This is to certify that the

dissertation entitled

CHARACTERIZATION OF THE RIBOSOMAL DNA
OF THE GENUS RHAGOLETIS
(DIPTERA: TEPHRITIDAE)

presented by

YUE MING

has been accepted towards fulfillment
of the requirements for

Ph. D. degree in ENTOMOLOGY

Gary L. Bush
Major professor

Date 27 OCT. 1995

CHARACTERIZATION OF THE RIBOSOMAL DNA OF THE GENUS *RHAGOLETIS* (DIPTERA: TEPHRITIDAE)

By

Yue Ming

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Entomology

1995

ABSTRACT

CHARACTERIZATION OF THE RIBOSOMAL DNA OF THE GENUS *RHAGOLETIS* (DIPTERA: TEPHRITIDAE)

By

Yue Ming

Although the Nearctic species of *Rhagoletis* have been well studied with respect to morphology and allozymes, there are still unanswered questions regarding the differentiation of sibling species, placement of certain species, and relationships among the existing species groups. In order to gain new information on these subjects, I have characterized the *Rhagoletis* ribosomal DNA, especially the non-coding spacers such as the internal transcribed spacers (ITS).

I have presented the ITS sequences of four North American sibling species of the *R. cingulata* species group. The inter-specific variation in this group is not significantly higher than the intra-specific variation. Consequently, the ITS sequences are of limited application for inferring phylogeny of the members in this group. However, several molecular markers in the ITS sequences have been described which can be potentially useful for differentiating some of the sibling species in this group. A few highly conserved secondary structure elements in the ITS regions have also been described and compared with those in *Drosophila*.

The ITS sequences have been obtained for eight additional *Rhagoletis* species, including *pomonella, cornivora, completa, juniperina, fausta, electromorpha, basiola*, and *striatella*, and a phylogenetic analysis was performed. This study indicated that *R.*

cornivora belongs to the *pomonella* group; *R. juniperina* was removed from the *tabellaria* group and may be more closely related to the *pomonella* group; close relatives of the *cingulata* group are more likely closely aligned with the *suavis* group rather than the *pomonella* group; *R. fausta* may be related with the *tabellaria* group; *R. basiola* and *R. striatella* ITS sequences are highly divergent from other species analyzed, indicating that *Rhagoletis* may not be monophyletic.

A genomic library for *R. pomonella* was constructed and several rDNA clones identified. Furthermore, the region containing the intergenic spacer and external transcribed spacer of rDNA from two *R. cingulata* flies of different host plants was PCR amplified and cloned. The regions were partially sequenced and found to be significantly divergent between the two *R. cingulata* of different host plants. Whether this observed divergence is related to the different fly host origins is an intriguing question worth further investigation.

DEDICATION

To the memory of my father, brother and nephew

ACKNOWLEDGEMENTS

I would like to thank my major professors Drs. Guy L. Bush and Frederick W. Stehr for their effort throughout all the different stages of my development as a scientist and person while at Michigan State University (MSU). Special thanks also to my Doctoral committee consisting of Drs. Alexander S. Raikhel, J. Mark Scriber and Donald O. Straney for their patience, encouragement and support. I thank Dr. Hugh M. Robertson for providing the PCR primers, and Drs. Stewart H. Berlocher, Randy Cooper, Ranjan Gupta, James Nugent, Jerry A. Payne, Adam Peters, Ron J. Prokopy, Gary J. Steck, George C. Steyskal, William J. Turner and John Wilterding for collecting the specimens used in this Dissertation. Special thanks to Dr. James J. Smith for helping me getting started in Dr. Bush's laboratory and for his technical assistance. I thank John Jenkins and Judith Sirota for their friendship over the years and for their helpful suggestions and specimen collections. I thank Dr. David R. Engelke from the University of Michigan (UM) for critically reading the secondary structure part in Chapter II and Dr. John P. Langmore for allowing me to use his laboratory computers at UM. I would also like to thank the Family Housing Community Center at UM for providing the computer and study facilities. I thank my family in China and my in-laws in Pullman, WA, especially my parents-in-law Mr. and Mrs. Krikor O. Bedoyan, for their support and encouragement that made my life at MSU more enjoyable. Very special thanks to my husband Jirair K. Bedoyan for his support, encouragement and assistance throughout all this time. His love and faith kept me on course. Last and most importantly, I thank God for giving me a healthy baby, Sarah M. Bedoyan, otherwise it would have been impossible for me to complete my Ph.D. program at MSU.

TABLE OF CONTENTS

IV.

IDEN

V.

SPACI

VI.

Tabl

Tabl

Tabl

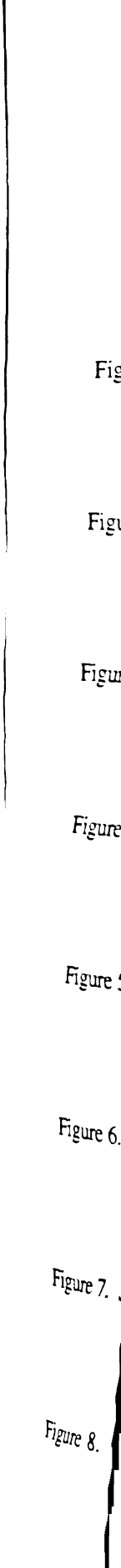
Table

Table

Table 7

Table 8.

LIST OF TABLES



LIST OF FIGURES

Figure 1

Figure 2

Figure 2

Figure 2

wide

1966

1991

Rhag

and to

new h

Morph

rapid s

the wil

newly e

has also

1974; 19

certain *R*

overview

chapter as

also offer

dissertatio

especially

species gr

CHAPTER I

INTRODUCTION

The genus *Rhagoletis* Loew currently has more than 60 described species, and is widely distributed over the Palearctic (Rohdendorf 1961; Kandybina 1972), Nearctic (Bush 1966; Berlocher and Bush 1982) and Neotropical regions (Foote 1981; Frias and Martines 1991). Because larvae of *Rhagoletis* feed in a wide variety of developing fruits, many *Rhagoletis* species are serious pests of fruits such as apples, cherries, blueberries, walnuts and tomatoes (Bush 1966). Many species in *Rhagoletis* have the ability to rapidly shift to new hosts, including introduced cultivated plants (Boller and Prokopy 1976). Morphologically, populations on the old and new hosts are often hard to distinguish. The rapid shifting of host plants contributes to the difficulty of controlling these pests because the wild hosts function as a reservoir for pest populations year after year. The presence of newly established host-associated populations and sympatric sibling species in *Rhagoletis* has also made the genus a model system for studying sympatric speciation (Bush 1969; 1974; 1975; 1992; 1994).

Over the last century, an extensive literature on the biology, ecology and control of certain *Rhagoletis* species, especially those in North America, has accumulated. A brief overview of a few outstanding features of the biology of *Rhagoletis* is presented in this chapter as they not only play important roles on the evolution of the flies themselves, but also offer clues that may be used to interpret the status of some taxa covered in my dissertation. A second section stresses the issues concerning host shifts in *Rhagoletis*—especially host race formation in two unique species groups, the *pomonella* and *cingulata* species groups. A third section presents the current taxonomy of North American

Rhag

genu

disse

Eggs

(Frick

24 hou

1992 a

which t

on the *A*

Prokopy

larval gr

varieties

addition,

indifferen

60°F; larv

for 4 hr (F

ground and

cm under t

Pupae and

Alt

has at least

Rhagoletis

continues w

Rhagoletis and discusses existing problems regarding the status of certain taxa in this genus. It is followed with an overview of the organization of the remaining parts of my dissertation.

Biology of *Rhagoletis*

Eggs and larvae

Rhagoletis eggs hatch within about a week after being laid underneath fruit skins (Frick et al. 1954). First instar larvae usually mine directly to the interior of the fruit within 24 hours after hatching (Frick et al. 1954), probably avoiding parasites as a result (Bush 1992 and references within). Larvae usually confine their feeding to the same fruit in which the eggs are laid and complete their development in about 8 to 40 days, depending on the *Rhagoletis* species and growth conditions (Ries 1935; Frick et al. 1954; Boller and Prokopy 1976). Fruit quality, such as sugar content and acidity, can dramatically influence larval growth rate and survival; larval mortality has been reported to be 100% in some varieties and species of apple and haws (Dean and Chapman 1973; Bush et al. 1989). In addition, temperature can also significantly affect larval development rate. In *R. indifferens*, for example, larval development takes about 10 days at 85°F vs. 35 days at 60°F; larval development ceases at 55°F and death occurs when larvae are exposed to 28°F for 4 hr (Frick et al. 1954). Mature larvae leave the fruit usually after fruit drops to the ground and burrow into the soil under the host tree. They pupate within a few days 4 - 10 cm under the ground (Boller and Prokopy 1976).

Pupae and Diapause

Although some Neotropical *Rhagoletis* are facultatively multivoltine—*R. tomatis* has at least five to six generations per year in Chile (Frias et al. 1991)—most temperate *Rhagoletis* are univoltine and their pupae undergo a winter diapause. Fly development continues when spring temperature and moisture increases. However, in some cases,

sm

cor

pro

cat

195

afte

seco

the

temp

haw

al. 19

units

al. 19

pomo

Eclosi

at spec

primari

1959).

elongati

Foote 19

emergen

of adult e

Rhagoleti

cerasi (Bo

small portion of pupae can remain in diapause in the soil for 2 to 5 winters before completing development (Boller and Prokopy 1976). Such delay in diapause termination provides a pupal reservoir and insures that the population will survive in case of catastrophe, such as the failure of their host plants to fruit (Boller and Prokopy 1976).

Small fractions of *R. pomonella* (Illingsworth 1912) and *R. indifferens* (Frick et al. 1954) populations, without undergoing pupa diapause, may complete development directly after a few weeks of pupation, resulting in a new generation of adults. However, the second adult generations are usually unable to oviposit before low temperatures arrive in the Fall. Diapause induction, as in other insects, is regulated by photoperiod and temperature (Prokopy 1968). Populations of *R. pomonella* adapted to apples and hawthorns respond differently to those diapause regulating factors (Prokopy 1968; Feder et al. 1993). Post-diapause regulation in *Rhagoletis* shows high correlation with thermal units accumulated over a developmental threshold temperature (Reissig et al. 1979; Feder et al. 1993). Post-diapause eclosion time of different host-associated populations in *R. pomonella* is significantly different and genetically programmed (Smith 1988a).

Eclosion and Adult Feeding

After over-wintering below the ground, *Rhagoletis* adults emerge from their puparia at specific times, in most cases, during the spring and summer. Adult emergence occurs primarily in the morning, probably stimulated by the rising morning temperature (Balduf 1959). After bursting the puparium, adult flies propel through the soil by contraction and elongation of the body and ptilinum in order to reach the ground surface (Christenson and Foote 1960). On average, females emerge a few days earlier than males; however, at peak emergence the sex ratio reaches equilibrium (Boller and Prokopy 1976). Synchronization of adult emergence with the fruit maturation of host plants has been demonstrated in many *Rhagoletis* species and has been proven an important trait in differentiating host races in *R. cerasi* (Boller and Bush 1974). Within two hours of emergence, most flies are capable of

flight a

1976).

yeast. b

structu

Prokop

al. 199

untreate

(Prokop

'grazing

Rhagole

extrude

ingestion

excess w

foraging

require c

1992), w

and Prok

Sc

instances,

movemen

host plant

flying seve

more by th

Prokopy 19

flight and feeding, spending little time on the ground before taking off (Boller and Prokopy 1976).

Rhagoletis adults feed on many different kinds of food, such as insect honeydew, yeast, bacteria and fungal spores. Nectar and plant liquid exuding from glandular structure, wounds and oviposition stings are probably additional food sources (Boller and Prokopy 1976). Bird droppings are also natural protein source for *Rhagoletis* (Prokopy et al. 1993). Bird droppings treated with antibiotics were significantly less attractive than untreated ones, indicating that bacteria may be involved in generating attractive volatile(s) (Prokopy et al. 1993). Carbohydrate obtained in the form of leachate by extensively 'grazing' on surface of host foliage can sustain fly longevity (Hendrichs et al. 1993a). *Rhagoletis* flies engorged with a great volume of dilute food, have been observed to extrude orally droplets of liquid crop contents ("bubbling") followed by subsequent re-ingestion (Hendrichs et al. 1992; 1993b). Through the bubbling behavior, flies eliminate excess water by evaporation to concentrate nutrients suspended in dilute solution while foraging for other resources (Hendrichs et al. 1992; 1993b). Both sexes of *R. pomonella* require carbohydrates, certain vitamins and amino acids for gonadal maturation (Bush 1992), which in most *Rhagoletis* species occurs within two weeks of emergence (Boller and Prokopy 1976).

Search for food in *Rhagoletis* is not restricted to the larval host plant but, in some instances, to various types of neighboring vegetation. Under normal crop conditions, movement associated with feeding is nondispersive, rarely taking individuals far from their host plants (Maxwell and Parsons 1968; Neilson 1971). Although *R. cerasi* is capable of flying several kilometers, most dispersive flights observed in *Rhagoletis* were influenced more by the availability of suitable fruit for oviposition than the search for food (Boller and Prokopy 1976).

Foraging

D

their larv

Prokopy

tree shap

host fruit

more tha

profound

Green et

chemical

The odors

to identify

straight-ch

host fruits

suggesting

from its ho

On

against the

visual stim

blend, do n

visual stim

of concentr

1993). Fe

structure, a

contact and

by oviposit

(Cmjar et a

Foraging for Mates and Oviposition Sites

During sexual maturation, both male and female *Rhagoletis* adults congregate on their larval host plants, where courtship, mating and oviposition occur (Bush 1969; Prokopy et al 1971). Flies search for host trees through visual cues such as foliage color, tree shape and tree size (Moericke et al. 1975), as well as olfactory cues from susceptible host fruit (Prokopy et al 1973). Flies appear to have difficulty locating trees at a distance of more than 1.6 m (Roitberg et al. 1982). Although model size, shape and color have profound effects on a fly's response to various models (Prokopy 1973a; 1973b; 1973c; Green et al. 1994), these visual cues are not host plant specific. Olfactory and contact chemical cues, however, play a more important role in the final detection of a correct host. The odors emanating from ripening host fruits provide specific attractants used by the flies to identify their correct host fruits (Prokopy et al. 1973). For example, a number of straight-chain esters (e.g., butyl hexanoate) isolated from extracts of volatile produced by host fruits of *R. pomonella* elicit highly selective behavioral responses by *R. pomonella*, suggesting that this fly is narrowly adapted to respond to specific compounds emanating from its hosts (Averill et al. 1988; Green et al. 1994).

Once on the host plant, flies detect the fruit on the basis of shape, contrast-color against the background and size (Boller and Prokopy 1976). In *R. pomonella*, if fruit visual stimulus is strong (e.g., red color), chemical stimuli such as synthetic apple volatile blend, do not increase the probability of finding fruit or fruit models; however, as the visual stimuli became progressively weaker (red to green to clear), fruit odor (irrespective of concentration) appears to aid flies during the fruit-finding process (Aluja and Prokopy 1993). Female flies inspect the fruit for oviposition on the basis of its size, surface structure, and stage of ripeness (Boller and Prokopy 1976) using chemical stimuli, such as contact and volatile stimulation with various chemicals associated with host fruits, received by ovipositor sensilla provide the fly with information about host suitability and/or quality (Crnjar et al. 1989). Female flies oviposit more often and remain longer on trees harboring

high vs

within a

also infl

less sear

apparent

Mating a

M

thus acts

as well as

are impor

group, wh

plant genu

cues, how

important

pomonella

have almos

and mate o

another oc

courtship a

1973a). Fu

group spec

Rhagoletis

apparently

females (P

Ma

laying pun

high vs low densities of fruit clusters (Roitberg et al. 1982), and will leave their host trees within a short time if they discover no fruit (Roitberg et al. 1982). The intertree distance also influences the foraging behavior of *R. pomonella* in the field. Flies generally invest less search time on a tree when neighboring trees are nearby than when farther away, apparently reducing travel costs (Roitberg and Prokopy 1982).

Mating and Oviposition

Mating in *Rhagoletis* occurs almost exclusively near or on host fruit; the host plant thus acts as an important site for courtship and mating (Bush 1969; Prokopy et al. 1971), as well as for larval development. Visual cues such as body coloration and wing pattern are important in courtship and species recognition, especially those from the *R. suavis* group, whose members have strikingly different color patterns but infest mainly the same plant genus *Juglans* (walnuts) (Bush 1966; Yokoyama and Miller 1994). These visual cues, however, are effective only at close range and they can not be considered as important reproductive isolating mechanisms in sibling species groups such as the *pomonella* and *cingulata* groups. Sibling species in the *R. pomonella* group, for instance, have almost always shifted to new hosts in the course of speciation. They therefore meet and mate on different hosts (Bush 1969; 1974). Even though different species meet one another occasionally visual cues seem not completely prevent them from attempting to courtship as different species *in copula* have been observed in nature (Prokopy and Bush 1973a). Furthermore, wing patterns and body coloration of most *pomonella* and *cingulata* group species are not distinguishable in almost all cases (Bush 1966). Males of several *Rhagoletis* species (*mendax, cingulata, tabellaria, pomonella* and *cornivora*) also are apparently unable to distinguish between the sexes and mount other males as often as females (Prokopy and Bush 1973a; Smith and Prokopy 1982; Smith 1984; 1985a; 1985b).

Male *Rhagoletis* are highly territorial. For example, male walnut flies guard egg-laying punctures on host walnut to increase access to females and defend these sites from

con

on t

(Pro

the p

gene

appr

Smit

fema

aphro

only

attract

both s

opposi

restrict

precopu

1973a;

of these

ripe rath

suavis (L

The waln

(Ries 193

hundred o

Superpara

than one fl

husks (La

conspecific and heterospecific males (Papaj 1994). The visual stimulus of a moving female on the same or nearby fruit elicits attention and initiates courtship by waiting males (Prokopy et al. 1971) which usually involves wing waving, posturing and 'pawing' with the prothoracic legs (Biggs 1972; Prokopy and Bush 1973a). In walnut flies, mating generally takes place as female initiates oviposition (Papaj 1994). Male *Rhagoletis* approach the female either directly or obliquely from the rear (Smith and Prokopy 1982; Smith 1984; 1985a; 1985b). If the female is receptive the male is allowed to mount onto females' abdomen usually by a jump or short flight (Prokopy and Bush 1973a).

Male *R. pomonella* secrete a pheromone that is assumed to function primarily as an aphrodisiac which he waffs to female as he waves his wings. This pheromone is active only over short distances in nature (Prokopy 1975). So far, no long distance sex attractants in *Rhagoletis* has been observed. Because attraction is short range, adults of both sexes must first find the correct host plant and locate fruit before they can meet the opposite sex. Therefore, host selection and mate recognition are directly correlated. The restriction of mating to the fruit of a specific host plant thus serves as an important precopulatory reproductive isolation in several species (Bush 1966; Prokopy and Bush 1973a; Feder et al. 1994) and has important implications in sympatric host race formation of these flies which will be discussed later.

Female flies of most *Rhagoletis* species lay only one egg at a time, usually in nearly ripe rather than immature fruit (Messina et al. 1991). However, walnut infesting flies, *R. suavis* (Loew) and *R. completa* for example, lay eggs in batches (Boyce 1934; Ries 1935). The walnut husk flies readily use sting holes made by conspecifics as oviposition sites (Ries 1935; Lalonde and Mangel 1994) and this probably accounts for the fact that a hundred or more eggs are not uncommon in a single puncture (Ries 1935). Superparasitism is probably a viable strategy because husks have sufficient food for more than one fly offspring and are difficult to parasitize initially due to the toughness of the husks (Lalonde and Mangel 1994). Also, the walnut husk contains very high levels of

juglone

protecti

deterring

least init

1992). I

parasitis

the ODP

Prokopy

been synt

of such sy

infestatior

tool (Aluj

Sir

in which h

closely rel

whose you

is supporte

pattern bet

genetic bas

Longevity

Ave

from 2 to 6

usually grea

adult longe

juglone—a very toxic substance. By feeding gregariously, larvae may receive more protective benefit in detoxifying this substance (Bush, personal communication).

After oviposition, females of several *Rhagoletis* species deposit oviposition deterring pheromone (ODP) on the fruit surface to reduce intraspecific larval competition at least initially (Crnjar and Prokopy 1982; Averill and Prokopy 1989; Aluja and Boller 1992). However, the benefit of host marking may be offset by increased risk of egg parasitism by wasps (Roitberg and Lalonde 1991). Receptor cells sensitive to extracts of the ODP have been identified in the tarsal D-sensilla of some *Rhagoletis* species (Crnjar and Prokopy 1982; Stadler et al. 1994), and isomers and derivatives of ODP for *R. cerasi* have been synthesized and bioassayed (Aluja and Boller 1992; Stadler et al. 1994). Application of such synthetic ODP in an experimental cherry orchard caused a tenfold reduction in fruit infestation suggesting the pheromone may potentially be useful as a fruit fly management tool (Aluja and Boller 1992; Stader et al. 1994).

Since *Rhagoletis* females select host fruits for oviposition and larvae have no choice in which host fruit they develop, successful foraging for oviposition sites may be more closely related to genetic fitness than is the successful foraging for food by other animals whose young may move between and select resources to which they are best adapted. This is supported by evidence which suggests that phenotypic differences in host response pattern between hawthorn and apple origin flies of *R. pomonella* have an underlying genetic basis (Prokopy et al. 1988; Feder et al. 1994).

Longevity

Average adult longevity in nature, although not yet established accurately, ranges from 2 to 6 weeks depending on the species (Boller and Prokopy 1976). Longevity is usually greater in cool weather; with light, humidity and food availability also effecting adult longevity (Boller and Prokopy 1976; Hendrichs et al. 1993a).

E

discovery

differenc

Rhagolet

1969; 19

described

1982; Ber

on Rosace

Snow; the

Cornaceae

recognize

(Smith 19

Marshall (

undescribe

1992) and

1993). Sp

new host f

group are

taxa now r

sympatric

Hybridiza

by earlier

studies (B

electroan

hybridizat

behaviora

Host Shifts in *Rhagoletis*

Extensive study of *Rhagoletis* biology over the past thirty years has led to the discovery of abundant sympatric sibling species which have little or no morphological differences. Speciation in species groups consisting mostly of sibling species in *Rhagoletis* has always been accompanied or preceded by a shift to a new host plant (Bush 1969; 1992; 1994). The *R. pomonella* species group, for example, consists of four described and at least two undescribed sibling species (Bush 1966; Berlocher and Bush 1982; Berlocher et al. 1993; Bush, personal comm.); these are the apple maggot or haw fly on Rosaceae, *R. pomonella* (Walsh); the snowberry fly on Caprifoliaceae, *R. zephyria* Snow; the blueberry fly on Ericaceae, *R. mendex* Curran; and the shrubby dogwood fly on Cornaceae, *R. cornivora* Bush. Recently two additional undescribed species have been recognized, the flowering dogwood fly whose larvae feed in the fruit of *Cornus florida* L. (Smith 1988b; Berlocher et al. 1993) and the sparkleberry fly on *Vaccinium arboreum* Marshall (Ericaceae) (Payne and Berlocher 1995). There also appears to be other undescribed species, such as those southern populations infesting wild plums (Bush 1966; 1992) and the spring population on mayhaw in eastern Texas (Berlocher and Enquist 1993). Speciation in the *pomonella* group has been accompanied by a shift to a radically new host family in almost every case (Bush 1969). However, members of the *pomonella* group are difficult to distinguish morphologically (Bush 1966; Westscott 1982), and many taxa now recognized as distinct species were originally considered as host races or sympatric subspecies by earlier authors (Bush 1966; Diehl and Prokopy 1986). Hybridization, oviposition-choice, ecological, and comparative serology studies carried out by earlier researchers (reviewed in Bush 1966; 1969) as well as more recent allozyme studies (Berlocher and Bush 1982; Feder, et al. 1989; Berlocher et al. 1993), electroantennogram studies on host odor recognition (Frey and Bush 1990), natural hybridization studies (Feder and Bush 1989a; Smith et al. 1993) and host associated behavioral differences (Bierbaum and Bush 1988) strongly support the view that the three

sympat

isolated

occurs

Minnes

pomone

Westcot

species

and Bus

hybridiz

pomonel

1990b).

populatio

Kaneshir

does not

hybridiza

discussed

than is no

novel reco

Be

plants such

communus

1992). In

(although

hawthorn

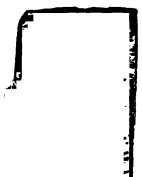
douglassi

quince (*Cy*

Some of th

sympatric eastern forms, *R. pomonella, R. mendax, R. cornivora* are reproductively isolated from one another and represent distinct sibling species. *Rhagoletis zephyria* which occurs primarily in the western United States and is sympatric only with *R. pomonella* in Minnesota (Bush 1966; Westcott 1982; McPheron 1990a; 1990b) differs slightly from *R. pomonella* in surstyli configuration, wing band ratio and ovipositor length (Bush 1966; Westcott 1982). Although *R. zephyria* is the most divergent morphologically of the four species (Bush 1969), it is the most closely related on the basis of allozyme data (Berlocher and Bush 1982; Berlocher et al. 1993). It is not surprising that a low level of interspecific hybridization between these two sibling species has been reported in areas where *R. pomonella* has recently been introduced into western North America (McPheron 1990a; 1990b). Such interspecific hybridization has been also noted between sympatric populations of *Drosophila heteroneura* and *D. silvestris* species in Hawaii (Carson and Kaneshiro 1989). Although F_1 and F_2 progeny are produced, interspecific hybridization does not appear to result in the loss of species identity. The role and outcome of hybridization between closely related animal species is an intriguing problem. As discussed by Bush (1992), interspecific hybridization may be more widespread in insects than is now realized, and in parasite insects a low level of hybridization may give rise to novel recombinant genotypes that facilitate the colonization of a new host.

Besides the above mentioned host plants, *R. pomonella*-like flies also infest other plants such as native plums (*Prunus* spp.), sour cherries (*P. cerasus* L.), pears (*Pyrus communus* L.), rose hips (*Rosa rugosa* Thumb.) and apricots (*P. armeniaca* L.) (Bush 1992). In addition, *R. pomonella* has also been reared from chokecherry (*P. virginiana* L.) (although rarely), sweet cherry (*P. avium* L.), mahaleb cherry (*P. maheleb* L.), ornamental hawthorn (*Crataegus monogyna* Jacquin and *C. mollis* Scheele), river hawthorn (*C. douglassi* Lindley), crabapple (*Malus* spp.), pyracantha (*Pyracantha coccinea* Roemer), and quince (*Cydonia oblonga* Miller) (Allred and Jorgensen 1993; and references within). Some of these host associated *Rhagoletis* populations, such as those southern populations



associated with wild plums, may represent undescribed species and others appear to be recently established host races, for example, those infesting sour cherry and rose hips (Bush 1966; 1992).

Of particular interest in the *pomonella* group is a new host race of *R. pomonella* which was established on introduced apples approximately 150 years ago from the original hawthorn (*Crataegus* sp.) infesting form (Bush 1966; 1969; 1974; 1975; Bush et al. 1989; Feder and Bush 1989b; Bush 1992). Over the years, extensive study has been carried out on the apple race of *R. pomonella* and it is found that the apple race is distinct from the original hawthorn race in several characteristics:

Host Preferences — Prokopy et al. (1988) have provided behavioral evidence suggesting significant differences in host response pattern between apple and hawthorn flies. With respect to choice of fruit for oviposition, female apple flies chose apples significantly more often than did hawthorn flies. Similarly, male apple flies tend to stay substantially longer on apples than male hawthorn flies. Feder et al. (1993; 1994) have observed, in the field mark-release-capture experiments, that *R. pomonella* tend to reproduce on the same host species in which larvae of the flies developed. This host fidelity, as a premating barrier between sympatric *R. pomonella* populations on apples and hawthorns, restricts gene flow to about 6% per generation (Feder et al. 1993; 1994). Frey and Bush (1990) also noticed a difference between the two host races in electrophysiological response to host odors, further supporting the conclusion that host preference is genetically-based although prior experience of adult *R. pomonella* effects their ability to find host fruit (Prokopy et al. 1994) and on their host preference behavior (Bush 1992). The two races of *R. pomonella* also show different learning ability to reject novel fruit species (Bush 1992; and references within).

Eclosion Time — It has been demonstrated that apple flies are genetically programmed to develop faster and emerge sooner after diapause is terminated than hawthorn flies (Smith 1988a; McPheron et al. 1988a; Feder et al. 1993). This difference in

emergence times between the races corresponds to the difference in fruit maturation between their apple and hawthorn hosts. This allochronic separation of the races accounts for part of the isolation of these two host races (Feder et al. 1993). This divergence of *R. pomonella* in eclosion time which is heritable (Smith 1988a) may substantially restrict gene flow among different host-associated populations and thereby contribute significantly to the initial divergence of new *R. pomonella* host races.

Allozyme Frequency — Several studies have demonstrated allozyme frequency differences between the apple and hawthorn populations (Feder et al. 1988; 1989; 1990a; 1990b; McPheron et al. 1988a). Allele frequency divergence is possibly linked to other loci involved with adaptation to apple and hawthorn, such as eclosion time, host fidelity and response to host odors (Bush 1992).

