinea* and *mesophragmatica*), arise from our study. This poorly resolved grouping (Figure 2-1) is the result of the consensus of two rather different arrangements of the taxa in this clade at the base of our tree. Although character weighting of the CO II partition slightly increases recovery of previously accepted monophyletic groups (Figure 2-5), successive weighting does not. Resolution in this area of the tree in an equal weighting framework will come from addition of more characters.

Analyses of the relative utility of the different partitions we used in this study indicate extreme saturation in the protein coding third positions and in transition changes, especially in the mitochondrial genes. These results suggest that mitochondrial protein coding regions are not adequate to resolve these basal relationships in an equal weighting analysis. On the other hand, addition of nuclear protein coding genes such as *hb* and Ef1α should be useful in deciphering relationships at the base of the phylogeny of this important group of drosophilids.

CHAPTER 3

ON CHROMOSOMES AND CROSSABILITY IN SEVERAL SPECIES GROUPS OF *DROSOPHILA*

Abstract

We place the classical *Drosophila* evolutionary biology tools of polytene chromosome inversion analysis and crossability tests in a molecular phylogenetic context. A simultaneous analysis framework was used to examine the congruence of the chromosomal inversion data with more recent DNA sequence data in four *Drosophila* species groups - the *melanogaster*, the *virilis*, the *repleta* and the Hawaiian *Drosophila* species groups. Polytene chromosome data are highly congruent with the DNA sequence data and when placed in a simultaneous analysis framework are shown to be more information rich than DNA sequence data. Crossability data can also be examined in a tree based context in these four species groups by examining phylogenetic distances that crossability can traverse. The comparisons we make allow us to detect aspects of crossability such as the importance of geographic affinity (sympatry and allopatry) that are characteristic of the species groups we include in this study.

Introduction

Species in the family Drosophilidae have been premier research subjects in evolutionary biology since Morgan first used *Drosophila melanogaster* as a genetic tool in the early part of the 20th Century. Classical approaches, such as polytene chromosome inversion analysis and studies of reproductive isolation via crossability experiments, have been commonplace in the examination of this family of flies. The chromosomal analyses have been used in two ways in evolutionary studies. The first use is as genetic markers (summarized in Powell, 1996; Powell and Krimbas, 1992) in which the chromosomal inversions are considered alleles and are utilized to examine gene flow and other population genetic parameters. The second use of polytene chromosomes in evolutionary studies is as tracers of phylogeny (Carson, 1972, 1987; Lemunier et al., 1986; Throckmorton, 1982; Wasserman, 1992). This approach has resulted in important and detailed chromosomal phylogenies for several groups of flies in the genus *Drosophila*. Lists of species for which polytene chromosome maps have been produced are available (Ashburner, 1989; Ashburner and Carson, 1983), and over 300 species of *Drosophila* have been examined at this level. In particular, chromosomal phylogenies for flies in the Hawaiian *Drosophila* (Carson, 1987), the *D. repleta* species group (Wasserman, 1992), the *D. virilis* species group (Throckmorton, 1982), and the *D. melanogaster* species group (Lemunier et al., 1986) exist. Cladistic analyses of the chromosomal data for these groups exist for the Hawaiian *Drosophila* (Kaneshiro et al., 1995) and the *melanogaster* species group (Lemunier and Ashburner, 1976). More recently, large amounts of DNA sequence information have been collected for these and many other species groups, yet no

detailed analysis of the overall utility of chromosomal inversion data or of their congruence with DNA sequence data has been done.

The ability to experimentally manipulate flies in this family also led to extensive examination both of crossability between different populations of single species of *Drosophila* and also of crossability between species. Crossability is defined as the ability of an intraspecific or an interspecific cross to produce offspring. There are different degrees of crossability that hinge on whether or not the offspring can reproduce or are sterile. Interspecific crosses have been done for two main reasons in the past. The first is to generate hybrids in polytene chromosome analysis and the second is to examine the interfertility of closely related forms in speciation studies. Crossability studies have shown that the ability of particular species to cross to one or more other species is the rule rather than the exception (Bock, 1984). Although crossability appears to be restricted to only closely related taxa, the degree of crossability and interfertility within and between species can be used as an indicator of reproductive isolation. Only minimal attempts at observing the phylogenetic distribution of crossability have been made (Powell and DeSalle, 1995) and this in only a single group of flies - the *D. obscura* species group.

The main objective of the present study is to place the chromosomal inversion data and the classical crossability data into a phylogenetic framework. With respect to chromosomes we examine the congruence of DNA sequence characters and chromosomal inversion data in four different species groups of Drosophilidae - the Hawaiian *Drosophila* species group, the *D. repleta* species group, the *D. melanogaster* species group and the *D. virilis* species group. Our analysis allows us to assess the relative contribution to phylogenetic hypotheses that the two different sources of character information make. With respect to the

crossability data we examine the average phylogenetic distance of crossable and uncrossable species pairs in the same four species groups used in our comparisons of polytene chromosomal inversion data to DNA sequence data. We also examine the impact of sympatry and allopatry on the phylogenetic distribution of crossability.

Materials and Methods

Data matrices: Four data matrices were constructed using DNA sequences and chromosomal inversion data from the literature (Table 1). The four species groups for which we have obtained inversion data represent the four major species groups for which such data exist. Chromosome inversion information for other smaller groups such as the *Antopocerus* species group (Hawaiian *Drosophila*) is available (Yoon and Richardson, 1976), however DNA sequence data were not obtainable for these groups. The chromosome inversion data for the *obscura* species group are not well interpreted (Lakovaara and Saura, 1982) and have been omitted from our analyses. The number of characters for each character partition is given in Table 2. Crossability data were also compiled for the same four species groups. Bock (1984) summarizes crossability in the Drosophilidae, and a quick survey of his review reveals that the four species groups upon which we focus in this paper are the ones with the greatest amount of crossability information in the family. The *cardini* and *obscura* groups are also extensively sampled for crossability, but DNA sequence information is not available for the former and the latter have been examined for crossability in a phylogenetic context in another publication (Powell and DeSalle, 1995).

Phylogenetic Trees: Phylogenetic trees were generated using PAUP 4.01b (Swofford, 1999). In order to assess the relative contribution of chromosomal

Table 3-1. Data Sources for the four data matrices used in this study. Yp=yolk protein, Adh=alcohol dehydrogenase, Sry=serendipity, nullo=nullo, 28S=28S rDNA, 5S=5S rDNA, Amy=α-amylase, g6pd=glucose-6-phosphate dehydrogenase, Co II=cytochrome oxidase subunit II, 16S=16S rDNA, hb=hunchback, ND1= NADH dehydrogenase subunit 1, bp=base pairs, Chromo=chromosome.

Species Group	DNA	No. of bp	Ref	Chromo Ref	Crossability Ref
Hawaiian *Drosophila*	yp	849	1	18,19	23
melanogaster	Adh	771	2,3,4,5,6,7	20	20
	Sry	1498	8		
	nullo	566	8		
	28S	342	9,10		
	5S	187	11,12		
	Amy	1568	13		

Table 3-1 (continued). Data Sources for the five data matrices used in this study. Yp=yolk protein, Adh=alcohol dehydrogenase, Sry=serendipity, nullo=nullo, 28S=28S rDNA, 5S=5S rDNA, Amy=α-amylase, g6pd=glucose-6-phosphate dehydrogenase, Co II=cytochrome oxidase subunit II, 16S=16S rDNA, hb=hunchback, ND1= NADH dehydrogenase subunit 1, bp=base pairs, Chromo=chromosome.

virilis	g6pd	1063	14,15	21	21
	Adh	988	16		
repleta	CO II	442	17	22	22,24,25
	16S	521	17		
	hb	526	17		
	ND 1	129	17		

Table 3-1 (continued).

1=Kambesyllis et al.(1995) 2=Coyne&Kreitman(1986) 3=Cohn&Moore(1988) 4=Ashburner (1990) 5=Moses et al.(1990) 6=McDonald&Kreitman(1991) 7=Jeffs et al.(1994) 8=Caccone et al.(1996) 9=Ruiz et al.(1991) 10=Pelandakis&Solignac(1993) 11=Samson&Wegnez(1984) 12=Paques et al.(1995) 13=Shibata &Yamazaki(1995) 14=Tominaga et al.(1995) 15=Kitagawa (1998) 16=Nurminsky et al.(1996) 17=Durando et al.(2000) 18=Carson(1972) 19=Gillespie et al.(1996) 20=Lemunier et al.(1986) 21=Throckmorton(1982) 22=Wasserman(1992) 23=Yang&Wheeler(1969) 24=Patterson(1947) 25=Marin et al.(1993)

inversion data and DNA sequence data we placed our analysis in a simultaneous analysis framework (Nixon and Carpenter, 1995; Kluge, 1989; Brower et al., 1996). Bootstrap values were computed using PAUP 4.01. Decay indices were computed using AUTODECAY (Erikson, 1997) and using the methods described in Baker et al. (1998). Significance of Incongruence Length Differences (ILDs; Farris et al., 1994; 1995) were calculated using the Partition Homogeneity Option in PAUP 4.01. Partitioned Bremer supports for the inversion partition and the DNA partition were calculated using the approaches outlined in Baker et al., (1998).

Examination of crossability in a phylogenetic context:: We examined the extent of crossability in a phylogenetic context in two ways. First, we obtained a parsimony tree for each of the four species groups examined in this study and used MacClade to display information on crossabilities (described below) on the tree. The mode of character reconstruction (ACCTRAN, DELTRAN, MINMAX) is not relevant here as we examined and used only the distribution of character states at the terminals. We then counted the number of nodes that were traversed for species that were crossable (see below) and that were uncrossable (see below). For instance, if sister taxa were crossable, then the number of nodes traversed for that particular cross was one node. If one of the sister taxa and its next closest relative were crossable, then the number of nodes traversed for that cross was one node (the node traversed leading to the sister taxon) plus one node (the single node traversed to the nearest taxon from the species that is being tested in the cross). See Figure 1. The average number of nodes that were crossed for both crossable and uncrossable classes was then calculated. Statistically significant differences of these averages were determined first by using F-tests to show whether or not the variances were equal and then by applying the appropriate t-test to determine whether or not the averages

maur-maur = 0 nodes
maur-sech = 1 nodes
maur-sim = 2 nodes

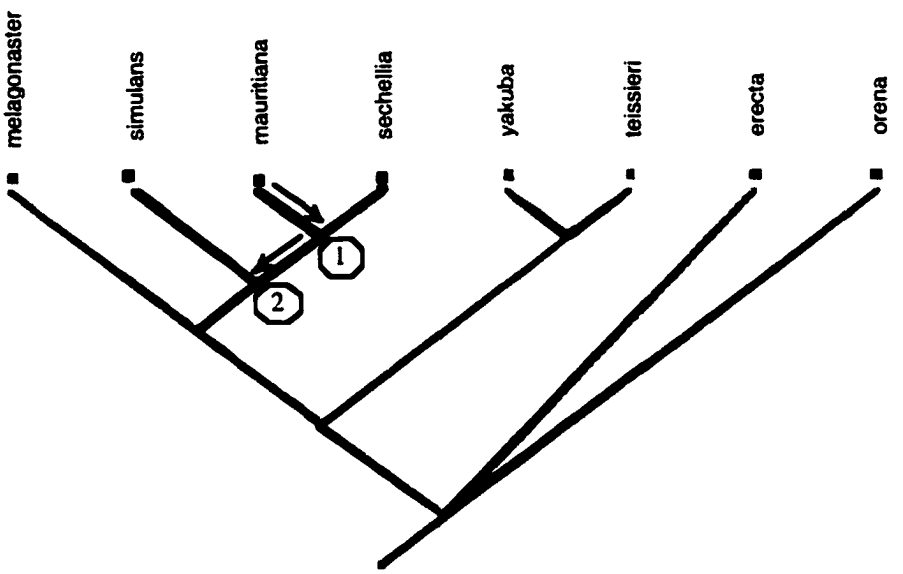


Figure 3-1. Method we used for counting number of nodes between species pairs. maur=mauritiana, sech=sechellia, sim=simulans. Numbers in octagons=number of nodes counted. Between D. mauritiana and itself no nodes are traversed. Between D. mauritiana and D. sechellia one node is traversed. Between D. mauritiana and D. simulans two nodes are traversed.

were significantly different. The second way of estimating extent of crossability was, once again, to obtain a parsimony tree for each of the four species groups and to use MacClade to display the crossabilities as scored above on the tree. We then used patristic distances calculated for that tree topology to determine the distance between crossable and uncrossable taxa on the tree. The average patristic distance that a particular class of crosses traversed was then calculated, and statistically significant differences were determined as above.

Data from crossability tables in the publications listed in Table 3-1 were mapped onto the terminals in the simultaneous analysis trees shown in Figures 3-2, 3-3, 3-4 and 3-5. In essence there are four major easily interpreted classes of observation that can be obtained from the crossability tables listed in Table 3-1. The first kind of observation is that a cross produces both fertile male and fertile female offspring. This class of crosses we called "fully crossable". The second class of crosses are those that produce either fertile males and sterile females or fertile females and sterile males, and we called these crosses "partially crossable". The third kind of cross produces viable, but sterile offspring. This type of crosses we called "crossable/sterile". The fourth kind of crosses are those that produce no offspring whatsoever, and we called these "fully uncrossable". The ability to pass genes between populations is most important in the evolutionary process. Even if only a small amount of genetic material is passed, this can be critical in determining whether two populations can diverge and speciate. Therefore, we reduced our four classes to two, using the ability to produce fertile offspring or its lack as criteria. Accordingly, we placed the "fully crossable" matings with the "partially crossable" matings and lumped the crosses that produced "crossable/sterile" offspring with the "fully uncrossable" matings. This approach emphasizes fertility as a factor in

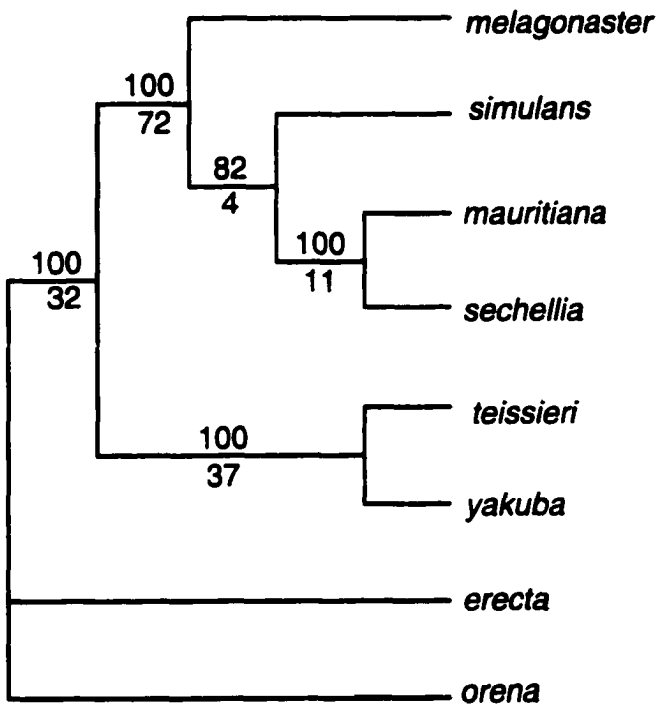


Figure 3-2. Cladograms showing the total evidence hypotheses for combined analyses of DNA sequences and chromosome for the melanogaster species group. The numbers above nodes indicates the bootstrap value and the number below indicates the Bremer index. < indicates a bootstrap value less than 50%.

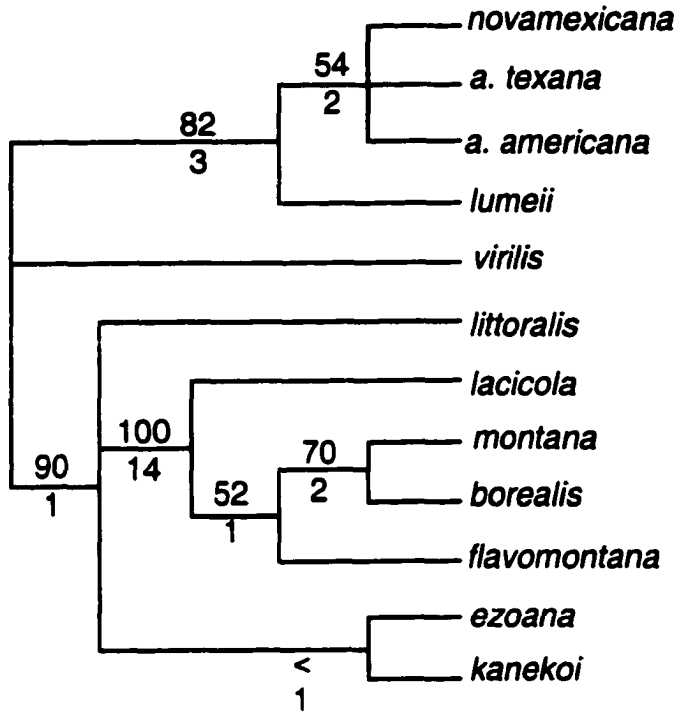


Figure 3-3. Cladograms showing the total evidence hypotheses for combined analyses of DNA sequences and chromosomes for the virilis species group. The numbers above nodes indicates the bootstrap value and the number below indicates the Bremer index. < indicates a bootstrap value less than 50%.

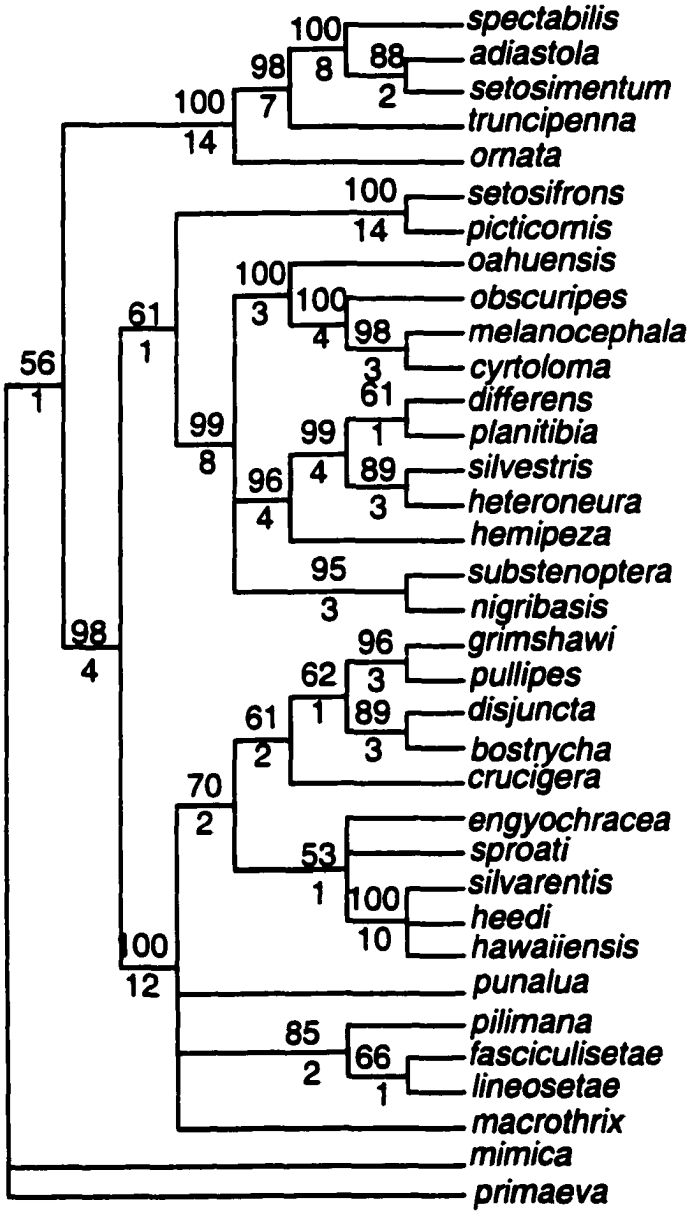


Figure 3-4. Cladograms showing the total evidence hypotheses for combined analyses of DNA sequences and chromosomes for the Hawaiian Drosophila species group. The numbers above nodes indicates the bootstrap value and the number below indicates the Bremer index. < indicates a bootstrap value less than 50%.