This evidence of genetically based difference between the population of *R. pomonella* associated with apple and haws is now sufficient to support the view that they represent genetically distinct host races (for host race criteria see Bush 1992). Because the two host races differ from each other in several biologically significant ways as the divergence is due to adaptations to different host plants, Bush (1969, 1974, 1992) has proposed that such adaptation might eventually lead to complete reproductive isolation without geographical isolation. To account for the rapid host race formation of *R. pomonella* and for the evolution of other sibling species in the *pomonella* group, Bush (1969; 1974; 1975) developed a model of sympatric host race formation based on genetic changes in host preference and host-based larval survival genes. The proposed host preference speciation (HPS) model suggested that recombination between new alleles of a host preference gene (also called habitat preference or host selection gene) and host-based larval survival gene (also called habitat-based fitness gene) will produce new genotypes that can colonize new host plants. Furthermore, gene flow reduction between the newly established and parental populations could be enhanced by other factors such as allochronic isolation on unrelated plants with different fruiting times (emergence patterns), conditioning

(associati

(Bush, 19

simulatio

a third, un

developm

conditions

1989). M

sympatric

based asso

simulation

genetically

linkage dis

fitness (*fit*)

and *fit* loci

process is c

conditions

generations

In th

parasites. I

pomonella

total of 438

As E

new host pl

those loci in

heterogenei

al. 1990b; F

great popul

(associative learning by induction), disruptive selection and semigeographic isolation (Bush, 1969, 1974 and 1975). The HPS model has later been supported by computer simulation (Diehl and Bush 1989), in which two unlinked loci influencing larval fitness and a third, unlinked locus involving habitat preference. Progress toward speciation (i.e., development of reproductive isolation) is likely to occur under a broad range of biological conditions when assortative mating is coupled with habitat preference (Diehl and Bush 1989). More recently, Johnson et al. (1995) have developed a multi-locus model for sympatric speciation in which habitat preference, habitat-based fitness and non-habitat based assortative mating genes are considered simultaneously. Using computer simulations, they demonstrate how, in organisms that mate within a preferred habitat, genetically based host preference initiates the process of sympatric speciation leading to linkage disequilibrium between the assortative mating gene (*asm*) loci and host-based fitness (*fit*) loci in diploid populations. Completion of linkage disequilibrium of the *asm* and *fit* loci yields no further interbreeding (gene flow), which implies the speciation process is complete. This process can occur sympatrically under a wide variety of conditions of selection pressure and gene penetrance, and may take less than 1000 generations.

In the case of *R. pomonella*, colonization on apples resulted in an escape from most parasites. In Washington State, for example, the average level of parasitism of *R. pomonella* on hawthorn (*C. monogyna*) is up to 90%; while no parasitoids emerged from a total of 4385 pupae reared from apple (Gut and Brunner 1994).

As Bush (1975) suggested, species groups having the potential for shifting to a new host plant might have substantially higher level of genetic polymorphism, especially at those loci involved with host adaptation. Indeed, *R. pomonella* has pronounced population heterogeneity in allozyme frequency (McPheron et al. 1988b; Feder et al. 1990a; Feder et al. 1990b; Feder and Bush 1989b; Berlocher and McPheron, unpublished data). Such a great population differentiation in *R. pomonella* may be related to its extreme flexibility in

diapause

capacity

times (lat

Jorgense

T

Rhagoleti

informatic

genetic ba

have subst

still, howe

possible w

established

distribution

repeated est

(Berlocher

The

be typical o

could, in pri

cingulata sp

indifferens,

turpiniae, de

(Staphyleac

Rhag

native *Prunu*

cultivated ch

diapause strategy and eclosion phenologies (Feder et al. 1993), resulting in its great capacity to adapt to many different native hawthorn species with a wide range of fruiting times (late April to early November in Texas, Berlocher and Enquist 1993; Allred and Jorgensen 1993).

The process of host shifts is the core of the model of sympatric speciation in *Rhagoletis* proposed by Bush in 1969. Over the years, an effort has been made to obtain information on various behavioral and ecological aspects of these flies and to examine the genetic basis of host selection and genetic differences for allozyme frequencies. The results have substantially clarified many points and placed the model on a firmer basis. There are still, however, some details which need to be established. Berlocher (1989) discussed a possible way in which a host race could arise in *R. pomonella.* If the apple race was established from a single colonization event and spread gradually through the apple distribution, a genetically homogeneous population should form, otherwise, independently repeated establishment on apples would result in several genetically distinct subpopulations (Berlocher 1989).

Relevance to speciation in the *R. cingulata* group

The pattern of sympatric speciation occurring in *R. pomonella* species group may be typical of many host-specific insects. The model proposed for *R. pomonella* group could, in principle, be applied to the members of *R. cingulata* group equally well. The *R. cingulata* species group consists of four native North American species: *R. cingulata*, *R. indifferens*, *R. osmanthi* and *R. chionanthi* (Bush, 1966) and one sub-tropical species, *R. turpiniae*, described recently from Mexico, infesting two species of *Turpinia* (Staphyleaceae) (Hernandez-Ortiz 1993).

Rhagoletis cingulata and *R. indifference* originally infested the fruits of different native *Prunus* (Rosaceae) and now both have established themselves on introduced cultivated cherries (*P. avium* and *P. cerasus*). The two species appear to be allopatrically

isolate

North A

needed

Rhagol

Osmant

species a

morphol

difficult.

even as h

W

species su

cherry), *P*

and *P. sal*

1966). In

and fruits

native host

introduced

to 5,000 fee

Normally, t

1975), even

Occasionall

zone with th

formed, but

Department

attack of thi

permanent e

1969; 1975

isolated from one another in the eastern (*R. cingulata)* and western parts (*R. indifferens)* of North America (Blanc and Keifer 1955; Bush 1969), although further investigation is needed to verify the situation in the central plain states where cultivated cherries are grown. *Rhagoletis chionanthi* and *R. osmanthi* have been reared from species of *Chionanthus* and *Osmanthus* (Oleaceae) respectively in southeastern USA, where the two olive-infesting species are sympatric with the eastern cherry fly, *R. cingulata* (Bush 1966). The minimal morphological differences among the four North American species make it extremely difficult, if not impossible, to differentiate between them. They had been generally treated even as host races or subspecies before Bush's 1966 revision.

Western cherry fruit fly, *R. indifferens* has been reported to infest several *Prunus* species such as its principal host, *P. emarginata* (Dougl.) D. Dietr. (wild pin or bitter cherry), *P. virginiana* L. var. *demissa* (Nutt.) Torr., *P. subcordata* Benth (Pacific plum) and *P. salicina* Lindl. (introduced Japanese plum) (reviewed in Frick et al. 1954 and Bush 1966). In California, the native bitter cherry grows at higher altitudes (3,500 to 9,000 feet) and fruits late in the Summer and Fall (Bush 1975). *R. indifferens* usually infests the native host in August (Bush 1975). Cultivated cherries, *P. avium* and *P. cerasus*, introduced to California 100-150 years ago, are grown mainly at relatively low altitudes (0 to 5,000 feet) and fruit much earlier than the native bitter cherry (Bush 1969; 1975). Normally, the cultivated cherries in California are not infested by *R. indifferens* (Bush 1975), even when they are completely surrounded by the native bitter cherry (Bush 1969). Occasionally, however, late maturing cultivated cherries, growing in the altitudinal overlap zone with the wild bitter cherry, become infested (Bush 1969; 1975). These newly formed, but highly localized, populations are often periodically eliminated by the California Department of Agriculture (Bush 1969; 1975). Usually the same infested area is free from attack of this fly the following year (Bush, 1975). This approach has effectively prevented permanent establishment of *R. indifferens* on commercial cherries in California (Bush 1969; 1975).

altitu

space

estab

popul

the ne

of finc

emerg

to low

two po

popula

1969; 1

emargin

other on

1975; Jo

cherries

(Bush 19

Si

host-asso

serotina),

(sweet che

the eastern

species, the

those on tha

therefore, ha

Since the majority of domestic cherries in California are allochronically and altitudinally semi-isolated from the wild bitter cherry, there is only a narrow window in space and time when a successful host shift can occur (Bush 1975). If permanent establishment of the fly population on the cultivated cherries were to be allowed, the population would probably become permanently established on California cherries. Within the newly established population, individual flies emerging earlier would have an advantage of finding a greater abundance of oviposition sites. Selection would favor individuals with emergence time shifted to an earlier date and the newly established population would spread to lower altitudes where cultivated cherries are more abundant (Bush 1975). Eventually two populations with different emergence times and host preferences would evolve.

In Oregon and Washington, *R. indifferens* apparently established permanent population on introduced domestic cherries (*P. avium* and *P. cerasus* L.) (Bush 1966; 1969; 1975). There appear to be two races coexisting in these areas, one on native *P. emarginata* at high altitude whose fruits mature from late July to early September, and the other on domestic cherries at low altitudes during late May and early July (Bush, 1969; 1975; Jones et al. 1991). The two *indifferens* populations from native host and cultivated cherries are almost completely allochronically isolated from one another north of California (Bush 1969; 1975).

Similar differences in emergence patterns have been observed between different host-associated populations of *R. cingulata*. In addition to its native host, black cherry (*P. serotina*), *R. cingulata* also now infests introduced cultivated cherries such as *P. avium* (sweet cherry), *P. cerasus* (sour cherry) and occasionally *P. mahaleb* (Mahaleb cherry) in the eastern United States. Since cultivated cherries mature earlier than native *Prunus* species, the majority of fruit in commercial orchards is semi-allochronically isolated from those on the wild host. The fly populations on cultivated cherries and wild black cherries, therefore, have different emergence times and if following the pattern of divergence in *R.*

pomc

races.

chionc

R. chic

beard)

(devilw

alternat

suggest

overlap

Osmanth

possibly

species s

species.

Osmanth

diverged

establishe

same moc

Osmanthu

have incre

the forma

formation

two host p

infested b

became ge

shifted. O

sympatric,

pomonella possibly different host preferences, showing some evidence of isolation as host races.

Allochronic isolation is even more pronounced between *R. osmanthi* and *R. chionanthi*, both infest native olives (Oleaceae) in southeastern United States (Bush 1966). *R. chionanthi* infests the fruit of *Chionanthus virginicus* (the fringe-tree or old man's beard) in the summer, while *R. osmanthi* attacks the fruit of *Osmanthus americanus* (devilwood) during midwinter (Bush 1966; 1969; 1975). Bush (1969) proposed three alternative explanations for the origin of the two olive infesting species. One explanation suggests that the host plants *Chionanthus* and *Osmanthus* may have considerably overlapped in fruiting time and both were infested by one *Rhagoletis* species. Later, *Osmanthus* shifted its fruiting time to cooler winter months in response to climatic changes, possibly occurring during the Pleistocene. During the process, the original olive infesting species split into two distinct, allochronically isolated races that eventually evolved into two species. Bush's second explanation suggests that a new host race may have established on *Osmanthus* from original *Chionanthus*-infesting population after the two host plants diverged in fruiting time. To me it seems also possible that a host race could become established when the two host plants had fruiting times broadly overlapping, following the same model for apple race formation in *R. pomonella*. Later, the fruiting time shift of *Osmanthus* in response to climatic changes, resulting in fly emergence time shifting, would have increased isolation of the host race from the parental population, eventually leading to the formation of two distinct, allochronically isolated species through sympatric host race formation. The third explanation proposed by Bush involves geographic isolation of the two host plants. Originally, the two host plants may have fruited at the same time and were infested by one olive-infesting species. Later, each host plant with its fly population became geographically isolated. Meanwhile, fruiting time of one host plant may have shifted. Once geographical contact was reestablished, the two fly populations, although sympatric, may have become allochronically isolated from each other. A fourth possibility

is th

Chic

spec

grou

broad

chion

Unite

distin

cingul

specie

cherrie

accomp

change

species

betwee

unequi

devised

DNA re

cingula

species

(1966),

tabellari

is that the *Chionanthus* population arose from a sympatric *cingulata* then a few adults of the *Chionanthus* population emerged during late Fall or Winter and established the *Osmanthus* species.

Therefore, the *cingulata* species group is similar in many ways to the *pomonella* group. It consists of sibling species with minimal morphological difference but with a broad range of host plants. The two olive-infesting species(*R. osmanthi* and *R. chionanthi*) and the cherry-infesting species (*R. cingulata*) are sympatric in the southeastern United States. Although slight differences in morphological characters do occur which distinguish the three species these characters are not consistently clear-cut. Also *R. cingulata* and *R. indifferens* are host specific on different but closely related native *Prunus* species, and have independently established populations on introduced sweet and sour cherries. These host associated populations show evidence of isolation as host races.

Speciation in the *cingulata* group, as in the *pomonella* group, has apparently been accompanied by a shift to a new host plant. Before the kinds and numbers of genetic changes that promote, accompany and follow the colonization of the members of *cingulata* species group of a new host can be established, an accurate means of distinguishing between host races or even closely related species is required. In the absence of unequivocal distinguishing morphological traits alternative means of identification must be devised. As a step towards resolving this problem, I have employed specific genomic DNA regions as molecular markers to resolve species and racial boundaries within the *R. cingulata* species group and explore the relationships of this group with other *Rhagoletis* species (see Chapter II).

Current Taxonomy of North American *Rhagoletis*

The North American *Rhagoletis* species were most recently monographed by Bush (1966), with 21 species segregated into seven species groups (*pomonella, cingulata, tabellaria, suavis, ribicola, striatella* and *alternata*) and one species , *R. fausta* , unplaced.

Bush c

genital

analysi

(1982)

electro

conser

pomon

morpho

Solanac

the clad

above r

(Cupress

Berlocha

describe

recently

tabellari

1966). H

species i

eastern N

(Bush, pe

Vaccinium

electromo

different (

Berlocher

the presenc

aedeagus, r

Bush classified the species groups mainly on the basis of structural similarities of the genitalia, chaetotaxy, wing venation, and karyotype. Since then the only phylogenetic analysis on North American *Rhagoletis* was the one conducted by Berlocher and Bush (1982), based on electrophoretic data. The main areas of agreement between the electrophoretic analysis and the conventional classification (Bush 1966) are the conservation of the *suavis* and *cingulata* groups and, in 2 out of 3 cladistic trees, the *pomonella* species groups. In addition, the species possessing the most ancestral morphological characteristics, such as *R. striatella*, a pest of husk tomatoes (*Physalis* sp., Solanaceae), and *R. basiola*, infesting fruit of *Rosa* (Rosaceae), branch from the base of the cladistic trees generated from electrophoretic data.

Despite the above congruence there are some areas of disagreement between the above mentioned studies. For example, *R. juniperina*, which infests *Juniperus* (Cupressaceae), is removed from the *tabellaria* group and placed with *cingulata* group in Berlocher and Bush (1982). The *tabellaria* group conventionally consists of 4 early described species: *tabellaria, juniperina, persimilis, and ebbettsi* (Bush 1966), plus a recently described species, *R. electromorpha* Berlocher (Berlocher 1984). Members of the *tabellaria* group share similarities in genitalia, wing pattern and body coloration (Bush 1966). Host plants of *persimilis* and *ebbettsi* are unknown. *R. tabellaria* is a wide ranging species infesting two *Cornus* species (Cornaceae), *C. stolonifera* and *C. amomum*, in eastern North America, and *C. stolonifera* in the north central and western North America (Bush, personal comm.). In the west a race or undescribed species is known infesting *Vaccinium* (Ericaceae) in western North America (Bush 1966; Bush, personal comm.). *R. electromorpha* Berlocher, the most recently described *tabellaria*-like species, infests two different *Cornus* species—*C. drummondi* and *C. racemosa* in Illinois (Berlocher 1980; Berlocher 1984). Morphologically, *R. electromorpha* is almost identical to *tabellaria*. Also the presence of gland-like tubular sac at the end of phallotheca, at the junction with aedeagus, relates *R. electromorpha* most closely with *R. tabellaria* with considerable

confid

tabella

differe

put the

and in

that *R.*

and Bus

Howeve

species

placeme

know wh

support

Rhagolet

Oedicare

However,

pointed o

fronto-orb

Another n

is the plac

However,

placement.

Therefore,

to further in

In ac

groups, such

among differ

confidence. In contrast, the morphology of *juniperina* is sufficiently different from *R. tabellaria* to suggest their relatively distant relationship. However, the great morphological difference between *juniperina* and the four North American members of the *cingulata* group put their close relationship in doubt. The status of *R. juniperina*, therefore, is debatable and in need of careful consideration.

Also, allozyme studies (Berlocher and Bush 1982; Berlocher et al. 1993) indicate that *R. cornivora* may not belong to the *pomonella* group. A recent mtDNA study (Smith and Bush, in preparation) also places *R. cornivora* outside of the *pomonella* group. However, the close morphological affinities between *R. cornivora* and the rest of the three species in the *pomonella* group made earlier *Rhagoletis* researchers hardly doubt the placement of *R. cornivora* in the *pomonella* group. Therefore, it would be interesting to know whether any DNA data, such as ribosomal DNA (rDNA) sequence analysis, will support the morphological implication regarding the relationship of *R. cornivora* to other *Rhagoletis* species. In addition, *R. striatella*, the tomato husk fly, is placed with genera *Oedicarena* and *Zonosemata* in Berlocher and Bush (1982) rather than within *Rhagoletis*. However, the placement of *striatella* with *Zonosemata* is not surprising as Bush (1966) pointed out, *R. striatella* shares similarity with *Zonosemata* in karyotype, number of lower fronto-orbital bristles, certain characteristics of male genitalia and host plant relationship. Another new insight gained from the electrophoretic analysis of Berlocher and Bush (1982) is the placement of previously unplaced species, *R. fausta*, with the *suavis group*. However, the recent mtDNA data (Smith and Bush, in preparation) does not support this placement. Instead, *R. fausta* forms a clade with *R. juniperina* according to mtDNA data. Therefore, the placement of *R. fausta* is currently still not completely resolved and subject to further investigation.

In addition to the uncertainty of placement of some species in certain species groups, such as *R. juniperina, cornivora, striatella* and *fausta*, the phylogenetic relationship among different species groups are not completely resolved either. For example, closest

relati

electr

mtDN

the *su*

region

still se

among

demon

1989).

Rhagole

cingula

and *ind*

represe

verifica

electron

with the

1966) o

(Berloc

possess

my rDN

been qu

phyloge

relatives of the *cingulata* group in North America is the *pomonella* group according to electrophoretic studies (Berlocher and Bush 1982; Berlocher et al. 1993), while based on mtDNA data (Smith and Bush, in preparation) the *cingulata* group is more closely related to the *suavis* group.

In summary, although the species of *Rhagoletis*, especially those from the Nearctic region, have been well studied and segregated into a number of species groups, there are still several unanswered questions regarding placement of certain species and relationships among the existing species groups. Furthermore, monophyly of the genus has not been demonstrated and its relationships to other *Carpomyina* are poorly understood (Norrbom 1989).

Dissertation Objectives

My research has two objectives 1) to estimate the phylogeny of North American *Rhagoletis* species and 2) to explore the genetic variation among the sibling species of the *cingulata* species group. The following ten species are analyzed in this study: *R. cingulata* and *indifferens* are representing the *cingulata* group; *R. pomonella* and *cornivora* representing the *pomonella* group with *cornivora*'s placement in this group deeming further verification using rDNA data; *R. completa* represents the well-defined *suavis* group; *R. electromorpha* and *juniperina* are from the *tabellaria* group with *juniperina's* relationship with the rest of the *tabellaria* group questionable; *R. fausta* was unplaced anywhere (Bush, 1966) or its placement is in disagreement from two previous independent molecular studies (Berlocher and Bush 1982; Smith and Bush, in preparation); and *R. basiola* which possesses some of the most ancient morphological characters and is used as outgroup for my rDNA phylogenetic analysis. The placement of *R. striatella* in the genus *Rhagoletis* has been questionable and will be further tested in this study using rDNA spacers.

I use internal transcribed spacers to explore the following problems: 1) What phylogeny do the rDNA spacer sequence data support? 2) Is the phylogenetic implication

from

morp

whic

cingu

molec

the di

compa

useful

phylog

spacer

such in

morpho

subdivid

bibliogra

one sectio

C

elements

cingulata

identified

Chapter III

among the

implications

morphology

pomonella ge

from the rDNA data congruent with any existing systematic of the genus based on morphology and allozyme data? 3) Are there any molecular markers in the rDNA spacers which can be used to differentiate sibling species of the morphologically indistinguishable *cingulata* complex? and if so, what phylogenetic relationship can be inferred from those molecular markers? 4) Are there any genetic polymorphism in the rDNA spacers among the different host-populations of *R. cingulata* and if so, how high is the level of variation compared to interspecific variation in *R. cingulata* group. In addition, I evaluate the usefulness of the secondary structure of internal transcribed spacers (ITS) in inferring phylogeny. Investigation on other rDNA spacer regions such as the external transcribed spacer (ETS) has also been conducted (Chapters IV and V) and the preliminary results from such investigation may form basis for future studies to gain insight into the phylogeny of morphologically indistinguishable species complex.

Organization of the Dissertation

The next four chapters are presented in scientific format, with each Chapter subdivided into an introduction, material and methods, results, discussion, and bibliography sections. In some cases the result and discussion sections are combined into one section. Chapter VI is a concluding summary.

Chapter II presents complete sequences and several constrained secondary structure elements of the rDNA ITS regions of the four North American sibling species of the *cingulata* group. In addition, molecular markers to differentiate those sibling species are identified and phylogenetic implications from those molecular markers are discussed. Chapter III makes use of the rDNA ITS regions to establish phylogenetic relationship among the ten representative *Rhagoletis* species noted above. The phylogenetic implications from the ITS regions are compared with those from earlier studies based on morphology and other molecular data. Chapter IV describes the construction of a *R. pomonella* genomic DNA library and identifies several clones which represent different

segments of the complete rDNA repeat unit. The research described in Chapter IV was performed at the beginning of my Ph. D program before the PCR technology was widely applied in molecular biology. Chapter V represents a partial characterization of the rDNA ETS region of two *R. cingulata* flies, one from native black cherry and the other from introduced sour cherry.

BIBLIOGRAPHY FOR CHAPTER I

Allred, D. I
the apple

Aluja, M., a
deploym
Entomol

Aluja, M., a
intratree
2696.

Averill, A.
S. Robin
control.

Averill, A.
response
222.

Balduf, W.
bionomic

Berlocher, S
Rhagolet
73:131-1

Berlocher, S
Tephritid
Entomol.

Berlocher, S
identifica
Hollande
39. Oxf

Berlocher, S
(Diptera:

Berlocher, S
fly, *Rhag*
66:51-59

Bibliography for Chapter I

Allred, D. B., and C. D. Jorgensen. 1993. Hosts, adult emergence, and distribution of the apple maggot (Diptera: Tephritidae) in Utah. Pan-Pacif. Entomol. **69**:236-246.

Aluja, M., and E. F. Boller. 1992. Host marking pheromone of *Rhagoletis cerasi*: Field deployment of synthetic pheromone as a novel cherry fruit fly management strategy. Entomol. Exp. Appl. **65**:141-147.

Aluja, M., and *R.* J. Prokopy. 1993. Host odor and visual stimulus interaction during intratree host finding behavior of *Rhagoletis pomonella* flies. J. Chem. Ecol. **19**:2671-2696.

Averill, A. L., and *R.* J. Prokopy. 1989. Host marking pheromones. Pp. 207-219 *in* A. S. Robinson, and G. Hooper, eds. Fruit flies, their biology, natural enemies and control. World Crop Pests 3A. Elsevier, Amsterdam.

Averill, A. L., W. H. Reissig, and W. L. Roelofs. 1988. Specificity of olfactory responses in the tephritid fruit fly, *Rhagoletis pomonella*. Entomol. Exp. Appl. **47**:211-222.

Balduf, W. V. 1959. Obligatory and facultative insects in rose hips, their recognition and bionomics. Ill. Biol. Monogr. No. **26**: 194 pp.

Berlocher, S. H. 1980. An electrophoretic key for distinguishing species of the genus *Rhagoletis* (Diptera: Tephritidae) as larvae, pupae, or adults. Ann. Entomol. Soc. Am. **73**:131-137.

Berlocher, S. H. 1984. A new North American species of *Rhagoletis* (Diptera: Tephritidae), with records of host plants of *Cornus*-infesting *Rhagoletis*. J. Kansas Entomol. Soc. **57**:237-242.

Berlocher, S. H. 1989. The complexities of host races and some suggestions for their identification by enzyme electrophoresis. Pp. 51-68 *in* H. D. Loxdale, and J. den Hollander, eds. Electrophoretic studies of agricultural pests. Syst. Assoc. Spec. Vol. **39**. Oxford University Press, Oxford.

Berlocher, S. H., and G. L. Bush. 1982. An electrophoretic analysis of *Rhagoletis* (Diptera: Tephritidae) phylogeny. Syst. Zool. **31**:136-155.

Berlocher, S. H., and M. Enquist. 1993. Distribution and host plants of the apple maggot fly, *Rhagoletis pomonella* (Diptera: Tephritidae) in Texas. J. Kansas Entomol. Soc. **66**:51-59.

Berloch
differ
comp

Bierbaur
accept
(Dipte
econor
No. 8

Biggs, J.
Can. E

Blanc, F.
Morpho
fruit fly

Boller, E.
Europea
paramet

Boller, E.
Ann. Re

Boyce, A.
8:363-5

Bush, G. L.
Psyche 7

Bush, G. L.
North Ar

Bush, G. L.
sibling sp
Nat. 103

Bush, G. L.
(Tephriti
Insects.

Bush, G. L.
in P. W.
Publishin

Bush, G. L.
(Diptera:

Bush, G. L.
Ecol. Evo

Bush, G. L.
Chilcote.

Berlocher, S. H., B. A. McPheron, J. L. Feder, and G. L. Bush. 1993. Genetic differentiation at allozyme loci in the *Rhagoletis pomonella* (Diptera: Tephritidae) species complex. Ann. Entomol. Soc. Am. **86**:716-727.

Bierbaum, T. J., and G. L. Bush. 1988. Divergence in key host examining and acceptance behaviors of the sibling species Rhagoletis mendax and R. pomonella (Diptera: Tephritidae). Pp. 26-55 *in* M. T. Aliniazee, ed. Ecology and management of economically important fruit flies. Agric. Exp. Sta. Oreg. State Univ. Special Report No. **830**.

Biggs, J. D. 1972. Aggressive behavior in the adult apple maggot (Diptera: Tephritidae). Can. Entomol. **104**:349-353.

Blanc, F. L., and H. H. Keifer. 1955. The cherry fruit fly in North America. Morphological differentiation between the eastern and western subspecies of the cherry fruit fly, *Rhagoletis cingulata* (Loew). Calif. Dept. Agric. Bull. **44**:77-88.

Boller, E. F., and G. L. Bush. 1974. Evidence for genetic variation in populations of the European cherry fruit fly, *Rhagoletis cerasi* (Diptera: Tephritidae) based on physiological parameters and hybridization experiments. Entomol. Exp. Appl. **17**:279-293.

Boller, E. F., and *R.* J. Prokopy. 1976. Bionomics and management of *Rhagoletis.* Ann. Rev. Entomol. **21**:223-246.

Boyce, A. M. 1934. Bionomics of the walnut husk fly, *Rhagoletis completa.* Hilgardia **8**:363-579.

Bush, G. L. 1965. The genus *Zonosemata* with notes on the cytology of two species. Psyche **72**:307-323.

Bush, G. L. 1966. The taxonomy, cytology, and evolution of the genus *Rhagoletis* in North America (Diptera: Tephritidae). Bull. Harvard Mus. Comp. Zool. **134**:431-562.

Bush, G. L. 1969. Mating behavior, host specificity, and the ecological significance of sibling species in frugivorous flies of the genus *Rhagoletis* (Diptera: Tephritidae). Am. Nat. **103**:669-672.

Bush, G. L. 1974. The mechanism of sympatric host race formation in the true fruit flies (Tephritidae). Pp. 3-23 *in* M. J. D. White, ed. Genetic Mechanisms of Speciation in Insects. Australian and New Zealand Book Co., Sydney.

Bush, G. L. 1975. Sympatric speciation in phytophagous parasitic insects. Pp. 187-205 *in* P. W. Price, ed. Evolutionary strategies of parasitic insects and mites. Plenum Publishing Corp., New York.

Bush, G. L. 1992. Host race formation and sympatric speciation in *Rhagoletis* fruit flies (Diptera: Tephritidae). Psyche **99**:335-358.

Bush, G. L. 1994. Sympatric speciation in animals: new wine in old bottles. Trends in Ecol. Evol. **9**:285-288.

Bush, G. L., J. L. Feder, S. H. Berlocher, B. A. McPheron, D. C. Smith, and C. A. Chilcote. 1989. Sympatric origins of *R. pomonella.* Nature **339**:346.

Carson, H. L., and K. Y. Kaneshiro. 1989. Natural hybridization between the sympatric Hawaiian species *Drosophila silvestris* and *Drosophila heteroneura*. Evolution **43**:190-203.

Christenson, L. D., and Foote, *R.* H. 1960. Biology of fruit flies. Ann. Rev. Entomol. **5**:171-192.

Crnjar, R. M and Prokopy, R. J. 1982. Morphological and electrophysiological mapping of tarsal chemoreceptors of oviposition-deterring pheromone in Rhagoletis pomonella flies. J. Insect Physiol. 28:393-400.

Crnjar, R., A. Angioy, P. Pietra, J. G. Stoffolano Jr., A. Liscia, and I. T. Barbarossa. 1989. Electrophysiological studies of gustatory and olfactory responses of the sensilla on the ovipositor of the apple maggot fly, *Rhagoletis pomonella* Walsh. Bull. Zool. **56**:41-46.

Dean, *R.* W., and P. J. Chapman. 1973. Bionomics of the apple maggot in eastern New York. Search Agric. (Geneva, N. Y.) **3**:1-62.

Diehl, S. R., and G. L. Bush. 1989. The role of habitat preference in adaptation and speciation. Pp. 345-365 *in* D. Otte and J. Endler, eds. Speciation and its consequences. Sinauer Assoc., Sunderland, MA.

Diehl, S. R., and *R.* J. Prokopy. 1986. Host-selection behavioral differences between the fruit fly sibling species *Rhagoletis pomonella* and *R. mendax* (Diptera: Tephritidae). Ann. Entomol. Soc. Am. **79**:266-271.

Feder, J. L., and G. L. Bush. 1989a. A field test of differential host usage between two sibling species of *Rhagoletis* fruit flies (Diptera: Tephritidae) and its consequences for sympatric models of speciation. Evolution **43**:1813-1819.

Feder, J. L., and G. L. Bush. 1989b. Gene frequency clines for host races of *Rhagoletis pomonella* in the midwestern United States. Heredity **63**:245-266.

Feder, J. L., C. A. Chilcote, and G. L. Bush. 1988. Genetic differentiation between sympatric host races of *Rhagoletis pomonella*. Nature **336**:61-64.

Feder, J. L., C. A. Chilcote, and G. L. Bush. 1989. Are the apple maggot, *Rhagoletis pomonella*, and blueberry maggot, *R. mendax*, distinct species? Implications for sympatric speciation. Entomol. Exp. Appl. **51**:113-123.

Feder, J. L., C. A. Chilcote, and G. L. Bush. 1990a. The geographic pattern of genetic differentiation between host associated populations of *Rhagoletis pomonella* (Diptera: Tephritidae) in the eastern United States and Canada. Evolution **44**:570-594.

Feder, J. L., C. A. Chilcote, and G. L. Bush. 1990b. Regional, local and microgeographic allele frequency variation between apple and hawthorn populations of *Rhagoletis pomonella* in western Michigan. Evolution **44**:595-608.

Feder, J. L., S. B. Opp, B. Wlazlo, K. Reynolds, W. Go, and S. Spisak. 1994. Host fidelity is an effective premating barrier between sympatric races of the apple maggot fly. Proc. Natl. Acad. Sci. USA **91**:7990-7994.

Feder, J. L., T. A. Hunt, and G. L. Bush. 1993. The effects of climate, host plant phenology and host fidelity on the genetics of apple and hawthorn infesting races of *Rhagoletis pomonella*. Entomol. Exp. Appl. **69**:117-135.

Foote, *R.* H. 1981. The genus *Rhagoletis* Loew south of the United States (Diptera: Tephritidae). U.S. Dep. Agric. Tech. Bull. **1607**, 75 pp.

Frey, J. E., and G. L. Bush. 1990. *Rhagoletis* sibling species and host races differ in host odor recognition. Entomol. Exp. Appl. **57**:123-131.

Frias, D. L., and H. Martinez. 1991. Taxonomic studies in *Rhagoletis tomatis* Foote (Diptera: Tephritidae). Acta Entomol. Chilena **16**:247-254.

Frick, K. E., H. G. Simkover, and H. S. Telford. 1954. Bionomics of the cherry fruit flies in eastern Washington. Washington Agric. Expt. Sta. Tech. Bull. **13**:1-66.

Green, T. A., *R.* J. Prokopy, and D. W. Hosmer. 1994. Distance of response to host tree models by female apple maggot flies, *Rhagoletis pomonella* (Walsh) (Diptera: Tephritidae): interaction of visual and olfactory stimuli. J. Chem. Ecol. **20**:2393-2413.

Gut, L. J., and J. F. Brunner. 1994. Parasitism of the apple maggot, *Rhagoletis pomonella*, infesting hawthorns in Washington. Entomophaga **39**:41-49.

Hendrichs, J., B. S. Fletcher, and *R.* J. Prokopy. 1993b. Feeding behavior of *Rhagoletis pomonella* flies (Diptera: Tephritidae): Effect of initial food quantity and quality on food foraging, handling costs, and bubbling. J. Insect Behavior **6**:43-64.

Hendrichs, J., C. *R.* Lauzon, S. S. Cooley, and *R.* J. Prokopy. 1993a. Contribution of natural food sources to adult longevity and fecundity of *Rhagoletis pomonella* (Diptera: Tephritidae). Ann. Entomol. Soc. Am. **86**:250-264.

Hendrichs, J., S. S. Cooley, and *R.* J. Prokopy. 1992. Post-feeding bubbling behavior in fluid-feeding Diptera: concentration of crop contents by oral evaporation of excess water R. Physiol. Entomol. **17**:153-161.

Hernandez-Ortiz, V. 1993. Description of a new *Rhagoletis* species from tropical Mexico (Diptera: Tephritidae). Proc. Entomol. Soc. Wash. **95**:418-424.

Illingworth, J. F. 1912. A study of the biology of the apple maggot (*Rhagoletis pomonella*), together with an investigation of methods of control. Cornell Univ. Ag R.. Expt. Sta. Bull. **324**:129-187.

Johnson, P. A., F. C. Hoppensteadt, J. J. Smith, and G. L. Bush. 1995. Conditions for sympatric speciation: a diploid model incorporating habitat fidelity and nonhabitat assortative mating, in press.

Jones, V. P., D. G. Alston, J. F. Brunner, D. W. Davis, and M. D. Shelton. 1991. Phenology of the western cherry fruit fly (Diptera: Tephritidae) in Utah (USA) and Washington (USA). Ann. Entomol. Soc. Am. **84**:488-492.

Kandybina, M. N. 1972. Fruit flies (Diptera: Tephritidae) of the Mongolian People's Republic. Entomol. Rev. **51**:540-545.

Lalonde, *R.* G., and M. Mangel. 1994. Seasonal effects on superparasitism by *Rhagoletis completa.* J. Animal Ecol. **63**:583-588.

Maxwell, C. W., and E. C. Parsons. 1968. The recapture of marked apple maggot adults in several orchards from one release point. J. Econ. Entomol. **61**:1157-1159.

McPheron, B. A. 1990a. Genetic structure of apple maggot fly (Diptera: Tephritidae) populations. Ann. Entomol. Soc. Am. 83:568-577.

McPheron, B. A. 1990b. Implications of genetic variation in western apple maggots for understanding biology. Pp. 37-49 *in R.* V. Dowell, L. T. Wilson, and V. P. Jones, eds. Apple maggot in the West: history, biology, and control. Univ. Calif. Div. Ag*R.* Nat. Res. Publ. **3341**.

McPheron, B. A., C. D. Jorgensen, and S. H. Berlocher. 1988b. Low genetic variability in a Utah cherry-infesting population of the apple maggot, *Rhagoletis pomonella.* Entomol. Exp. Appl. **46**:155-160.

McPheron, B. A., D. C. Smith, and S. H. Berlocher. 1988a. Genetic differences between host races of *Rhagoletis pomonella.* Nature **336**:64-66.

Messina, F. J., D. G. Alston, and V. P. Jones. 1991. Oviposition by the western cherry fruit fly (Diptera: Tephritidae) in relation to host development. J. Kansas Entomol. Soc. **64**:197-208.

Moericke, V., *R.* J. Prokopy, S. Berlocher, and G. L. Bush. 1975. Visual stimuli associated with attraction of *Rhagoletis pomonella* (Diptera: Tephritidae) flies to trees. Entomol. Exp. Appl. **18**:497-507.

Neilson, W. T. A. 1971. Dispersal studies of a natural population of apple maggot adults. J. Econ. Entomol. **64**:648-653.

Norrbom, A. L. 1989. The status of *Urophora acuticornis* and *U. sabroskyi* (Diptera: Tephritidae). Entomol. News **100**:59-66.

Papaj, D. *R.* 1994. Oviposition site guarding by male walnut flies and its possible consequences for mating success. Behavioral Ecol. Sociobiol. **34**:187-195.

Payne, J. A., and S. H. Berlocher. 1995. Phenological and electrophoretic evidence for a new blueberry-infesting species in the Rhagoletis pomonella sibling species complex. Entomol. Exp. Appl. **75**:183-187.

Prokopy, *R.* J. 1968. Influence of photoperiod, temperature, and food on initiation of diapause in the apple maggot. Can. Entomol. **100**:318-329.

Prokopy, *R.* J. 1975. Mating behavior in *Rhagoletis pomonella* (Diptera: Tephritidae). V. Virgin female attraction to male odor. Can Entomol. **107**:905-908.

Prokopy, R. J., and D. R. Papaj. 1988. Learning of apple fruit biotypes by apple maggot flies. J. Insect Behav. **1**:67-74.

Prokopy, *R.* J., and G. L. Bush. 1973a. Mating behavior of *Rhagoletis pomonella* (Diptera: Tephritidae). IV. Courtship. Can. Entomol. **105**:873-891.

Prokopy, *R.* J., and G. L. Bush. 1973b. Oviposition by grouped and isolated apple maggot flies. Ann. Entomol. Soc. Am. **66**:1197-1200.

Prokopy, *R.* J., and G. L. Bush. 1973c. Ovipositional responses to different sizes of artificial fruit by flies of *Rhagoletis pomonella* species group. Ann. Entomol. Soc. Am. **66**:927-929.

Prokopy, *R.* J., E. W. Bennett, and G. L. Bush. 1971. Mating behavior in *Rhagoletis pomonella* (Diptera: Tephritidae). I. Site of assembly. Can. Entomol. **103**:1405-1409.

Prokopy, *R.* J., S. *R.* Diehl, and S. S. Cooley. 1988. Behavioral evidence for host races in *Rhagoletis pomonella* flies. Oecologia **76**:138-147.

Prokopy, *R.* J., S. S. Cooley, L. Galarza, C. Bergweiler, and C. *R.* Lauzon. 1993. Bird droppings compete with bait sprays for *Rhagoletis pomonella* (Walsh) flies (Diptera: Tephritidae). Can. Entomol. **125**:413-422.

Prokopy, *R.* J., V. Moericke, and G. L. Bush. 1973. Attraction of apple maggot flies to odor of apples. Environ. Entomol. **2**:743-749.

Reissig, W. H., J. Barnard, *R.* W. Weires, E. H. Glass, and *R.* W. Dean. 1979. Prediction of apple maggot fly emergence from thermal unit accumulation. Environ. Entomol. **8**:51-54.

Ries, D. T. 1935. Biological study of the walnut husk fly (*Rhagoletis suavis* Loew). Mich. Acad. Sci. Arts Letters XX:717-723.

Rohdendorf, B. B. 1961. Palaearctic fruit-flies (Diptera: Tephritidae) of the genus *Rhagoletis* Loew and closely related genera. Entomol. Rev. **40**:89-102.

Roitberg, B. D., and *R.* G. Lalonde. 1991. Host marking enhances parasitism risk for a fruit-infesting fly *Rhagoletis basiola.* Oikos **61**:389-393.

Roitberg, B. D., and R. J. Prokopy. 1982. Influence of intertree distance on foraging behaviour of *Rhagoletis pomonella* in the field. Ecol. Entomol. **7**:437-442.

Roitberg, B. D., J. C. van Lenteren, J. J. M. van Alphen, F. Galis, and *R.* J. Prokopy. 1982. Foraging behaviour of *Rhagoletis pomonella.* a parasite of hawthorn (*Crataegus viridis*), in nature. J. Animal Ecol. **51**:307-325.

Smith, D. C. 1984. Feeding, mating, and oviposition by *Rhagoletis cingulata* (Diptera: Tephritidae) flies in nature. Ann. Entomol. Soc. Am. **77**:702-704.

Smith, D. C. 1985a. General activity and reproductive behavior of *Rhagoletis cornivora* (Diptera: Tephritidae) flies in nature. J. New York Entomol. Soc. **93**:1052-1056.

Smith, D. C. 1985b. General activity and reproductive behavior of *Rhagoletis tabellaria* (Diptera: Tephritidae) flies in nature. J. Kansas Entomol. Soc. **58**:737-739.

Smith, D
asynch

Smith, D
fruit p
Soc. 9

Smith, D
Tephr

Smith, D
verific
corniv

Stadler, E
the hos
natural

Westcott,
(Walsh
Pan-Pa

Yokoyam
free and
Zealand

Smith, D. C. 1988a. Heritable divergence of *Rhagoletis pomonella* host races by seasonal asynchrony. Nature **336**:66-67.

Smith, D. C. 1988b. Reproductive differences between *Rhagoletis* (Diptera: Tephritidae) fruit parasites of *Cornus amomum* and *C. florida* (Cornaceae). J. New York Entomol. Soc. **96**:327-331.

Smith, D. C., and *R.* J. Prokopy. 1982. Mating behavior of *Rhagoletis mendax* (Diptera: Tephritidae) flies in nature. Ann. Entomol. Soc. Am. **75**:388-392.

Smith, D. C., S. A. Lyons, and S. H. Berlocher. 1993. Production and electrophoretic verification of F-1 hybrids between the sibling species *Rhagoletis pomonella* and *R. cornivora*. Entomol. Exp. Appl. **69**:209-213.

Stadler, E., B. Ernst, J. Hurter, and E. Boller. 1994. Tarsal contact chemoreceptor for the host marking pheromone of the cherry fruit fly, *Rhagoletis cerasi*: responses to natural and synthetic compounds. Physiol. Entomol. **19**:139-151.

Westcott, R. L. 1982. Differentiating adults of apple maggot, *Rhagoletis pomonella* (Walsh) from snowberry maggot, *R. zephyria* Snow (Diptera: Tephritidae) in Oregon. Pan-Pacif. Entomol. **58**:25-30.

Yokoyama, V. Y., and G. T. Miller. 1994. Walnut husk fly (Diptera: Tephritidae) pest-free and preovipositional periods and adult emergence for stone fruits exported to New Zealand. J. Econ. Entomol. **87**:747-751.

CHAPTER II

rDNA INTERNAL TRANSCRIBED SPACERS 1 AND 2 OF THE *RHAGOLETIS CINGULATA* SPECIES GROUP: SEQUENCE, CONSTRAINED SECONDARY STRUCTURES, AND PHYLOGENETIC IMPLICATIONS

Introduction

The genus *Rhagoletis* (Diptera: Tephritidae) is composed of several economically and biologically important species groups. Speciation in some of these species groups have been accompanied by colonizing new host plants. The *R. pomonella* species group, for example, has emerged as a model system for studying host race formation and sympatric speciation (Bush 1993). The apple and hawthorn races of *R. pomonella* are morphologically indistinguishable, but biologically they show several genetically-based character differences (see chapter 1). Host shifts to introduced plants have also occurred in other, less well studied, North American *Rhagoletis* species, but their biological status deems further investigation. Of particular interest are members of the *Rhagoletis cingulata* species group which have undergone rapid host shifts but are also difficult to distinguish using morphological characters alone.

The *R. cingulata* species group consists of four native North American species: *R. cingulata, R. indifferens, R. osmanthi* and *R. chionanthi* (Bush, 1966) and one sub-tropical species, *R. turpiniae*, described recently from Mexico, infesting two species of *Turpinia* (Staphyleaceae) (Hernandez-Ortiz 1993). The four North American sibling species show little or no morphological difference, but have a wide range of host plants. *R. cingulata* and *R. indifferens* are serious cherry pests in the eastern and western United States, respectively. These two species originally infest different native wild cherries

(*Prunus spp.*), and now both have established themselves on introduced cultivated cherries such as sweet cherries and sour cherries. For each of the two species, populations on the old native host and introduced cherries appear to be semi-allochronically isolated due to different emergence times which are synchronous with the maturation of their different respective host fruit. The two host-associated populations of *R. indifferens*, on native pin cherries and introduced commercial cherries are also semi-geographically isolated from each other because cultivated cherries are usually grown at considerably lower altitude than the wild cherry in the western United States. Therefore, two host races, with different emergence patterns and different host associations, exist for each of the cherry fly species. The other two sibling species in the *cingulata* group, *R. osmanthi* and *R. chionanthi*, infest native olives (Oleaceae) in southeastern United States where they are sympatric with *R. cingulata* (Bush 1966). *R. chionanthi* infests the fruits of the fringe-tree, *Chionanthus virginicus*, which fruits in the summer; while the larvae of *R. osmanthi* are found in devilwood, *Osmanthus americanus*, which fruits during midwinter (Bush 1966; 1969; 1975). Bush (1969) proposed, as in the case of the formation of the apple race of *R. pomonella*, host races adapted to the two olive species could have established themselves when the two host plants had broadly overlapping fruiting times . When the fruiting time of *Osmanthus* shifted, probably in response to climatic changes, the pattern of fly emergence time also shifted. As a consequence, isolation of the host race from the parental population would have been increased, eventually leading to the formation of two distinct, allochronically isolated species through sympatric host race formation specialized on different host plants.

The *cingulata* species group is, therefore, similar in many ways to the *pomonella* group. It consists of sibling species infesting various plants but with minimal overlapping morphological differences. The morphological characters that distinguish two olive-infesting species (*R. osmanthi* and *R. chionanthi*) and the cherry-infesting species (*R. cingulata*) which are sympatric in the southeastern United States are thus not clear-cut. The

allopatric

species. I

some evi

pomonell

the kinds

colonizat

an accura

In the abs

identifica

nuclear rD

and racial

this group

Se

made in de

variation s

substantial

it has both

phylogene

tandemly r

The entire

which is th

Sollner-W

28S) and r

related tax

rDNA an

transcribe

18S and 5.

allopatric *R. cingulata* and *R. indifferens* which utilize as hosts different native *Prunus* species, have also established populations on introduced sweet and sour cherries that show some evidence of isolation as host races. Speciation in the *cingulata* group, as in the *pomonella* group, has probably been accompanied by a shift to a new host plant. Before the kinds and numbers of genetic changes that promote, accompany and follow the colonization of host plants by the members of *cingulata* species group can be established, an accurate means of distinguishing closely related species or even host races is required. In the absence of unequivocal distinguishing morphological traits alternative means of identification must be devised. As a step towards resolving this problem, I have employed nuclear rDNA internal transcribed spacers (ITS) as molecular markers to investigate species and racial boundaries within the *R. cingulata* species group and explore the relationships of this group with other *Rhagoletis* species.

Selecting the sequence to be analyzed is probably the most important decision to be made in designing a DNA analysis in phylogenetic studies because the level of sequence variation should be sufficient to display enough variation but not too much that there is substantial homoplasy of nucleotide substitution. Nuclear rDNA is unique in the sense that it has both highly variable and conserved regions, providing information across a broad phylogenetic spectrum (Hillis and Davis 1986). In eukaryotes, rDNA is composed of tandemly repeated transcriptional units separated from each other by intergenic spacers. The entire unit is transcribed by RNA polymerase I as a single 45S precursor molecule, which is then processed to yield mature 18S, 5.8S and 28S rRNAs (Hadjiolov 1985; Sollner-Webb and Tower 1986). The highly conserved coding regions (18S, 5.8S and 28S) and relatively fast-evolving spacers allow investigation of both distantly and closely related taxa. In addition, the highly conserved coding regions flanking the spacers make rDNA an excellent system for PCR amplification and analysis. For instance, the internal transcribed spacers (ITS1 and ITS2) are located between the well-conserved coding regions 18S and 5.8S, and 5.8S and 28S, respectively. Even though the sequence of *Rhagoletis*

rDNA is not known, PCR primers can be designed based on either highly conserved sequences or known sequences of closely related taxa. Because of the above mentioned features of rDNA, the rDNA spacers have recently become an attractive source of phylogenetic characters for differentiating populations (Nazar et al. 1991; Bakker et al. 1992; Kooistra et al. 1992; O'Donnell 1992; Gardes and Bruns 1993; Fritz et al. 1994; Volger and DeSalle 1994) and for phylogenetic analysis (Lee and Taylor 1991; Baldwin 1992; Pleyte et al. 1992; Wesson et al. 1992; Wingfield et al. 1994). Because of the relatively rapid rate at which new mutants are fixed in rDNA spacers, these regions may distinguish closely related species that otherwise show little genetic divergence (Brown et al. 1972; Furlong and Maden 1983; Tautz et al. 1987; Porter and Collins 1991). In addition, ITS1 and ITS2 RNAs in yeast *Saccharomyces cerevisiae*, have been shown to function independently and are important for the processing of the pre-rRNA to the mature forms (Musters et al. 1990; van der Sande et al. 1992).

A secondary-structure model for *S. cerevisiae* ITS2, based on chemical and enzymatic probing, has been proposed (Yeh and Lee 1990). The ITS regions have a high propensity of forming secondary structures in several other organisms as well (Kupriianova et al. 1989). Some of these conserved potential secondary structures in ITS are presumed to be functionally important. Therefore, within the ITS sequences several regions may be relatively constrained and not free-evolving as in the case of *Drosophila* (Schlotterer et al. 1994). Having those conserved secondary structures as partial alignment guides will increase the accuracy of aligning homologous regions rather than only similar regions which might be a consequence either of common ancestry or of chance (Olsen and Woese 1993). Because it is very important to compare aligned homologous regions in phylogenetic study, secondary structure analysis has become essential when we extract phylogenetic information from rDNA sequences (Wesson et al. 1992; Schlotterer et al. 1994).

cond

Rhag

speci

level

well

chara

and e

close

const

based

free e

those

systen

condit

expan

uncert

variati

phylog

charac

provide

species

Biologic

the Unite

My immediate goals in this study are to 1) obtain suitable primers and PCR reaction conditions for amplifying, for the first time, the ITS regions of rDNA in the genus *Rhagoletis*; 2) determine the sequences of ITS regions of the 4 North American sibling species in the *cingulata* group and those of *R. pomonella* for comparison; 3) examine the level of ITS sequence polymorphism among different individual flies of each species, as well as among different host-associated populations of *R. cingulata*; 4) establish molecular characters which can be used for distinguishing the sibling species in the *cingulata* group and evaluate the usefulness of the ITS sequences in the phylogenetic analysis of those closely related species and/or host-associated population in this group; and 5) identify constrained potential secondary-structure elements in ITS of *Rhagoletis* using an analysis based on the principle of positional covariance in addition to the computer-based minimum free energy method. Some conserved secondary-structure elements will be compared with those from *Drosophila*. I also infer a phylogenetic relationship and investigate the systematic status of taxa in the *cingulata* group. The suitable PCR primers and reaction conditions determined in this study will be employed in future phylogenetic analysis of an expanded number of taxa in the genus *Rhagoletis*, especially those whose placement is uncertain or in question as mentioned in Chapter I. The level of intra- and inter-specific variation discovered here will help evaluate the usefulness of the ITS regions in future phylogenetic analysis of other taxa in the genus *Rhagoletis*. Identification and further characterization of nucleotide changes within specific secondary structural elements may provide additional insight to the mode of evolutionary divergence of certain *Rhagoletis* species and functional significance of the ITS during processing of precursor rRNA.

Materials and Methods

Biological Material

All species were collected during 1988-90 from various locations and host plants in the United States of America (Table 1). Larvae emerged from field infested fruit and were

Tabl

Spec
R. ci
R. ci
R. ci
R. ci
R. ci
R. ci
R. ch
R. ch
R. os
R. os
R. ind
R. ind
R. ind
R. pon
R. pon
R. pon
R. cor
R. jun
R. faus

Note.—

Table 1. Collection Sites and Host Plants of *Rhagoletis* Species Used in This Study

Species	Sample	Sex	Host Plant (Common Name)	Location (USA)
R. cingulata	RC1	M	*Prunus avium* (sweet cherry)	Traverse City, MI
R. cingulata	RC2	F	*Prunus avium* (sweet cherry)	Traverse City, MI
R. cingulata	RC3	M	*Prunus cerasus* (sour cherry)	Hart, MI
R. cingulata	RC4	F	*Prunus cerasus* (sour cherry)	Hart, MI
R. cingulata	RC5-I2	M	*Prunus serotina* (black cherry)	Roselake, MI
R. cingulata	RC5-I1	F	*Prunus serotina* (black cherry)	Roselake, MI
R. chionanthi	RK1	M	*Chionanthus virginicus* (fringe-tree)	Perry, GA
R. chionanthi	RK2	F	*Chionanthus virginicus* (fringe-tree)	Perry, GA
R. osmanthi	RO1	M	*Osmanthus americanus* (wild tea-olive)	Alligator Lake, FL
R. osmanthi	RO2	F	*Osmanthus americanus* (wild tea-olive)	Alligator Lake, FL
R. indifferens	RI1	M	*Prunus cerasus* (sour cherry)	Pullman, WA
R. indifferens	RI2	M	*Prunus cerasus* (sour cherry)	Pullman, WA
R. indifferens	RI3	F	*Prunus cerasus* (sour cherry)	Pullman, WA
R. pomonella	RP1	M	*Crataegus* spp. (hawthorn)	E. Lansing, MI
R. pomonella	RP2	M	*Crataegus* spp. (hawthorn)	E. Lansing, MI
R. pomonella	RP3	ND	*Malus pumula* (apple)	Door Co., WI
R. cornivora	RCo1	M	*Cornus amomum* (dogwood berries)	E. Lansing, MI
R. juniperina	RJ1	M	*Juniperus virginiana* (E. red cedar)	Dixon Springs, IL
R. fausta	RF1	M	*Prunus cerasus* (sour cherry)	Fish Creek, WI

Note.— ND = Not Determined

all

sto

22

em

Spe

DN

and

140

35R

regi

(ma

(ma

regi

prim

(fina

of ea

and 1

Cetus

sec; 3

6 min

visuali

from t

Rad), a

allowed to pupate in fine, moist vermiculite. Pupae were sifted from the vermiculite and stored at 4° C for at least 5 months. Pupae were then removed from the cold and held at 22° C under a 15 hr light and 9 hr dark cycle to terminate diapause. Within 2-7 days after emergence most adult flies were frozen at -70° C for subsequent genomic DNA isolation. Specimens from each collection were pinned for species identification.

DNA Isolation and Amplification

Total genomic DNA was isolated from individual flies as described by Procunier and Smith (1993). The ITS1 region was amplified using primer 1406F 5'CCTTTGTACACACCGCCCGT (matching the 3' end of 18S) and primer 35R 5'AGCTRGCTGCGTTCTTCATCGA (matching the 5' end of 5.8S). The ITS2 region alone was amplified using primer 108F 5'GAACATCGACHHKTYGAACGCA (matching the 3' end of 5.8S) and primer 52R 5'GTTAGTTTCTTTTCCTCCSCT (matching the 5' end of 28S). Amplification of the ITS1 and ITS2 regions as a combined region on one DNA fragment was performed for *R. indifferens*, using the 1975F and 52R primers. Amplification by the polymerase chain reaction (PCR) was carried out in 25 µl (final volume) containing 10 mM Tris-HCl, pH 8.5, 50 mM KCl, 3 mM $MgCl_2$, 375 µM of each dNTP, 0.1-0.4 µM primer 1975F (or 108F) and 0.1-0.4 µM primer 35R (or 52R), and 1.25-2.50 units of Ampli *Taq* DNA polymerase (Perkin Elmer Cetus) with 5-20 ng genomic DNA. Amplification parameters were 92° C for 3 min 10 sec; 30 cycles each at 92° C for 15 sec, 65° C for 15 sec and 72° C for 2 min; and 72° C for 6 min 10 sec. Amplified DNA was subjected to electrophoresis on a 1.0 % agarose gel and visualized with ethidium bromide. Bands containing the DNA of interest were excised from the gel and the DNA purified using the Prep-A-Gene DNA purification matrix (Bio-Rad), according to the manufacturer's instructions.

Cloning a

Th

direct inse

transforma

the manuf

DNA purif

randomly

species. D

Sanger et a

^{35}S-dATP

to determin

additional

different sp

used for th

5'A(AT)(A

DNA Seque

The

Wisconsin

alignment a

with the co

adjusted. E

pair of taxa

(1969) and

software ve

comparison

several perc

Cloning and Sequencing

The TA Cloning kit (Invitrogen) was used in a one-step cloning strategy for the direct insertion of the purified PCR products into a plasmid vector, followed by transformation into competent cells. Plasmid vector and competent cells were supplied by the manufacturer. Plasmid DNA was purified from individual clones using the Magic-Prep DNA purification kit (Promega), following the manufacturer's instructions. One clone was randomly picked from each fly and, in general, several flies were sequenced for each species. DNA sequencing was performed according to the chain-termination method of Sanger et al. (1977), and using the Sequenase Version 2.0 DNA sequencing kit (USB) and ^{35}S-dATP (Amersham). The same primers used in the amplification reactions were used to determine the DNA sequence in both directions. Once a stretch of DNA was sequenced additional primers were employed to complete the sequencing of the ITS regions from the different species: Primers 35R-GB27 5'ACC(CT)AAACATTTTCAAGT(CT)GCG was used for the ITS1 regions; and primer 108F-GB25 5'A(AT)(AG)(AG)AATC(AT)(CT)AGTATTCCC was used for the ITS2 regions.