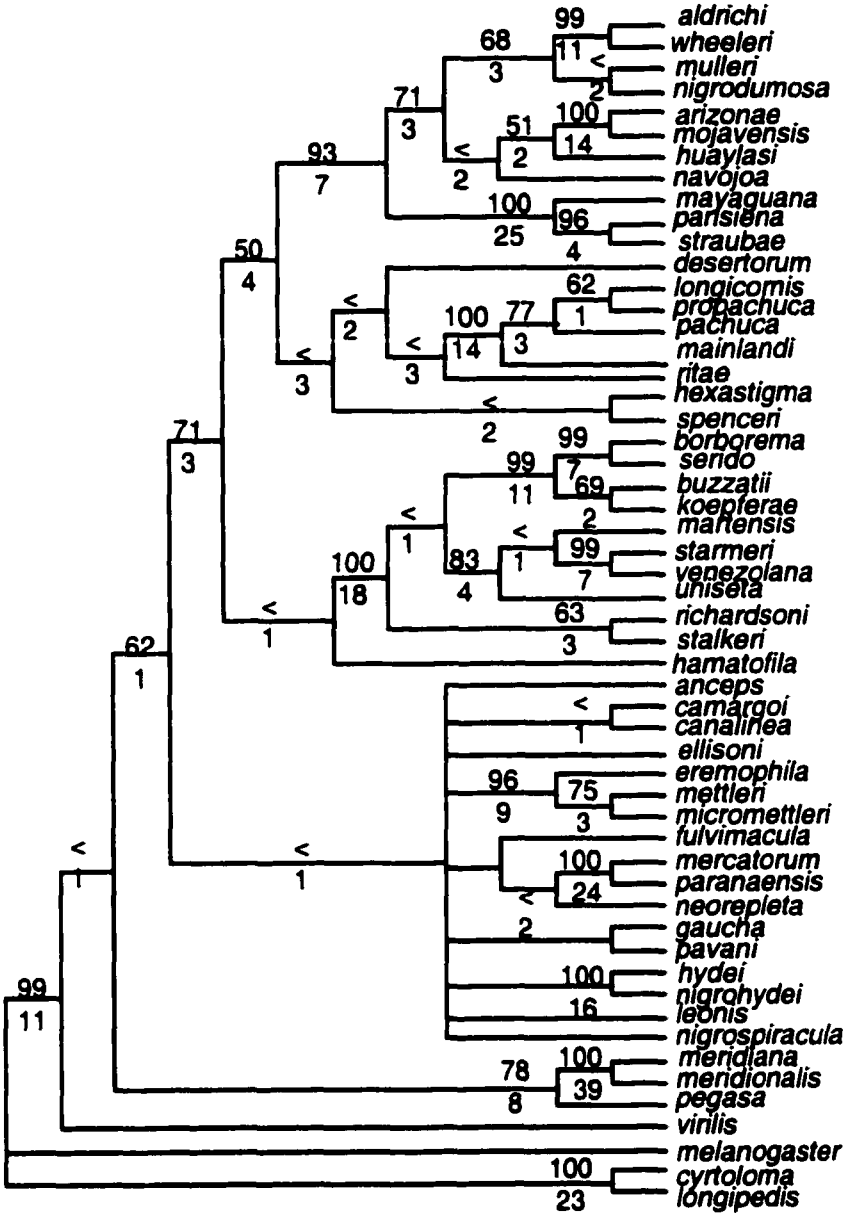


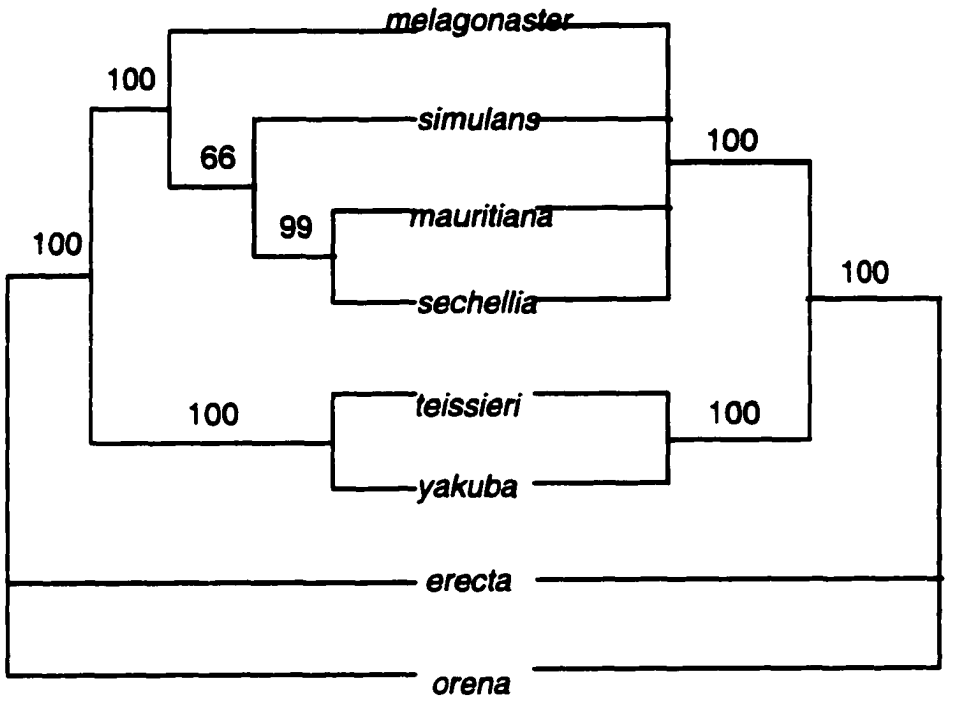
Figure 3-5. Cladograms showing the total evidence hypotheses for combined analyses of DNA sequences and chromosomes for the repleta species group. The numbers above nodes indicates the bootstrap value and the number below indicates the Bremer index. < indicates a bootstrap value less than 50%.

crossability and fertility of offspring as an aspect of isolation between species. No correlation of these kinds of crosses with a specific mating isolation mechanism,such as a premating or a postmating mechanism, can be made using the crossability tables since observations of copulations, which would be necessary to discriminate between pre- and post-mating mechanisms, are not usually included in the crossability tables (Table 3-1). Information on the sympatry or allopatry for each species pair where crossability data was available was also collected from the papers listed in Table 3-1. We used the sympatry and allopatry data to further subdivide crosses to examine for phylogenetic effects on crossability as a result of sympatry.

Results and Discussion

Chromosomal inversions

Table 3-2 shows the results of phylogenetic analysis of the inversion and DNA sequence character partitions separately and in combination. Tree topologies of the chromosomal and DNA sequence character partitions were very similar (Figures 3-6, 3-7, 3-8 and 3-9). The major difference upon direct examination of the molecular and inversion cladograms in Figures 3-6, 3-7, 3-8 and 3-9 was the lack of resolution for inversions compared to DNA sequences. The similarity in topology of these trees was borne out by the ILD analyses (HD=0.007 [NS], mel=0.000 [NS], repl=0.022 [NS] and vir=0.046 [NS], where none of the ILD measures was greater than 0.05. These results indicated that there was less than a 5% increase in length of the simultaneous analysis cladogram due to combining the DNA and inversion partitions. Figures 3-2, 3-3, 3-4 and 3-5 show the simultaneous analysis of the four data sets with Bremer support indices and bootstrap values on nodes.



DNA Sequences Inversions

Figure 3-6. Cladograms as described in the text for the melanogaster species group Numbers above branches indicate the bootstrap values for the nodes to the right of the number. The trees on the left are for DNA sequences only and the trees on the right are for chromosomal inversion data.

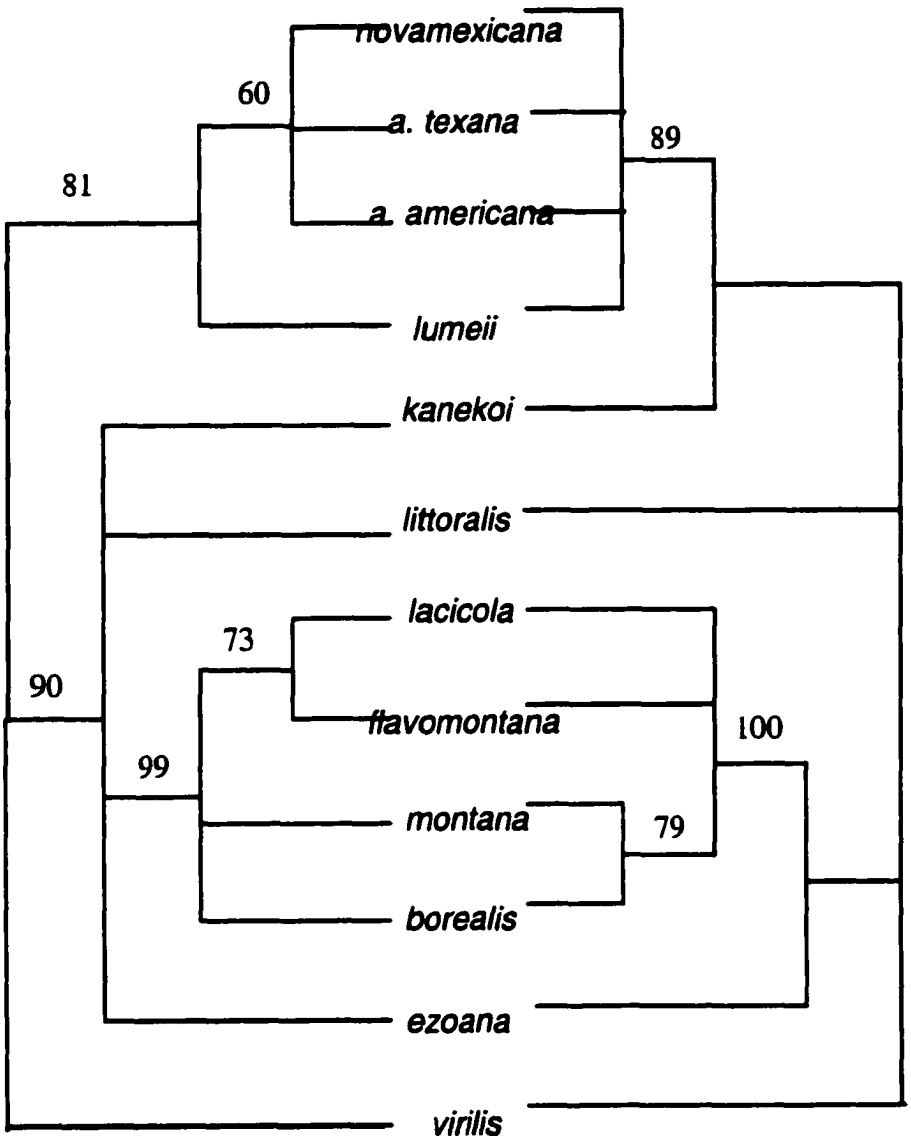
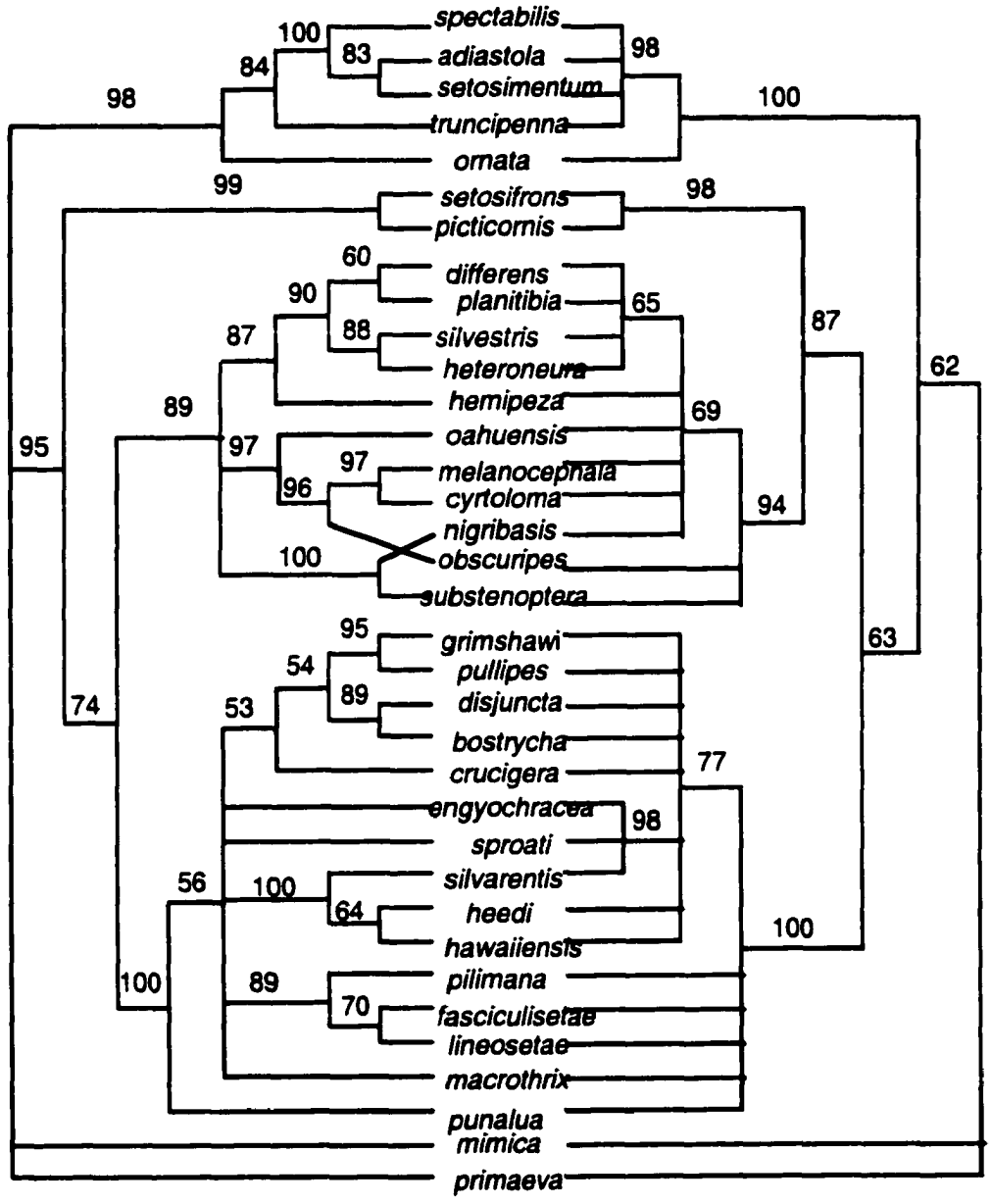


Figure 3-7. Cladograms as described in the text for the virilis species group Numbers above branches indicate the bootstrap values for the nodes to the right of the number. The trees on the left are for DNA sequences only and the trees on the right are for chromosomal inversion data.



DNA Sequences Inversions

Figure 3-8. Cladograms as described in the text for the Hawaiian Drosophila species group. Numbers above branches indicate the bootstrap values for the nodes to the right of the number. The trees on the left are for DNA sequences only and the trees on the right are for chromosomal inversion data.

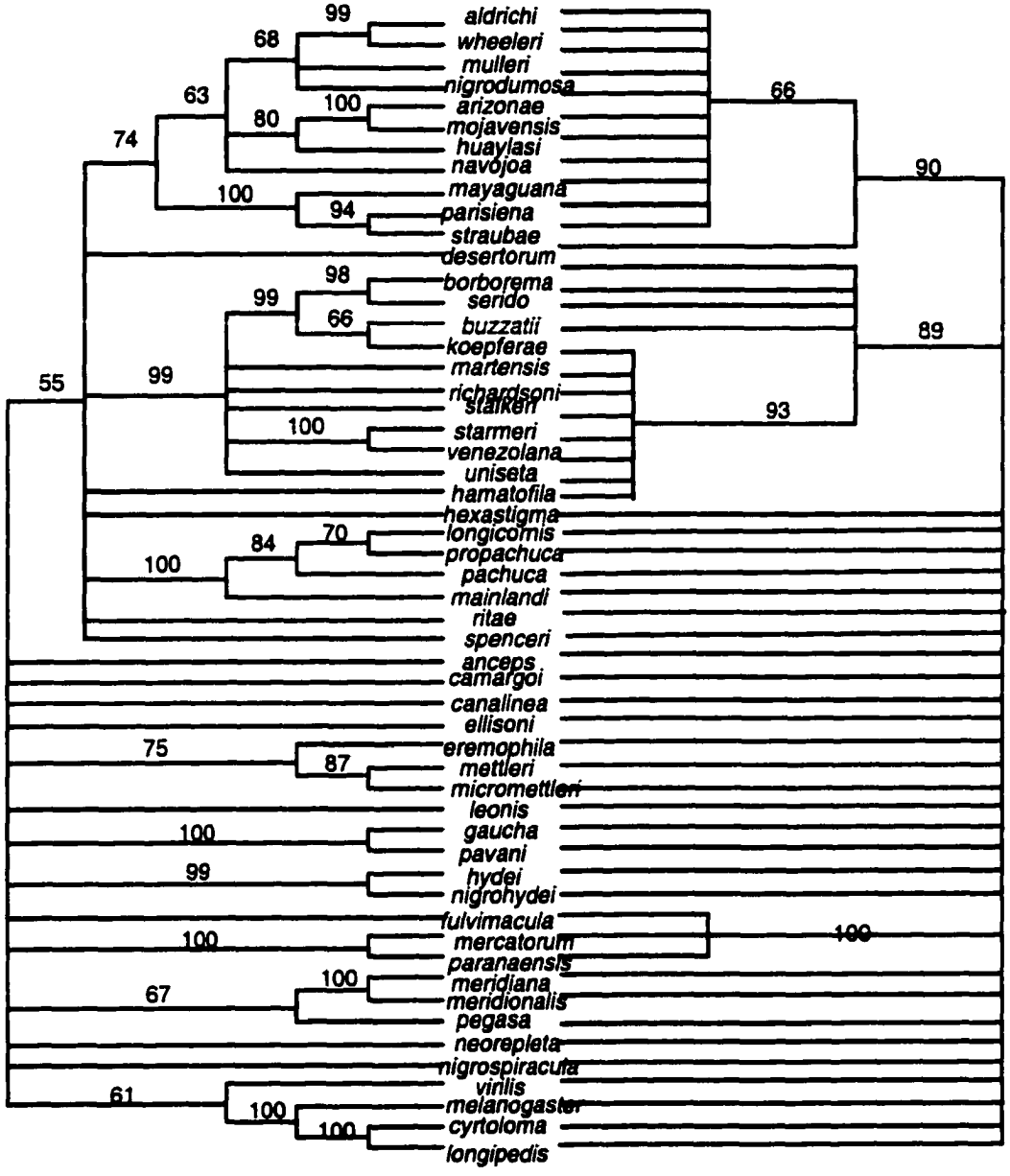


Figure 3-9. Cladograms as described in the text for the repleta species group. Numbers above branches indicate the bootstrap values for the nodes to the right of the number. The trees on the left are for DNA sequences only and the t rees on the right are for chromosomal inversion data.

In general, the agreement of chromosomal inversion topology with DNA sequence topology was extremely good. The number of nodes in the trees that disagreed (as assessed by a negative partitioned Bremer support value) for both data partitions was extremely low (see the " # of - nodes" column in Table 3-3). In all species groups examined here there was at least one node that was negative for partitioned Bremer support of the molecular partition, indicating that the molecular data are in conflict with the simultaneous analysis hypothesis for these nodes. Both the *melanogaster* and the *virilis* group chromsomal data showed complete agreement with the simultaneous analysis tree, while three nodes in the *repleta* group tree and one node in the Hawaiian *Drosophila* tree had negative partitioned Bremer supports for the inversion partition. While there were many nodes that had zero partitioned Bremer support for the inversion partition (see the "# of 0 nodes" column in Table 3), the total support rendered by the inversion data to the simultaneous analysis tree was relatively large ("total BS" column in Table 3-3). To standardize the total partitioned Bremer support contribution of each partition to the simultaneous analysis tree we divided the total Bremer support by the number of phylogentically informative characters in that character partition and by the minimum steps of the simultaneous analysis tree ("corrected total BS" column in Table 3-3). Using both kinds of correction (Table 3-3) resulted in inversions showing higher corrected values relative to the DNA sequence characters in seven out of eight comparisons. This result is most likely due to the considerably higher consistency of the chromosomal inversion data. Only when the total Bremer support values of both the inversion data and the molecular data were standardized by the minimum steps of the simultaneous analysis for the *repleta* group did we find that both data partitions contributed equally to the simultaneous analysis tree.

Table 3-3. Results of partitioning Bremer support to molecular and inversion character sources. Each entry in the table is listed as values calculated for inversions / molecular data. The number of nodes (# of nodes), the number of negative nodes (#of - nodes) and the # number of zero plus negative nodes (# of 0 nodes) are given in the first three columns. "% nodes 0 or -" indicates the percent of total nodes that are either zero or negative Bremer support values. "Total BS" is the total Bremer support summed over all nodes in the cladogram. Corrected total BS is the corrected total Bremer support where Bremer support has been divided by the number of informative characters in the character partition and by the minimum steps for each partition (see text).