DNA Sequence, Structure and Phylogenetic Analysis

The PILEUP and FOLDRNA programs in the GCG package of the University of Wisconsin Genetics Computer Group (UWGCG package, version 8.0) were used for alignment and secondary-structure calculations, respectively. Alignments were done first with the computer (gap weight = 3.00 and gap length weight = 0.20) and then manually adjusted. Estimates of the percent nucleotide substitutions per nucleotide site between each pair of taxa and their standard errors were determined by the methods of Juke and Cantor (1969) and Tamura (1992) using Molecular Evolutionary Genetics Analysis (MEGA) software version 1.01. All gap sites were removed from the subset data during pairwise comparisons. Because several flies from each species were sequenced, which resulted in several percent substitution estimates for pairs of the same species, a weighted average of

the data points was calculated taking into account the standard error values for each pair of taxa by the following equation:

$$\mu = \frac{\Sigma(x_i/\sigma_i^2)}{\Sigma(1/\sigma_i^2)}$$

where x_i is the Jukes-Cantor estimate for each pairwise comparison and σ_i its standard error estimate with μ the average number of substitutions per nucleotide site. The standard error of the weighted average value was determined by taking the square root of the reciprocal of the denominator in the above equation. Subsequently, the average percent nucleotide substitution (number of nucleotide substitutions per 100 nucleotide sites) was obtained by multiplying μ and its standard error value by 100. Phylogenetic analysis was accomplished by the maximum parsimony method in which all uninformative characters were ignored using the programs in PAUP version 3.1 by D. L. Swofford (University of Illinois, Champaign, IL). Uninformative characters were ignored and gaps were considered as missing data. Taxas RC2 with its ITS1 sequence and RI3 with its ITS2 sequence were excluded in the PAUP analysis because their corresponding ITS2 and ITS1 sequences, respectively, were not determined in this study (see Table 2) and equivalent taxas are required when combining informative characters from both ITS sequences in a PAUP analysis. The exhaustive search option was employed to find the most parsimonious tree(s). The three *R. pomonella* flies were taken as outgroup and made a monophyletic sister group to ingroup (i.e., rooted).

Results

DNA Amplification

The PCR amplification of the ITS1 and ITS2 or a combined region containing the two spacers was successful. Most of the amplification products were visualized as a single

sharp ba

purifica

Sequenc

group a

the ITS

bp of rD

defined

sequence

(using 14

within th

99 nt seq

observed

observed.

Th

conserved

insertion/d

group and

ITS1 appe

which are

Th

nucleotide

and 3, and

most cases

and Tamur

transition/t

sharp band on agarose gels as in Figure 1. Some of them do not even need further purification and can be directly used for cloning.

Sequence Analysis

The complete ITS1 and ITS2 sequences of the 4 sibling species in the *cingulata* group and of *R. pomonella* are presented in Figures 2 and 3, respectively. In addition to the ITS sequences presented in Figures 2 and 3, I also sequenced approximately 50 to 200 bp of rDNA coding regions. The boundaries between the ITS and the coding regions were defined by comparing the *Rhagoletis* sequences with published *Drosophila melanogaster* sequences (Tautz et al. 1988). For ITS1, of the 182 nt sequenced within the 3' of the 18S (using 1406F), I found 2 insertion/deletions and one substitution; of the 54 nt sequenced within the 5' end of 5.8S (using 35R) no variation was found. Similarly, for ITS2, of the 99 nt sequenced within the 3' end of 5.8S (using 108F) only 2 substitutions were observed; of the 80 nt sequenced within the 5' end of 28S (using 52R) no variation was observed.

The sequences of the different members in the *cingulata* species group are highly conserved with very few nucleotide changes. However, there are considerable insertion/deletion differences in both ITS1 and ITS2 between the *R. cingulata* species group and *R. pomonella*. Furthermore, the region between nucleotide positions 206-300 in ITS1 appears to be quite variable between the three individual flies of *R. pomonella*, two of which are from hawthorns and the third from apples.

The average percent nucleotide substitutions (nucleotide substitutions per 100 nucleotide sites) for the ITS sequences were calculated based on alignments in Figures 2 and 3, and are presented in Table 3. The results from the pairwise comparisons were, in most cases, identical between two different statistical approaches; Jukes and Cantor (1969) and Tamura (1992). The Tamura (1992) approach compensates for biases in transition/transversion rates and G+C content in addition to compensating for multiple hits

Ta

Spe

R. c
R. c
R. c
R. c
R. c
R. c
R. c
R. o
R. o
R. in
R. in
R. in
R. po
R. po
R. po
R. co
R. ju
R. fa

Note.

Table 2. A-T Content and Length of the Analyzed *Rhagoletis* ITS Sequences

Species	Sample	ITS1		ITS2	
		A-T Content (%)	Length (nt)	A-T Content (%)	Length (nt)
R. cingulata	RC1	79.8	660	82.7	555
R. cingulata	RC2	79.8	657	ND	ND
R. cingulata	RC3	79.9	662	82.7	555
R. cingulata	RC4	79.7	661	82.7	554
R. cingulata	RC5-I1&2	80.1	652	82.9	556
R. chionanthi	RK1	80.1	658	82.6	553
R. chionanthi	RK2	80.1	659	82.8	554
R. osmanthi	RO1	80.0	657	82.4	553
R. osmanthi	RO2	80.0	657	82.7	554
R. indifferens	RI1	79.7	659	82.7	555
R. indifferens	RI2	79.7	660	82.8	557
R. indifferens	RI3	ND	ND	82.8	557
R. pomonella	RP1	80.4	684	82.0	471
R. pomonella	RP2	79.2	653	81.7	471
R. pomonella	RP3	80.1	674	81.7	475
R. cornivora	RCo1	80.7	652	81.2	482
R. juniperina	RJ1	81.0	683	82.7	557
R. fausta	RF1	80.0	624	81.7	527

Note.— ND = Not Determined

Figure 1. PCR amplification products from *Rhagoletis* ITS2 using primers 108F and 52R. PCR products were analyzed on a 0.8% agarose gel by electrophoresis. The 123 bp DNA ladder (lanes M), *Pst*I (lane M') and *Hin*dIII (lane M") digests of λ DNA were used as molecular size markers. Lanes 1-9 correspond to *pomonella*, *completa*, *electromorpha*, *cornivora*, *striatella*, *fausta*, *basiola*, *indifferens* and *cingulata*. The genomic DNA used in this analysis were prepared from male flies except for *R. cingulata*.

Figure

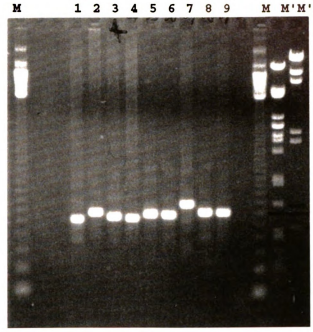


Figure 1

Figure 2. ITS1 sequences and alignment for *Rhagoletis cingulata* species group and *R. pomonella*. Identical nucleotides are denoted by dashes, and gaps are denoted by dots. Phylogenetically informative characters are denoted by asterisks. Conserved computer generated secondary structure domains are boxed, whereby the loop regions of the stem-loop structures (see Fig. 5) are separately boxed and labeled. The stem regions are indicated. The labeling of the structural domains corresponds to the numbering in Fig. 5.

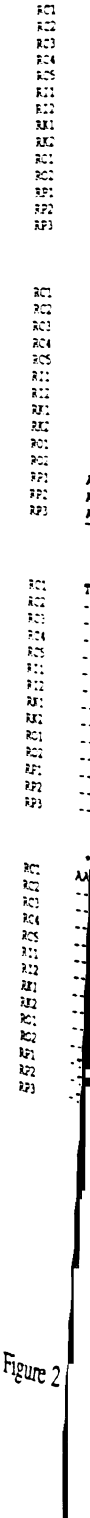


Figure 2

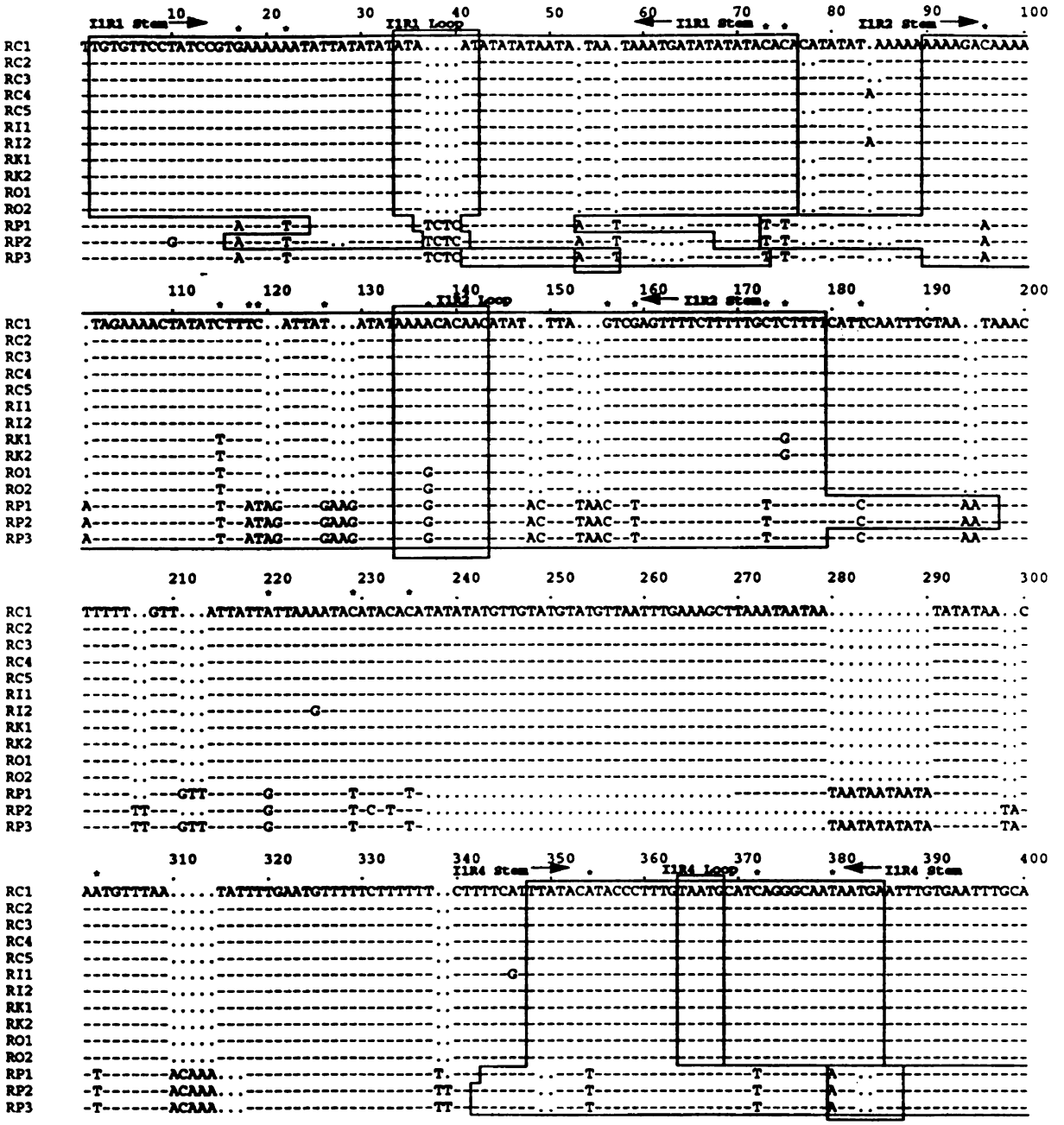

I1R1 Stem
I1R1 Loop
I1R1 Stem
I1R2 Stem
I1R3 Loop
I1R2 Stem
I1R4 Stem
I1R4 Loop
I1R4 Stem
RC1
RC2
RC3
RC4
RC5
RI1
RI2
RK1
RK2
RO1
RO2
RP1
RP2
RP3

Figure 2

Figure 2 (co

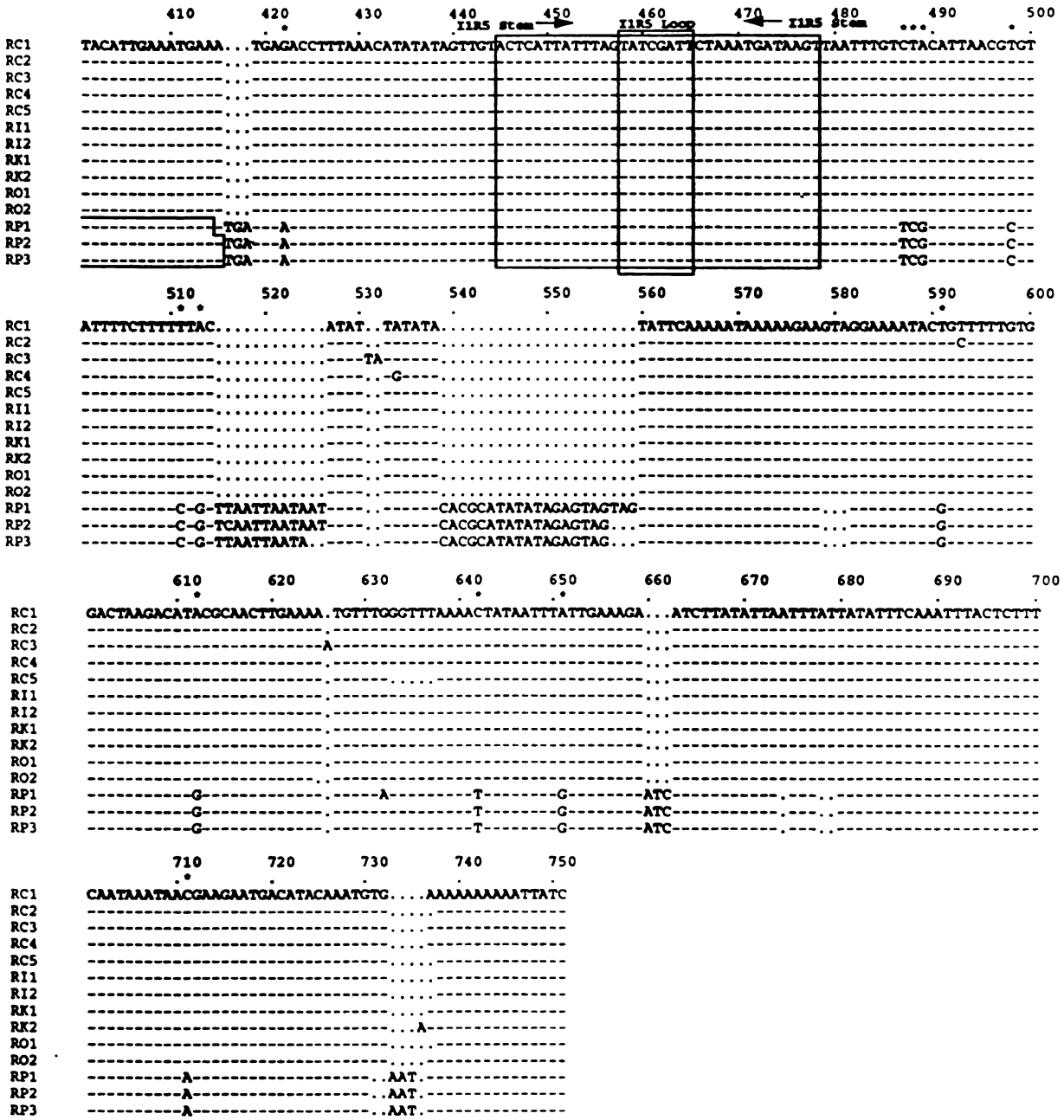


Figure 2 (cont'd)

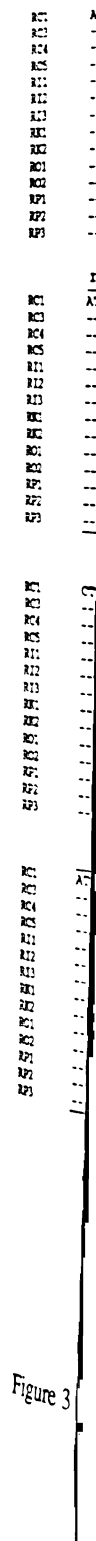


Figure 3. ITS2 sequences and alignment for *Rhagoletis cingulata* species group and *R. pomonella*. Notations same as in legend of Figure 2.

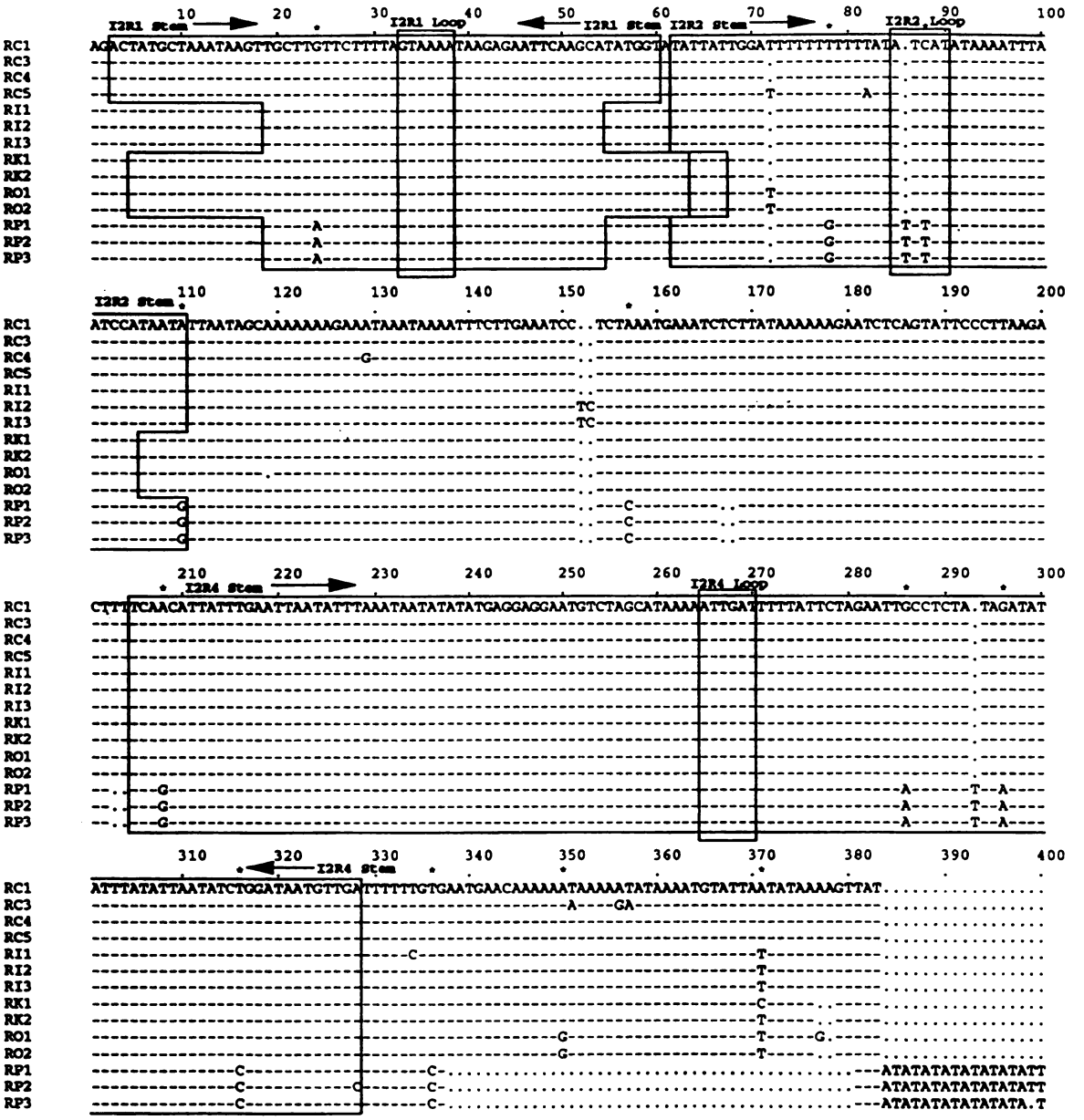

I2R1 Stem
I2R1 Loop
I2R1 Stem
I2R2 Stem
I2R2 Loop
I2R2 Stem
I2R4 Stem
I2R4 Loop
I2R4 Stem
RC1
RC3
RC4
RC5
RI1
RI2
RI3
RK1
RK2
RO1
RO2
RP1
RP2
RP3

Figure 3

Figure 3 (c

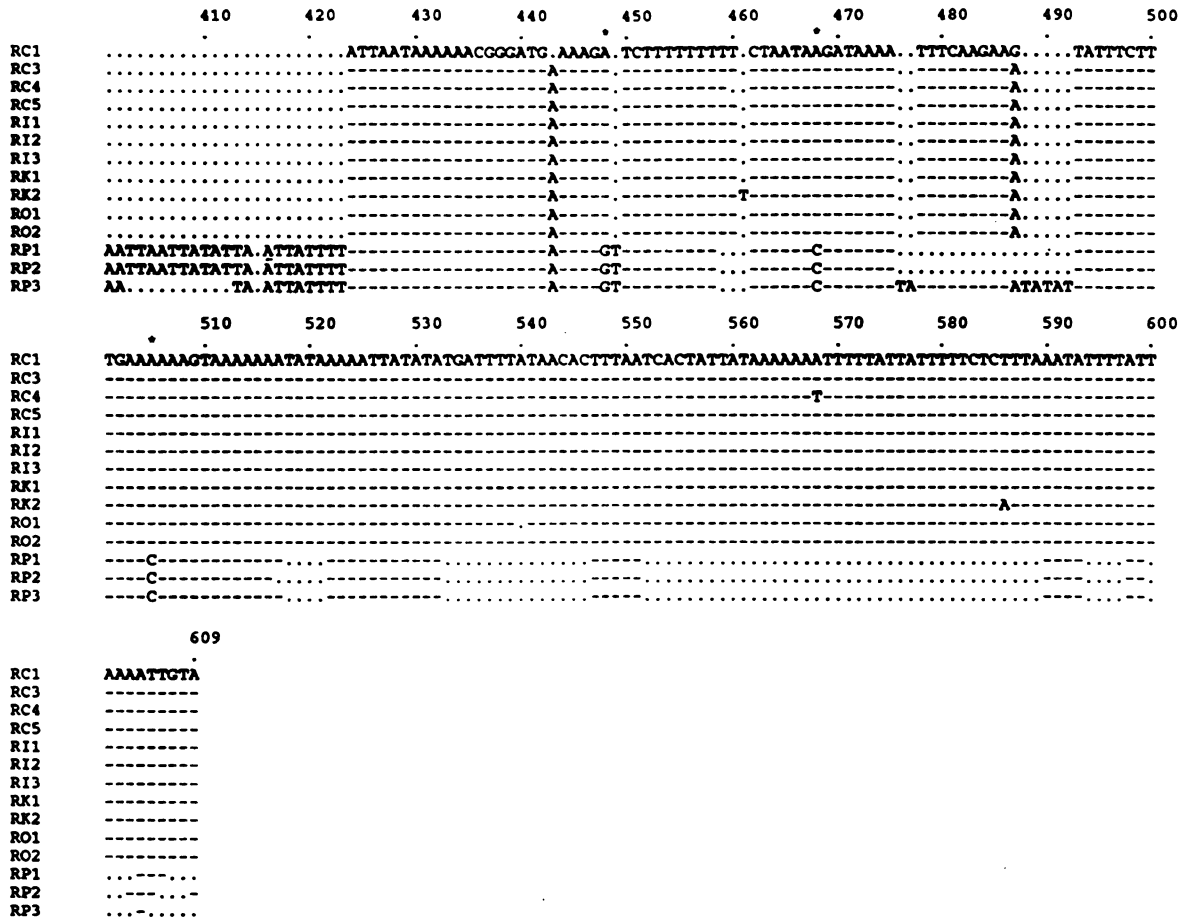


Figure 3 (cont'd)

Table 3. Average Percent Nucleotide Substitutions (Nucleotide Substitutions per 100 Nucleotide Sites) in the ITS sequences for the *Rhagoletis cingulata* Species Group and *R. pomonella* in Pairwise Comparisons

ITS1	1	2	3	4	5
1. cin	0.18 ± 0.07 (10)				
2. ind	0.21 ± 0.06 (10)	0.35 (1)			
3. chi	0.40 ± 0.08 (10)	0.52 ± 0.15 (4)	0.00 (1)		
4. osm	0.40 ± 0.08 (10)	0.52 ± 0.15 (4)	0.35 ± 0.12 (4)	0.00 (1)	
5. pom	5.77 ± 0.27 (15)	5.88 ± 0.42 (6)	5.70 ± 0.42 (6)	5.32 ± 0.40 (6)	0.20 ± 0.11 (3)
ITS2					
1. cin	0.27 ± 0.11 (6)				
2. ind	0.27 ± 0.09 (12)	0.24 ± 0.17 (3)			
3. chi	0.24 ± 0.12 (8)	0.24 ± 0.17 (6)	0.00 (1)		
4. osm	0.24 ± 0.12 (8)	0.24 ± 0.17 (6)	0.00 (4)	0.00 (1)	
5. pom	3.33 ± 0.26 (12)	3.29 ± 0.30 (9)	3.21 ± 0.36 (6)	3.21 ± 0.36 (6)	0.24 ± 0.17 (3)

Note.—cin = *R. cingulata*; ind = *R. indifferens*; chi = *R. chionanthi*; osm = *R. osmanthi*; and pom = *R. pomonella*. Number in the brackets represents the number of pairs analyzed. All gap sites were removed from the subset data before the pairwise comparisons.

as in the

from the

0.07% in

the diffe

values fo

observed

represen

1994). F

substitut

represen

DNA am

temperat

range fr

about 0.

Because

and Mg-

the error

after 30

that, as a

plasmid

provided

no differ

DNA sli

amplifie

simple re

as in the Jukes and Cantor (1969) approach. Therefore, here I only present the results from the Jukes and Cantor (1969) approach.

Among the different *R. cingulata* flies the level of genetic variation was 0.18 ± 0.07% in ITS1 and 0.27 ± 0.11% in ITS2. Similarly, the level of genetic variation among the different *R. pomonella* flies was 0.20 ± 0.11 in ITS1 and 0.24 ± 0.17 in ITS2. The values for the level of within-species variation in the ITS sequences are comparable to those observed in a recent study of five different isofemale *Drosophila melanogaster* lines, representing five different individuals from different countries (Schlotterer and Tautz 1994). From Table I of Schlotterer and Tautz (1994) I calculated the percent nucleotide substitutions of the ITS sequences between isofemales lines (i.e., intra-specific assay representing different individuals) to be 0.10 ± 0.07.

The fidelity of *Taq* polymerase is highly dependent on the conditions used during DNA amplification — especially dNTP and Mg^{+2} concentrations, and annealing temperature (Gelfand and White 1990). The average nucleotide mutation rate per cycle can range from 1.7×10^{-4} (at 1.5 mM each dNTP, 10 mM $MgCl_2$ and 37° C annealing) to about 0.5×10^{-5} (at 200 μM each dNTP, 1.5 mM $MgCl_2$ and 54 to 55° C annealing). Because the conditions used for amplification in this study are at the lower end for dNTP and Mg^{+2} concentrations as well as higher annealing temperature, I *conservatively* estimate the error frequency from DNA amplification to be ≤3 mutation per length of ITS1 or ITS2 after 30 cycles. This is within the limits of deviation found in Table 3. It is noteworthy that, as a control, when the *D. melanogaster* ITS sequences were amplified by PCR from plasmid pDm238 (which contains the complete *D. melanogaster* rDNA repeat unit; provided by Dr. G. A. Dover) and sequenced using the conditions in this study, I detected no difference from published results (Tautz et al. 1988; Schlotterer et al. 1994). Significant DNA slippage-induced length variation is observed when simple repeating sequences are amplified by various DNA polymerases *in vitro* (Schlotterer and Tautz 1992). Variations at simple repeat loci (e.g., length expansions) have also been observed *in vivo* and appear to

arise from DNA slippage synthesis by DNA polymerases as well (see Tautz and Schlotterer 1994; and references within). Therefore, it seems possible that some of the observed length variation in the simple repeats found in ITS of *Rhagoletis* may be attributed to the *in vitro* amplification step before cloning and not an inherent variation among the populations or species studied. Other researchers have also attributed this type of sequence variation to be the product of slippage events in other ITS sequences (Wesson et al. 1992; Vogler and DeSalle 1994). This is partially the reason why gaps were excluded in the calculation of nucleotide substitution rates. However, other length expansions in simple repeats, such as those in *R. pomonella* (206-300 in ITS1 and 384-423 in ITS2), may reflect inherent heterogeneity within *R. pomonella* because I do not observe a random distribution of such large length expansions of such di- and tri-nucleotide motifs (e.g., TA or TAA) among species of the *cingulata* group.