SET	# of nodes	# of - nodes	# of 0 nodes	% nodes 0 or -	total BS	corrected total BS by # of. PI chars	corrected total BS by min. # steps
HD	32	1/1	21/6	0.72/0.22	34/91	0.98 /0.63	0.37/0.29
Mel	5	0/1	2/0	0.4/0.0	47/109	1.3 /0.5	0.77/0.17
Repl	44	3/1	25/3	0.57/0.09	30.5/269.5	1.6 /0.6	0.37/0.37
Vir	9	0/1	4/4	0.44/0.55	11/12	0.5 /0.2	0.16/0.06

We also computed the consistency indices of the inversion and DNA partitions for the four data matrices using the simultaneous analysis trees for each as a constraint. The consistency indices for the inversion partitions were considerably higher than those for the DNA partitions. Previous surveys of the patterns and distributions of consistency indices with taxon number indicate that in general the CI decreases with the number of taxa in a study (Sanderson and Donoghue, 1992). Figure 10 shows a plot of the CI versus number of taxa for both the inversion and the DNA partitions for all four data matrices. This figure demonstrates that inversions were highly consistent in all four cladograms and that they did not show the characteristic lowering of consistency index with number of taxa that most character data show. However, the DNA partitions did show the depression of consistency index value with number of taxa. Together these analyses suggest that there is a high degree of agreement among classical chromosomal data and more recent DNA sequence data and that inversion data are extremely consistent with simultaneous analysis hypotheses of relationship in these groups of *Drosophila*.

Tests concerning crossability:

Several authors have examined crossability in various species groups of the Drosophilidae (reviewed by Bock, 1984). In addition, the ability to cross among and between *Drosophila* species has been examined from a distance based perspective (Coyne and Orr, 1989; 1997). Our analysis took a tree topology based approach to crossability. As explained above we scored crossability as present only in those crosses in which fertile offspring were produced for either males or females or both, emphasizing fertility of offspring as a determinant of crossability.

Since most of the studies that examined crossability in the past have archived information about the male or female direction of the crosses (that is, they

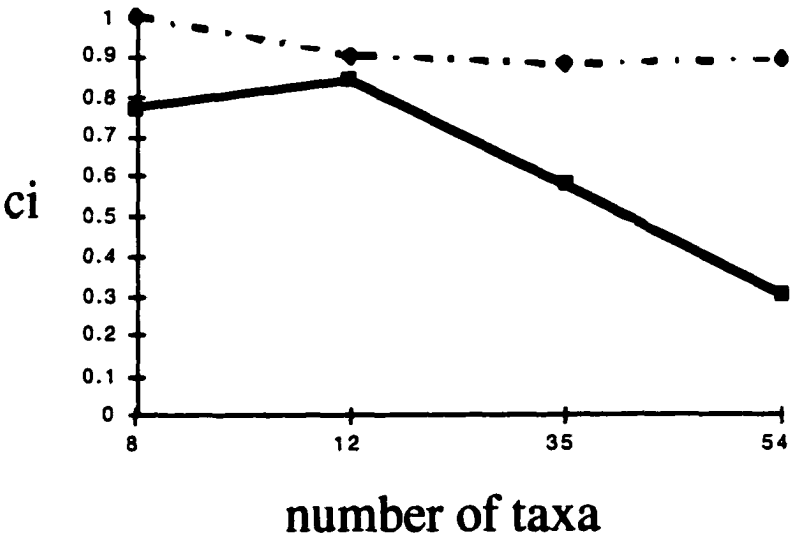


Figure 10. Plot of the consistency index of chromosomal inversion partition (dotted line) and the DNA sequence partition (solid line) when forced to fit the parsimony tree versus the number of taxa in the data set. The melanogaster species group analysis has eight taxa, the virilis species group analysis has 12 taxa, the Hawaiian Drosophila species group has 35 taxa and the repleta species group has 54 taxa.

have compiled data regarding the results of attempts to mate males of one species with females of various other species and to mate females of one species with males of various other species), we first tested whether there were differences between male and female patterns for crossability. This was done by simply asking if the patterns of crossability in the male direction were different from those in the female direction using the same node counting methods outlined above. All of the averages for comparisons of node crossing and patristic distance analysis were not significantly different (Table 3-4) for the male and female comparisons as determined by first performing F-tests to ascertain whether the variances were equal or unequal and then applying the appropriate t-tests of the means. Consequently, we present results for only the female crosses in the remainder of this paper.

Next we examined the patterns of crossability among these taxa. If cladogenesis results in complete reproductive isolation, then the general trend should be that the crossability of a species should not extend past the nearest node of that species in a phylogenetic tree. Previous authors have recognized the ability to cross among many of these taxa as being restricted to only closely related taxa (Bock, 1984). While this suggestion is based on the relationships of these flies, no formal attempts at quantifying this general perception have been made using a tree based approach. Our approach allowed us to quantify how extensive Bock's (1984) observation is in these flies. We counted the number of nodes traversed and calculated the patristic distance between two species where crossability occurs. In a similar fashion we counted nodes and summed patristic distances for species pairs that were uncrossable. For each of the four groups we determined whether or not the phylogenetic distribution of crossability was significantly different from that of uncrossability ($p. < .001$) by, again, performing F-tests of the variances and then applying the appropriate t-tests of the means from number of nodes crossed and

Table 3-4. Results obtained from t-test comparisons for direction of crosses. F = female, M = male, PAT. DIST. = Patristic distance, FERT. = fully or partially fertile, STER. = sterile, VAR = variance, d.f. = degrees of freedom, SIG. = significance.

	FvM NODES		FvM PAT. DIST.	
	FERT. F_1	STER. OR NO F_1	FERT. F_1	STER. OR NO F_1
MELANOGASTER				
VAR (F-TEST)	EQUAL	EQUAL	UNEQUAL	EQUAL
d.f. (T-TEST)	25	83	14	83
SIG. (T-TEST)	NS	NS	NS	NS
VIRILIS				
VAR (F-TEST)	EQUAL	EQUAL	EQUAL	EQUAL
d.f. (T-TEST)	90	110	90	110
SIG. (T-TEST)	NS	NS	NS	NS
HAWAIIAN *Drosophila*				
VAR (F-TEST)	UNEQUAL	EQUAL	UNEQUAL	EQUAL
d.f. (T-TEST)	39	232	36	232
SIG. (T-TEST)	NS	NS	NS	NS

patristic distance traversed between taxa. Figure 3-11 shows the average number of nodes and patristic distances over which crossability extends for the four species groups. In all comparisons the number of nodes and the patristic distances traversed by crossable taxa were significantly less than those traversed by uncrossable taxa (Table 3-5). The *repleta* group showed the greatest extent of crossability with an average of 2.2 nodes spanned by crossable taxa, while the *melanogster* group exhibited the lowest average number of nodes spanned at 0.8 nodes. Figure 3-11 shows the general tendency in all of the species groups for crossability to be restricted to only very close relatives.

The effects of sympatry and allopatry on crossability

Another concern that can be addressed with the phylogenetic approach is the role of geographic affinity of species in the patterns of crossability. We can ask if crossability is enhanced or depressed by sympatry of taxa in the phylogenetic trees we constructed. Geographic affinity of species in the divergence process is important because one might expect crossability to be severely depressed in sympatry relative to the ability to cross in allopatry. Ideas about reinforcement that suggest that hybridization of species in sympatry should be selected against are the basis for this expectation. In other words, the number of nodes and patristic distances traversed for sympatric comparisons should be significantly lower than those for allopatric comparisons. We approached this question by asking if crossable species in sympatry traversed more nodes or greater patristic distances than those in allopatry. Once again F-tests were performed to determine equality or inequality of variances and the appropriate t-tests were applied to the means. Figure 3-12 and Table 3-6 show that in the *melanogaster, virilis* and *repleta* species groups the number of nodes traversed in sympatric species pairs was significantly

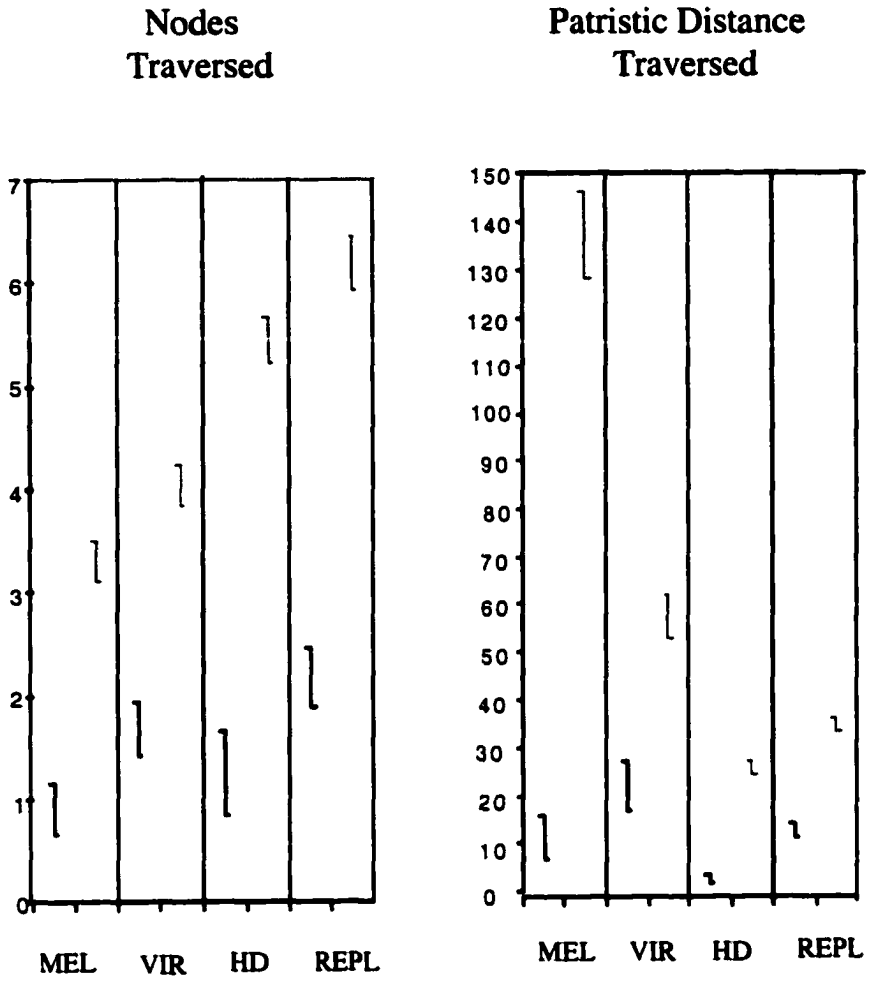


Figure 3-11. Plots showing the 95% confidence ranges for the number of nodes traversed and for patristic distances traversed in comparisons made between "crossable" (heavy lines) and "uncrossable" (light lines) species pairs. MEL=melanogaster, VIR=virilis, HD=Hawaiian Drosophila, REPL=repleta.

Table 3-5. Results obtained from t-test comparisons for fertile F_1 or sterile or no F_1 in four species groups of Drosophilidae. PAT. DIST. = Patristic distance, FERT. = fully or partially fertile, STER. = sterile, VAR = variance, d.f. = degrees of freedom, SIG. = significance.

	FERT. F_1 v STER. OR NO F_1 - NODES	FERT. F_1 v STER. OR NO F_1 - PAT.DIST.
MELANOGASTER		
VAR(F-TEST)	EQUAL	UNEQUAL
d.f.(T-TEST)	55	54
SIG.(T-TEST)	<.001	<.001
VIRILIS		
VAR(F-TEST)	EQUAL	EQUAL
d.f.(T-TEST)	100	100
SIG.(T-TEST)	<.001	<.001
HAWAIIAN *Drosophila*		
VAR(F-TEST)	EQUAL	UNEQUAL
d.f.(T-TEST)	128	116

Table 3-5 continued.

SIG.(T-TEST)	<.001	<.001
REPLETA		
VAR(F-TEST)	EQUAL	UNEQUAL
d.f.(T-TEST)	164	155
SIG.(T-TEST)	<.001	<.001

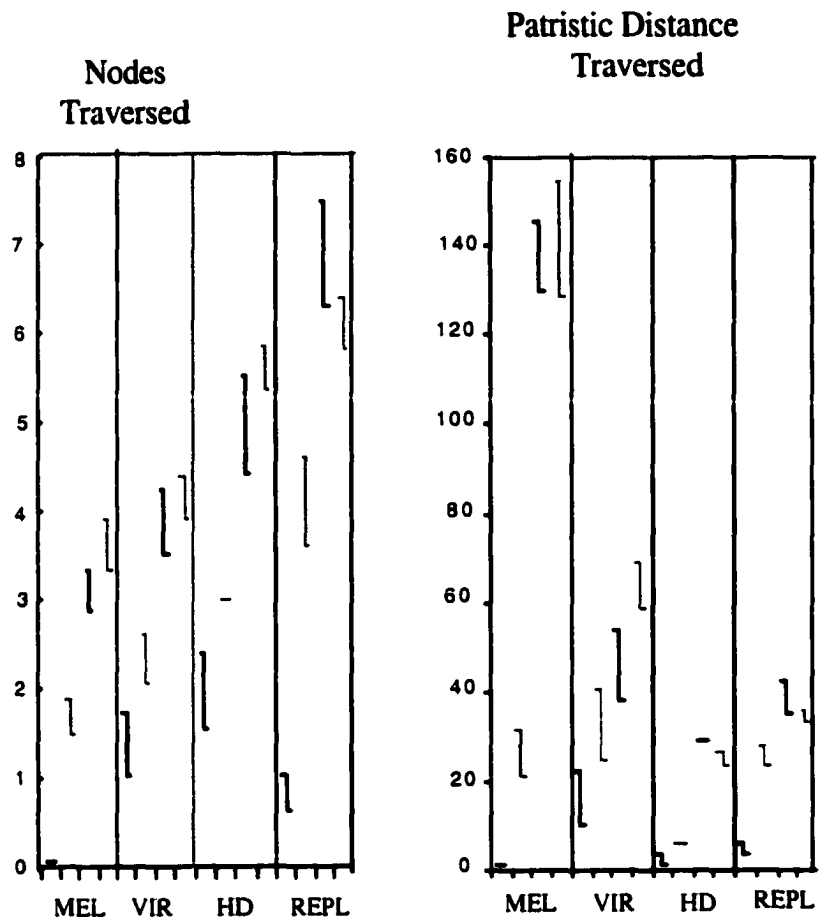


Figure 3-12. Plots showing the 95% confidence ranges for the number of nodes traversed and for partistic distances traversed in comparisons made between sympatric (heavy lines) and allopatric (light lines) crosses. The first two plots in each of the four boxes are for comparisons where the species were crossable and the last two are for uncrossable comparisons. MEL = melanogaster species group, VIR = virilis species, HD = Hawaiian Drosophila species, REPL = repleta species group.

Table 3-6. Results obtained from t-test comparisons for sympatry and allopatry comparisons. S = sympatric, A = allopatric, PAT. DIST. = Patristic distance, FERT. = fully or partially fertile, STER. = sterile, VAR = variance, d.f. = degrees of freedom, SIG. = significance.

	S v A	NODES	S v A	PAT. DIST
	FERT. F_1	STER. OR NO F_1	FERT. F_1	STER. OR NO F_1
MELANOGASTER				
VAR(F-TEST)	EQUAL	EQUAL	UNEQUAL	EQUAL
d.f.(T-TEST)	12	41	5	41
SIG.(T-TEST)	<.001	NS	<.005	NS
VIRILIS				
VAR(F-TEST)	UNEQUAL	EQUAL	EQUAL	EQUAL
d.f.(T-TEST)	33	64	34	64
SIG.(T-TEST)	<.05	NS	NS	NS
HAWAIIAN *Drosophila*				
VAR(F-TEST)	EQUAL	EQUAL	EQUAL	EQUAL

Table 3-6 continued.

d.f.(T-TEST)	22	103	22	103
SIG.(T-TEST)	NS	NS	NS	NS
REPLETA				
VAR(F-TEST)	UNEQUAL	EQUAL	UNEQUAL	EQUAL
d.f.(T-TEST)	44	79	52	79
SIG.(T-TEST)	<.001	NS	<.001	NS

smaller than the number of nodes traversed in allopatric pairs ($p<.001$ for all three species groups), supporting the general idea that crossability should be lower in sympatric pairs. Figure 3-12 and Table 3-6 also show that the Hawaiian *Drosophila* species group was the only species group in which the number of nodes crossed for both sympatric and allopatric crossable species was not significantly different. For patristic distances the results showed a significant difference for both sympatric and allopatric crossable species for the *melanogaster* ($p<.005$) and *repleta* ($p<.001$) species groups. However, no significant difference was found for patristic distance values between sympatric and allopatric crossable species for either the *virilis* or the Hawaiian *Drosophila* species groups. The lack of significant difference between crossable species pairs in sympatry and allopatry exhibited by the Hawaiian *Drosophila* species group in both number of nodes traversed and patristic distance traversed is probably due to the fact that the sample size was not adequate to test for statistical significance in this species group. However, when one looks at Figure 6, a discernible trend emerges for all groups: species pairs that are crossable traverse shorter phylogenetic distances in sympatry than in allopatry. In addition, it is of note that uncrossable species pairs, which can be thought of as controls in this analysis, reveal no significant differences whether in sympatry or in allopatry. This result is as one would expect since absence of the ability to produce fertile offspring should be randomly distributed among sympatry and allopatry.

Conclusion

Classical *Drosophila* studies have used chromosmal inversions and crossability of taxa to understand phylogeny and speciation. More recent DNA sequence information can be combined with these classical data to make inferences about phylogeny and species divergence. The approach we have taken here is to

combine the chromosomal inversion data with DNA sequence data to examine some of the classical notions of *Drosophila* evolution. Our results suggest that there is a great deal of congruence among DNA sequence data and chromosomal inversion data. Although this result is reassuring, there are still some areas of the phylogenies of the four species groups examined here that are not congruent. These areas are indicative of poor phylogenetic signal from one or both of the kinds of data - DNA and chromosomal inversion data. Chromosomal inversion data are much more information rich as assessed in the present study, and this is probably due to the higher consistency of these characters.

Crossability studies in the Drosophilidae have contributed to our understanding of the divergence process in these flies. Due to the ease with which these experiments can be performed in the laboratory, a great number of interspecific crosses have been attempted by researchers (Bock, 1984). We place these crossability results in a phylogenetic context in this study in order to examine the role of phylogenetic relatedness to the crossability patterns of these species groups. We are able to examine two possible effects of phylogeny on crossability - male or female asymmetry and geographic affinity. We conclude the following: 1) at this level of examination there is no difference in the direction of the cross on the phylogenetic approach we have taken. Results for male crosses are statistically non-significant from results for females. This observation can be contrasted with lower level comparisons (Coyne and Orr, 1989; 1997) where asymmetries of mating relationships are observed with respect to males and females. Our higher level approach that examines patterns within and across species groups is coarse grained and obscures the fine grained observations on mating asymmetries made in other studies. 2) In agreement with common notions (Bock, 1984), more closely related taxa are more likely to cross than are distantly related taxa. While this

observation may at first seem trivial, we have been able to characterize and confirm this general qualitative statement in the four species groups we examined within a phylogenetic context. 3) Geographic affinity appears to be an important factor in the divergence of species in all four species groups. One might be expected to find that hybridization of species in sympatry is selected against, thus reinforcing reproductive isolation.

CHAPTER 4

RELATIONSHIPS AMONG AND WITHIN THE SPECIES OF THE *DROSOPHILA MAYAGUANA* SUBCLUSTER: A MOLECULAR APPROACH

Introduction

The *mayaguana* species subcluster, consisting of the species *Drosophila mayaguana* (Vilela, 1983); *D. straubae* (Heed and Grimaldi, 1991); and *D. parisiena* (Heed and Grimaldi, 1991), is part of the *mulleri* species cluster of the *D. repleta* species group. The *repleta* species group, part of the *D. virilis-repleta* radiation (Throckmorton, 1982), is one of the largest species groups in the genus *Drosophila*, consisting of 89 described and 2 undescribed species.

D. mayaguana is found throughout the Caribbean region on Mayaguana Island, Concepcion Island, Great Inagua, Grand Cayman, Jamaica, Hispaniola and Tortola. *D. straubae* occurs on Hispaniola, Cuba and the small island of Navassa. *D. parisiena* is found on Hispaniola and Cuba together with *D. straubae* and also exists in Jamaica (Wasserman, 1982a, 1992; Heed and Grimaldi, 1991; Wasserman and Wasserman, 1992). *D. mayaguana* is thought to be ancestral to the other two species in the subcluster. Like most of the other species in the *mulleri* cluster, it is homosequential - that is, it is monomorphic for identical gene orders - and is closely associated with the cactus, *Opuntia*. However, *D. straubae* and *D. parisiena* are polymorphic for newer inversions (Wasserman, 1992; Wasserman and Wasserman, 1992). While these two species are difficult to identify based on morphological characters, they are both cytologically and ecologically distinct. The more derived *D. parisiena* is found on the columnar cactus, *Stenocereus*, while *D. straubae* breeds on *Opuntia*, *Cephalocereus* and *Pelocereus* (Heed and Grimaldi, 1991; Wasserman and Wasserman, 1992). Nonetheless, Wasserman (1992; Wasserman and Wasserman, 1992) has shown that crosses between these two species produce not only fertile F_1 females, but also fertile F_1 males in both directions. It is of great interest to note that populations of *D. parisiena* found in

Jamaica show noticeable differences in morphology from those found elsewhere (Heed and Grimaldi, 1991). This accords well with the evidence from allozyme electrophoresis done by Heed et al. (1990) in which the two Jamaican populations of *D. parisiena* form a monophyletic sister group to the other *D. parisiena* populations. In addition, Heed and Grimaldi (1991) found several pronounced morphological differences between the Jamaican *D. parisiena* and those found elsewhere, such as a greater number of setae on the epandrial lobes and fewer and less prominent spicules and scales on the distiphallus. Furthermore, Wasserman and Wasserman (1992) have found that the Jamaican populations of *D. parisiena* exhibit a great deal of cytological difference from the populations of *D. parisiena* found on Hispaniola. The Jamaican population is homozygous for a new terminally situated X chromosome inversion, Xz, while the population from Hispaniola has the standard X chromosome sequence.

The morphological and cytological work, along with the mating and allozyme studies, have been useful not only in the reconstruction of phylogenetic relationships among the species, but also in the examination of their geographic distributions. These investigations have also helped to elucidate patterns of relationship among populations within species and assess the potential and actual amount of gene exchange taking place between species.

In order to gain a greater understanding of the species in the *mayaguana* subcluster at both the species level and the population level, we have chosen to use DNA sequence information. For this study we have sequenced four nuclear genes (*Acetylcholinesterase* (Fournier et al., 1989), *hunchback* (Treier et al., 1989), *mastermind* (Bettler et al., 1996; Newfield et al., 1991), and *vestigial* (Williams et al., 1994)) and the mt A+T rich region (Lewis et al., 1994). We have assessed the relative contribution of each gene partition to the cladistic hypothesis we generated

for this study. In addition, we have examined the data from a Population Aggregation Analysis (PAA) perspective in order to search for characters that might diagnose either geographic units or species. Furthermore, we have taken a population genetics approach to the data by calculating AMOVAs (analyses of molecular variance) for each locus as a means to examine percent variability among species, among populations within species and within populations. We also calculated F_{ST} values to determine the amount of gene flow between populations.

Materials and Methods

Flies and DNA sequences: Table 4-1 lists both field collected specimens and lab cultured flies used in this study. Flies collected in the field were identified by examination of the male genitalia. In particular, we differentiated between *D. straubae* and *D. parisiena* by inspecting the surstyli. Heed and Grimaldi (1991) found that the inner margin of the surstylus of *D. straubae* is almost straight while that of *D. parisiena* is markedly crescentic. This character proved to be an extremely valuable diagnostic. We chose two Caribbean strains of *D. mulleri* as outgroups to root our taxa because chromosomal data suggest that this species gave rise to *D. mayaguana*, the most primitive of the three species in the *D. mayaguana* subcluster (Wasserman, 1992).

DNA was isolated from single and multiple flies using the methods outlined in Vogler et al. (1993). We have chosen to sequence fragments from five genes as follows.

(1) The *Acetylcholinesterase* (*Ace*) locus:

The *Acetylcholinesterase* gene codes for the enzyme that mediates the hydrolysis of the neurotransmitter, acetylcholine. This gene is organized into

Table 4-1. List of populations and strains used in this study. Designations marked with an * denote lab stocks. Others refer to field collected specimens.

Designation	Species	Strain Number	Locale
PA1 *	*D. parisiena*	950.25 (10 pairs)	Monte Cristi DR
PA2 *	*D. parisiena*	951.51G (1 pair)	Barahona DR
PA3 *	*D. parisiena*	961.1	Guantanamo Bay CU
PA4 *	*D. parisiena*	961.65B&J (2 pairs)	Sigua Beach CU
PA5 *	*D. parisiena*	951.55N,N2,O2,M2 (4pairs)	Barahona DR
PA6 *	*D. parisiena*	903.7 ORV-3	Mirabelais HA
PA7 *	*D. parisiena*	940.1	Monte Cristi DR
PA8 *	*D. parisiena*	961.65A	Sigua Beach CU
PA9 *	*D. parisiena*	951.55F2,G2 (2FF)	Barahona DR
PA10 *	*D. parisiena*	902.9	Gonaives HA
PA11 *	*D. parisiena*	ORV-24	Port Henderson JA
PA1-XXX	*D. parisiena*	980	Airport RD., Kingston JA
PA2-XXX	*D. parisiena*	981	Monte Cristi DR
PA2A-XXX	*D. parisiena*	981	Monte Cristi DR (under *Opuntia monoliformis*)
PA3-XXX	*D. parisiena*	982	20 km NW of Barahona DR
PA4-XXX	*D. parisiena*	983	27 km NE of Barahona DR
ST2 *	*D. straubae*	966.3	Guantanamo Bay CU
ST3 *	*D. straubae*	966.2	Guantanamo Bay CU
ST4 *	*D. straubae*	961.60A,B,C (3 pairs)	Sigua Beach CU
ST5 *	*D. straubae*	922.2A (1 pair)	Navassa Island
ST6 *	*D. straubae*	ORV-1	Fond Parisien HA
ST2-XXX	*D. straubae*	981	Monte Cristi DR

Table 4-1 continued.

ST4-XXX	*D. straubae*	983	27 km NE of Barahona DR
MA26C *	*D. mayaguana*		Skyline DR., CU
MA26D *	*D. mayaguana*		Concepcion Island
MA26E *	*D. mayaguana*		Grand Cayman
MA4-XXX	*D. mayaguana*	983	27 km NE of Barahona DR
MU1 *	*D. mulleri*	927.3	Cayman Brac
MU2 *	*D. mulleri*	942.3	Gonaives HA

ten exons with nine introns spaced between them (Fournier et al., 1989). Because introns are usually held to be less constrained by selective pressures, and, therefore, able to change at a more rapid rate, we have chosen to sequence one of the introns of length 180 bases.

(2) The A+T rich region of the mitochondrial DNA:

The circular, double-stranded mitochondrial DNA molecule in the genus *Drosophila* varies in size from 16-19.5 kilobases (kb) and encodes 13 polypeptides and the 22 transfer RNAs (tRNAs) and two ribosomal RNAs (rRNAs) needed for mitochondrial translation, with little or no intergenic spacing. It contains a long non-coding region which is composed of predominantly deoxyadenylate and thymidylate residues. This region can be divided into two segments: one, which is highly conserved and another, which is divergent. We have chosen to analyze a segment of approximately 320 bases that has been shown to exhibit a high degree of variability (Lewis et al., 1994).

(3) The *hunchback* (*hb*) locus:

The *hunchback* gene belongs to a class of developmental loci called gap genes. These genes participate in the development of segmental units in the embryo. The *hunchback* gene is comprised of two exons (one short, the other long) with an intron between them. The longer exon contains a hypervariable region, in which one finds poly-glutamine (CAA or CAG) repeats, of about 150 bases (Treier et al., 1989). We have elected to investigate this region and and the flanking regions on both sides, a fragment of approximately 440 bases.

(4) The *mastermind* (*mam*) gene:

The *mastermind* locus codes for a highly repetitive nuclear protein that functions in neural development. It contains a rather variable segment comprised of poly-glutamine (CAA or CAG) repeats (Bettler et al., 1996; Newfield

et al., 1991). We have elected to sequence a fragment of about 190 bases from this region.

(5) The *vestigial* (*vg*) locus:

The *vestigial* gene is necessary for the formation of the entire wing in *Drosophila*. This is a large gene, with exons comprising more than 16 kilobases (Williams et al., 1994). We have included a 5' nontranslated region of about 255 bases in this study.

The polymerase chain reaction (PCR) was used to amplify fragments for sequencing. PCR products were cleaned using Geneclean kits (BIO 101) and sequenced directly using automated sequencing methods with the ABI dye terminator system on an ABI 373 machine. Sequences were then corrected and edited using the SEQUENCHER software (GeneCodes Corp., 1995).

Data matrices and cladistic analysis: All sequences were compiled into NEXUS files after alignment of individual gene partitions. The A+T rich mtDNA alignments were trivial as very few indels occurred in this gene region for all the data in our matrix. The four nuclear genes (*Ace*, *hb*, *mam* and *vg*) all contain variable numbers of short tandem repeats. However, the regions flanking these repeat segments contained no indels. Therefore, alignments in the flanking regions were trivial. For all of the genes, we then aligned the di- or tri- nucleotide repeats in tandem and added gaps where necessary after the repeated regions. Cladistic analysis (we use the term cladistic in place of phylogenetic as we are exploring a tokogenetic system; in other words, we are examining taxa below the level of the species (Hennig, 1966)) was accomplished using PAUP 4.01b (Swofford,1999). We analyzed each of the five genes separately and in combination in order to investigate the interaction of the various gene regions in the cladistic analysis. Because the taxa in each of the

individual data partitions were not completely overlapping, we chose taxa which had been sequenced for three or more of the five genes for our simultaneous analysis, for a total of 40 taxa. Each of the individual data partitons was analyzed using all of the taxa for which we had sequence data. Distance matrices were calculated using PAUP 4.01b and redundant taxa were removed before analysis and were re-included after the analysis as polytomies with the taxa to which they were identical. We also analyzed a data set that combined the A+T rich mtDNA and the *hb* gene using 44 taxa for which we had complete sequence for both genes. In addition, we generated Bootstrap values (Felsenstein, 1985) using PAUP 4.01b. Bremer support values (Bremer, 1988; 1994) were calculated using the AUTODECAY program (Eriksson, 1997). Furthermore, we calculated partitioned Bremer support values for the various character partitions using the methods outlined in Baker and DeSalle (1997) and Baker et al. (1998).

Population aggregation analysis: Tables of diagnostic sites for population aggregation analysis were compiled using PAUP 4.01b (Swofford, 1999) and MacClade 3.01 (Maddison and Maddison, 1992) using the match first function which places a dot (.) in all character cells displaying the same character state as the first taxon. All sites were excluded with the exception of those that diagnosed either species or geographic populations within species.

AMOVA and F_{ST} values: The AMOVA (Analyses of MOlecular Variance) feature in the program Arlequin ver 1.1 (Schneider et al., 1997) was used in order to determine the percent of the total variability that could be ascribed to among species variability, among populations within species variability and within population variability. In addition, we used this program to calculate F_{ST} values in order to

ascertain the amount of gene flow occurring between populations. The program also calculated the significance of the between population F_{ST} values at a significance level of 0.05.

Results and Discussion

Cladistic analysis of the mayaguana subcluster using the combined data set

We obtained 2364 most parsimonious trees for the analysis of the 38 ingroup taxa and two outgroup taxa (two Caribbean populations of *D. mulleri*). There was a total of 132 characters out of 1376 characters that were phylogenetically informative in the combined data matrix. The strict consensus of the 2364 most parsimonious trees (CI=0.66, RI=0.73, L=239 steps) is shown in Figure 1 (refer to Table 2 for key to locale numbers appearing in rectangles or ovals to the right of taxon names). The consensus tree places the taxa belonging to the species *D. mayaguana* together in a single clade with the Concepcion Island population (MA26D) as ancestral to the polytomy containing the Grand Cayman (MA26E) and Skyline Drive, Cuba (MA26C) populations and the members of the population located at 27 km northeast of Barahona, Dominican Republic (MA-4-XXX). The populations belonging to the species *D. straubae* and *D. parisiena*, on the other hand, are intermixed throughout the tree and limited resolution is detected. We found little affinity either by species or by geographic locale. While several populations of a single species may be found grouped together throughout the tree, no clear pattern of relationship emerges. For example, ST2 (Guantanamo Bay, Cuba) and ST4 (Sigua Beach, Cuba) are seen to be sister taxa with ST3 (Guantanamo Bay, Cuba) as sister to them and ST6 (Fond Parisien, Haiti) as sister to all three. However, PA2 (Barahona, Dominican Republic) is shown as sister to

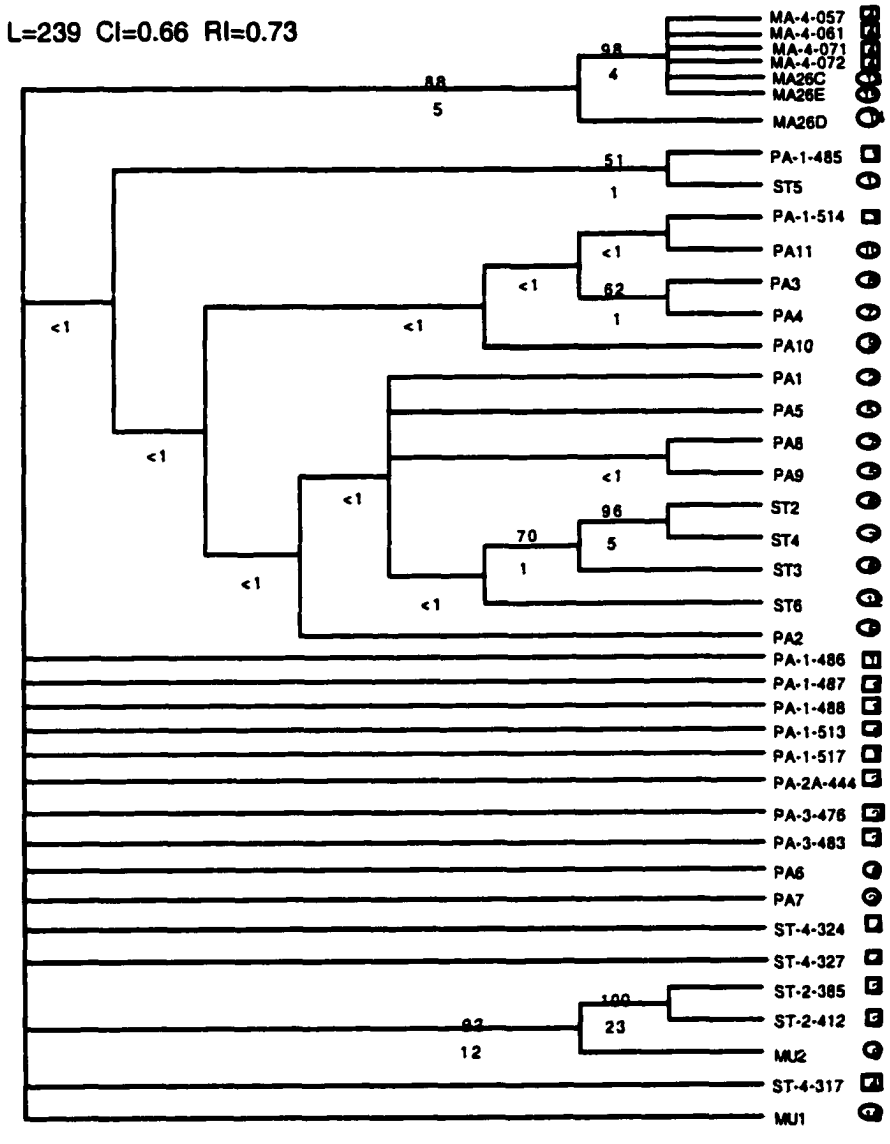


Figure 4-1. Strict consensus of 2364 most parsimonious trees from combined analysis of all the data equally weighted. Numbers to the right in rectangles denote wild caught specimens, numbers to the right in ovals denote lab stocks. Bremer support values are provided below each node and bootstrap values above each node.

Table 4-2. Key to locale number found to the right of taxon names in Figure 4-1.

Number in Rectangle or Oval	LOCALE
1	Kingston, JA
2	Monte Cristi, DR
3	20 km NW of Barahona, DR
4	27 km NE of Barahona, DR
5	Barahona, DR
6	Guantanamo Bay, CU
7	Sigua Beach, CU
8	Mirabelais, HA
9	Gonaives, HA
10	Port Henderson, JA
11	Navassa Island
12	Fond Parisien, HA
13	Skyline DR., CU
14	Concepcion Island
15	Grand Cayman
16	Cayman Brac

all four of these populations of *D. straubae*. PA3 (Guantanamo Bay, Cuba) and PA4 (Sigua Beach, Cuba) emerge as sister taxa while another *D. parisiena* from Sigua Beach, Cuba (PA8) is associated with PA9 (Barahona, Dominican Republic) as its sister taxon. The PA3-PA4 clade is seen as sister to a clade comprised of PA-1-514 (Kingston, Jamaica) and PA11 (Port Henderson, Jamaica). However, another *D. parisiena* taxon from Kingston, Jamaica (PA-1-485) forms a clade with ST5, a member of the species *D. straubae*, from Navassa Island. Both Jamaica and Navassa are areas of allopatry for these species, and their chromosomal constitution is very different so it seems highly unlikely that these two taxa would group together The remaining members of the Kingston, Jamaica population of *D. parisiena* (PA-1-XXX) are found in an unresolved state at the base of the tree along with members of both other *D. parisiena* populations and *D. straubae* populations. The most highly supported clade in the tree comprises the two *D. straubae* individuals from the population collected from Monte Cristi, Dominican Republic (ST-2-XXX; BP=100, BS=23). However, the flies belonging to the species *D. mulleri* from Gonaives, Haiti are seen as sister to this group. The support for this node is relatively high (BP=92, BS=12). It is interesting to note that the *D. straubae* populations from Monte Cristi and from Gonaives, Haiti still possess the standard chromosomes that are found in the six homosequential species (of which *D. mulleri* is one), but are not found anywhere else in *D. straubae* (Wasserman and Wasserman, 1992).

Separate analyses of the individual gene partition

We wished to estimate the relative contribution of each of the gene partitions to the simultaneous analysis (SA) hypothesis, so we analyzed each gene separately and calculated partitioned Bremer support values. Figure 4-2 presents the strict

consensus trees obtained for the individual data sets after removal of redundant taxa. These trees present a spurious picture of our results so we re-included all taxa that had been removed as polytomies with the taxa to which they had been found identical in the distance matrices so that we could gain a clearer representation of what the data had to say. Figures 4-3 through 4-7 show the strict consensus trees generated for the various gene partitions with all taxa included. Overall we uncovered very little structure from any of the gene partitions. The tree generated from the *ace* partition (Figure 4-3) exhibits a marked lack of structure and resolution as does the tree produced by analysis of the *mam* locus (Figure 4-6). Only the *hunchback* gene (Figure 4-5) recovers *D. mayaguana* as as monophyletic group, albeit not a highly resolved one. Two of the gene trees, *hb* (Figure 4-5) and *vg* (Figure 4-7) group the two Guantanamo Bay *D. straubae* (ST2 and ST3) together. The *mam* locus does unite the taxa from the Monte Cristi *D. straubae* population (ST-2-XXX), but there is no support for this node. The A+T rich mtDNA (Figure 4-4) recovers one highly supported clade comprised of the Monte Cristi *D. straubae* taxa (ST-2-XXX; BP=100, BS=24). Table 4-3 presents the tree statistics for the individual data partitions and for the combined analysis. The A+T rich mtDNA data set appears to contain least amount of homoplasy (CI=0.94, RI=0.95) of all of the data partitions while *hb* contains the greatest amount of homoplasy (CI=0.59, RI=0.71).