The inter-specific variation in the *cingulata* group was not significantly higher than the intra-specific polymorphism, indicating that the ITS regions do not display sufficient variation distinguishing closely-related members of this group. However, the divergence in the ITS sequences between the *cingulata* group and *R. pomonella* were significantly high (about 5.8% *vs.* 0.2% intra-specific variation in ITS1 for *R. pomonella*). Therefore, the ITS regions may provide information on the phylogeny of taxa from different species groups of the genus *Rhagoletis* rather than on relationships among members of the *cingulata* group.

The length and A-T content of the ITS regions for the *R. cingulata* species group and *R. pomonella* are shown in Table 2. On average the ITS1 sequences of the *R. cingulata* group were about 100 nt longer than the ITS2 sequences (658 ± 3 nt for ITS1 vs. 555 ± 1 nt for ITS2), while in *R. pomonella* ITS1 is about 200 nt longer than ITS2. In addition, ITS1 of *R. pomonella* showed considerable length heterogeneity ranging from 653 to 684 nucleotides among three flies, implying the potential usefulness for future genetic study on host-associated populations of *R. pomonella* species. The ITS of the

cingulata species group had a very high A-T content, with 79.9 ± 0.2% (average ± sample standard deviation) in ITS1 and 82.7 ± 0.2% in ITS2. The high A-T content found for the *Rhagoletis* ITS sequences is comparable to that of *Drosophila* (approx. 75% in ITS1 and 79% in ITS2 from Schlotterer et al. 1994) and *Cicindela* beetles (approx. 79% in ITS1 from Volger and DeSalle 1994), but much higher than is found in *Aedes* mosquitoes (approx. 42% in ITS1 and 47% in ITS2 from Wesson et al. 1992). The A-T content of *Rhagoletis* ITS exceeds that found for noncoding DNA in *Drosophila*, which is about 60% (Moriyama and Hartl 1993), but is less than the 96% A-T content found in the 4,601 bp A + T region of *D. melanogaster* mitochondrial DNA (Lewis et al. 1994). The significance of high A-T content found in the ITS sequences is not clear but in the case of *Rhagoletis*, it may have some relation to a recent observation that *R. pomonella* rDNA clusters are located at the periphery of fibrillar centers in the nucleolus (Procunier and Smith 1993) or may be related to the organization of rDNA gene clusters into heterchromatin as suggested for *Drosophila* by Schlotterer et al. (1994). High A-T content DNA is known to be associated with the nuclear matrix (or nuclear scaffold), which may affect the processes of transcription and replication (van Holde 1989). *Drosophila* histone gene clusters, tandemly repeated about 100-fold, were found to be periodically attached to type I nuclear scaffold via A-T rich sequences lying in the spacers between histone H1 and histone H3 (Mirkovitch et al. 1984).

Molecular Markers and Phylogeny

Thirty-four and 15 informative characters were obtained from ITS1 and ITS2, respectively (asterisks, Figures 2 and 3). Forty-four out of the total 49 characters group the *cingulata* species group as a separate cluster distinct from *R. pomonella*. The remaining 5 characters (#s 115, 137 and 175 in ITS1; #s 350 and 371 in ITS2) were used to determine relationships among the members of the *cingulata* species group. Because of the limited number of informative characters, a phylogenetic analysis on the combined ITS1

and

two

flies

rete

two

4A.

Figu

and

were

those

positi

differ

these

covere

C in th

two oli

Rhagol

175,

other 7

cingulat

which is

olive-inf

within a

markers.

60 nts fr

reaction

cingulat

and ITS2 data was performed. Two most parsimonious trees were found (Figure 4). The two trees were basically the same except for the outcome of the two *R. osmanthi* individual flies (RO1 and RO2). Both trees have same length 50, consistency index (CI) 1.00 and retention index (RI) 1.00. Both a strict and 50% majority-rule consensus analyses of the two most parsimonious trees gave me a tree with the same topology as the tree in Figure 4A. On the other hand, the semi-strict consensus tree had the same topology as the tree in Figure 4B. The trees indicate that *R. osmanthi* is the most ancestral species in this group and *R. cingulata* forms a derived clade with *R. indifferens.* It should be noted that the trees were based on a very limited number of characters and all phylogenetic implications from those trees are tentative and subject to further investigation. In ITS1 there are three positions (#s 115, 137 and 175) potentially useful as molecular markers for distinguishing different members of the *cingulata* species group (Table 4). Nucleotide composition at these three positions were compared with several other *Rhagoletis* species which are covered in more details in the following Chapter. Position 115 is composed of the residue C in the two cherry-infesting species (*cingulata* and *indifferens*, total 7 flies), but T in the two olive-infesting species (*osmanthi* and *chionanthi*, total 4 flies), as well as in 4 other *Rhagoletis* species (*pomonella, cornivora, juniperina* and *fausta*, total 6 flies). At position # 175, *R. chionanthi* is the only *Rhagoletis* species studied which has residue G, while the other 7 *Rhagoletis* species presented here, including the remaining three members of the *cingulata* group, have C at this position. Position #175, together with position # 137 which is A in *chionanthi* and G in *osmanthi,* possibly can be used to distinguish the two olive-infesting species from one another. These three informative positions in ITS1 lie within a very well aligned region, adding confidence to their potential use as molecular markers. In addition the three positions are tightly packed in a relative short fragment (only 60 nts from #115 to 175) and the sequence can be easily obtained in a one-step sequencing reaction or in an automatic sequencer. Between the two cherry infesting species (*R. cingulata* and *R. indifferens*) there was also a characteristic nucleotide position (position

Table 4. Potential Molecular Markers in ITS1 for Distinguishing Members of the *R. cingulata* Species Group

	R. cingulata species group				Other *Rhagoletis* species			
Position	cin (5)	ind (2)	chi (2)	osm (2)	pom (3)	cor (1)	jun (1)	fau (1)
115	C	C	T	T	T	T	T	T
137	A	A	A	G	G	A	A	A
175	C	C	G	C	C	C	C	C

Note.—cin = *R. cingulata*; ind = *R. indifferens*; chi = *R. chionanthi*; osm = *R. osmanthi*; pom = *R. pomonella*; cor = *R. cornivora*; jun = *R. juniperina*; and fau = *R. fausta*. The numbers in brackets represent the number of flies sequenced to determine the type of nucleotide at the specified position. The nucleotide composition at the equivalent positions for *R. cornivora*, *R. juniperina* and *R. fausta* ITS1 sequences were taken from data in Chapter III.

Figure 4. Phylogeny inferred from the combined ITS1 and ITS2 sequences from *Rhagoletis cingulata* species group and *R. pomonella*. Two most parsimonious trees were obtained using PAUP. All uninformative characters were ignored and gaps in the alignments were treated as missing data. The number of character-state changes along each branch are indicated.

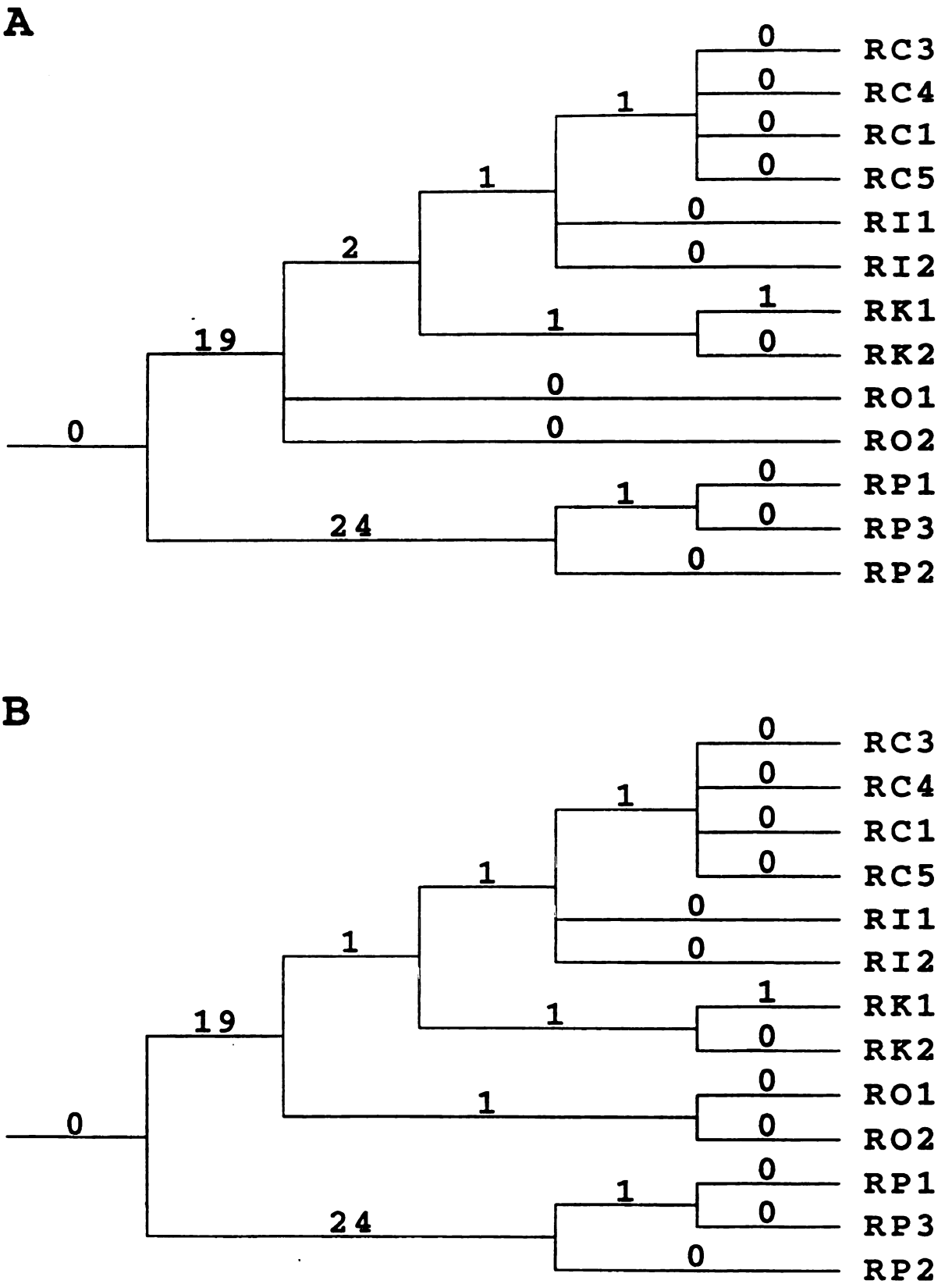

A
0
RC3
0
RC4
1
0
RC1
0
RC5
1
0
RI1
2
0
RI2
1
RK1
1
0
RK2
19
0
RO1
0
0
RO2
0
RP1
1
0
RP3
24
0
RP2
B
0
RC3
0
RC4
1
0
RC1
0
RC5
1
0
RI1
1
0
RI2
1
RK1
19
1
0
RK2
0
RO1
1
0
RO2
0
RP1
1
0
RP3
24
0
RP2

Figure 4

#371 i

further

subsuit

Second

ITS1 S

and pre

cingula

Figure

(Figure

structur

(12). R s

loop str

the struc

studied

Furtherm

changes

FOLDR

same pa

stem reg

deletion

species

U:G pai

and othe

stem-loo

Figure 2

#371 in ITS2; Figure 3) which distinguishes between them. However, this position needs further verification because it seems dimorphic in *R. chionanthi,* although this is transition substitution rather than transversion.

Secondary Structure Analysis:

ITS1 Secondary Structures

Several secondary structural domains in ITS1 were determined using FOLDRNA and presented in Figure 2. The computer generated secondary structure models for the *cingulata* species group are generally similar and represented by that of *R. osmanthi* in Figure 5A. For comparison, the secondary structure of *R. pomonella* is also shown (Figure 5B). The nomenclature used to describe the structural elements of *Rhagoletis* ITS structure models was as follows: The first two symbols refer to either ITS1 (I1) or ITS2 (I2), R stands for *Rhagoletis* and the last symbol refers to the number of the *major* stem-loop structure from 5' to 3' of the ITS sequences (Figures 2, 3 and 5). In the ITS1 two of the structural elements (I1R4 and I1R5) were found to be highly conserved among all the studied *Rhagoletis* species, including 4 non-*cingulata* group species (Figures 6 and 7). Furthermore, the I1R4 structural element was associated with several compensatory changes between the *cingulata* species group and the other *Rhagoletis* species studied. The FOLDRNA assigned the base-pair A354:U380 for the *cingulata* species group and this same pair covaries with U354:A380 in *R. pomonella* and *R. cornivora* suggesting that a stem region may indeed exist (Fig. 7A). In addition, two compensatory deletions/insertions were found; A350:U383 and U349:G384 base pairs in the *cingulata* species group are absent in *R. pomonella* (Fig. 7A, boxed nucleotides). The non-canonical U:G pair (found in I1R4 by FOLDRNA) is common among the 16S rRNA of (eu)bacteria and other RNA helices (Gutell et al. 1994; and references within). Although the I1R4 stem-loop structure for *R. pomonella* is outlined according to FOLDRNA calculations in Figure 2, it is more likely that the loop region for this species is the same as

Figure 5. Typical computer generated secondary-structure models for *Rhagoletis* ITS1 (A and B) and ITS2 (C and D). Panels A and C correspond to *R. osmanthi*, while Panels B and D correspond to *R. pomonella*. Several of the domains indicated in Figs. 2 and 3 are boxed and labeled. Free energy values determined by the FOLDRNA program in kcal/mol are indicted for each structure.

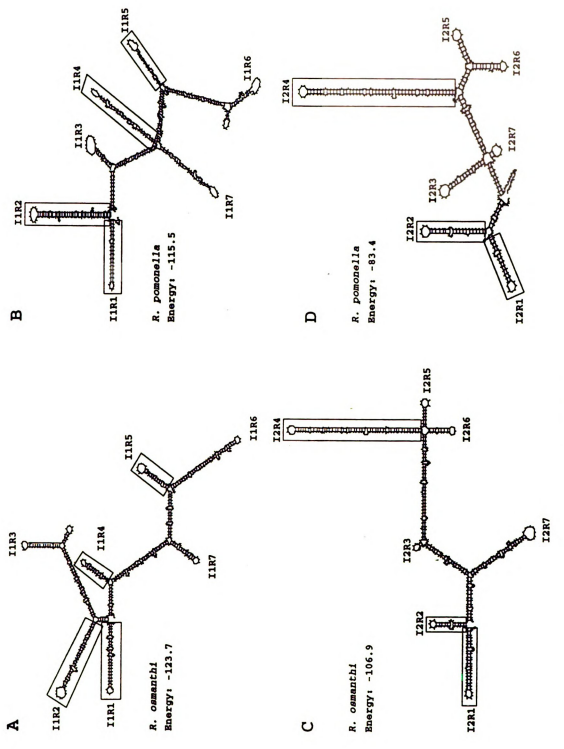

A
I1R3
I1R4
I1R5
I1R6
I1R2
I1R1
I1R7
R. oamanthi
Energy: -123.7
B
I1R2
I1R3
I1R4
I1R5
I1R6
I1R1
I1R7
R. pomonella
Energy: -115.5
C
I2R4
I2R5
I2R6
I2R3
I2R2
I2R7
I2R1
R. oamanthi
Energy: -106.9
D
I2R4
I2R5
I2R6
I2R3
I2R7
I2R2
I2R1
R. pomonella
Energy: -83.4

Figure 5

Figure 6. Sequences and alignment of specific structural domains found in the *Rhagoletis* ITS sequences. The numbers in brackets show the number of flies sequenced. The proposed loop regions for the secondary structural domains are boxed and labeled (see Figs. 2, 3 and 7). The sequences in Panels A, B and C correspond to domains I1R4, I1R5 and I2R4, respectively, and in two cases (Panels A and B) sequences adjacent to the domains (as defined in Figs. 2, 3 and 7) are shown. Asterisks in Panel C correspond to identical nucleotides found at those positions among 8 *Drosophila* species (see text). Gaps and identities are denoted by dots and dashes, respectively. Nucleotide positions are numbered by following the numbering system of Figs. 2 and 3. Furthermore, to remain consistent with the numbering system in Figs 2 and 7, the nucleotide positions in Panel A correspond only to *cingulata* species group and *R. pomonella* (i.e., gaps are not numbered). R stands for A or G residues. R. cin = *R. cingulata*, R. ind = *R. indifferens*, R. chi = *R. chionanthi*, R. osm = *R. osmanthi*, R. pom = *R. pomonella*, R. cor = *R. cornivora*, R. jun = *R. juniperina* and R. fau = *R. fausta*.

A

R. cin
R. ind
R. chi
R. osm
R. pom
R. cor
R. jun
R. fau

B

R. cin
R. ind
R. chi
R. osm
R. pom
R. cor
R. jun
R. fau

C

R. cin
R. ind
R. chi
R. osm
R. pom
R. cor
R. jun
R. fau

Figure 6

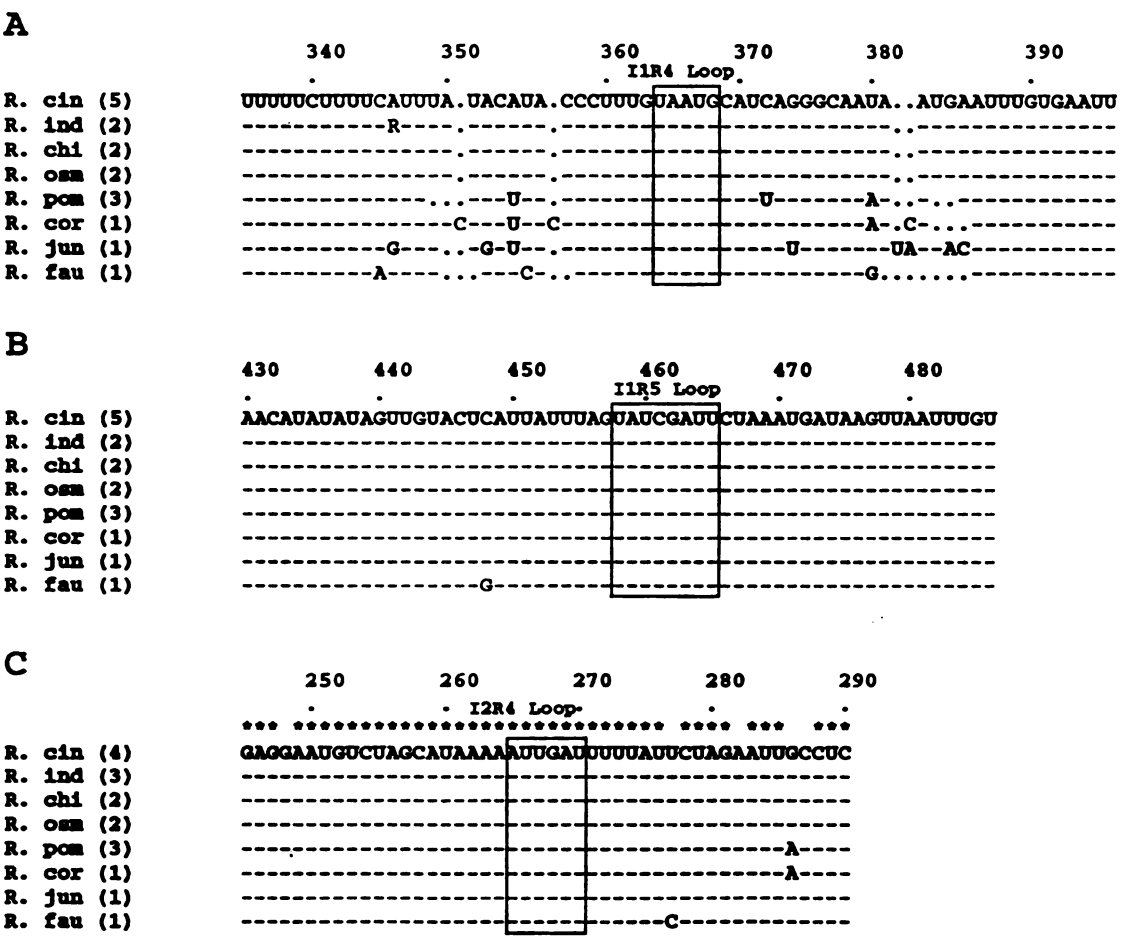
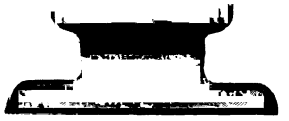

A
340
350
360
370
380
390
I1R4 Loop
R. cin (5) UUUUUCUUUUCAUUUA.UACAUA.CCCUUUGUAAUGCAUCAGGGCAAUA..AUGAAUUUGUGAAUU
R. ind (2) ----------R----.------.------------------------..---------------
R. chi (2) ----------.------.------------------------..---------------
R. osm (2) ----------.------.------------------------..---------------
R. pom (3) ----------...---U--.---------------U-------A-..-..------------
R. cor (1) ----------C---U--C---------------------A-.C-...-----------
R. jun (1) ----------G---..-G-U--.---------------U-------UA--AC----------
R. fau (1) ----------A----...---C-..---------------G.......----------
B
430
440
450
460
470
480
I1R5 Loop
R. cin (5) AACAUAUAUAGUUGUACUCAUUAUUUAGUAUCGAUUCUAAAUGAUAAGUUAAUUUGU
R. ind (2) ----------
R. chi (2) ----------
R. osm (2) ----------
R. pom (3) ----------
R. cor (1) ----------
R. jun (1) ----------
R. fau (1) ----------G----------
C
250
260
270
280
290
I2R4 Loop
R. cin (4) GAGGAAUGUCUAGCAUAAAAAUUGAUUUUUAUUCUAGAAUUGCCUC
R. ind (3) ----------
R. chi (2) ----------
R. osm (2) ----------
R. pom (3) ----------A----
R. cor (1) ----------A----
R. jun (1) ----------
R. fau (1) ----------C----------

Figure 6

Figure 7. Secondary-structure models for specific domains in the *Rhagoletis* ITS sequences. Proposed canonical (Watson-Crick) base pairs are connected by lines, and non-canonical U:G pairs are connected by filled dots. Nucleotide positions are marked with a tick mark and numbered every 10th position; the first and last positions for each structural element are numbered. Thick arrows and symbols with double arrowheads represent nucleotide or nucleotide pair replacements at those positions. Nucleotides associated with thick arrowheads denote additions at the specified position among specific *Rhagoletis* species (nucleotides not circled) and *Drosophila* species (circled nucleotides). Panel A represents the secondary structure model for domain I1R4 in ITS1 for the *R. cingulata* species group. Nucleotides in bold are invariant among all the *Rhagoletis* species surveyed in this study. Several nucleotide or nucleotide pair replacements (symbols with double arrowheads) or additions (thick arrowhead) are shown for *R. pomonella* and *R. cornivora* only (see Fig. 6A and text). The boxed nucleotide pairs are absent among *R. pomonella*. Panel B represents the secondary structure model for domain I1R5 in ITS1 (see Figs. 5A and 5B). Nucleotides in bold are invariant among all the *Rhagoletis* species surveyed in this study. Panel C represents the secondary structure model for a partial region of domain I2R4 in ITS2 (see Figs. 3, 5C and 5D). Nucleotides in bold are invariant among all the *Rhagoletis* species surveyed in this study and among 8 other species of *Drosophila* (see text) and correspond to those with asterisks in Figure 6. Circled nucleotides or nucleotide pairs correspond to those found among specific *Drosophila* species (see text). The boxed nucleotide replacement found for *Rhagoletis* was absent in the *Drosophila* species.



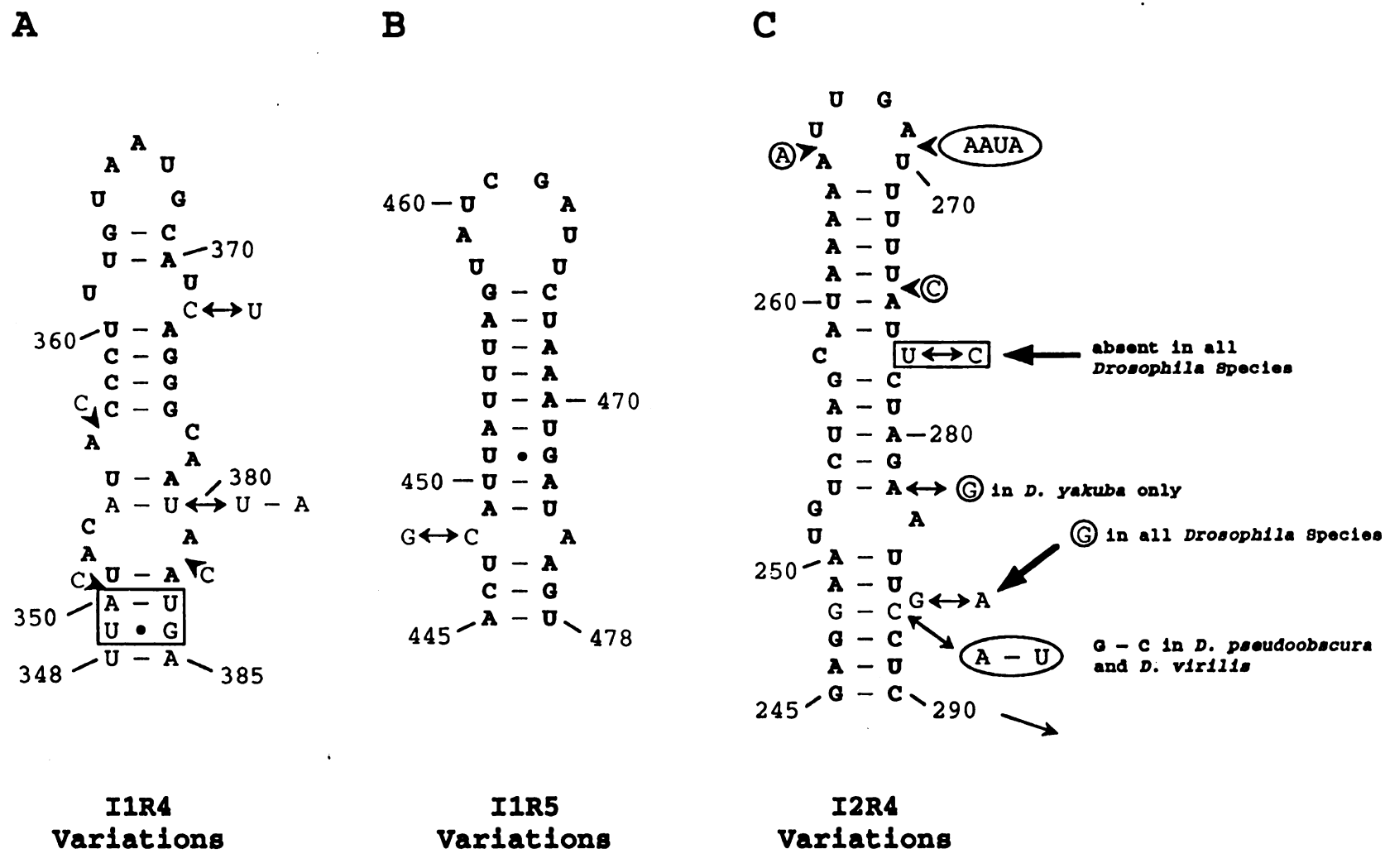

A
B
C
AAUA
270
260
absent in all
Drosophila Species
280
in D. yakuba only
in all Drosophila Species
250
G ↔ A
G – C in D. pseudoobscura
and D. virilis
A – U
245
290
460
470
450
445
478
370
380
360
350
348
385
I1R4
Variations
I1R5
Variations
I2R4
Variations

Figure 7

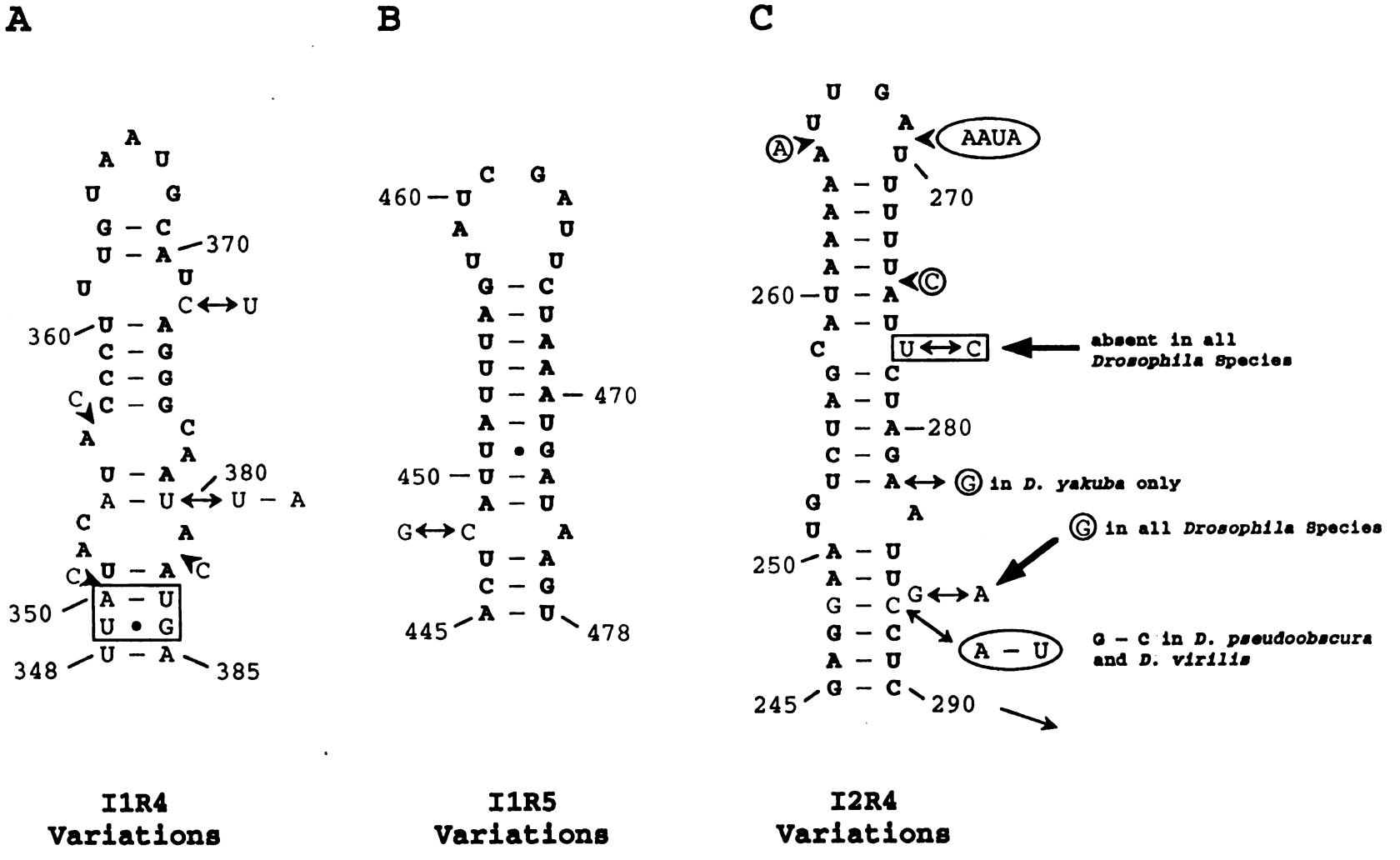

A
B
C
AAUA
270
260
280
absent in all Drosophila Species
U ↔ C
in D. yakuba only
in all Drosophila Species
G ↔ A
A – U
G – C in D. pseudoobscura and D. virilis
290
245
250
360
370
380
385
348
350
460
470
450
445
478
C ↔ U
U ↔ A
G ↔ C
I1R4 Variations
I1R5 Variations
I2R4 Variations

Figure 7

that for the *R. cingulata* species group (i.e.UAAUG) because of the above mentioned compensatory changes. Apart from the three compensatory changes, there was only one C to T transition (from the *cingulata* group to *R. pomonella*) found at position 372 (Figures 6A and 7A), which is a relatively less significant position regarding secondary structure because of its location at a bulge (i.e., non base-pairing region; Figure 7A). *R. cornivora* sequence for the I1R4 region appears to diverge more than that of *R. pomonella*; for example, three insertions of C residues after positions 350, 356 and 381 (Figures 6A and 7A) are noted — again, at or near bulges. Actually, the insertion of C residues may even enhance the stem structure because additional non-canonical base pairs can be potentially formed (Figure 7A, the C residues with arrowheads). In two other *Rhagoletis* species (*juniperina* and *fausta*), the potential loop motif in I1R4 was still preserved, however, the flanking region, especially those forming the lower part of the stem structure show more divergence in these two species (Figure 6A). The above observations in I1R4 are compatible with the electrophoretic results which indicated that the closest relatives of the *cingulata* group are the members of the *pomonella* group which consists of several sibling species such as *R. pomonella, R. mendax, R. zephyria* and *R. cornivora* (Berlocher and Bush 1982; Berlocher et al 1993).