Since the results of the individual data analyses exhibited such poor resolution, it is troublesome to ascertain to what degree individual genes support relationships that appear in the combined analysis. Partitioned Bremer support provides one means for evaluating the contribution of different genes to the SA topology. Table 4-4 gives the results of partitioning the Bremer support for each of

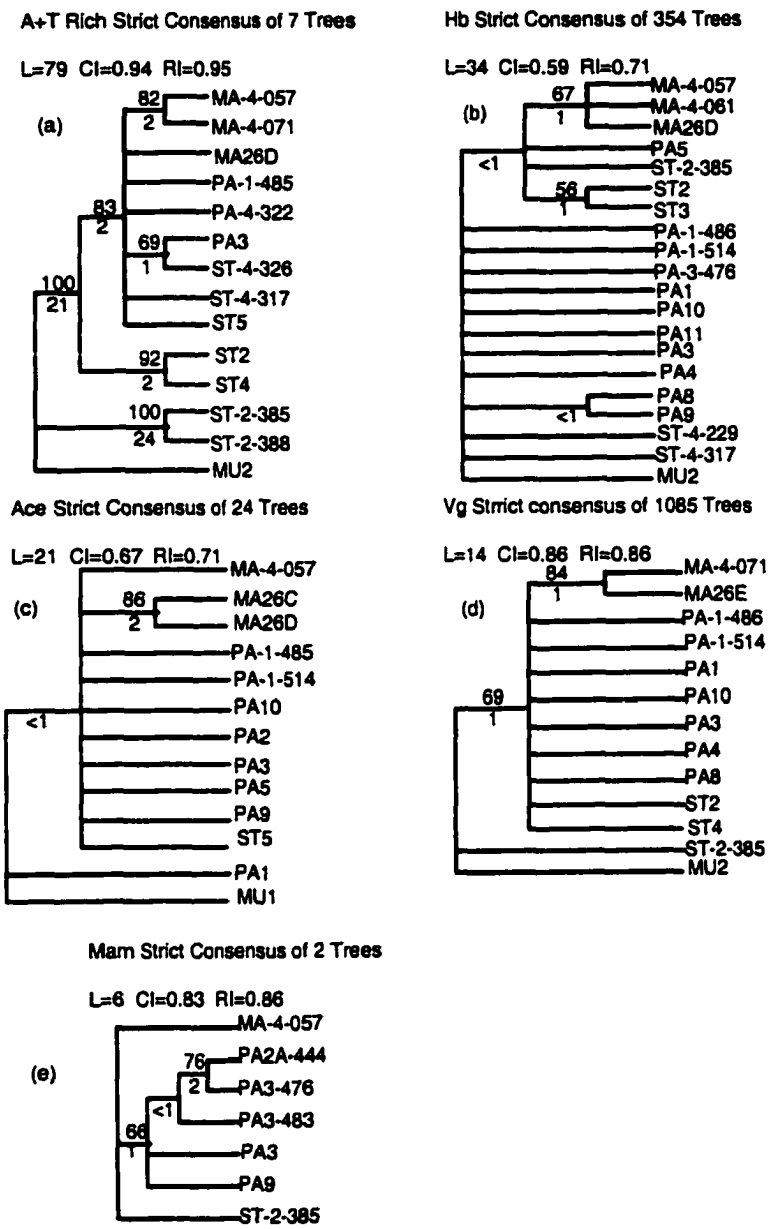


Figure 4-2. Strict consensus of most parsimonious trees from analyses of individual gene partitions before the re-inclusion of redundant taxa. Bremer support values are provided below each node and bootstrap values above each node.

Ace

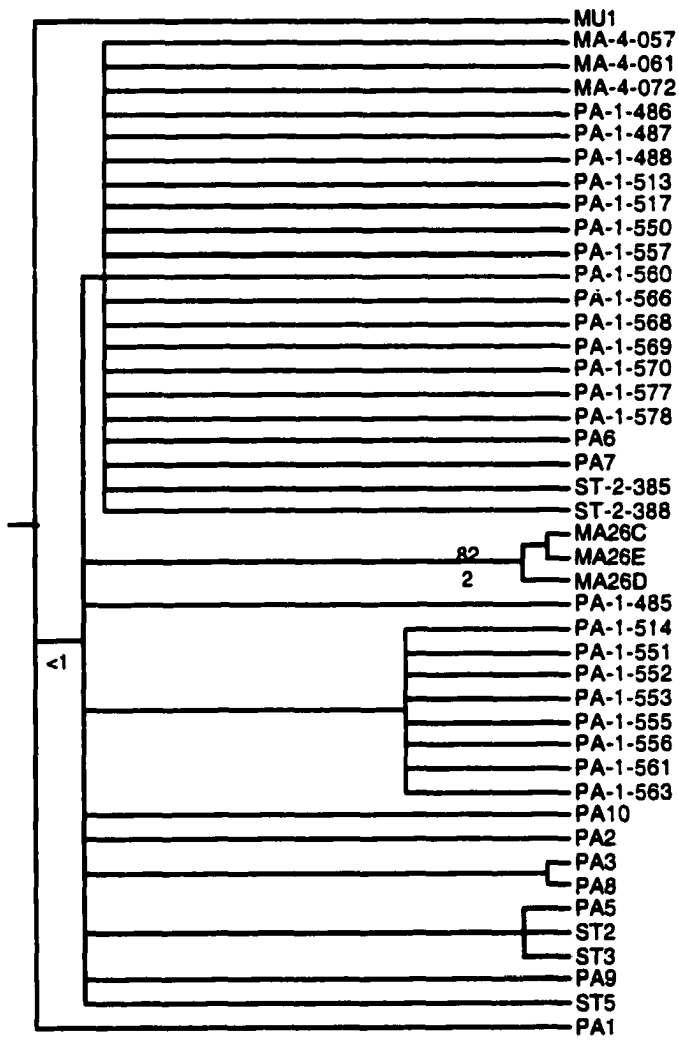


Figure 4-3. Strict consensus of 7 most parsimonious trees from analysis of the Ace gene partition after the re-inclusion of redundant taxa. Bremer support values are provided below each node and bootstrap values above each node.

A+T rich mtDNA

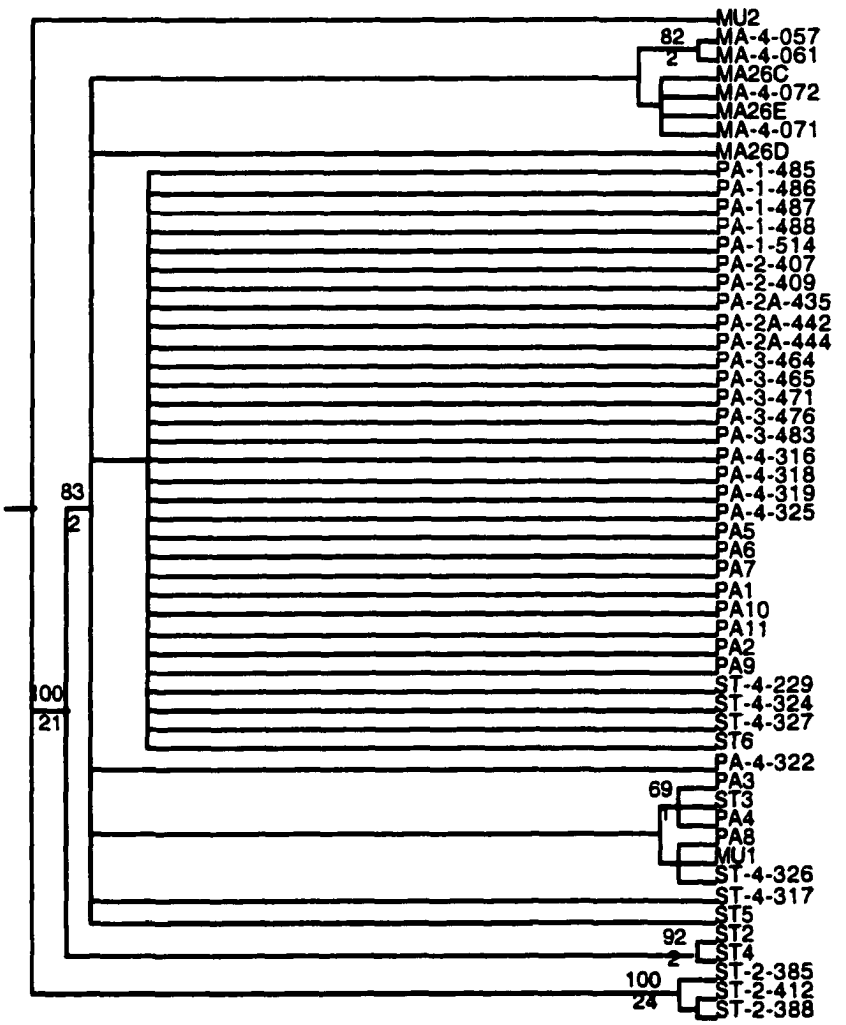


Figure 4-4. Strict consensus of 24 most parsimonious trees from analysis of the A+T rich mtDNA gene partition after the re-inclusion of redundant taxa. Bremer support values are provided below each node and bootstrap values above each node.

hb

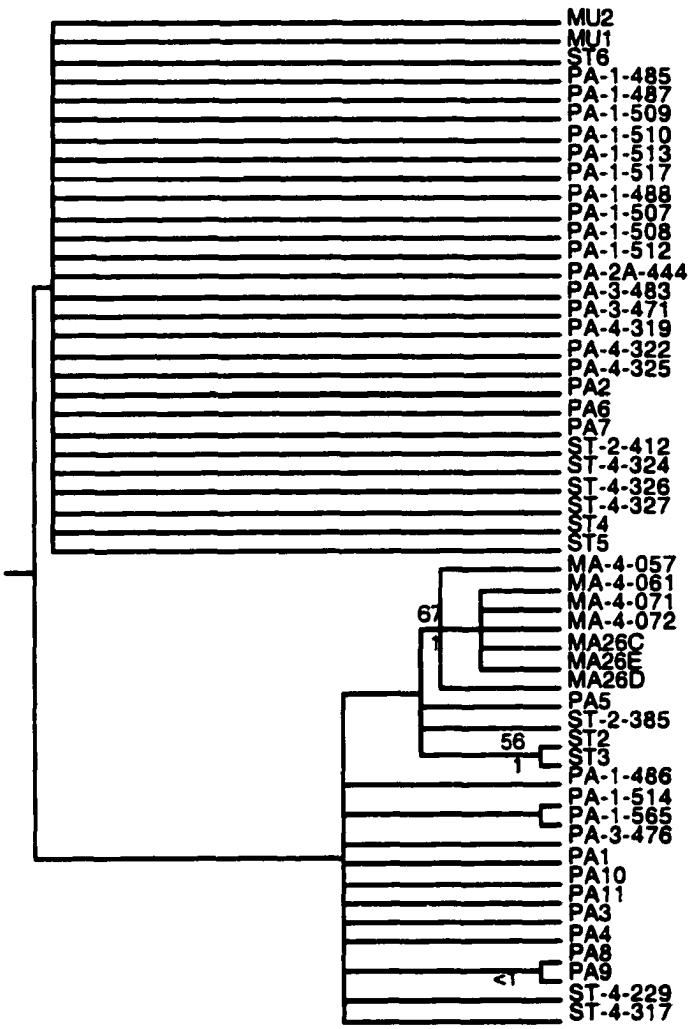


Figure 4-5. Strict consensus of 354 most parsimonious trees from analysis of the hb gene partition after the re-inclusion of redundant taxa. Bremer support values are provided below each node and bootstrap values above each node.

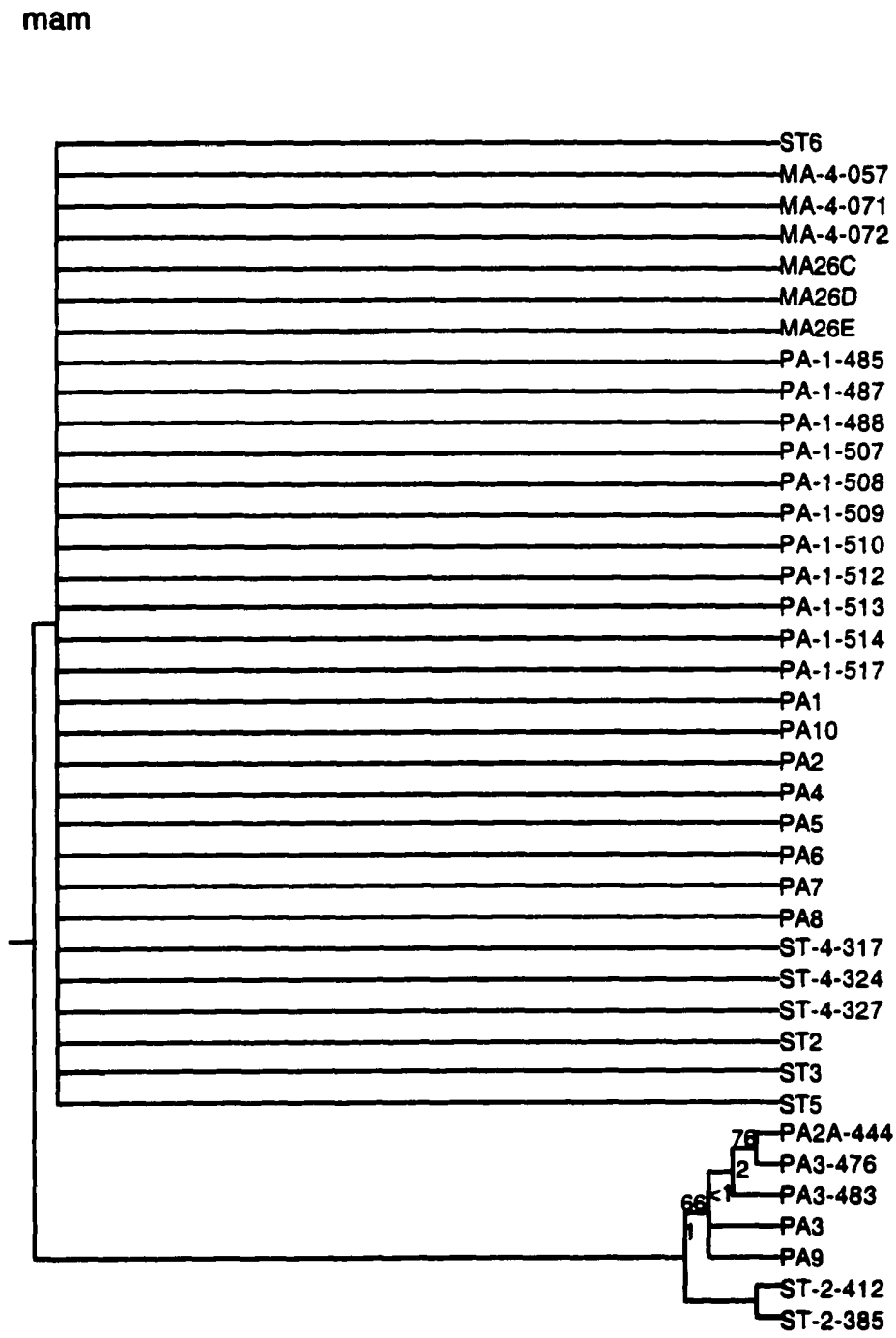


Figure 4-6. Strict consensus of 2 most parsimonious trees from analysis of the mam gene partition after the re-inclusion of redundant taxa. Bremer support values are provided below each node and bootstrap values above each node.

vg

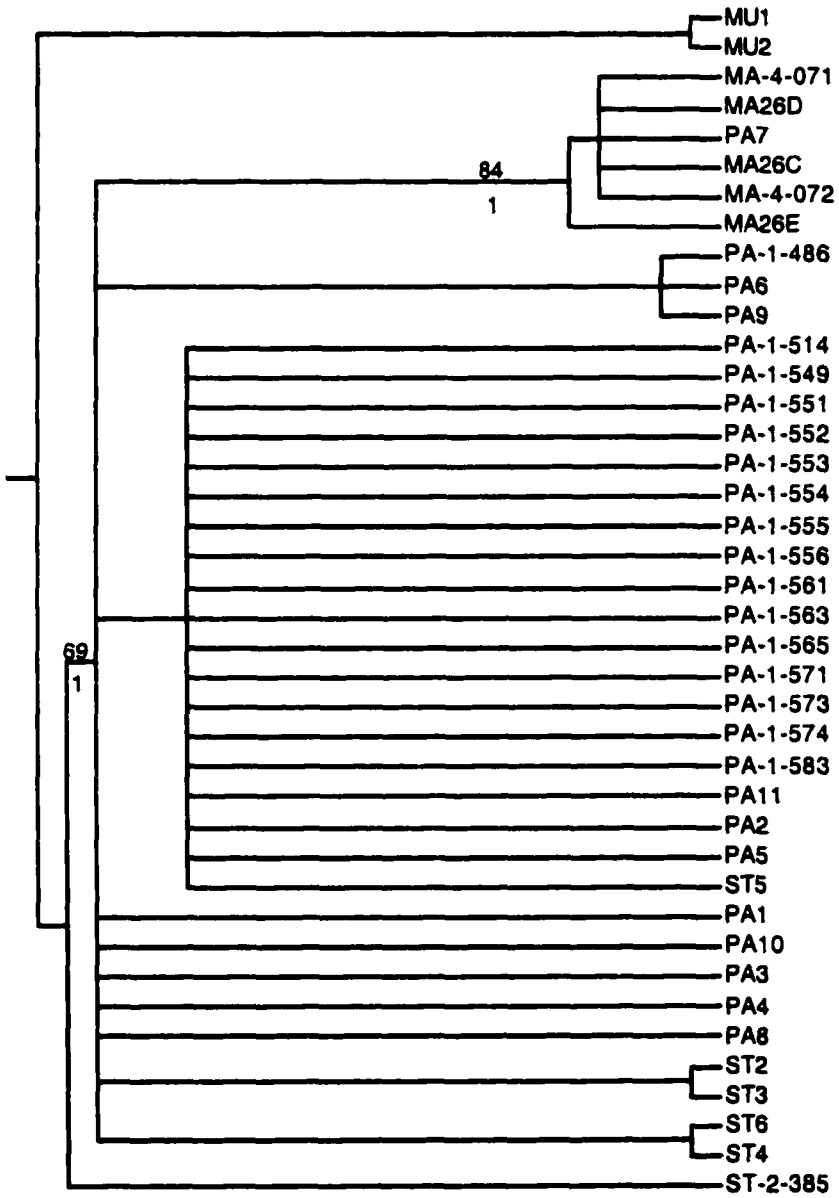


Figure 4-7. Strict consensus of 1085 most parsimonious trees from analysis of the vg gene partition after the re-inclusion of redundant taxa. Bremer support values are provided below each node and bootstrap values above each node.

Table 4-3. Tree statistics for individual and combined data partitions. Tot. no. chars. = total number of characters in the data partition, PI = number of phylogenetically informative characters in the data partition, No. trees = number of most parsimonious trees obtained in the analysis, Steps = length of most parsimonious tree, CI = consistency index, RI = retention index.

	Tot. no. chars.	No. PI	No. Trees	Steps	CI	RI
Ace	180	10	24	21	0.67	0.71
A+T Rich	317	62	7	79	0.94	0.95
Hb	438	17	354	34	0.59	0.71
Mam	187	5	2	6	0.83	0.86
Vg	254	8	1085	14	0.86	0.86
A+T and Hb	755	80	2804	125	0.76	0.83
Total	1376	132	2364	239	0.66	0.73

Table 4-4 . Results of partitioning Bremer support for each of the data sets across the simultaneous analysis tree (Figure 4-1). Summed PBS is the total Bremer support summed over all nodes in the cladogram. Corrected Summed BS is the standardized summed Bremer support where Bremer support has been divided by the number of informative characters in the character partition and by the minimum number of steps for each partition.

		Corrected Summed BS	
Gene	Summed PBS	by # of PI	by min. # steps
Ace	-13.9	-1.39	-0.32
A+T Rich mtDNA	59.5	0.96	0.61
Hb	18.6	1.09	0.32
Mam	2.3	0.46	0.10
Vg	-5.8	-0.73	-0.48

the data sets across the simultaneous analysis tree. The A+T rich mtDNA partition contributes 59.5 units of Bremer support - almost all of the 61 summed Bremer support units for the simultaneous analysis tree. The *hb* gene contributes only 18.6 Bremer support units, while *mam* provides a mere 2.3 units of Bremer support. In contrast, both *Ace* (summed PBS=-13.9) and *vg* (summed PBS=-5.8) furnish "negative support". In other words, these two genes exhibit a high degree of conflict with the SA tree. In fact, out of 17 nodes on the SA tree, *Ace* contributes a zero or a negative Bremer support value to 16 nodes while *vg* contributes a zero or a negative Bremer support value to 9 nodes.

We also performed a cladistic analysis on 44 taxa that had complete sequences for both the A+T rich mtDNA region and the *hb* region. Figure 4-8 shows the strict consensus of 2804 trees obtained for this analysis (CI=0.76, RI=0.83, L=125 steps). This analysis recovers both the *D. mayaguana* species clade (BP=95, BS=4) and the clade consisting of the Monte Cristi *D. straubae* individuals (BP=100, BS=24). However, no new information as to the relatioinship of the taxa to each other is revealed.

Population Aggregation Analysis (PAA)

Population aggregation analysis (Davis and Nixon, 1992) is a useful tool for for determining the limits of phylogenetic species. This approach is used to differentiate between attributes that are characters (attributes that are invariable, i.e., fixed within a terminal lineage or a monophyletic group) and attributes that are traits (attributes that occur in some but not all representatives of a terminal lineage or

A+T rich mtDNA and hb

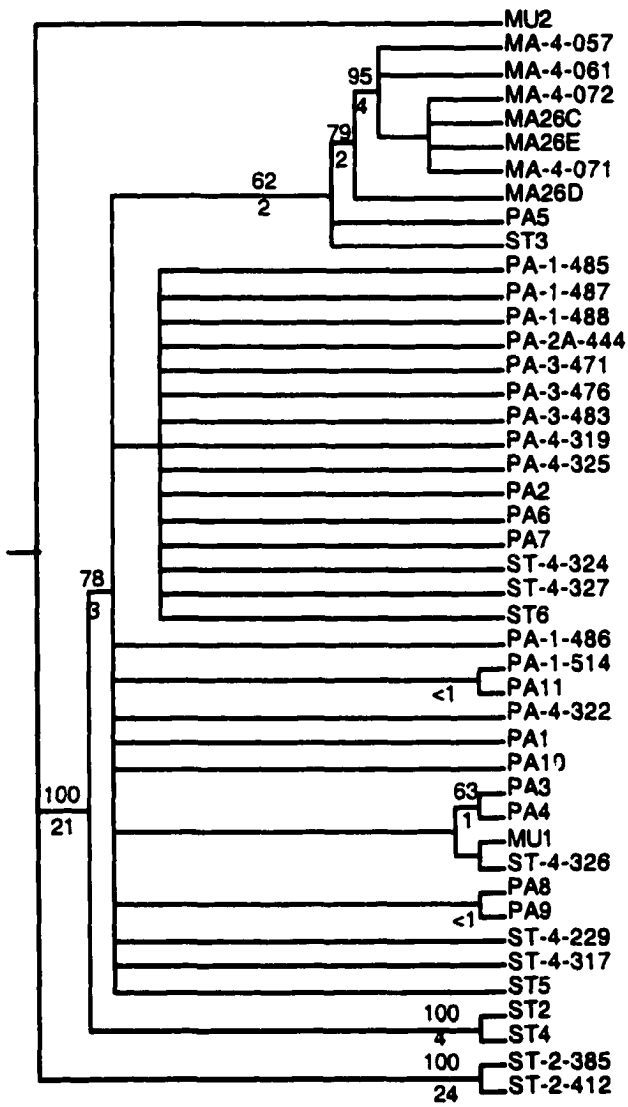


Figure 4-8. Strict consensus of 755 most parsimonious trees from analysis of the Ace gene partition after the re-inclusion of redundant taxa. Bremer support values are provided below each node and bootstrap values above each node.

monophyletic group). Only characters can be used to diagnose populations or species as they alone provide evidence that gene flow no longer occurs between groups of populations or species. For our analysis we generated tables of diagnostic characters for each of our data partitions. Table 4-5 shows diagnostic sites found in the *Ace* gene. Character 22 diagnosed the *D. straubae* population from Navassa Island (ST5) while character 27 diagnosed *D. parisiena* (PA3) from Guantanamo Bay, Cuba. Character 129 showed diagnostic sites for both the *D. parisiena* population found in Gonaives, Haiti (PA10) and the members of *D. straubae* from Guantanamo Bay, Cuba (ST2 and ST3). The A+T rich mtDNA partition (Table 4-6) revealed a total of 48 diagnostic characters for the individuals of *D. straubae* that were caught in Monte Cristi, Dominican Republic (ST-2-385, ST-2-388, ST-2-412) and one character that diagnosed *D. mayaguana* from Concepcion Island (MA26D). *Hunchback* (Table 4-7) was the only gene that contained a character that aggregated all of the populations of a species into a single unit. Character 339 diagnosed the species *D. mayaguana* (MA-4-057, MA-4-061, MA-4-071, MA-4-072, MA26C, MA26D, MA26E). In addition, characters 246 and 312 were diagnostic for the Concepcion Island population of *D. mayaguana* (MA26D) and character 122 was diagnostic for *D. parisiena* from Guantanamo Bay, Cuba (PA3). Fourteen sites diagnosing the Monte Cristi *D.straubae* population (ST-2-385 and ST-2-412) were found in the *mastermind* locus (Table 4-8) while site 153 was diagnostic for *D. straubae* from Guantanamo Bay (ST2 and ST3) and site 168 delimited the Navassa population of *D. straubae* (ST5). The *vestigial* gene (Table 4-9) contained sites that diagnosed each of the populations of *D. straubae* . Sites 23, 129, 131 and 169 were diagnostic for the population from Fond Parisien, Haiti (ST6), site 167 diagnosed Navassa Island *D. straubae* (ST5), sites 135 and 137 defined the population from Sigua Beach, Cuba (ST4) and site 40

Table 4-5. *Ace* diagnostic sites. Sequences are numbered sequentially form #1 corresponding to the first base position of the amplified region. An asterisk (*) indicates samples that are being diagnosed, with the diagnosing characters underlined. Sites are listed above in the order that they appear in the sequence. DNA sequences correspond to IUPAC format. A dot (.) indicates that the sequence matches the DNA symbol above.

	1 222 179
MA-4-057	CAC
MA-4-061	...
MA-4-072	...
MA26C	..?
MA26D	...
MA26E	..?
PA-1-485	...
PA-1-486	...
PA-1-487	...
PA-1-488	...
PA-1-513	...
PA-1-514	...
PA-1-517	...
PA-1-550	...
PA-1-551	...
PA-1-552	...
PA-1-553	...
PA-1-555	...
PA-1-556	...
PA-1-557	...
PA-1-560	...
PA-1-561	...
PA-1-563	...
PA-1-566	...
PA-1-568	...
PA-1-569	...
PA-1-570	...
PA-1-577	...
PA-1-578	...
PA1	..T
PA10*	..G
PA2	...
PA3*	.CT
PA5	..?
PA6	...
PA7	...

Table 4-5 (continued). *Ace* diagnostic sites. Sequences are numbered sequentially form #1 corresponding to the first base position of the amplified region. An asterisk (*) indicates samples that are being diagnosed, with the diagnosing characters underlined. Sites are listed above in the order that they appear in the sequence. DNA sequences correspond to IUPAC format. A dot (.) indicates that the sequence matches the DNA symbol above.

	1 222 179
PA8	..T
PA9	..?
ST-2-385	...
ST-2-388	...
ST2#	..A
ST3#	..A
ST5*	T.?

Table 4-6. A+T rich mtDNA diagnostic sites. Sequences are numbered sequentially form #1 corresponding to the first base position of the amplified region. An asterisk (*) indicates samples that are being diagnosed, with the diagnosing characters underlined. A pound sign (#) indicates more than one sample from a population or species is being diagnosed. Sites are listed above in the order that they appear in the sequence. DNA sequences correspond to IUPAC format. A dot (.) indicates that the sequence matches the DNA symbol above.

```
                                  1111111112222222222222223333333
                  2222335555555889002267782223334456779990000011
                  7456923013578917978357345457014186815623701567 45

MA-4-057          GTAAATGGCGTTTAATAGGCTGGAAAATTTCACAATAAAAAAACA-TT
MA-4-061          ...........................................-..
MA-4-071          ...........................................-..
MA-4-072          ...........................................-..
MA26C             ...........................................-..
MA26D*            ..................................C........-..
MA26E             ...........................................-..
PA-1-485          ...........................................-..
PA-1-486          ...........................................-..
PA-1-487          ...........................................-..
PA-1-488          ...........................................-..
PA-1-514          ...........................................-..
PA-2-407          ...........................................-..
PA-2-409          ...........................................-..
PA-2A-435         ...........................................-..
PA-2A-442         ...........................................-..
PA-2A-444         ...........................................-..
PA-3-464          ...........................................-..
PA-3-465          ...........................................-..
PA-3-471          ...........................................-..
PA-3-476          ...........................................-..
PA-3-483          ...........................................-..
PA-4-316          ...........................................-..
PA-4-318          ...........................................-..
PA-4-319          ...........................................-..
PA-4-322          ...........................................-..
PA-4-325          ...........................................-..
PA1               ...........................................-..
PA10              ...........................................-..
PA11              ...........................................-..
```

Table 4-6 (continued). A+T rich mtDNA diagnostic sites. Sequences are numbered sequentially form #1 corresponding to the first base position of the amplified region. An asterisk (*) indicates samples that are being diagnosed, with the diagnosing characters underlined. A pound sign (#) indicates more than one sample from a population or species is being diagnosed. Sites are listed above in the order that they appear in the sequence. DNA sequences correspond to IUPAC format. A dot (.) indicates that the sequence matches the DNA symbol above.

```
                                  1111111112222222222222223333333
                 2222335555555889002267782223334456779990000011
                 7456923013578917978357345457014186815623701567 45
PA2              ...........................................-..
PA3              ...........................................-..
PA4              ...........................................-..
PA5              ...........................................-..
PA6              ...........................................-..
PA7              ...........................................-..
PA8              ...........................................-..
PA9              ...........................................-..
ST-2-385#        ACTTTCCCATCAATTGGTAAAATTTTTAAAATTTCATCTTTTTTTCAG
ST-2-388#        ACTTTCCCATCAATTGGTAAAATTTTTAAAATTTCATCTTTTTTTCAG
ST-2-412#        ACTTTCCCATCAATTGGTAAAATTTTTAAAATTTCATCTTTTTTTCAG
ST-4-229         ...........................................-..
ST-4-317         ...........................................-..
ST-4-324         ...........................................-..
ST-4-326         ...........................................-..
ST-4-327         ...........................................-..
ST2              ......T....................................-..
ST3              ...........................................-..
ST4              ......T....................................-..
ST5              ...........................................-..
ST6              ...........................................-..
```

Table 4-7. *hb* diagnostic sites. Sequences are numbered sequentially form #1 corresponding to the first base position of the amplified region. An asterisk (*) indicates samples that are being diagnosed, with the diagnosing characters underlined. A pound sign (#) indicates more than one sample from a population or species is being diagnosed. Sites are listed above in the order that they appear in the sequence. DNA sequences correspond to IUPAC format. A dot (.) indicates that the sequence matches the DNA symbol above.

	1233 2413 2629
MA-4-057#	CGGT
MA-4-061#	....
MA-4-071#	....
MA-4-072#	....
MA26C#	....
MA26D*#	.TA.
MA26E#	....
PA-1-485	...C
PA-1-486	...C
PA-1-487	...C
PA-1-488	...C
PA-1-507	...C
PA-1-508	...C
PA-1-509	...C
PA-1-510	...C
PA-1-512	...C
PA-1-513	...C
PA-1-514	...C
PA-1-517	...C
PA-1-565	...C
PA-2A-444	...C
PA-3-471	...C
PA-3-476	...C
PA-3-483	...C
PA-4-319	...C
PA-4-322	...C
PA-4-325	...C
PA1	...C
PA10	...C
PA11	...C
PA2	...C
PA3*	A..C

Table 4-7 (continued). *hb* diagnostic sites. Sequences are numbered sequentially form #1 corresponding to the first base position of the amplified region. An asterisk (*) indicates samples that are being diagnosed, with the diagnosing characters underlined. A pound sign (#) indicates more than one sample from a population or species is being diagnosed. Sites are listed above in the order that they appear in the sequence. DNA sequences correspond to IUPAC format. A dot (.) indicates that the sequence matches the DNA symbol above.

	1233 2413 2629
PA4	...C
PA5	...C
PA6	...C
PA7	...C
PA8	...C
PA9	...C
ST-2-385	T..C
ST-2-412	...C
ST-4-229	...C
ST-4-317	...C
ST-4-324	...C
ST-4-326	...C
ST-4-327	...C
ST2	...C
ST3	...C
ST4	...C
ST5	...C
ST6	...C

Table 4-8. *mam* diagnostic sites. Sequences are numbered sequentially form #1 corresponding to the first base position of the amplified region. An asterisk (*) indicates samples that are being diagnosed, with the diagnosing characters underlined. A pound sign (#) indicates more than one sample from a population or species is being diagnosed. Sites are listed above in the order that they appear in the sequence. DNA sequences correspond to IUPAC format. A dot (.) indicates that the sequence matches the DNA symbol above.

	111 4455566666678566 1406901234516338
MA-4-057	GAACTGCACAGACACG
MA-4-071	....
MA-4-072	....
MA26C	....
MA26D	....
MA26E	....
PA-1-485	....
PA-1-487	....
PA-1-488	............G..
PA-1-507	............G..
PA-1-508	....
PA-1-509	....
PA-1-510	....
PA-1-512	............G..
PA-1-513	............G..
PA-1-514	....
PA-1-517	....
PA2A-444	....
PA3-476	............G..
PA3-483	............G..
PA1	....
PA10	....
PA2	....
PA3	............G..
PA4	....
PA5	....
PA6	............???
PA7	....
PA8	....
PA9	............G..
ST-2-385#	AGGGCCATGCAGG..A
ST-2-412#	AGGGCCATGCAGG..A
ST-4-317	............G..

Table 4-8 (continued). *mam* diagnostic sites. Sequences are numbered sequentially form #1 corresponding to the first base position of the amplified region. An asterisk (*) indicates samples that are being diagnosed, with the diagnosing characters underlined. A pound sign (#) indicates more than one sample from a population or species is being diagnosed. Sites are listed above in the order that they appear in the sequence. DNA sequences correspond to IUPAC format. A dot (.) indicates that the sequence matches the DNA symbol above.

	111 4455566666678566 1406901234516338
ST-4-324	.............G..
ST-4-327	.............G..
ST2#	.............T..
ST3#	.............T..
ST5*	.............G.
ST6	....

Table 4-9. *vg* diagnostic sites. Sequences are numbered sequentially form #1 corresponding to the first base position of the amplified region. An asterisk (*) indicates samples that are being diagnosed, with the diagnosing characters underlined. A pound sign (#) indicates more than one sample from a population or species is being diagnosed. Sites are listed above in the order that they appear in the sequence. DNA sequences correspond to IUPAC format. A dot (.) indicates that the sequence matches the DNA symbol above.

```
                  11111111111111122222
              22233452223333666788990012 5
              36801051391457379201357170 72

MA-4-071      GCAACAATACCCGCGGCGGCTGTTTAGC
MA-4-072      ............................
MA26C         ............................
MA26D         ............................
MA26E*        .G..........................
PA-1-486      .....-.A....................
PA-1-514      .....-......................
PA-1-549      .....-......................
PA-1-551      .....-......................
PA-1-552      .....-......................
PA-1-553      .....-......................
PA-1-554      .....-......................
PA-1-555      .....-......................
PA-1-556      .....-......................
PA-1-561      .....-......................
PA-1-563      .....-......................
PA-1-565      .....-......................
PA-1-571      .....-......................
PA-1-573      .....-......................
PA-1-574      .....-......................
PA-1-583      .....-......................
PA1           ............................
PA10*         .....-..G.....C........CG...
PA11          .....-......................
PA2           .....-......................
PA3*          .....-.......G..............
PA4           .....-.........C..T.........
PA5           .....-......................
PA6*          .....-.G....................
PA7           ............................
PA8           ............................
PA9           .....-......................
```

Table 4-9 (continued). *vg* diagnostic sites. Sequences are numbered sequentially form #1 corresponding to the first base position of the amplified region. An asterisk (*) indicates samples that are being diagnosed, with the diagnosing characters underlined. A pound sign (#) indicates more than one sample from a population or species is being diagnosed. Sites are listed above in the order that they appear in the sequence. DNA sequences correspond to IUPAC format. A dot (.) indicates that the sequence matches the DNA symbol above.

```
                    11111111111111122222
          22233452223333666788990012 5
          36801051391457379201357170 72

ST-2-385*  ..GCA.T....T.....TCAGAG..TAT
ST2#       .....T.....................
ST3#       .....T.....................
ST4*       .......----.CT.............
ST5*       .....-.........A...........
ST6*       A........TT.....T...........
```

delimited the Guantanamo Bay, Cuba population (ST2 and ST3). In addition, there were 14 characters that diagnosed the Monte Cristi, Dominican Republic population (ST-2-385). Position 121 was diagnostic for *D. parisiena* from Mirabelais, Haiti, position 139 diagnosed Guantanamo Bay, Cuba *D. parisiena* and sites 123, 163, 201 and 207 diagnosed *D. parisiena* from Gonaives, Haiti. Moreover, there was one site (site 26) that was diagnostic for *D. mayaguana* from Grand Cayman (MA26E).

It appears that the same groups continue to be recovered from these data partitions. For example, the Guantanamo Bay, Cuba *D. straubae* were diagnosed by three of the data sets (*Ace*, *mam* and *vg*) while the Monte Cristi *D. straubae* were aggregated by A+T rich mtDNA, *mam* and *vg*. *D. parisiena* from Guantanamo Bay were also diagnosed by three of the five genes (*Ace*, *hb* and *vg*). While our sample sizes are rather small, the fact that the data continue to assemble the same groups repeatedly suggests that these groups may, in fact, be distinct from the others in our analysis. On the other hand, all of the data sets failed to exhibit any unifying characters for any of the field collected specimens except for the Monte Cristi *D. straubae*.

A population genetics approach

We calculated F_{ST} (fixation index) values using the AMOVA option in the program Arlequin ver 1.1 (Schneider et al., 1997). This feature was also used to generate Analyses of Molecular Variance (AMOVA, Excoffier et al., 1992) which we used to identify the percent sequence variation attributable to variation between species, the percent sequence variation attributable to variation among populations

within a species and the percent sequence variation attributable to variation within populations.

Table 4-10 presents the key to the row and column headings used in the tables of F_{ST} values for the various gene partitions (Tables 4-11 through 4-15). Values in the F_{ST} tables printed in boldface type are significant at the 0.05 level. A value of 0.00 would imply that there is no between population variation and that the total variation can be atributed to within population sequence differences. A value of 1.00 suggest that there are no sequence differences within populations, and that all variation comes from between populations. In our study, however, members of many of the geographic locales are represented by lab stocks, and we only generated one sequence per gene per lab culture. This produces misleading results as there can be no sequence differences "between" one sequence. Therefore, for the remainder of our discussion of F_{ST} values, we shall refer only to those locales containing individuals that had been collected in the field (designated M1, P1, P2, P3, P4, S1 and S2 in the tables).

Table 4-11 shows the F_{ST} values calculated for the *Ace* data partition. No significant values emerge for any of the field collected populations. Looking at the values for the A+T rich mtDNA data set Table 4-12) we note a great many F_{ST} s between the wild caught populations that are significant. Between the NE of Barahona population of *D. mayaguana* (M1) and all four of the *D. parisiena* populations (P1, P2, P3, P4) we find F_{ST} s ranging from 0.69 to 0.95. Most interesting is the fact that the lowest value (0.69) was calculated between M1 and P4 both of which occur NE of Barahona. These results imply that in the comparison of M1 with P4 about 69% of the variation can be attributed to between population differences while in the comparisons of M1 with the P1, P2 and P3 95% of the variation can be attributed to between population differences. The

comparisons of the M1 population with the two populations of *D. straubae* show values of .88 with S1 (the Monte Cristi population) and .98 with S2 (the NE of Barahona population). This result is opposite to what we saw when we compared M1 to the *D. parisiena* populations. Comparisons of the *D. parisiena* populations with the NE of Barahona *D. straubae* yielded a range of between population variation of from 95-99%. When the two *D. straubae* populations were compared, 97% of the variation was found to be attributable to between population differences. The values produced for the *hb* partition (Table 4-13) present a somewhat different picture. When M1 was compared to each of the other populations, the F_{ST} values ranged from a low of 0.48 to a high of 0.92. Comparisons with the *D. parisiena* populations showed the highest values while comparisons with the two *D. straubae* populations showed the lowest, implying greater between population (and, in this case, between species) differences in the *mayaguana* versus *parisiena* comparisons than in the *mayaguana* versus *straubae* comparisons. When the Jamaican *D. parisiena* were compared to conspecifics the F_{ST} s were 0.67 and 0.54. However, when they were compared to the NE of Barahona *D. straubae*, the F_{ST} value was only 0.36 indicating that the preponderance of variation came from within population differences. The only significant F_{ST} values we obtained for the *mam* data set (Table 4-14) occurred between the Jamaican *D. parisiena* population and the Monte Cristi populations of both *D. parisiena* and *D. straubae*. The *D. parisiena* and *D. straubae* comparison revealed that most of the variation was attributable to between population differences (F_{ST} =0.95) while the conspecific comparison yielded a value of 0.66, indicating that about two-thirds of the variation came from between species. Finally, if we look at Table 4-15 we see only one significant comparison for the *vg* gene partition. This shows that 89% of the total variation is attributable to differences between M1 and P1.