The I1R5 structural element, like the I1R4, was also highly conserved among the 8 *Rhagoletis* species surveyed here (Figures 6B and 7B). The extensive numbers of canonical base pairs along the proposed stem (Fig. 7B) and the common non-canonical U451:G472 pair (determined by FOLDRNA; Gutell et al. 1994) leaves almost no doubt that the proposed secondary structure indeed exists. Actually the I1R5 appears even more constrained than the I1R4 because there is no sequence variation in this region of more than 50 nucleotides among all the 8 *Rhagoletis* species, except the C residue was replaced with G in *R. fausta* at position 448 (Figure 6B), which does not significantly alter the secondary structure because of its location at a bulge (Figure 7B).

ITS2 Secondary Structures

Several secondary structural domains in ITS2, determined using FOLDRNA, are shown in Figure 3. The computer generated secondary structure models for *R. osmanthi* and *R. pomonella* are shown in Figures 5C and 5D, respectively. The structural domains at the 5' end of ITS2, including the potential stem-loop structures I2R1, I2R2, and I2R4, appear to be highly conserved between the *cingulata* species group and *R. pomonella* (Figures 3, 5C and 5D). The loop of the I2R4 structural element (AUUGAU) and the remaining sequence from position 245 to 290 was actually conserved even among several other *Rhagoletis* species (Figures 6C and 7C). The FOLDRNA did not pair positions 258 and 277 in I2R4, however, a non-canonical pair of C258:U277 (C:C in *fausta*) can possibly exist because Y:Y (i.e., pyrimidine: pyrimidine) non-canonical pairing, although rare, have been observed in various helices (Gutell et al. 1994). In *R. pomonella* and *R. cornivora* the G at position 286 was replaced by A and this change may not affect the outcome of the secondary structure because this position forms at a bulge in the proposed model (Figures 6C and 7C).

A secondary structural element, described as D3 in *Drosophila* by Schlotterer et al. (1994), appears to be equivalent to the I2R4 of *Rhagoletis* in this study (Figures 6C and 7C). The sequences between positions 215-234 and 250-269 in ITS2 of *D. melanogaster* (nucleotide positions numbers are as in Fig. 1B of Schlotterer et al. 1994) matches almost perfectly with sequences between positions 245-264 and 271-290 in ITS2 of *Rhagoletis* (Figure 6C, letters with asterisks; Figure 7C, nucleotides in bold), except for minor transition and insertion/deletions. There are also compensatory changes in the I2R4 between *Rhagoletis* and *Drosophila*. The computer predicted base-pairing G248:C287 in *Rhagoletis* was replaced by A:U (218:266) in 6 *Drosophila* species (*sechelia*, *simulans*, *mauritiana*, *melanogaster*, *orena* and *yakuba*, position numbers for *Drosophila* as in Figure 1B of Schlotterer et al. 1994). However, in *D. pseudobscura* and *D. virilis*, which

diverged much earlier from the above mentioned six *Drosophila* species, the same pairing position remains G:C as in the *Rhagoletis* species surveyed here (Figure 7C).

Discussion

Although the primers were designed based on coding sequences of distant relatives of *Rhagoletis*, such as *Drosophila*, the PCR amplification of *Rhagoletis* rDNA spacers were successful. The PCR reaction conditions determined in this study yielded the desired products. The same primers and reaction conditions will be useful guides for future application of the rDNA spacers in *Rhagoletis* studies.

The ITS regions are thought of as being universal fast-evolving genomic DNA regions suitable for resolving closely related species that otherwise show little genetic divergence (Brown et al. 1972; Furlong and Maden 1983; Tautz et al. 1987; Porter and Collins 1991). In the case of *Rhagoletis*, however, the ITS regions seem to display a limited amount of genetic variation and only a few informative characters were available for inferring a phylogenetic relationship among the closely related sibling species in the *cingulata* group. Some of the observed sequence variations could be attributed to *in vitro* slippage synthesis events in simple repeat motifs. Such variations, however, generate phylogenetically uninformative (autapomorphic and homoplastic) characters, which do not affect the results of phylogenetic analysis. Other observed slippage mutations in simple repeats may have an *in vivo* basis, created probably in regions of low selective (structural and/or functional) constraints. This interpretation has been suggested by other researchers as well (Tautz and Schlotterer 1994; and references within). A few key nucleotide positions, however, have been described in ITS1 that could be useful to distinguish some of the morphologically indistinguishable species in this group. By designing primers from the conserved regions of ITS1 adjacent to positions 115 and 137, and taking advantage of current PCR and sequencing technology, one could in a short time differentiate between the cherry and olive infesting species of insects. Moreover, the considerable divergence

discovered between the *cingulata* group and *R. pomonella* in this study (Figures 2 and 3; Table III) indicates that the ITS sequences may provide valuable information in the phylogenetic study of taxa from different species groups of the genus *Rhagoletis*.

Although a secondary-structure model of minimum free energy can be produced with the FOLDRNA, the real secondary structure may, on the basis of base-pairing and other constraints, have a different free energy value (Zucker and Stiegler 1981; Gutell 1993; Gutell et al. 1994; and references within). For the past decade, a comparative approach based on the concept of positional covariance has been applied to elucidate the *Escherichia coli* 16S and 23S rRNA higher-order structure and identify functionally important elements in these molecular structures (for review see Gutell et al. 1994). In this study, using the FOLDRNA program and some principles of comparative analysis based on compensatory nucleotide changes, a number of constrained secondary structural elements in ITS of several *Rhagoletis* species have been described. One of them, namely I2R4, is highly conserved between *Rhagoletis* and *Drosophila* indicating the possible functional importance of this stem-loop structure given the estimated divergence time between the families Tephritidae and Drosophilidae which ranges from 77 MYR (million years ago; Kwiatowski et al. 1994) to 90 MYR (Collier and MacIntyre 1977) and even to 123 MYR (Beverley and Wilson 1984). The *Drosophila* radiation is estimated to be between 40 to 62 MYA (Beverley and Wilson 1984; Spicer 1988; Kwiatowski et al. 1994), at least 15 MYR after the divergence of Tephritidae and Drosophilidae. Phylogenetic studies of *Drosophila* based on the ITS sequence comparisons (Schlotterer et al. 1994) implies that *D. psuedoobscura* and *D. virilis* , both with G:C at a specific pairing position in I2R4, diverged much earlier than the 6 *Drosophila* species with A:U at the same paring position. This evolutionary divergence suggests that the G:C pair in the secondary structural element of ITS2 (248:287 in Figure 7C) among the 8 known *Rhagoletis* species and the two older *Drosophila* species is a primitive condition, which later may have mutated to A:U among the 6 *Drosophila* species sometime during their common evolutionary

history. However, convergent evolution by independent substitution events (A:U to G:C) having occurred in *Drosophila* and *Rhagoletis* can also be envisioned and can not be completely ruled out.

Among the ITS1 structural elements investigated, the proposed I1R5 stem-loop structure was highly conserved among all the 8 *Rhagoletis* species presented in Figures 6B and 7B, while another such element, I1R4, demonstrated high conservation only among the *cingulata* species group and diverges considerably beyond this species group (Figures 6A and 7A). This observation suggests the existence of differential levels of constraint throughout the length of ITS1, presumably due to its functional or higher order structural role(s). Furthermore, the variation in the I1R4 suggests that the *cingulata* species group may be more closely related to the members in the *pomonella* group rather than to *R. juniperina* and *R. fausta*, as indicated by electrophoretic data (Berlocher and Bush 1982; Berlocher et al 1993).

To obtain further information on the phylogeny of the *cingulata* group, a faster evolving DNA may be more satisfactory, such as those coding for alcohol dehydrogenase which show considerable divergence among sibling species of *Drosophila* (Bodmer and Ashburner 1984) and other non-coding rDNA spacers, i.e., external transcribed spacers and intergenic spacers (see chapter IV and V)

In conclusion, I have presented the complete ITS sequences of the 4 members of the *cingulata* species group as well as that of *R. pomonella*. The *Rhagoletis* ITS sequences are highly A-T rich. I found low levels of interspecific ITS variation in the *cingulata* species group, implying that ITS sequences are of limited application in phylogenetic analysis of host-associated populations and/or closely related sibling species. A few molecular markers have been described and can be potentially useful for distinguishing the olive infesting species (*R. osmanthi* and *R. chionanthi*) from the two cherry flies and for differentiating between the two olive infesting flies. The high sequence divergence found between the *cingulata* group and *R. pomonella* indicates that the ITS regions can provide

better resolution for studying phylogenetic relationship of taxa from different *Rhagoletis* species groups. The overall computer generated secondary structure model of the ITS was presented for the *cingulata* species group and *R. pomonella*. Several highly conserved secondary structural elements were determined by the FOLDRNA and comparative analysis. One such element (I2R4) seems to be conserved even among two distant families, Tephritidae and Drosophilidae, which may have diverged in the mid to late Cretaceous period (65 to 130 million years ago). The conserved ITS secondary structural elements should be a valuable guide for accurate alignment of taxa from different species groups of *Rhagoletis*. The mode of ITS evolutionary divergence should also be useful in future investigations of the structure, function and processing of precursor rRNA. Furthermore, future studies on the ITS of other *Rhagoletis* species will allow us to elucidate the degree of functional and structural constraints on the ITS sequences in *Rhagoletis*.

BIBLIOGRAPHY FOR CHAPTER II

Bibliography for Chapter II

Bakker, F. T., J. L. Olsen, W. T. Stam, and C. van den Hoek. 1992. Nuclear ribosomal DNA internal transcribed spacer regions (ITS1 and ITS2) define discrete biogeographic groups in *Cladophora albida* (Chlorophyta). J. Phycol. **28**:839-845.

Baldwin, B. G. 1992. Phylogenetic utility of the internal transcribed spacers of ribosomal DNA in plants: an example from the Compositae. Mol. Phylogenet. Evol. **1**:3-16.

Berlocher, S. H., and G. L. Bush. 1982. An electrophoretic analysis of *Rhagoletis* (Diptera: Tephritidae) phylogeny. Syst. Zoo. **31**:136-155.

Berlocher, S. H., B. A. McPheron, J. L. Feder, and G. L. Bush. 1993. Genetic Differentiation at allozyme loci in the Rhagoletis pomonella (Diptera: Tephritidae) species complex. Ann. Entomol. Soc. Am. **86**:716-727.

Beverley, S. M., and A. C. Wilson. 1984. Molecular evolution in *Drosophila* and the higher diptera II. A time scale for fly evolution. J. Mol. Evol. **21**:1-13.

Bodmer, M., and M. Ashburner. 1984. Conservation and change in the DNA sequences coding for alcohol dehydrogenase in sibling species of *Drosophila*. Nature **309**:425-430.

Brown, D. D., P. C. Wensink, and E. Jordan. 1972. A comparison of the ribosomal DNA's of *Xenopus laevis* and *Xenopus mulleri*: the evolution of tandem genes. J. Mol. Biol. **63**:57-73.

Bush, G. L. 1966. The taxonomy, cytology, and evolution of the genus *Rhagoletis* in North America (Diptera: Tephritidae). Bull. Mus. Comp. Zool. **134**: 431-562.

Bush, G. L. 1969. Sympatric host race formation and speciation in frugivorous flies of the genus *Rhagoletis* (Diptera: Tephritidae). Evolution **23**:237-251.

Bush, G. L. 1975. Sympatric speciation in phytophagous parasitic insects. Pp. 187-205 *in* P. W. Price, ed. Evolutionary strategies of parasitic insects and mites. Plenum Publishing Corp., New York.

Bush, G. L. 1993. Host race formation and sympatric speciation in *Rhagoletis* fruit flies (Diptera: Tephritidae). Psyche **99**: 335-357.

Collier, G. E., and R. J. MacIntyre. 1977. Microcomplement fixation studies on the evolution of a-glycerophosphate dehydrogenase within the genus Drosophila. Proc. Natl. Acad. Sci. USA **74**:684-688.

Feder, J. L., and G. L. Bush. 1989. Gene frequency clines for host races of *Rhagoletis pomonella* in the midwestern United States. Heredity **63**:245-266.

Feder, J. L., C. A. Chilcote, and G. L. Bush. 1990a. The geographic pattern of genetic differentiation between host associated populations of *Rhagoletis pomonella* (Diptera: Tephritidae) in the eastern United States and Canada. Evolution **44**:570-594.

Feder, J. L., C. A. Chilcote, and G. L. Bush. 1990b. Regional, local and microgeographic allele frequency variation between apple and hawthorn populations of *Rhagoletis pomonella* in western Michigan. Evolution **44**:595-608.

Fritz, G. N., J. Conn, A. Cockburn, and J. Seawright. 1994. Sequence analysis of the ribosomal DNA internal transcribed spacer 2 from populations of *Anopheles nuneztovari* (Diptera: Culicidae). Mol. Biol. Evol. **11**:406-416.

Furlong, J. C., and B. E. H. Maden. 1983. Patterns of major divergence between internal transcribed spacers of ribosomal DNA in *Xenopus borealis* and *Xenopus laevis*, and of minimal divergence within ribosomal coding regions. EMBO J. **2**:443-448.

Gardes, M., and T. D. Bruns. 1993. ITS primers with enhanced specificity for basidiomycetes: application to the identification of mycorrhizae and rusts. Mol. Ecol. **2**:113-118.

Gelfand, D. H., and T. J. White. 1990. Thermostable DNA polymerases. Pp. 129-141 in M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White, eds. PCR protocols a guide to methods and applications. Academic Press Inc., New York.

Gutell, R. R. 1993. The simplicity behind the elucidation of complex structure in ribosomal RNA. Pp. 477-488 *in* K. H. Nierhaus, ed. The translational apparatus. Plenum Press, New York.

Gutell, R. R., N. Larsen, and C. R. Woese. 1994. Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. Microbiol. Rev. **58**:10-26.

Hadjiolov, A. A. 1985. The nucleolus and ribosome biogenesis. Springer Verlag, New York.

Hernandez-Ortiz, V. 1993. Description of a new *Rhagoletis* species from tropical Mexico (Diptera: Tephritidae). Proc. Entomol. Soc. Wash. **95**:418-424.

Hillis, D. M., and S. K. Davis. 1986. Evolution of ribosomal DNA: fifty million years of recorded history in the frog genus *Rana*. Evolution **40**:1275-1288.

Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules. Pp. 21-132 *in* H. N. Munro, ed. Mammalian protein metabolism. Academic Press, New York.

Kooistra, W. H. C. F., W. T. Stam, J. L. Olsen, and C. van den Hoek. 1992. Biogeography of *Cladophoropsis membranacea* (Chlorophyta) based on comparisons of nuclear rDNA ITS sequences. J. Phycol. **28**:660-668.

Kupriianova, N. S., M.I. Timofeeva, and A. A. Baev. 1989. Secondary structure of pre-RNA 41S spacers and its possible role in determining the processing sites. Mol. Biol. (Mosk) **23**:1611-1615.

Kwiatowski, J., D. Skarecky, K. Bailey, and F. J. Ayala. 1994. Phylogeny of *Drosophila* and related genera inferred from the nucleotide sequence of the Cu,Zn *Sod* gene. J. Mol. Evol. **38**:443-454.

Lee, S. B., and J. W. Taylor. 1991. Phylogeny of five fungal-like protoctistan Phytophthora species inferred from the internal transcribed spacers of ribosomal DNA. Mol. Biol. Evol. **8**: 620-640.

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**:523-538.

McPheron, B. A., D. C. Smith, and S. H. Berlocher. 1988. Microgeographic genetic variation in the apple maggot *Rhagoletis pomonella*. Genetics **119**:445-451.

Mirkovitch, J., M. -E. Mirault, and U. K. Laemmli. 1984. Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold. Cell **39**:223-232.

Moriyama, E. N., and D. L. Hartl. 1993. Codon usage bias and base composition of nuclear genes in *Drosophila*. Genetics **134**:847-858.

Musters, W., K. Boon, C. A. F. M. van der Sande, H. van Heerikhuizen, and R. J. Planta. 1990. Functional analysis of transcribed spacers of yeast ribosomal DNA. EMBO J. **9**:3989-3996.

Nazar, R. N., X. Hu, J. Schmidt, D. Culham, and J. Robb. 1991. Potential use of PCR-amplified ribosomal intergenic sequences in the detection and differentiation of verticillium wilt pathogens. Physiol. Mol. Plant Pathol. **39**:1-11.

O'Donnell, K. 1992. Ribosomal DNA internal transcribed spacers are highly divergent in the phytopathogenic ascomycete *Fusarium sambucinum* (*Gibberella pulicaris*). Curr. Genet. **22**:213-220.

Olsen, G. L., and C. R. Woese. 1993. Ribosomal RNA: a key to phylogeny. FASEB **7**:114-123.

Pleyte, K. A., S. D. Duncan, and R. B. Philips. 1992. Evolutionary relationships of the salmonid fish genus Salvelinus inferred from DNA sequences of the first internal transcribed spacer (ITS1) of ribosomal DNA. Mol. Phylogenet. Evol. **1**:223-230.

Porter, C. H., and F. H. Collins. 1991. Species-diagnostic differences in a ribosomal DNA internal transcribed spacer from the sibling species *Anopheles freeborni* and *Anopheles hermsi* (Diptera: Culicidae). Am. J. Trop. Med. Hyg. **45**:271-279.

Procunier, W. S., and J. J. Smith. 1993. Localization of ribosomal DNA in *Rhagoletis pomonella* (Diptera: Tephritidae) by *in situ* hybridization. Insect Mol. Biol. **2**:163-174.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **12**:5463-5467.

Schlotterer, C., and D. Tautz. 1994. Chromosomal homogeneity of *Drosophila* ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. Curr. Biol. **4**:777-783.

Schlotterer, C., and D. Tautz. 1994. Slippage synthesis of simple sequence DNA. Nucleic Acids Res. **20**:211-215.

Schlotterer, C., M. Hauser, A. von Haeseler, and D. Tautz. 1994. Comparative evolutionary analysis of rDNA ITS regions in *Drosophila*. Mol. Biol. Evol. **11**:513-522.

Sollner-Webb, B., and J. Tower. 1986. Transcription of cloned eukaryotic ribosomal RNA genes. Annu. Rev. Biochem. **55**:801-830.

Spicer, G. S. 1988. Molecular evolution among some *Drosophila* species groups as indicated by two-dimensional electrophoresis. J. Mol. Evol. **27**:250-260.

Swofford, D. L. 1993. PAUP: phylogenetic analysis using parsimony, version 3.1.1. University of Illinois, Champaign.

Tamura, K. 1992. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C content biases. Mol. Biol. Evol. **9**:678-687.

Tautz, D., and C. Schlotterer. 1994. Simple sequences. Curr. Opin. Genet. Dev. **4**:832-837.

Tautz, D., C. Tautz, D. Webb, and G. A. Dover. 1987. Evolutionary divergence of promoters and spacers in the rDNA family of four *Drosophila* species: implications for molecular coevolution in multigene families. J. Mol. Biol. **195**:525-542.

Tautz, D., J. M. Hancock, D. A. Webb, C. Tautz, and G. A. Dover. 1988. Complete sequences of the rRNA genes of Drosophila melanogaster. Mol. Biol. Evol. **5**:366-376.

Torres, R. A., M. Ganal, and V. Hemleben. 1990. GC balance in the internal transcribed spacers ITS1 and ITS2 of nuclear ribosomal RNA genes. J. Mol. Evol. **8**:669-686.

van der Sande, C. A. F. M., M. Kwa, R. W. van Nues, H. van Heerikhuizen, H. A. Raue, and R. J. Planta. 1992. Functional analysis of internal transcribed spacer 2 of *Saccharomyces cerevisiae* ribosomal DNA. J. Mol. Biol. **223**:899-910.

van Holde, K. E. 1989. Chromatin. *In* A. Rich, ed. Springer series in molecular biology. Springer-Verlag, New York.

Vogler, A. P., and R. DeSalle. 1994. Evolution and phylogenetic information content of the ITS-1 region in the tiger beetle *Cicindela dorsalis*. Mol. Biol. Evol. **11**:393-405.

Wesson, D. M., C. H. Porter, and F. H. Collins. 1992. Sequence and secondary structure comparisons of ITS rDNA in mosquitoes (Diptera: Culicidae). Mol. Phylogenet. Evol. **1**:253-269.

Wingfield, B. D., W. S. Grant, J. F. Wolfaardt, and M. J. Wingfield. 1994. Ribosomal RNA sequence Phylogeny is not congruent with ascosopore morphology among species in *Ceratocystis* Sensu Stricto. Mol. Biol. Evol. **11**:376-383.

Yeh, L. C., and J. C. Lee. 1990. Structural analysis of the internal transcribed spacer 2 of the precursor ribosomal RNA from *Saccharomyces cerevisiae*. J. Mol. Biol. **211**:699-712.

Yokota, Y., T. Kawata, Y. Iida, A. Kato, and S. Tanifuji. 1989. Nucleotide sequences of the 5.8S rRNA gene and internal transcribed spacer regions in carrot and broad bean ribosomal DNA. J. Mol. Evol. **29**:294-301.

Zucker, M., and P. Stiegler. 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Res. **9**:133-148.

CHAPTER III

PHYLOGENETIC IMPLICATIONS FROM ANALYSIS OF INTERNAL TRANSCRIBED SPACER REGIONS IN THE rDNA OF TEN *RHAGOLETIS* SPECIES

Introduction

Nearctic *Rhagoletis* were comprehensively reviewed by Bush (1966), with 21 species segregated into seven species groups (*pomonella, cingulata, tabellaria, suavis, ribicola, striatella* and *alternata*). Bush's revision was based mainly on morphology and karyotype analysis. Since then attention has been given to the mode of host race formation and sympatric speciation in one of the *Rhagoletis* species groups (i.e. the *pomonella* group). Currently, the only phylogenetic analyses of this group are those based on allozyme studies (Berlocher and Bush 1982; Berlocher et al 1993), and mtDNA COII (Smith and Bush, unpublished data) and morphological characters (Jenkins, in preparation). The results from the allozyme studies have some congruence with the conventional classification of Bush (1966), such as the conservation of the *suavis*, *cingulata*, and to some extent, *pomonella* species groups. However, there are a few discrepancies between the previous studies — even among the biochemical and molecular studies (i.e., the allozyme and mitochondrial DNA studies). To date, the placement of several species in this genus, such as *R. juniperina, cornivora, striatella, fausta* and *basiola*, is still uncertain. In addition, the phylogenetic relationships among different species groups are yet to be completely resolved. For example, the close relatives of the *cingulata* group in North America may either be the *pomonella* group, according to

electrophoretic studies (Berlocher and Bush 1982; Berlocher et al. 1993), or the *suavis* group, according to mtDNA COII data (Smith and Bush, unpublished data).

Although the *Rhagoletis* species of the Nearctic region have been well documented and segregated into a number of species groups, the phylogeny of *Rhagoletis* in the Nearctic region is not well established as there are still several unanswered questions regarding placement of certain species and relationships among the existing species groups. Furthermore, monophyly of the genus has not been demonstrated and its relationship to the closely related genus *Carpomyina* is poorly understood (Norrbom 1989). Since the genus *Rhagoletis* contains many economically important pest species and some of *Rhagoletis* species groups are believed to speciate sympatrically through host shifting, a reliable phylogenetic framework is not only required for understanding the evolutionary history of this genus and testing evolutionary theories such as speciation, but also required for providing accurate information for appropriate control of those pest species, some of which show little or no morphological differences. Therefore, in order to get further insight into the phylogeny of this group, I have sequenced the internal transcribed spacers (ITS1 and ITS2) of the nuclear rDNA of 7 additional *Rhagoletis* species.

The taxa covered in this ITS sequence analysis are the following nine species: *R. cingulata* and *indifferens* represent the *cingulata* group; *R. pomonella* and *cornivora* conventionally belong to the *pomonella* group, but *cornivora*'s placement in this group is questionable and needs verification; *R. completa* represents the well-delimited *suavis* group; *R. electromorpha* and *juniperina* are placed in the *tabellaria* group with *juniperina's* relationship with the rest of the *tabellaria* group yet to be resolved; *R. fausta* is unplaced (Bush, 1966) or its placement in previous independent studies (Berlocher and Bush 1982; Smith and Bush, unpublished data) does not agree; and *R. basiola* which appears to possess some of the most primitive morphological characters, is used as outgroup in this phylogenetic analysis.

The results of this study show both congruence and disagreement with earlier studies. In addition, there are some points which are not recovered in previous analyses. The ITS sequence analyses strongly indicate that 1) *R. cornivora* belongs to the *R. pomonella* group, 2) *R. juniperina* is removed from the *R. tabellaria* group and may be more closely related to the *R. pomonella* group, 3) a possible relationship between *R. fausta* and the *R. tabellaria* group is indicated, 4) the close relatives of the *R. cingulata* species group are more likely the members of the *R. suavis* group, and 5) *R. basiola* is indeed quite distinct from the other *Rhagoletis* species.

In addition to the above phylogenetic implications, the usefulness of the ITS regions in phylogenetic studies and its feasibility in this particular study are discussed. Some sequence features characteristic of *Rhagoletis* rDNA ITS are compared with published ITS information from other insects.

Materials and Methods

Biological Material

All species were collected during 1988-90 from various locations and host plants in the United States of America (Table 5). Larvae emerged from field infested fruit and were allowed to pupate in fine vermiculite. Pupae were sifted from the vermiculite and stored at 4° C for at least 5 months. Pupae were then removed from the cold and held at 22° C under a 15 hr light and 9 hr dark cycle to terminate diapause. Within 2-7 days after emergence most adult flies were frozen at -70° C for subsequent genomic DNA isolation. Specimens from each collection were pinned for species identification.

DNA Isolation and Amplification

Total genomic DNA was isolated from individual male flies as described by Procunier and Smith (1993). The ITS1 region was amplified using primer 1975F 5'TAACAAGGTTTCCGTAGGTG (matching the 3' end of 18S) and primer 35R

Table 5. Collection Sites and Host Plants of the *Rhagoletis* Species Used

Species	Location	Host Plant (Common Name)
1. *R. cingulata*	Hart, MI	*Prunus cerasus* (sour cherry)
2. *R. indifferens*	Pullman, WA	*Prunus cerasus* (sour cherry)
3. *R. completa*	Grand Junction, CO	*Juglans nigra* (black walnut)
4. *R. juniperina*	Dixon Springs, IL	*Juniperus virginiana* (E. red cedar)
5. *R. fausta*	Fish Creek, WI	*Prunus cerasus* (sour cherry)
6. *R. electromorpha*	Meridian Twshp., MI	*Cornus foemina*
7. *R. pomonella*	E. Lansing, MI	*Crataegus mollis* (hawthorn)
8. *R. cornivora*	E. Lansing, MI	*Cornus amomum* (dogwood berries)
9. *R. basiola*	Montrose, CO	*Rosa* spp.

5'AGCTRGCTGCGTTCTTCATCGA (matching the 5' end of 5.8S). The ITS2 region was amplified using primer 108F 5'GAACATCGACHHKTYGAACGCA (matching the 3' end of 5.8S) and primer 52R 5'GTTAGTTTCTTTTCCTCCSCT (matching the 5' end of 28S). Amplification of the ITS1 and ITS2 regions as a combined region on one DNA fragment was performed for *R. basiola*, *R. completa*, *R. fausta*, *R. juniperina*, *R. cingulata*, and *R. indifferens*, using the 1975F and 52R primers. The ITS1 and ITS2 regions from *R. cornivora* and *R. pomonella* were amplified as separate pieces but from the same individual fly. From *R. electromorpha*, I was able to amplify only the ITS2 region. Amplification conditions were as described in Chapter II. Amplified DNA was subjected to electrophoresis on a 1.0% agarose gel and visualized with ethidium bromide. Bands containing the DNA were excised from the gel and the DNA purified using the Prep-A-Gene DNA purification matrix (Bio-Rad), according to the manufacturer's instructions.

Cloning and Sequencing

The TA Cloning kit (Invitrogen) was used in a one-step cloning strategy for the direct insertion of the purified PCR products (see earlier sections) into a plasmid vector, followed by transformation into competent cells. Plasmid vector and competent cells were supplied by the manufacturer. Plasmid DNA was purified from individual clones using the Magic-Prep DNA purification kit (Promega), following the manufacturer's instructions. DNA sequencing was performed according to the chain-termination method of Sanger et al. (1977), and using the Sequenase Version 2.0 DNA sequencing kit (USB) and ^{35}S-dATP (Amersham). The same primers used in the amplification reactions were employed to determine the DNA sequence in both directions. Once a certain stretch of DNA was sequenced additional primers were employed to complete the sequencing of the whole specific region from the different species: Primers 35R-GB27 5'ACC(CT)AAACATTTTCAAGT(CT)GCG (for all the species) and 1975F-GB28 5'AAATAAGCCAAACAAAGGAG (for *basiola* alone) were used for the ITS1 regions;

and primer 108F-GB25 5'A(AT)(AG)(AG)AATC(AT)(CT)AGTATTCCC was used for the ITS2 regions for all the species. The ITS sequences of *R. cingulata*, *R. indifferens* and *R. pomonella* were taken from an extensive study of intraspecific polymorphism's in the ITS region (RC4, RI1 and RP1 respectively in Chapter I).

DNA Sequence and Phylogenetic Analysis

The PILEUP program in the GCG package of the University of Wisconsin Genetics Computer Group (UWGCG package, version 7.0) was used to align the ITS sequences. Alignments were done first with the computer (gap weight = 3.00 and gap length weight = 0.20). They were then manually adjusted taking into account, in some cases, secondary structural constraints as determined using the FOLDRNA program in GCG and/or by a comparative analysis approach of looking for compensatory substitutions between taxa (Chapter I).

Because several secondary structural and/or potential functional elements may constrain *Rhagoletis* ITS1 and ITS2 sequence evolution differently in different parts of these molecules, two different approaches were used for inferring a *Rhagoletis* phylogeny for the nine species surveyed in this study. One approach is based on the analysis of pairwise-distance measures and another on the analysis of discrete molecular characters.

Phylogenetic analyses by parsimony methods were carried out using the programs in PAUP version 3.1.1 (Swofford 1993). The exhaustive search option was used to generate the most parsimonious tree(s). In addition, throughout these analyses, all uninformative characters were ignored and gaps were treated as missing data. Bootstrap 50% majority-rule consensus trees using the branch and bound option in PAUP were constructed from 500 replicates (using seed number 1). The percentage of times that a group of taxa appeared as a clade in the bootstrapped parsimony trees, the bootstrap confidence limits (BCL), were indicated at the internal nodes of the trees.

Phylogenetic analyses by the distance matrix methods were performed using programs in the Molecular Evolutionary Genetics Analysis (MEGA) version 1.01. Estimates of the average percent nucleotide substitutions (number of nucleotide substitutions per 100 nucleotide sites) between each pair of taxa were determined using the method of Kimura (1980) and the percent sampling standard deviations (SD) calculated. All gap and missing information sites were removed in the pairwise comparisons. The Kimura matrix of distances was used to produce phenograms by the neighbor-joining method (Saitou and Nei 1987) using MEGA. As with the parsimony approach mentioned above, bootstrapped distance trees were generated from 100 replications (with random seed number 1), and the BCL in each tree were indicated. With the latter bootstrap analyses, gaps were removed only in the pairwise comparisons. In calculating transition to transversion ratios using MEGA all gap sites were removed from the subset data. Calculations were performed by first employing the Kimura two-parameter distance option in MEGA.

Results

ITS1 Sequence Analysis

The ITS1 sequences of 7 *Rhagoletis* species are presented in Figure 8A. At least three clones for each of these species were obtained but for some species only one clone was sequenced due to cost and time constraints. The remaining clones were preserved at -70° C for future examination of within-individual variation. Similarly, three ITS1 clones for both *R. basiola* and *R. striatella* were obtained and one clone for each species was sequenced. However, their sequences are too diverged to align with the other *Rhagoletis* species (Figure 8B). Because of the inability to PCR amplify the ITS1 region of *R. electromorpha*, the ITS1 sequence for this species was not included in this study.

The length of the ITS1 regions studied here ranged from 625 to 922 bp, with *basiola* about 250bp longer than the other surveyed species (Table 6). The ITS1 of

Figure 8. *Rhagoletis* ITS1 sequences and alignment. A) The complete ITS1 sequences of the species surveyed except for the longer *R. basiola* sequences (Table 6). Phylogenetically informative characters are denoted by asterisks. Identities are denoted by dashes, and gaps are denoted by dots. B) The complete *R. basiola* sequence compared with the partial sequence of *R. striatella*. The alignment was done using GAP in GCG with the gap weight = 1.00 and gap weight length = 0.10. The complete length of *R. striatella* remains to be determined. bas = *R. basiola*, str = *R. striatella*, and ND = not determined.

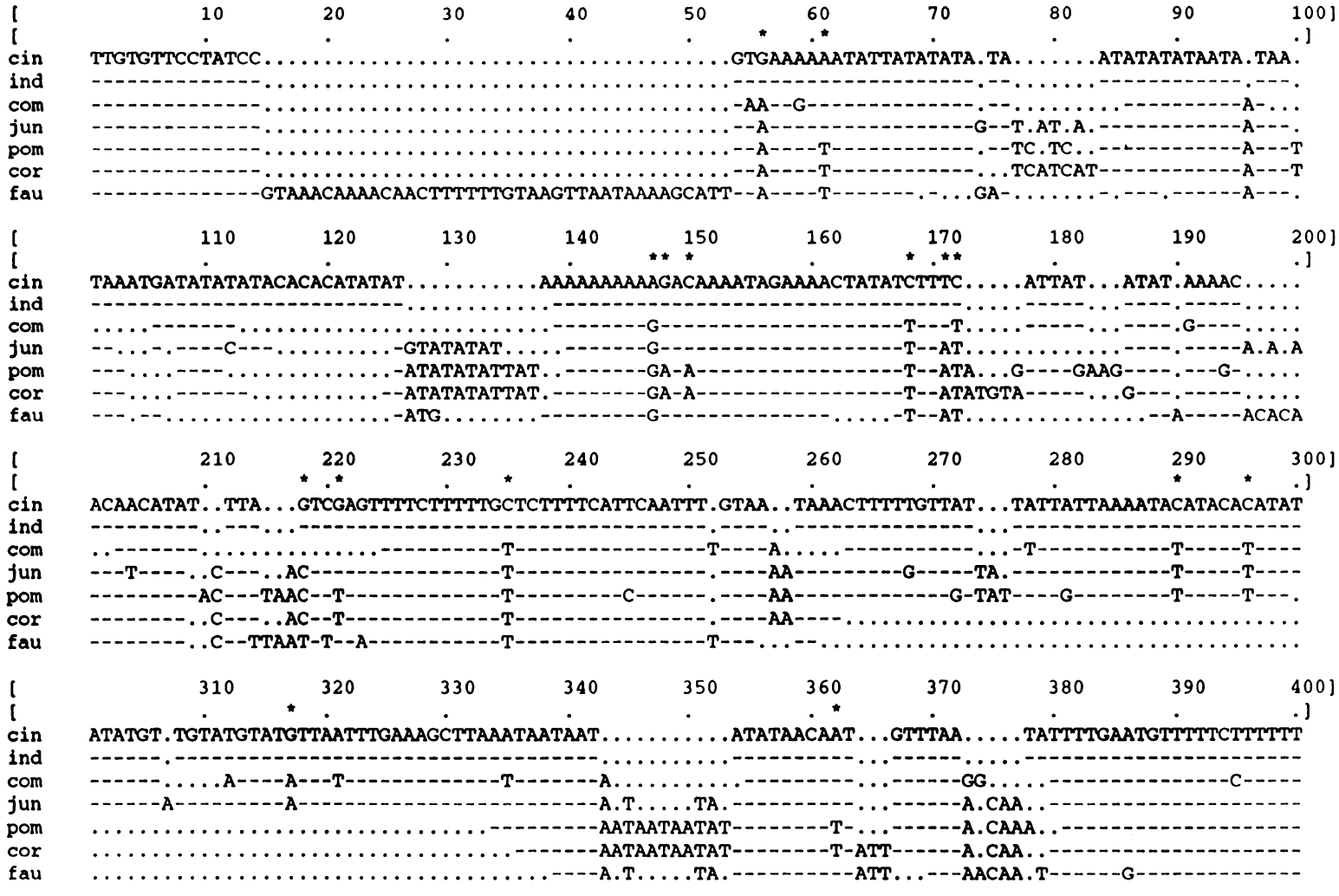

[10 20 30 40 50 60 70 80 90 100]
cin TTGTGTTCCTATCC.....GTGAAAAAATATTATATATA.TA.....ATATATATAATA.TAA.
ind ----------.....----------.--.....----------.---.
com ----------.....-AA--G----------.--.....----------A-...
jun ----------.....--A----------G--T.AT.A.----------A---.
pom ----------.....--A----T----------.--TC.TC..----------A---T
cor ----------.....--A----T----------...TCATCAT----------A---T
fau ----------GTAAACAAAACAACTTTTTTGTAAGTTAATAAAAGCATT--A----T-------.-...GA-.....-.---.------A---.
[110 120 130 140 150 160 170 180 190 200]
cin TAAATGATATATATACACACATATAT.....AAAAAAAAAAGACAAAATAGAAAACTATATCTTTC.....ATTAT...ATAT.AAAAC.....
ind ----------.....----------.....----------.....-----.-----.....
com .....-------.....-------G----------T---T.....-----...-----.G----.....
jun --...-.----C---.....-GTATATAT.....-------G----------T--AT.....----.-----A.A.A
pom ---....----.....-ATATATATTAT..-------GA-A----------T--ATA...G----GAAG----.---G-.....
cor ---....------.....--ATATATATTAT.--------GA-A----------T--ATATGTA------...G---.-----.....
fau ---.--.....-ATG.....--------G----------.....T--AT.....--A-----ACACA
[210 220 230 240 250 260 270 280 290 300]
cin ACAACATAT..TTA...GTCGAGTTTTCTTTTTGCTCTTTTCATTCAATTT.GTAA..TAAACTTTTTGTTAT...TATTATTAAAATACATACACATAT
ind ---------..---...----------.-----..----------...----------
com ..-------.....----------T----------T----A.....----------...-T----------T-----T----
jun ---T----..C---..AC----------T----------.----AA---------G-----TA.----------T-----T----
pom ---------AC---TAAC--T----------T---------C------.----AA----------G-TAT----G---------T-----T---.
cor ---------..C----..AC--T----------T----------.----AA----.....
fau ---------..C--TTAAT-T--A----------T----------T---...--.....
[310 320 330 340 350 360 370 380 390 400]
cin ATATGT.TGTATGTATGTTAATTTGAAAGCTTAAATAATAAT.....ATATAACAAT...GTTTAA.....TATTTTGAATGTTTTTCTTTTTT
ind ------.----------.....----------...------.....----------
com ------.....A----A---T----------T-------A.....----------...------GG.....----------C-----
jun ------A---------A----------A.T.....TA.----------...------A.CAA..----------
pom .....---------AATAATAATAT--------T-...------A.CAAA..----------
cor .....-------AATAATAATAT--------T-ATT------A.CAA..----------
fau .....----A.T.....TA.----------ATT...---AACAA.T------G----------

Figure 8A

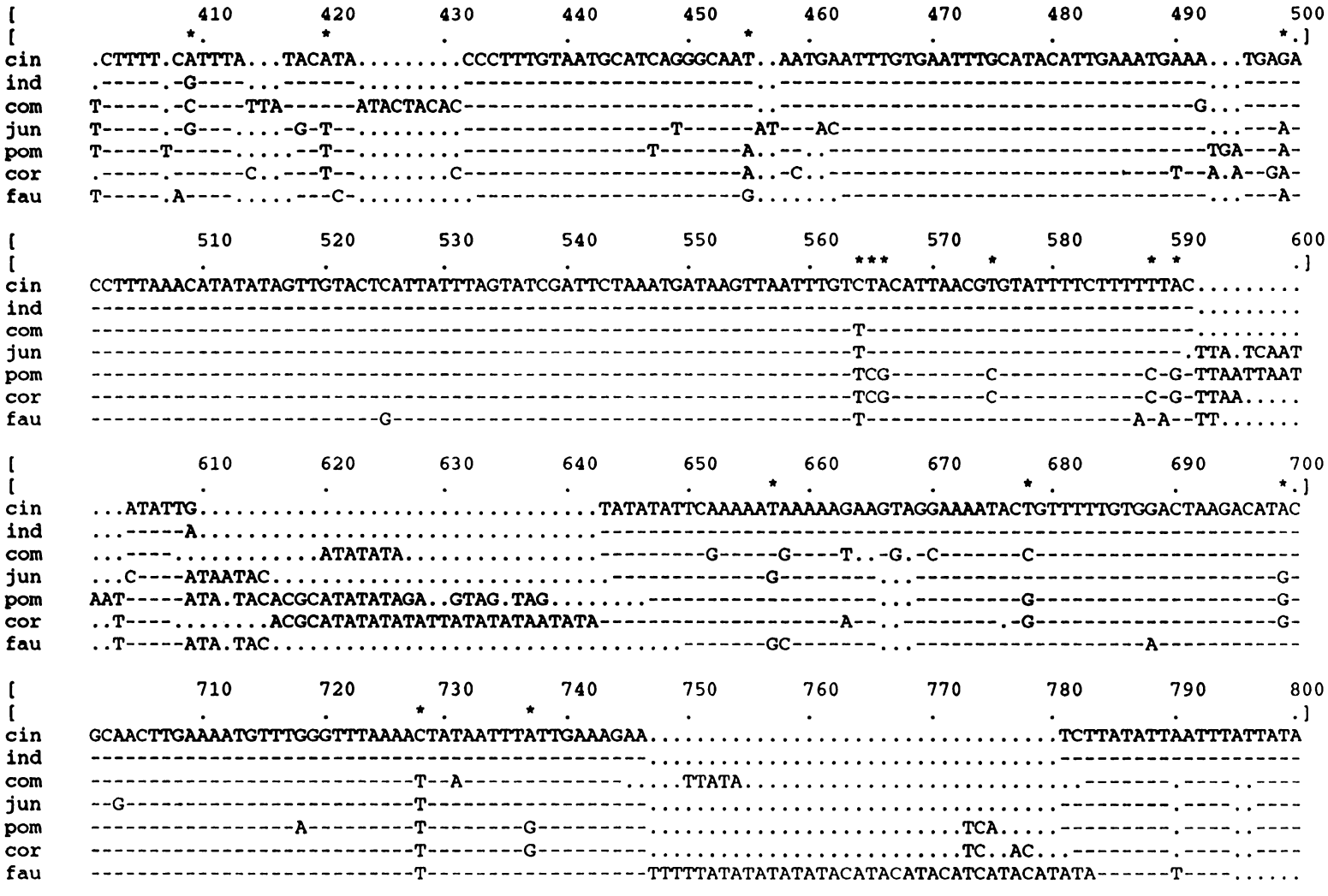


Figure 8A (cont'd)

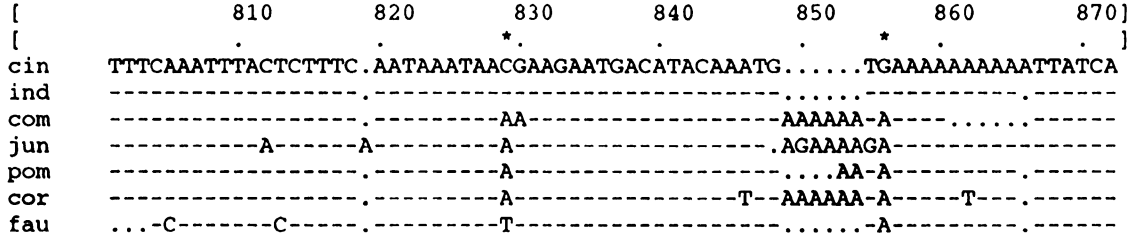


Figure 8A (cont'd)


bas 1 TTGTGTTCCTATCCGTAAAAAAAAATATATAATATAATACAATATTATAT 50
str 1 TTGTATTCTTATCCGTGAAAAAGCA.AGGAAATATAATA.TATA.TATAT 47
51 TACATGTTTGAACAAAAAAATAAAAATTATTCTGGTTCACGTCCCGCCAG 100
48 TATATAT....ACAAAAAAAAAAAAA..... 69
101 AGTAATTATATGCTTGTCTTTAAGATTAAGCCACGCCATATGTAATGCCT 150
70 ..:.....AAAAT.....GGACA....AAAT.... 83
151 CATGACAATACATCTAACGTATTGTGGATTTGCATACATTGGAAAATACC 200
84 ..TGA....AAATCGAA.GTA.....ATCT..ATA.....TAAAACAC. 112
201 AATAAAAAAAGAACAAAAAACAAAAATTTTGGTTTTGTACTTTTCATTCA 250
113 .ATATAAAAAG.....CATATATTTTCTTTTTGTTCTTTTCACTCA 152
251 ATTTGTGTAAAAACATTTGTTTTTGACATAACTTTGAGTGTTTTTCTTTT 300
153 AATCTTGT...TATATTGATTTATAACATAATATTGAGTGTTTTTCCTTT 199
301 TTTCTTT...ACATACATTGAAAAAGACAAAATCAAATGATATGTTTGAA 347
200 TTNGTTTTCGAAATGCATTAAAAAA...CATGCCATGTG.TATGTGTGAC 245
348 AATAA....GCCAA.....aCAAAGGAGAAA.TATC.....GTT 376
246 AATAAATTTGCATATATCTTTAAGAAGAAGAAACTTTCTTCTTGTGTGTT 295
377 C....TT....ACATAAAATACAATATA.TATATGTATT..AGAGA.... 411
296 CCTCTTTG (ND) AAATAGGAAATGCATTAAAAAGATGCA ND
412 .....ACGAAAATTAT.TTCTCACAGTACTAATTGTAATTAGTA..C 450
ND GGGTGCTGGCGAGNGCTATATGCAAAAAGGNATAAAAATAATTTGNATTT ND
451 AA..GAAGTGTGCATGC.AAAAATTATAATTGTAAACAA...AATGAATG 494
ND AATTGAAGGGTGCCGGCTGAGCCCTATA..TGCAAAAAACCTACAGAAGG ND
495 AAACGAATGTATATTTTTGTA...TAGCCACACTATTATGAAATTTTNAA 541
ND ATAACAATTATTGTTTTCCTATCCTAAGAAAACTAAGGCRAAaATTCGAA ND
542 T...ATGTGGAAAATAT.CA.....TATTTTCGCACATTATGCTTGTCT 581
ND TTGCAT.TAAAAAAAATGCAGGGTGCTGGCTGAGCCCTATATGC..... ND
582 CAAAGATTAAGCCACG.AATATGTAATGCTTCATCACAATATATGTAAAT 630
ND CAAA....AGGCAAAGAAATATGCAATAAATtAT....AT.TATTTATAT ND

Figure 8B

Figure 8B (cont'd)

Table 6. A-T Content and Length of the *Rhagoletis* ITS Sequences Analyzed

Species	ITS1		ITS2	
	A-T Content (%)	Length (nt)	A-T Content (%)	Length (nt)
1. *R. cingulata*	79.8	662	82.7	557
2. *R. indifferens*	79.7	660	82.7	558
3. *R. completa*	80.4	634	81.2	541
4. *R. juniperina*	81.0	684	82.7	557
5. *R. fausta*	80.0	625	81.7	527
6. *R. electromorpha*	ND	ND	81.0	501
7. *R. pomonella*	80.5	686	82.0	478
8. *R. cornivora*	80.7	653	81.2	482
9. *R. basiola*	75.3	922	76.6	642

Note.— ND = Not Determined

R. fausta had two major insertions of approximately 30-40 bp long (positions 15-53 and 747-780 in Figure 8A). The first insertion had long stretches of A and T residues while the second insertion had long stretches of T immediately followed by $(AT)_5$ and then by $(ACAT)_5$ repeats. *R. fausta* also had one major deletion of approximately 80 bp long, shared almost completely with *R cornivora* (positions 263-335 in Figure 8A) and partially with *R. pomonella* (positions 300 -333 in Figure 8A). The sequence differences of the ITS1 were generally distributed evenly from 5' to 3'. However, a number of relatively conserved regions were found, such as positions 500-563. The A-T content of the ITS1 regions were high (about 80%; Table 6). The high AT concentration was partially contributed by long stretches of A and T such as regions around positions 140 $(A)_{7-10}$, 224 $(T)_9$,388 $(T)_{12}$ and 857 $(A)_{10}$. Simple direct repeats involving A and T residues were also common in the ITS1. For instance, the AAT motif was repeated 5-6 times around position 340 in *R. pomonella* and *R. cornivora*, and AT motif occurred approximately 14 times within positions 620-650 of *R. cornivora*. A number of other, less frequent motifs were also detected (e.g., TGTA and CATA).

The average number of nucleotide substitutions per 100 nucleotide sites for pairwise comparisons and transition to transversion ratios were calculated based on the alignment in Figure 8 using three different statistical approaches (i.e., the Jukes and Cantor 1969, Kimura 1980 and Tamura 1992 approaches). Because the results from these three statistical approaches were not significantly different, only the values from the Kimura (1980) approach are shown in Tables 7 and 8. In comparison with the intraspecific nucleotide substitution rates obtained in Chapter II (0.00 to 0.35 average percent nucleotide substitutions), the interspecific nucleotide substitutions are substantially higher (2.99 to 7.88) except for the two sibling species from the *cingulata* group (i.e., *R. cingulata* and *R. indifferens*). Because generally only one clone was sequenced for each surveyed species, the level of intraspecific variation in ITS1 could not be determined. However, the intraspecific variation of *R. cingulata*, calculated from 10 pairs, was 0.18 ± 0.07 average

percent nucleotide substitutions (see Chapter II) and the values for several other *Rhagoletis* species, including 3 *R. pomonella* individuals, were similarly low (see Chapter II). Therefore, it is reasonable to assume that the intraspecific variations for those unknown *Rhagoletis* species are also low and negligible when comparing interspecific substitution rates for phylogenetic analysis. The transition to transversion ratios for the pairwise sequence comparisons are presented in Table 8. These ratios range from 0.94 to 3.04, with transitions alone contributing 100% of the substitutions observed between *R. cingulata* and *R. indifferens*. In principle, a transition to transversion ratio of 0.5 is the expected rate, if there were no bias at all, because there are twice as many ways that a transversion (A to T, A to C, G to T and G to C) can occur than a transition (A to G and T to C). Some of the transition to transversion ratios found in this study are relatively higher than those in *Drosophila* (Schloterer et al. 1994).

ITS 2 Sequence Analysis

The complete ITS2 sequences of 9 *Rhagoletis* species are presented in Figure 9. Two additional taxas (*R. electromorpha* and *R. basiola*) have been included in comparison with the ITS1 sequence analysis in the earlier section (Figure 8). The length of the ITS2 varies from 478 to 642 bp (Table 6). Among the aligned sequences *R. basiola* was 84-164 bp longer than the other surveyed species, in part due to the long insertion/deletion found between positions 487 to 572. Generally, the ITS2 was about 100-300bp shorter than the respective ITS1 sequences. Like in ITS1, the ITS2 region had a high A-T content (81.3 ± 1.9% A-T content; Table 6). Long stretches (3 to 11 bp) of A or T residues were predominant in the ITS2 sequences; for example, the $(A)_4$ sequence occurred at least six times in *R. cingulata* ITS2. Long stretches of TA simple repeats were also common, such as the $(TA)_9$ around position 380 in *R. basiola*. The 5' half of the ITS2 was relatively conserved among the surveyed *Rhagoletis* species. Actually one region (249-295) was found to be highly conserved even between *Rhagoletis* and *Drosophila*, suggesting this

Table 7. Average Percent Nucleotide Substitutions (Nucleotide Substitutions per 100 Nucleotide Sites) for the *Rhagoletis* ITS1 and ITS2 Sequence Pairwise Comparisons

Average Percent Nucleotide Substitution, Calculated by the Kimura Method (1980), for ITS1 (above Diagonal) and partial ITS2 (below the Diagonal) for Pairwise Analysis (gaps excluded only in pairwise comparisons)

	1	2	3	4	5	6	7	8	9
1. cin	-	0.30	5.22	5.12	7.35	ND	6.76	6.35	ND
2. ind	0.31	-	5.22	4.79	7.37	ND	6.76	6.55	ND
3. com	2.24	1.91	-	6.14	7.11	ND	7.09	6.79	ND
4. jun	1.59	1.27	1.27	-	6.67	ND	5.51	5.91	ND
5. fau	5.21	4.88	4.89	3.88	-	ND	7.88	7.79	ND
6. ele	3.22	2.89	2.90	2.23	3.89	-	ND	ND	ND
7. pom	3.55	3.21	3.20	2.24	5.89	3.88	-	2.99	ND
8. cor	3.23	2.90	2.89	2.24	5.57	3.88	1.59	-	ND
9. bas	8.03	7.64	7.62	6.81	9.20	8.00	8.34	9.13	-

Note.— cin = *R. cingulata*; ind = *R. indifferens*; com = *R. completa*; jun = *R. juniperina*; fau = *R. fausta*; ele = *R. electromorpha*; pom = *R. pomonella*; cor = *R. cornivora*; and bas = *R. basiola*. ND = not determined.

Table 8. Transition to Transversion Ratios for the *Rhagoletis* ITS1 and partial ITS2 Sequences

Transition/Transversion Ratios, Calculated by the Kimura Method (1980), for ITS1 (above Diagonal) and partial ITS2 (below the Diagonal) Where All Gaps Are Excluded

	1	2	3	4	5	6	7	8	9
1. cin	-	NS	2.76	1.48	1.20	ND	2.75	1.98	ND
2. ind	NS	-	2.76	1.36	1.29	ND	2.88	2.09	ND
3. com	1.35	1.01	-	0.94	0.67	ND	1.74	1.25	ND
4. jun	4.06	3.03	0.00	-	0.47	ND	1.31	0.94	ND
5. fau	1.02	0.87	0.29	0.37	-	ND	1.25	0.96	ND
6. ele	1.53	1.27	0.28	0.40	0.57	-	ND	ND	ND
7. pom	3.59	3.06	0.60	1.01	0.67	0.84	-	3.04	ND
8. cor	4.11	3.59	0.81	1.35	0.79	1.02	NS	-	ND
9. bas	1.51	1.38	0.76	0.92	0.80	0.70	1.13	1.23	-

Note.— The numbers corresponding to the different taxa are as in Tables 5-7. NS = not shown; corresponds to cases where transitions = 100% and transversions = 0%. ND = not determined.

region may be functionally important (for details see Chapter II). In contrast, the 3' half of the ITS2 region showed considerable sequence divergence making this region rather difficult to align unambiguously, especially the region beyond about position 340.