Figure 4-9 shows the percent of the total variation contributed by variation among species, variation among populations within species and variation within populations. The *mam* gene reveals the greatest among populations within a species variation (>75%) with almost no between species variation. The A+T rich region is similar (>65% of the differentiation due to among population differences), but with somewhat more among species variation (>25%). Both *vg* and *Ace* show large percentages of between species variation (*vg* >57% and *Ace* >45%). However, these two regions differ greatly in the amount of variation attributable to within population differences (*vg* =10%, *Ace* >40%). The hunchback gene reveals the greatest amount of within population variation (>45%) with the variation contributed by among species differences to be >38%. We saw no pattern of variation emerging from these genes. The *hb* and *Ace* data sets appear to behave in a similar fashion with low variation among populations within a species and much higher values for both among species differences and within population differences. However, comparison of the behavior of the other three genes in terms of the partitioning of variation reveals no similarity.

Conclusions

The present study suggests that the species belonging to the *D. mayaguana* subcluster, especially *D. parisiena* and *D. straubae* are the products of a very recent speciation. The lack of resolution in our tree (Figure 4-1) and the intermixing of the two species throughout the tree indicate that complete divergence has not occurred between these two species. This conclusion is supported by each of the individual gene trees (Figures 4-3 through 4-7), which also present a similar scenario. In addition, with the exception of the single character in the *hb* gene that diagnosed the

Table 4-10. Key to species and locale designations used in tables of F_{ST} values.

Population Designation	Species and Locale
M1	*D. mayaguana* 27 km NE of Barahona DR
M2	*D. mayaguana* Skyline Dr. CU *
M3	*D. mayaguana* Concepcion Island *
M4	*D. mayaguana* Grand Cayman *
P1	*D. parisiena* Airport Rd., Kingston JA
P2	*D. parisiena* Monte Cristi DR
P3	*D. parisiena* 20 km NW of Barahona DR
P4	*D. parisiena* 27 km NE of Barahona DR
P5	*D. parisiena* Monte Cristi DR *
P6	*D. parisiena* Barahona DR *
P7	*D. parisiena* Gonaives HA
P8	*D. parisiena* Fort Henderson JA *
P9	*D. parisiena* Guantanamo Bay CU *
P10	*D. parisiena* Sigua Beach CU *
P11	*D. parisiena* Mirabelais HA *
S1	*D. straubae* Monte Cristi DR
S2	*D. straubae* 27 km NE of Barahona DR
S3	*D. straubae* Guantanamo Bay CU *
S4	*D. straubae* Sigua Beach CU *
S5	*D. straubae* Navassa Island *
S6	*D. straubae* Fond Parisien HA *
MU1	*D. mulleri* Cayman Brac *
MU2	*D. mulleri* Gonaives HA *

Table 4-11. F_{ST} values between populations of species in the *mayaguana* subcluster calculated for the *Ace* gene. Values in **boldface** type are significant at the 0.05 level.

	M1	M2	M3	M4	P1	P5	P6	P7	P9	P10	P11	S1	S3	S5
M2	**1.00**													
M3	**1.00**	0.00												
M4	**1.00**	0.00	0.00											
P1	0.10	**0.41**	**0.41**	0.41										
P5	**0.35**	-0.43	-0.43	-0.43	0.04									
P6	0.58	**0.38**	**0.38**	**0.38**	0.18	-0.14								
P7	1.00	1.00	1.00	1.00	**0.23**	-0.25	-0.43							
P9	1.00	1.00	1.00	1.00	0.40	0.17	0.23	1.00						
P10	1.00	1.00	1.00	1.00	0.22	0.00	0.00	1.00	1.00					
P11	1.00	1.00	1.00	1.00	**-0.17**	-1.00	0.09	1.00	1.00	1.00				
S1	0.00	1.00	1.00	1.00	-0.03	0.00	0.42	1.00	1.00	1.00	0.00			
S3	**1.00**	1.00	1.00	1.00	0.26	0.00	-0.20	1.00	1.00	1.00	1.00	1.00		
S5	1.00	1.00	1.00	1.00	0.31	-0.11	**0.17**	1.00	1.00	1.00	1.00	1.00	1.00	
MU1	1.00	1.00	1.00	1.00	0.90	0.81	0.87	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 4-12. F_{ST} values between populations of species in the *mayaguana* subcluster calculated for the A+T Rich mt DNA.

Values in **boldface** type are significant at the 0.05 level.

	M1	M2	M3	M4	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	S1	S2
M2	-0.33																
M3	0.88	1.00															
M4	-0.33	0.00	1.00														
P1	**0.95**	1.00	1.00	1.00													
P2	**0.95**	1.00	1.00	1.00	0.00												
P3	**0.95**	1.00	1.00	1.00	0.00	0.00											
P4	**0.69**	0.55	0.07	0.55	0.00	0.00	0.00										
P5	**0.91**	1.00	1.00	1.00	0.00	0.00	0.00	-0.29									
P6	**0.93**	1.00	1.00	1.00	0.00	0.00	0.00	-0.13	0.00								
P7	0.88	1.00	1.00	1.00	0.00	0.00	0.00	-1.00	0.00	0.00							
P8	0.88	1.00	1.00	1.00	0.00	0.00	0.00	-1.00	0.00	0.00	0.00						
P9	**0.88**	1.00	1.00	1.00	1.00	1.00	1.00	-0.17	1.00	1.00	1.00	1.00					
P10	0.90	0.87	0.71	0.87	0.86	**0.86**	0.86	0.22	0.67	0.77	0.33	0.33	-1.00				
P11	0.89	1.00	1.00	1.00	0.00	0.00	0.00	-1.00	0.00	0.00	0.00	0.00	1.00	0.33			
S1	**0.88**	**0.96**	**0.96**	**0.96**	**0.99**	**0.99**	**0.99**	**0.95**	0.98	**0.98**	**0.96**	**0.96**	**0.96**	**0.97**	**0.96**		
S2	**0.98**	0.82	0.50	0.82	0.08	0.08	0.08	-0.02	-0.20	-0.05	-0.83	-0.83	0.08	0.24	-0.83	**0.97**	
S3	0.84	0.00	-0.64	0.00	**0.47**	**0.47**	**0.47**	0.18	0.00	0.25	-1.00	-1.00	-1.00	0.00	-1.00	**0.92**	0.29
S4	0.64	1.00	1.00	1.00	1.00	1.00	1.00	0.69	1.00	1.00	1.00	1.00	1.00	0.89	1.00	**0.96**	0.87
S5	0.94	1.00	1.00	1.00	1.00	1.00	1.00	-0.40	1.00	1.00	1.00	1.00	1.00	**0.60**	1.00	**0.96**	0.31
S6	0.85	1.00	1.00	1.00	0.00	0.00	0.00	-1.00	0.00	0.00	0.00	0.00	1.00	0.33	0.00	**0.96**	-0.83
MU1	**0.91**	1.00	1.00	1.00	1.00	1.00	1.00	0.18	1.00	1.00	1.00	1.00	1.00	-1.00	1.00	**0.96**	0.31
MU2	0.99	1.00	1.00	1.00	1.00	1.00	1.00	**0.94**	1.00	1.00	1.00	1.00	1.00	0.98	1.00	**0.96**	**0.98**

Table 4-12. Continued.

	S3	S4	S5	S6	MU1
S4	**-1.00**				
S5	**-0.64**	1.00			
S6	**-1.00**	1.00	1.00		
MU1	**-0.64**	1.00	1.00	1.00	
MU2	0.81	1.00	1.00	1.00	1.00

Table 4-13. F_{ST} values between populations of species in the *mayaguana* subcluster calculated for the *hb* gene. Values in **boldface** type are significant at the 0.05 level.

	M1	M2	M3	M4	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	S1
M2	-1.00															
M3	0.82	1.00														
M4	-1.00	0.00	1.00													
P1	**0.92**	0.93	0.95	0.93												
P2	0.85	1.00	1.00	1.00	0.58											
P3	**0.91**	1.00	1.00	1.00	**0.67**	1.00										
P4	**0.87**	**0.89**	0.93	**0.89**	**0.54**	-3333.3	0.25									
P5	0.81	0.60	0.74	0.60	0.64	-2.00	0.00	0.06								
P6	**0.71**	0.48	**0.66**	0.48	**0.72**	-0.38	0.03	0.22	0.03							
P7	0.87	1.00	1.00	1.00	**0.77**	1.00	1.00	0.50	-1.00	-0.22						
P8	0.89	1.00	1.00	1.00	0.74	1.00	**1.00**	0.80	0.33	0.15	1.00					
P9	0.90	1.00	1.00	1.00	0.78	1.00	1.00	0.82	0.40	0.48	1.00	1.00				
P10	0.80	0.56	0.69	0.56	**0.63**	-0.33	**0.31**	0.33	-0.08	-0.04	-0.60	-0.33	0.00			
P11	0.85	1.00	1.00	1.00	0.58	0.00	1.00	-3333.3	-1.00	-0.38	1.00	1.00	1.00	-0.33		
S1	**0.48**	-0.64	-0.29	-0.64	0.68	-1.00	0.25	0.22	-0.02	0.21	-0.93	-0.80	-0.80	-0.02	-1.00	
S2	**0.62**	0.38	0.57	0.38	**0.36**	-1.00	-0.06	-0.11	-0.12	0.06	-0.44	-0.08	0.13	0.00	-1.00	0.25
S3	**0.70**	**0.23**	**0.44**	**0.23**	**0.86**	-0.25	**0.71**	**0.57**	0.47	0.41	**0.29**	**0.44**	0.50	0.50	0.00	0.10
S4	0.87	1.00	1.00	1.00	0.88	1.00	1.00	**0.67**	0.25	0.15	1.00	1.00	1.00	**0.27**	1.00	-0.69
S5	0.87	1.00	1.00	1.00	0.58	1.00	1.00	0.50	-0.20	**0.08**	1.00	1.00	1.00	-0.14	1.00	-1.00
S6	0.85	1.00	1.00	1.00	0.88	1.00	1.00	**0.67**	0.33	**0.27**	1.00	1.00	1.00	0.33	1.00	-0.74
MU1	0.85	1.00	1.00	1.00	**0.26**	0.00	1.00	-1.00	-1.00	-0.22	1.00	1.00	1.00	-0.60	0.00	-1.08
MU2	0.83	1.00	1.00	1.00	**0.05**	0.00	1.00	-1.00	-1.00	-0.38	1.00	1.00	1.00	-1.00	0.00	-1.08

Table 4-13. Continued.

	S2	S3	S4	S5	S6	MU1
S3	**0.37**					
S4	-0.08	**-0.43**				
S5	-0.53	**0.29**	1.00			
S6	-0.04	**-1.00**	0.00	1.00		
MU1	-0.86	**0.17**	1.00	1.00	1.00	
MU2	-1.17	**0.17**	1.00	1.00	1.00	0.00

Table 4-14. F_{ST} values between populations of species in the *mayaguana* subcluster calculated for the *mam* gene. Values in **boldface** type are significant at the 0.05 level.

	M1	M2	M3	M4	P1	P2	P3	P5	P6	P7	P9	P10
M2	0.00											
M3	0.00	0.00										
M4	0.00	0.00	0.00									
P1	0.01	-0.57	-0.57	-0.57								
P2	1.00	1.00	1.00	1.00	**0.66**							
P3	0.77	0.33	0.33	0.33	0.33	0.60						
P5	0.00	0.00	0.00	0.00	**-0.10**	1.00	0.67					
P6	0.00	-1.00	-1.00	-1.00	-0.07	0.50	0.20	-0.20				
P7	0.00	0.00	0.00	0.00	**-0.57**	1.00	0.33	0.00	-1.00			
P9	1.00	1.00	1.00	1.00	**-0.04**	1.00	-1.00	1.00	**-0.33**	1.00		
P10	0.00	0.00	0.00	0.00	-0.01	1.00	0.67	0.00	-0.20	0.00	1.00	
P11	0.00	0.00	0.00	0.00	-8.40	1.00	-3.4E+07	0.00	-6.7E+04	0.00	0.00	0.00
S1	1.00	1.00	1.00	1.00	**0.95**	1.00	0.97	1.00	0.94	1.00	1.00	1.00
S2	1.00	1.00	1.00	1.00	0.28	1.00	0.25	1.00	**0.33**	1.00	0.00	1.00
S3	1.00	1.00	1.00	1.00	**0.42**	1.00	0.67	1.00	0.37	1.00	1.00	1.00
S5	1.00	1.00	1.00	1.00	0.62	1.00	0.71	1.00	0.50	1.00	1.00	1.00
S6	0.00	0.00	0.00	0.00	**-0.57**	1.00	0.33	0.00	-1.00	0.00	1.00	0.00

Table 4-14 continued.

	P11	S1	S2	S3	S5
S1	1.00				
S2	0.00	1.00			
S3	0.00	1.00	1.00		
S5	0.00	1.00	1.00	1.00	
S6	0.00	1.00	1.00	1.00	1.00

Table 4-15. F_{ST} values between populations of species in the *mayaguana* subcluster calculated for the *vg* gene. Values in **boldface** type are significant at the 0.05 level.

	M1	M2	M3	M4	P1	P5	P6	P7	P8	P9	P10	P11	S1	S3	S4	S5	S6	MU1
M2	0.00																	
M3	0.00	0.00																
M4	1.00	1.00	1.00															
P1	**0.89**	0.88	0.88	**0.92**														
P5	0.00	-1.00	-1.00	0.00	0.49													
P6	0.45	**0.14**	**0.14**	0.40	**0.29**	-0.20												
P7	1.00	1.00	1.00	1.00	**0.94**	0.56	0.60											
P8	1.00	1.00	1.00	1.00	-1.00	-1.00	-1.00	1.00										
P9	1.00	1.00	1.00	1.00	0.78	0.00	0.00	1.00	1.00									
P10	0.40	-0.20	-0.20	0.14	0.78	0.00	0.16	0.50	-0.20	0.14								
P11	1.00	1.00	1.00	1.00	**0.75**	-1.00	-0.50	1.00	1.00	1.00	-1.00							
S1	1.00	1.00	1.00	1.00	0.98	0.87	**0.87**	1.00	1.00	1.00	0.80	1.00						
S3	1.00	1.00	1.00	1.00	0.78	0.00	0.08	1.00	1.00	1.00	0.00	0.00	1.00					
S4	1.00	1.00	1.00	1.00	0.92	0.33	0.40	1.00	1.00	1.00	0.14	1.00	1.00	1.00				
S5	1.00	1.00	1.00	1.00	0.78	0.00	0.00	1.00	1.00	1.00	0.00	1.00	1.00	1.00	1.00			
S6	1.00	1.00	1.00	1.00	0.92	0.33	0.40	1.00	1.00	1.00	0.14	1.00	1.00	1.00	1.00	1.00		
MU1	1.00	1.00	1.00	1.00	0.98	0.87	**0.88**	1.00	1.00	1.00	0.80	1.00	1.00	1.00	1.00	1.00	1.00	
MU2	1.00	1.00	1.00	1.00	0.99	0.88	**0.88**	1.00	1.00	1.00	0.81	1.00	1.00	1.00	1.00	1.00	1.00	1.00

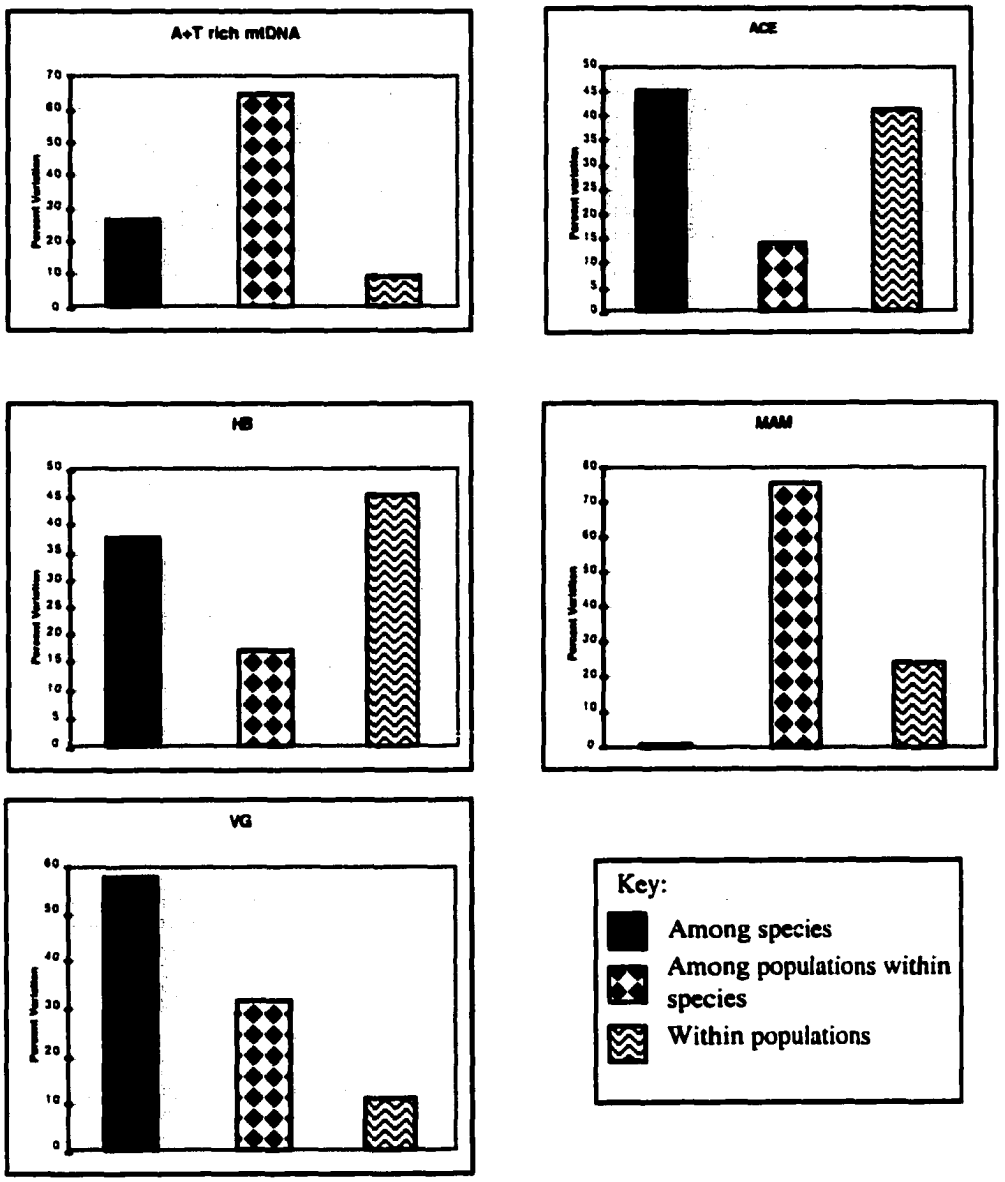


Figure 4-9. Percent of the total variation attributable to among species differences, among populations within species differences and within species differences for each of the individual data partitions.

species *D. mayaguana*, there were no characters that aggregated samples by species and very few that even grouped together any of the geographic populations (Tables 4-5 through 4-9). While, on the whole, the F_{ST} values show the greatest amount of variation to arise from between population differences, we also find some anomalous results, such as the fact that approximately 64% of the total sequence difference in the comparison of *D. parisiena* from Kingston, Jamaica and *D. straubae* from NE of Barahona, Dominican Republic using the*hb* genecan be attributed to within population differences (Table 4-13). When one factors in the paucity of morphological differentiation in combination with the production of fertile hybrids, one might conclude that these two taxa are both members of the same species. However, the cytological and ecological differentiation argue against this.

In order to more fully explore the relationships among the flies in this species subcluster, especially the relationships of the different geographic populations of *D. parisiena* and *D. straubae*, it will be necessary to ascertain exactly how much interspecific mating is taking place in nature and also to examine additional DNA sequences from these flies. Augmenting our data set with additional mitochondrial DNA sequences and several more protein coding genes may shed some light on this absorbing question.

References

Allard, M. W, J. S. Farris, and J. M. Carpenter. 1999. Congruence among mammalian mitochondrial genes. Cladistics 15:75-84.

Ashburner, M. 1989. Drosophila A Laboratory Handbook. Cold Spring Harbor Press, New York.

Ashburner, M. 1990. Genbank accession number X54118.

Ashburner, M. and H.L. Carson. 1983. A checklist of maps of polytene chromosomes of drosophilids. Dros. Inf. Serv. 59: 148-151.

Baker, R. H. and R. DeSalle. 1997. Multiple sources of character information and the phylogeny of Hawaiian Drosophilids. Syst. Biol. 46:654-673.

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

Barker, J.S.F. 1982. Population genetics of *Opuntia* breeding *Drosophila* in Australia. Pp. 209-224 in Ecological Genetics and Evolution: the Cactus-Yeast-*Drosophila* Model System (J.S.F. Barker and W.T. Starmer, eds). Academic Press. New York.

Barker, J.S.F. 1990. Experimental analysis of habitat selection and maintenance of

genetic variation. Pp. 161-176 in Ecological and Evolutionary Genetics of *Drosophila* (J.S.F. Barker, W.T. Starmer and R. J. MacIntyre, eds.). Plenum Press. New York.

Bettler, D., S. Pearson and B. Yedvobnick. 1996. The nuclear protein encoded by the*Drosophila* neurogenic gene *mastermind* is widely expressed and associates with specific chromosomal regions. Genetics 143:859-875.

Bloomer, P., and T. M. Crowe. 1998. Francolin phylogenetics: Molecular, morphobehavioral, and combined evidence. Mol. Phylogenet. Evol. 9:236-254.

Bock, I. 1984. Interspecific hybridization in the genus *Drosophila*. Evol. Biol. 18: 41-70.

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., R. DeSalle, and A. Vogler. 1996. Gene trees, species trees, and systematics: A perspective. Annu. Rev. Ecol. Syst. 27:423-450.

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 data. Mol. Biol. Evol. 13 (9): 1224-1232.

Carson, H.L. 1970. Chromosomal tracers of evolution. Science 168: 1414-1418.

Carson, H.L. 1987. Tracing ancestry with chromosomal sequences. Trends in Ecol and Evol 2: 203--207.

Carson, H.L. and K.Y. Kaneshiro. 1976. Drosophila of Hawaii: systematics and ecological genetics. Ann. Rev. Ecol. Syst. 7:311-345.

Carson, H.L. and A.R. Templeton.1984. Genetic revolutions in relation to speciation phenomenon: the founding of new populations. Ann Rev. Ecol. Syst. 15:97-131.

Cho, S., A. Mitchell, J. C. Regier, C. Mitter, R. W. Poole, T. P. Friedlander and S. Zhao. 1995. A highly conserved nuclear gene for Low-level phylogenetics: *Elongation Factor-1a* recovers morphology-based tree for Heliothine moths. Mol. Biol. Evol. 12:650-656.

Cohn, V.H. and G. P. Moore. 1988. Organization and evolution of the alcohol dehydrogenase gene in *Drosophila.* Mol. Biol. Evol. 5: 154-166.

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

Coyne, J. and A. Orr. 1989. Patterns of speciation in *Drosophila*. Evolution 43:362- 381.

Coyne, J. and A. Orr. 1997. Patterns of speciation in *Drosophila* revisited. Evolution 51:295-303.

Danforth, B. N., and S. Ji. 1998. Elongation Factor-1α occurs as two copies in bees: Implications for phylogenetic analysis of EF-1α sequences in insects. Mol. Biol. Evol. 15:225-235.

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

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., A. K. Williams, and M. George. 1993. Isolation and characterization of animal mitochondrial DNA. Pp. 176-204. In E. A. Zimmer, T. J. White, R. L. Cann, and A. C. Wilson (eds.), Methods in Enzymology. Vol. 24. Academic Press. New York.

Eriksson, T. 1997. AutoDecay Version 2.9.8 (Hypercard stock disk distributed by the author). Stockholm: Botoniska Institutionen. Stockholm University.

Etges, W.J. 1990. Direction of life history evolution in *Drosophila mojavensis*. Pp. 37-56 in Ecological and Evolutionary Genetics of *Drosophila* (J.S.F. Barker, W.T. Starmer and R. J. MacIntyre, eds.). Plenum Press. New York.

Etges, W.J. and W.B. Heed. 1987. Sensitivity to larval density in populations of *Drosophila mojavensis*: Influences of host plant variation on components of fitness. Oecologia 71:375-381.

Etges, W.J. and C.S. Klassen. 1989. Influences of atmospheric ethanol on adult *Drosophila mojavensis*: Altered metabolic rates and increases in fitness among populations. Physiol. Zool. 62:170-193.

Excoffier, L., Smouse, P. and Quattro, J. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitichondrial DNA restriction data. Genetics 136:343-359.

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

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

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

Fogelman, J.C. and J.R. Abril. 1990. Ecological and evolutionary importance of host plant chemistry. Pp. 121-144 in Ecological and Evolutionary Genetics of *Drosophila* (J.S.F. Barker, W.T. Starmer and R. J. MacIntyre, eds.). Plenum Press. New York.

Fontevila, A., M. Wasserman, C. Pla, L. Pilares, R. de Armengol, M. P. Suyo, A. Sanchez, J. Vasquez, A. Ruiz, and J. L. Garcia. 1990. Description and evolutionary relationships of two species of the *Drosophila mulleri* cluster (Diptera: Drosophilidae). Ann. Entomol. Soc. Am. 83:444.

Fournier, D., F. Karch, J.-M. Bride, L. Hall, J.-B. Berge and P. Spierer. 1989. *Drosophila melanogaster* acetylcholinesterase gene: structure, evolution and mutations. J. of Mol. Biology 210:15-22.

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

Gibson, T., D. Higgins and J. Thompson. 1994. ClustalX. Program and documentation available at ftp://ftp.ebi.ac.uk/pub/software/mac/clustalw/clustalx.

Heed, W. B. 1971. Host plant specificity and speciation I: Hawaiian *Drosophila*. Taxon 20: 115-121.

Heed, W.B. and D. A. Grimaldi. 1991. revision of the morphocryptic, Caribbean *mayaguana species* subcluster in the *Drosophila repleta* group (Diptera: Drosophilidae). American Museum Novitates. 2999.

Heed, W. B., A. Sanchez, R. Armengol and A. Fontdevila. 1990. Genetic differentiation among island populations and species of cactophilic *Drosophila* in the West Indies. pp. 447-490 in Ecological and Evolutionary Genetics of *Drosophila* (J.S.F. Barker, W.T. Starmer and R. J. MacIntyre, eds.). Plenum Press. New York.

Helm-Bychowski, K., and J. Cracraft. 1993. Recovering phylogenetic signal from DNA sequences: Relationships within the corvine assemblage (Class Aves) as inferred from complete sequences of the mitochondrial DNA cytochrome-b gene. Mol. Biol. Evol. 10:1196-1214.

Hennig, W. 1966. Phylogenetic Systematics. University of Illinois Press. Urbana.

Jeffs, P.S., E. C. Holmes, and M. Ashburner. 1994. The molecular evolution of the alcohol dehydrogenase and alcohol dehydrogenase-related genes in the *Drosophila melanogaster* species subgroup. Mol. Biol. Evol. 11 (2): 287-304.

Kambysellis, M.D., K. F. Ho, E. M. Craddock, F. Piano, M. Parisi and J. Cohen. 1995. Patterns of ecological shifts in the diversification of

Hawaiian *Drosophila* inferred from a molecular phylogeny. Current Biology 5:1129-1139.

Kaneshiro, K.Y., R. Gillespie andH. L. Carson. 1995. Chromosomes and male genitalia of Hawaiian *Drosophila*: tools for interpreting phylogeny and geography. Pp. 52-72 in Hawaiian Biogeography: Evolution on a Hotspot, eds. W. L. Wagner and V. A. Funk. Smithsonian Institution Press, Washington DC.

Kitagawa, H. 1998. Genbank accession numbers AB019507, AB019546-AB019550.

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

Knight, A., and D. P. Mindell. 1993. Substitution bias, weighting of DNA sequence evolution, and the phylogenetic position of fea's viper. Syst. Biol. 42:18-31.

Koepfer, H.R. 1987. Selection for sexual isolation between geographical forms of *Drosophila mojavensis*. I. Interaction between the selected forms. Evolution 41.

Krimbas, C. and J.R. Powell. 1992. Drosophila Inversion Polymorphism. CRC Press, Boca Raton FL.

Kwiatowski, J. and F.J. Ayala, (1999). Phylogeny of *Drosophila* and related genera: conflict between molecular and anatomical analyses. Mol. Phyl.Evol. (in press).

Lakovaara, S. and A. Saura, 1982. Evolution and speciation in the *Drosophila obscura* species group. Pp. 1-59 in The genetics and Biology of *Drosophila* Vol 3b, eds. M. Ashburner, H.L. Carson and J.N. Thompson. Academic Press, New York.

Lemunier, F.and M. Ashburner. 1976. Relationships in the melanogaster species subgroup of the genus *Drosophila* (*Sophophora*). II. Phylogenetic relationships of six species based upon polytene banding sequences. Proc. R. Soc. Lond. B. 193: 257-294.

Lemeunier, F., J. David, L. Tsacas and M. Ashburner. 1986. The *melanogaster* species group. Pp. 147-256 in The Genetics and Biology of *Drosophila* (M. Ashburner et al., eds.), vol 3e. Academic Press, London and Orlando.

Lewis, D.L., C.L. Farr, A.L Farquhar and L.S. Kaguni. 1994. Sequence, organization, and evolution of the A + T region of *Drosophila melanogaster* mitochondrial DNA. Mol. Biol. Evol. 11(3):523-538.

Maddison, W.P. and D.R. Maddison. 1992. *MacClade version 3.* Sinauer Associates, Inc. Sunderland MA.

Markow, T.A. 1982. Mating systems of cactophilic *Drosophila*. Pp. 273-290 in Ecological Genetics and Evolution: the Cactus-Yeast-*Drosophila* Model System (J. S. F. Barker and W. T. Starmer, eds.). Academic Press. New York.

Marin, I., A. Ruiz, C. Pla and A. Fontdevila. 1993. Reproductive relationships among ten species of the *Drosophila repleta* group from South America and the West Indies. Evolution. 47:1616-1624.

Martin, A. 1995. Mitochondrial DNA sequence evolution in sharks: Rates, patterns, and phylogenetic inferences. Mol. Biol. Evol. 12:1114-1123.

Martin, A. P., and E. Bermingham. 1998. Systematics and evolution of lower Central American cichlids inferred from analysis of cytochrome b gene sequences. Mol. Phylogenet. Evol. 9:192-203.

McDonald, J.H. and M. Kreitman. 1991. Adaptive protein evolution at the Adh locus in *Drosophila*. Nature 351 (6328): 652-654.

Moses K., U. Heberlein, M. Ashburner. 1990. The Adh gene promoters of *Drosophila melanogaster* and *Drosophila orena* are functionally conserved and share features of sequence structure and nuclease-protected sites. Mol. Cell. Biol. 10(2):539-48.

Murphy, W. J., and G. E. Collier. 1997. A molecular phylogeny for Aplocheiloid fishes (Atherinomorpha, Cyprinodontiformes): The role of vicariance and the origins of annualism. Mol. Biol. Evol. 18:790-799.

Newfield, S.J., Smoller, D.A. and Yedvobnick, B. 1991. Interspecific comparison of the unusually repetitive *Drosophila* locus *mastermind.* J. Mol. Evol. 32:415-420.

Nixon, K. and J. Carpenter, 1996. On simultaneous analysis. Cladistics 12: 221-241.

Nurminsky,D.I., E. N. Moriyama, E. R. Lozovskaya and D. L. Hartl. 1996. Molecular phylogeny and genome evolution in the *Drosophila virilis* species group duplications of the alcohol dehydrogenase gene. Mol. Biol. Evol. 13 (1): 132-149.

Paques F., M. L. Samson , P. Jordan and M. Wegnez. 1995. Structural evolution of the *Drosophila* 5S ribosomal genes. J. Mol. Evol. 41(5): 615-21.

Patterson, J.T. 1947. Sexual isolation in the *mulleri* subgroup. Univ. Tex. Pub. 4720: 32-40.

Patterson, J.T. and W.S. Stone. 1952. Evolution in the Genus *Drosophila.* MacMillan. New York.

Pelandakis, M. and M. Solignac. 1993. Molecular phylogeny of *Drosophila* based on ribosomal RNA sequences. J. Mol. Evol. 37 (5): 525-543.

Powell, J.R. 1996. The Drosophila Model. Oxford University Press, Oxford.

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

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

Richardson, R.H., M.E. Richardson and P.E. Smouse. 1975. Evolution of electrophoredtic mobility in the *Drosophila mulleri* complex. Pp. 533-545 in Isozymes IV: Genetics and Evolution. (C.L. Markert, ed.). Academic Press. New York.

Richardson, R.H. and P. E. Smouse. 1976. Patterns of molecular variation. I. Interspecific comparison of electromorphs in the *Drosophila mulleri* complex. Biochem. Genet. 14:447-465.

Richardson, R.H., P.E. Smouse and M.E. Richardson. 1977. Patterns of molecular variation. II. Associations of electrophoretic mobility and larval substrate within species of the *Drosophila mulleri* complex. Genetics 85:141-154.

Ruiz, A. and M. Wasserman. 1993. Evolutionary cytogenetics of the *Drosophila buzzatii* species complex. Heredity 70:582-596.

Ruiz Linares, A., J. M. Hancock and G. A. Dover. 1991. Secondary structure constraints on the evolution of *Drosophila* 28S ribosomal RNA expansion segments. J. Mol. Biol. 219 (3): 381-390.

Russo, C.A.M., N. Takezaki and M. Nei. 1995. Molecular phylogeny and divergence times of Drosophilid species. Mol. Biol. Evol. 12:391-404.

Samson, M.L. and M. Wegnez. 1984. The 5S ribosomal genes in the *Drosophila melanogaster* species subgroup. Nucleotide sequence of a 5S unit from *Drosophila simulans* and *Drosophila teissieri* . Nucleic Acids Res. 12 (2): 1003-1014.

Sanderson, M. and M. Donoghue. 1992. Patterns of variation in levels of homoplasy. Evolution 43: 1781-1795.

Schneider, S., J.M. Kueffer, D. Roessli and L. Excoffier. 1997. *Arlequin ver 1.1* Geneva, Switzerland. URL: http:/anthropologie.unige.ch/arlequin.

Shibata,H. and T. Yamazaki. 1995. Molecular evolution of the duplicated *Amy* locus in the *Drosophila melanogaster* species subgroup: concerted evolution only in the coding region and an excess of nonsynonymous substitutions in speciation.

Simon, C., F. Frati, A., Beckenbach, B. Crespi, H., Liu, and P. Flook. 1994. Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Amer. 87: 651-710.

Simons, A. M., and R. L. Mayden. 1998. Phylogenetic relationships of the Western North American Phoxinins (Actinopterygii: Cyprinidae) as inferred from mitochondrial 12S and 16S ribosomal RNA sequences. Mol. Phylogenet. Evol. 9:308-329.

Spicer, G.S. 1995. Phylogenetic utility of the mitochondrial cytochrome oxidase gene: molecular evolution of the *Drosophila buzzatii* species complex. J. Mol. Evol. 41:749-759.

Spicer, G.S. and S. Pitnick. 1996. Molecular systematics of the *Drosophila hydei* subgroup as inferred from mitochondrial DNA sequences. J. Mol. Evol. 43:281-286.

Sullivan, D.T., P.W. Atkinson, C.A. Bayer and Marilyn Menotti-Raymond. 1990. The evolution of Adh expression in the *repleta* group of *Drosophila*. pp. 407-418 in Ecological and Evolutionary Genetics of *Drosophila* (J.S.F. Barker, W.T. Starmer and R. J. MacIntyre, eds.). Plenum Press. New York.

Swofford, D., 1998. PAUP*: Phylogenetic analysis using parsimony (* and other methods), 4.0 beta version. Sinauer, Sunderland MA.

Swofford, D. L. 1999. PAUP 4.01: Phylogenetic analysis using parsimony (and other methods), version 4.01. Sinauer.

Throckmorton, L. H. 1982a. Pathways of evolution in the genus *Drosophila* and the founding of the *repleta* group. Pp. 33-48 in Ecological Genetics and Evolution: The Cactus-Yeast-*Drosophila* Model System (J. S. F. Barker and W. T. Starmer, eds.). Academic Press. New York.

Throckmorton, L.H. 1982b. The *virilis* species group. Pp. 227-296 in The Genetics and Biology of *Drosophila* (M. Ashburner et al., eds.), vol 3b. Academic Press, London and New York.

Tominaga H. and S. Narise. 1995. Sequence evolution of the Gpdh gene in the *Drosophila virilis* species group. Genetics 141 (1): 223-236.

Treier, M., C. Pfeifle, and D. Tautz. 1989. Comparison of the gap segmentation gene*hunchback* between *Drosophila melanogaster* and *Drosophila virilis* reveals novel modes of evolutionary change. The EMBO Journal.

Vilela, C. R. 1983. A revision of the *Drosophila repleta* species group (Diptera, Drosophilidae). Rev. Brasil. Entomol. 27:1-114.

Vogler, A.P., R. DeSalle, T. Assmann, C. B. Knisley, and T. D. Schultz. 1993. Molecular population genetics of the endangered tiger beetle *Cicindela dorsalis*. (Coleoptera: Cicindelidae). Ann. Entomol. Soc. Am. 86:142-152.

Ward, C.L. and W.S. Stone. 1952. Studies in the *repleta* group: the *melanopalpa* subgroup. Univ. Texas Publ. 5204.

Wasserman, M. 1954. Cytological studies on the *repleta* group. Univ. Texas Publ. 5422.

Wasserman, M. 1960. Cytological and phylogenetic relationships in the *repleta* group of the genus *Drosophila*. Proc. Natl. Acad. Sci. 46.

Wasserman, M. 1962a. Cytological studies of the *repleta* group of the genus *Drosophila*. IV. The *hydei* subgroup. Univ. Texas Publ. 6205.

Wasserman, M. 1962b. Cytological studies of the *repleta* group of the genus *Drosophila*. VI. The *fasciola* subgroup. Univ. Texas Publ., 6205.

Wasserman, M. 1982a. Evolution of the *repleta* group. Pp. 61-139 in The Genetics and Biology of *Drosophila* (M. Ashburner, H.L. Carson and J.N. Thompson, eds.). Academic Press. New York.

Wasserman, M. 1982b. Cytological evolution in the *Drosophila repleta* species group. pp. 49-64 Ecological Genetics and Evolution: the Cactus-Yeast-*Drosophila* Model System .(J. S. F. Barker and W. T. Starmer,eds.). Academic Press. New York.

Wasserman, M. 1992. Cytological evolution of the *Drosophila repleta* species group. pp. 455-552. In C. B. Krimbas and J. R. Powell (eds.), Drosophila Inversion Polymorphism. CRC Press. Boca Raton, FL.

Wasserman, M. and F. Wasserman. 1992. Inversion polymorphism in island species of*Drosophila*. Pp. 351-381 in Evolutionary Biology Vol. 26 (M.K. Hecht, B. Wallace and R.J. MacIntyre, eds.). Plenum Press. New York.

Wharton, L.T. 1942. Analysis of the *repleta* group of *Drosophila*. Univ. Texas Publ. 4228.

Wharton, L.T. 1944. Interspecific hybridization of the *repleta* group. Univ. Texas Publ. 4445.

Williams, J.A., S.W. Paddock, K. Vorwerk and S.B. Carroll. 1994. Organization of wing formation and induction of a wing-patterning gene at the dorsal/ventral compartment boundary. Nature Vol. 368:299-305.

Yang, Z. 1998. On the best evolutionary rate for phylogenetic analysis. Syst. Biol. 47:125-133.

Yang, H.Y. and M.R. Wheeler. 1969. Studies on interspecific hybridization within the picture-winged group of endemic Hawaiian *Drosophila*. Univ. Texas Publ. 6918:133-170.

Yoder, A. D., M. Cartmill, M. Ruvolo, K. Smith, and R. Vilgalys. 1996. Ancient single origin for Malagasy primates. Proc. Natl. Acad. Sci. 93:5122-5126.

Yoon, Y.S. and R. Richardson. 1976. Evolution of the Hawaiian Droosphilidae II. Patterns and rates of chromosome evolution in the *Antopocerus* phylogeny. Genetics 83: 827-843.

Zouros, E. 1973. Genic differentiation associated with the early stages of speciation in the*mulleri* subgroup of *Drosophila*. Evolution 27:601-621.