The average percent nucleotide substitutions and transition to transversion ratios were calculated according to the alignment in Figure 9 and presented in Table 7 and Table 8, respectively. Since the region beyond the position 340 was difficult to align unambiguously this region was excluded in the above calculations. Except for the *cingulata-indifferens* pair, the interspecific percent nucleotide substitution rates (1.27-9.20) were significantly higher than the intraspecific variations observed for *R. cingulata*, *R. indifferens* and *R. pomonella* (0-0.24 percent nucleotide substitutions; see chapter II). The transition to transversion ratios varied from 0.00 to 4.11, with transition alone contributing 100% of the substitutions in the *cingulata-indifferens* and *pomonella-cornivora* pairs of taxa. The ratios in the ITS2 were lower, in most cases, than those measured for ITS1; the ratios for most pairs among *completa*, *juniperina*, *fausta* and *electromorpha* were lower than 0.5.

Phylogenetic Analyses

A total of 32 informative characters were obtained from the ITS1 (asterisks in Figure 8A) and 19 from the ITS2 (asterisks and crosses in Figure 9) aligned sequence data. Only one most parsimonious tree for the ITS1 sequence data was found with *R. fausta* taken as the outgroup (Figure 10A). The tree is of length 40 with consistency index (CI) 0.925 and retention index (RI) 0.919. Bootstrap analysis using PAUP gave 100% support for the branches of *R. cingulata* and *R. indifferens*, and *R. pomonella* and *R. cornivora*; 88% support for the clade of *R. cingulata-R.indifferens* and *R. completa*; and 61% support for the clade of *R. pomonella-R. cornivora* and *R. juniperina* (Figure 10B). The same tree was obtained using the NJ distance approach in MEGA. The bootstrap showed, again, 100% support for the *R. cingulata* and *R. indifferens* branch, 98% support for the

Figure 9. *Rhagoletis* ITS2 sequences and alignment. A) The complete ITS2 sequences of the species surveyed except for *R. striatella*. Phylogenetically informative characters are denoted by asterisks and crosses. Notations same as in legend of Figure 8. B) The *R. basiola* sequence compared with the sequence of *R. striatella*. The alignment was performed as in Figure 8B. Notations are as in Figure 8.

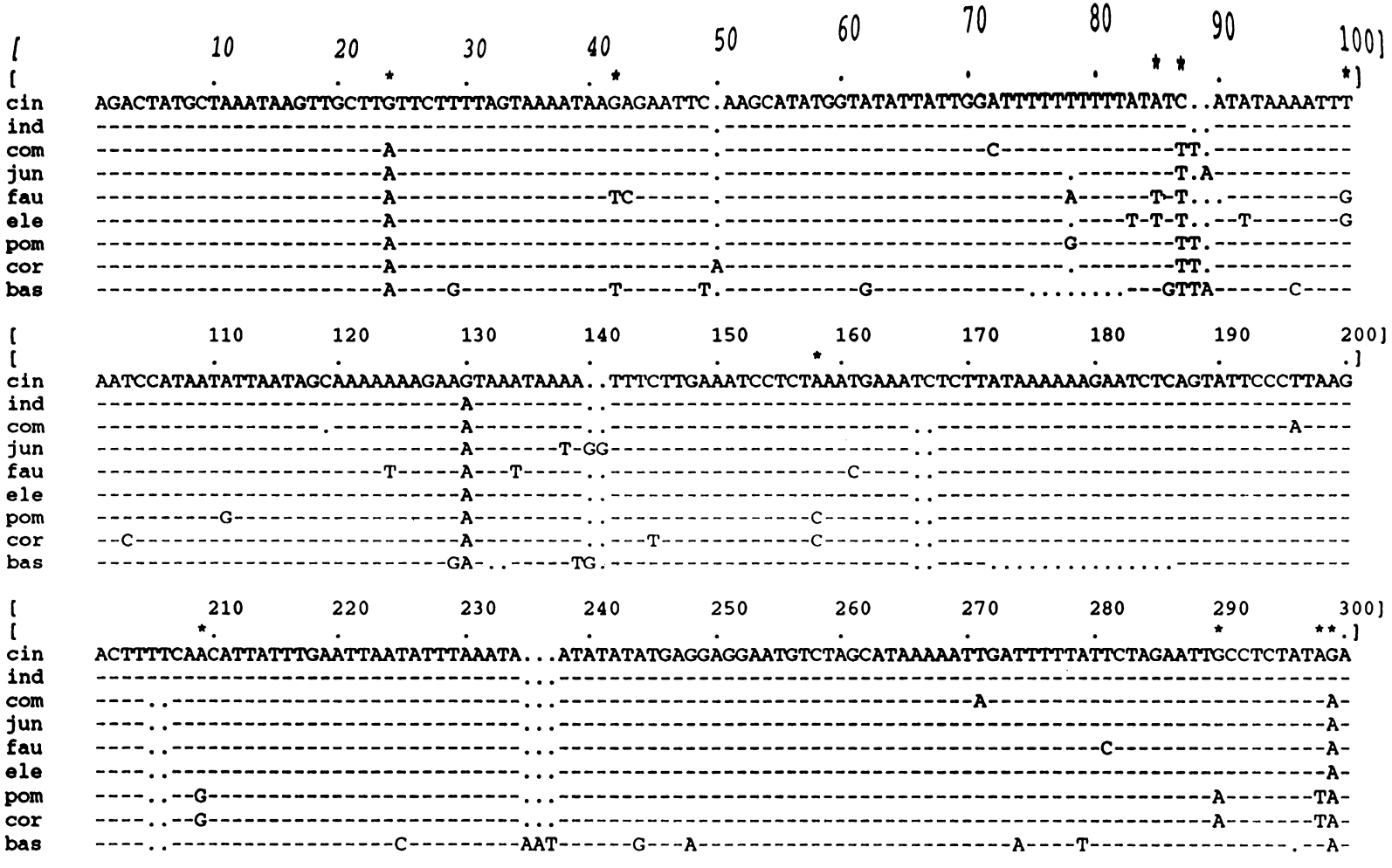

10 20 30 40 50 60 70 80 90 100
110 120 130 140 150 160 170 180 190 200
210 220 230 240 250 260 270 280 290 300
cin AGACTATGCTAAATAAGTTGCTTGTTCTTTTAGTAAAATAAGAGAATTC.AAGCATATGGTATATTATTGGATTTTTTTTTTATATC..ATATAAAATTT
ind
com
jun
fau
ele
pom
cor
bas
cin AATCCATAATATTAATAGCAAAAAAAGAAGTAAATAAAA..TTTCTTGAAATCCTCTAAATGAAATCTCTTATAAAAAAGAATCTCAGTATTCCCTTAAG
ind
com
jun
fau
ele
pom
cor
bas
cin ACTTTTCAACATTATTTGAATTAATATTTAAATA...ATATATATGAGGAGGAATGTCTAGCATAAAAATTGATTTTTATTCTAGAATTGCCTCTATAGA
ind
com
jun
fau
ele
pom
cor
bas

Figure 9A

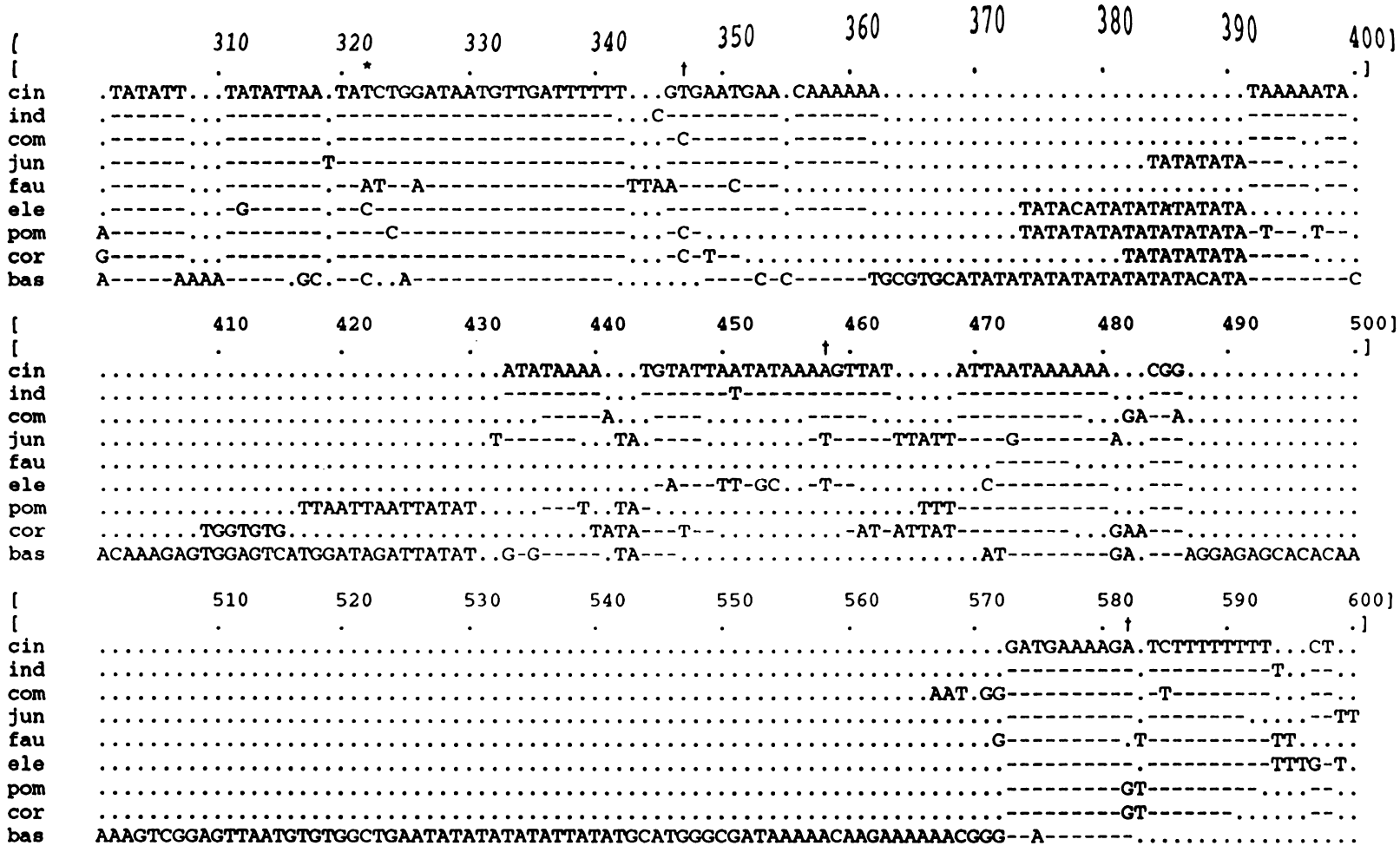


Figure 9A (cont'd)

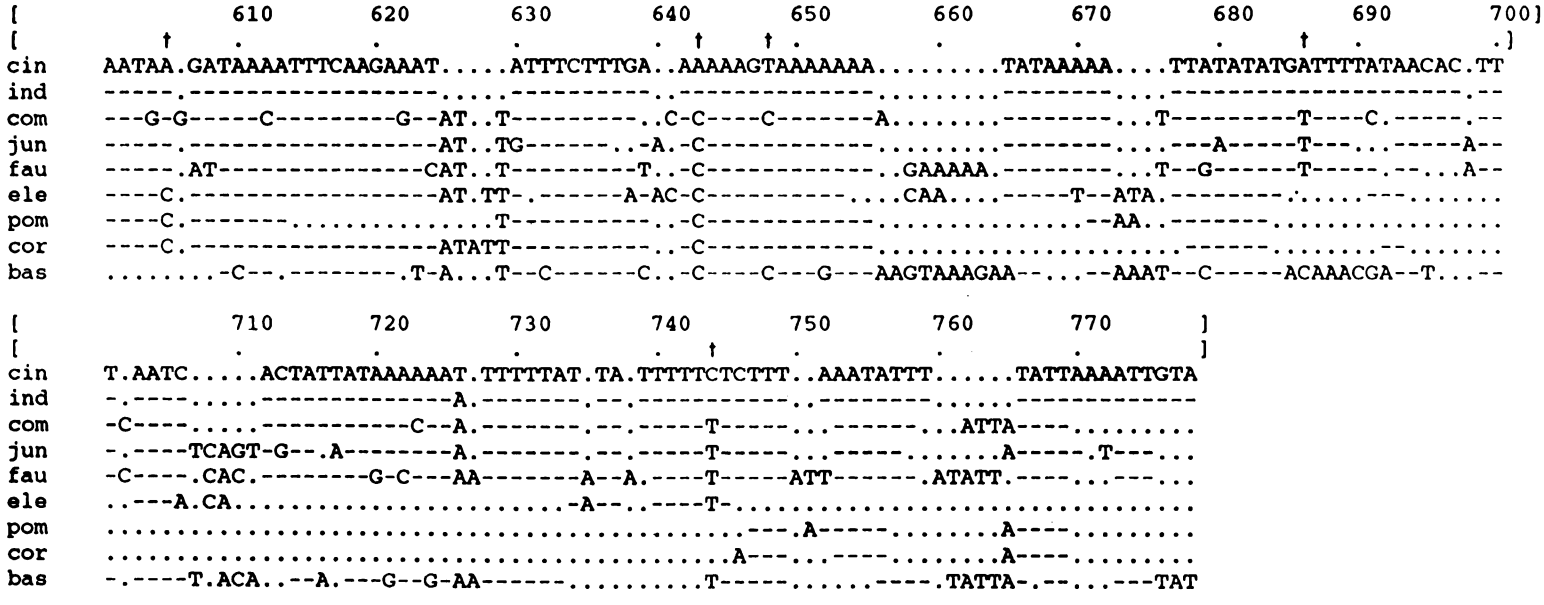


Figure 9A (cont'd)


bas 1 AGACTATGCTAAATAAGTTGCTTATTCTGTTAGTAAAATAATAGAATTTA 50
str 1 AGACTATGCTAAATAAGTTGCTTATTC.....AT.TTGGAATTCA 39
51 AGCATATGGTGTATTATTGGATTATAGTTAATATAACATTTAATCCATAA 100
40 AGCACATGGTATGTTATTGG...ATCGTT....TAA.....AATCCATAA 77
101 TATTAATAGCAAAA..AAAGAGATA....TAAATGTTTCTTGAAA..TCC 142
78 TATTAATAGCAAAAGGAAAGATATATACGTAAATATTTCTTGAAATGTCC 127
143 TCTAAATGA.....AATCTTAGTATTCCCTTAAGACTTCAACATTAT 184
128 TCAAAAAAAAAAAAAAAGAATCATAGTATTCCCGAAAAACGTCAACGTTAT 177
185 TTGAAT.....TAACATTTAAATAAATATATATGTGAAGAGGAATGTCTA 229
178 TTGCATTTATATACAATATATATATGTATATATGAAAAGAGGAATGTCTA 227
230 GCATAAAAATTGAA.TTTTTTTCTAGAATTGCCTC...TTAAAATATATA 275
228 GCATAAAACTTTAACTTTTTTTCTAGAATTGCCTCCTTTTATTATATATT 277
276 AAATATATGCTACAGATAATGTTGATTTTTAATGCACCAAAAATGCGTGC 325
278 ACACAAA....GAAAAAAATATATATATAAAAAGCAAGGAATAT.TGATA 322
326 ATATATATATATATATATACATATAAAAA....TACACAAAGAGTGGAGT 371
323 AGATATGTATATATATATAAATATACATATGTTTAAACAGTGAAAGGATT 372
372 CATGGATAGATTATATGTGTAAAATATGTATATAAAAAAGACGGAGGAGA 421
373 TCGGGAT.GAAAAGATTTTT...TTAGTTTCATGAAATA....TATTATA 414
422 GCACACAAAAAGTCGGAGTTAATGTGTGGCTGAATATATATATATTATAT 471
415 TCTTTC..AAAGT.GGTG....TGTGTGTGTGCGTGTGTGTGT.GTGTGT 456
472 GCATGGGCGATAAAAACAAGAAAAAACGGGGAAGAAAAGATCAATTTCAA 521
457 GGGTGGGTGGT.....TATTAA 473
522 GATTATATCTCTTTGCACAAAGCAAAGAAAAAGTAAAGAAATAAAAATTT 571
474 CCTAACCCCTCCTAAC.CACCACAACCTAAA..TAAA.ACACATATATTT 519
572 CTATATACAAACGAAATTTTAATCTACATAAATAGAAGAAATTTTTATTC 621
622 TTTATTTTATTATTTATTTAT 642

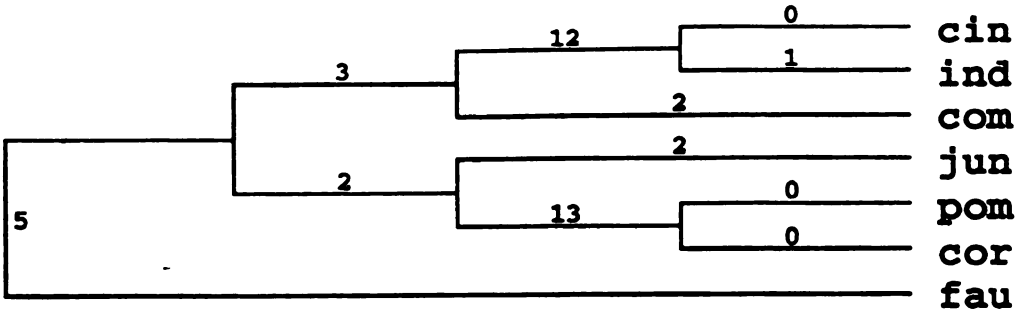
Figure 9B

R. pomonella and *R. cornivora* branch, 57% support for the clade of *R. cingulata-R.indifferens* and *R. completa* and 54% support for the clade of *R. pomonella-R. cornivora* and *R. juniperina* (Figure 10C).

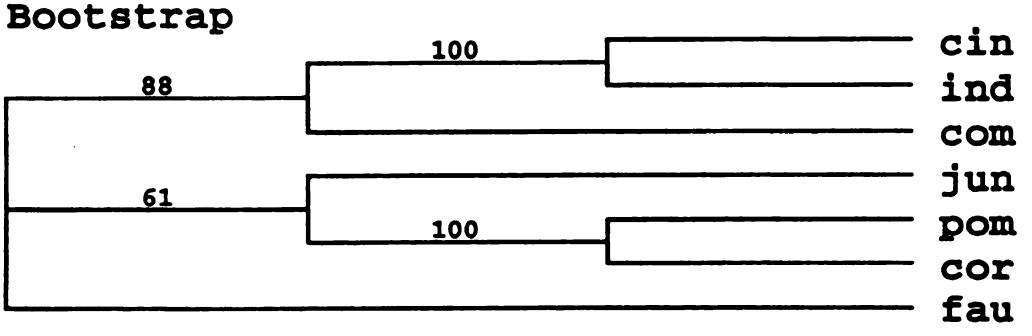
For the ITS2, since the 3' half was so highly divergent and difficult to align with confidence, I chose only the informative characters between positions 1 to 340 (11 out of the total 19 PAUP informative characters; asterisks in Figure 9) in the search of a most parsimonious tree rooted with *R. basiola* as the outgroup. An exhaustive search gave only one most parsimonious tree with length 13, CI = 0.923 and RI = 0.909 (Figure 11A). Of the 8 excluded character sites (crosses in Figure 9), three of them grouped *R. cingulata* and *R. indifferens* together, one grouped *R. pomonella* with *R. cornivora*, and one grouped *R. cingulata, R. indifferens* and *R. completa* together. However, almost all of them had significant number of gaps and their reliability was dependent on the sequence alignment. Bootstrap analysis using PAUP of the ITS2 sequence data (up to position 336) showed 97% support for the *R. cingulata* and *R. indifferens* branch, 100% support for the *R. pomonella* and *R. cornivora* branch, and 73% support for the *R. fausta* and *R. electromorpha* branch (Figure 11B); placements of *R. juniperina* and *R. completa* were unresolved. A similar bootstrap analysis in PAUP but using instead the whole ITS2 sequence showed 100% support for the *R. cingulata* and *R. indifferens* branch, 100% support for the *R. pomonella* and *R. cornivora* branch, and 61% support for the *R. fausta* and *R. electromorpha* branch (Figure 11C). Again, *R. completa* and *R. juniperina* remained unresolved. A similar tree, but with higher resolution, was obtained from a NJ approach using ITS2 sequences up to position 340 (Figure 11D). The clade of *R. fausta* and *R. electromorpha* became a sister group to the other taxa which had *R. juniperina* forming a tritomy with two other clades. The first clade included an earlier branched *R. completa* and the derived *cingulata-indifferens* pair, while the second had *R. pomonella* and *R. carnivora.* A bootstrap test of the NJ tree showed 95% support for the *R. cingulata* and *R. indifferens* branch, 96% support for the *R. pomonella* and *R. cornivora* branch, and

Figure 10. Phylogenetic analyses of *Rhagoletis* using the ITS1 sequence alignment. a) This single most parsimonious tree (consistency index (CI) = 0.925 and retention index (RI) = 0.919) was obtained for the ITS1 sequence alignment (Figure 8A) using the exhaustive search option in PAUP. All uninformative characters were ignored and gaps in the alignment were treated as missing data. The number of character-state changes along each branch are indicated. b) 50% majority-rule bootstrap consensus tree (using the branch and bound option in PAUP) after 100 replications with random seed number set at 1. All uninformative characters and gaps were treated as in part a. The bootstrap confidence limits (BCLs) are indicated. c) Bootstrap distance tree analysis on the ITS1 sequence alignment (Figure 8A) using the Kimura (1980) matrix of distances and the neighbor-joining (NJ; Saitou and Nei 1987) tree generating options in MEGA (see Material and Methods) with 100 replications and random seed number set at 1. BCls are indicated. The scale corresponds to 0.06 average percent nucleotide substitutions.

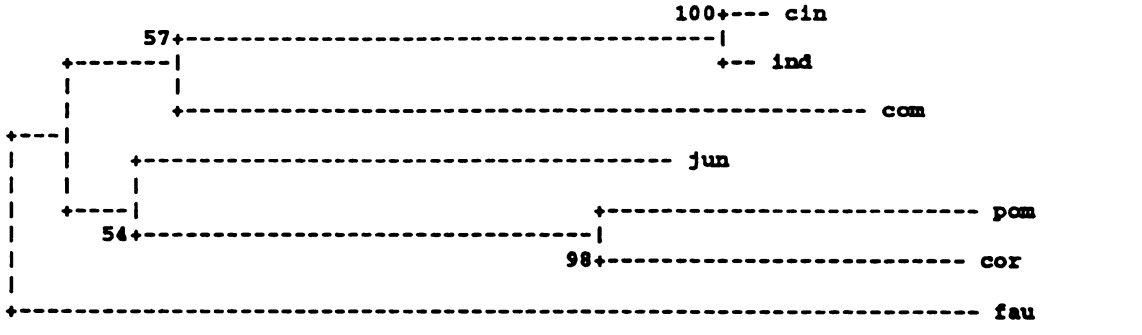
A



B



C



Scale: each - is about equal to 0.06 percent nucleotide substitutions

Figure 10

Figure 11. Phylogenetic analyses of *Rhagoletis* using the ITS2 sequence alignment. a) This single most parsimonious tree (CI = 0.923 and RI = 0.909) was obtained for the ITS2 sequence alignment (Figure 9A) using the exhaustive search option in PAUP. All characters beyond position 340 (Figure 9A) were excluded in this analysis. Other parameters were as indicated in the legend for Figure 10A. The number of character-state changes along each branch are indicated. b) 50% majority-rule bootstrap consensus tree (from PAUP) with the parameters set as in the legend for Figure 10B. All characters beyond position 340 (Figure 9A) were excluded in this analysis. BCLs are indicated. c) Same as in part b except the complete ITS2 sequences were used in this analysis. d) Bootstrap distance tree analysis on the ITS2 sequence alignment (Figure 9A) as indicated in Figure 10C. The scale corresponds to 0.07 average percent nucleotide substitutions.

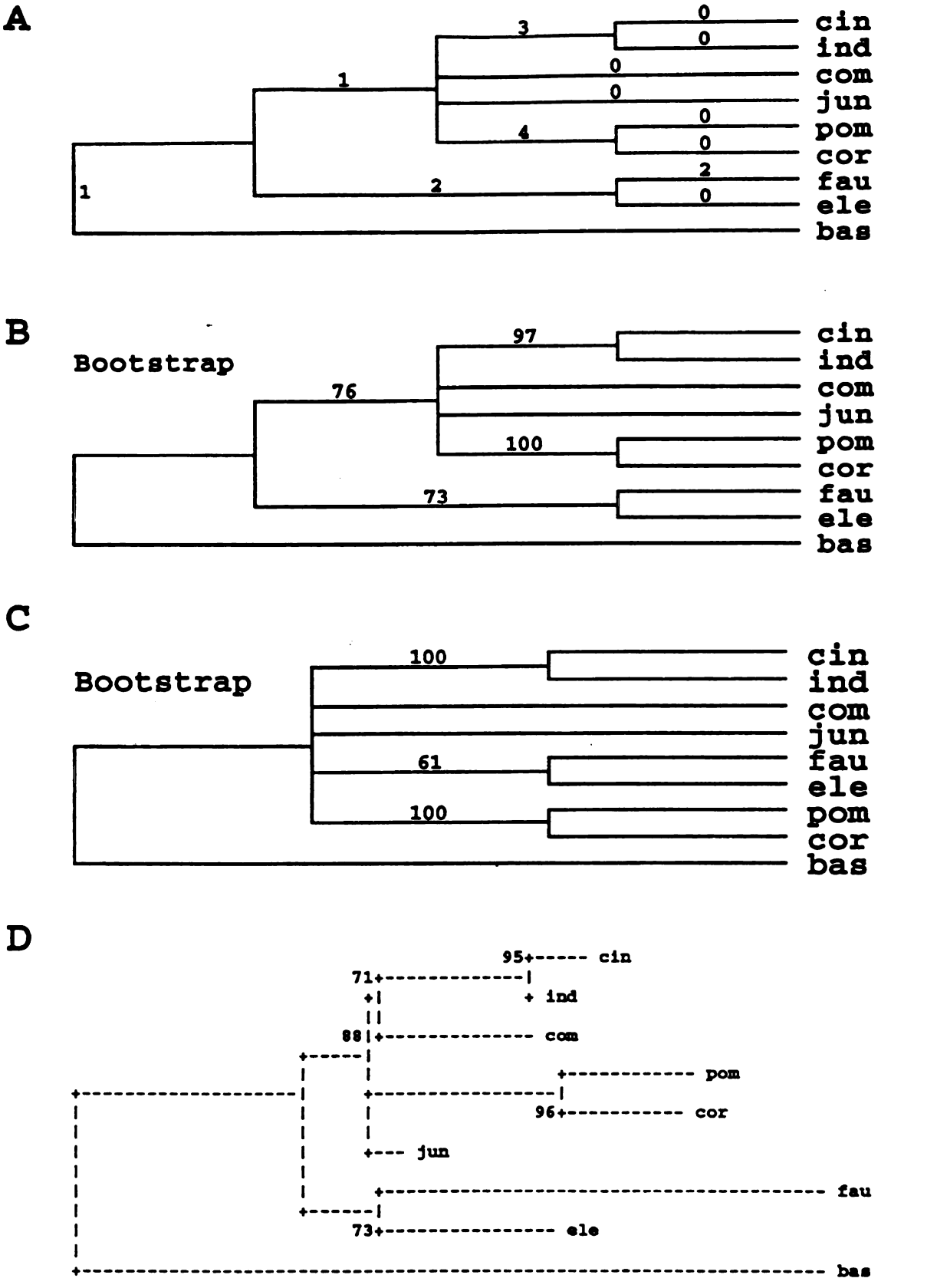
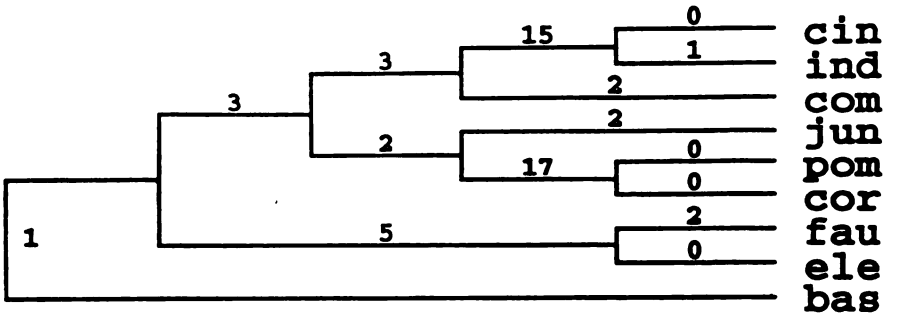

A
3
0
0
cin
ind
1
0
com
0
jun
0
4
pom
0
cor
2
1
2
fau
0
ele
bas
B
Bootstrap
97
cin
ind
76
com
jun
100
pom
cor
73
fau
ele
bas
C
Bootstrap
100
cin
ind
com
jun
61
fau
ele
100
pom
cor
bas
D
95
cin
71
ind
88
com
pom
96
cor
jun
fau
73
ele
bas
Scale: each - is approximately equal to 0.07 percent nucleotide substitutions

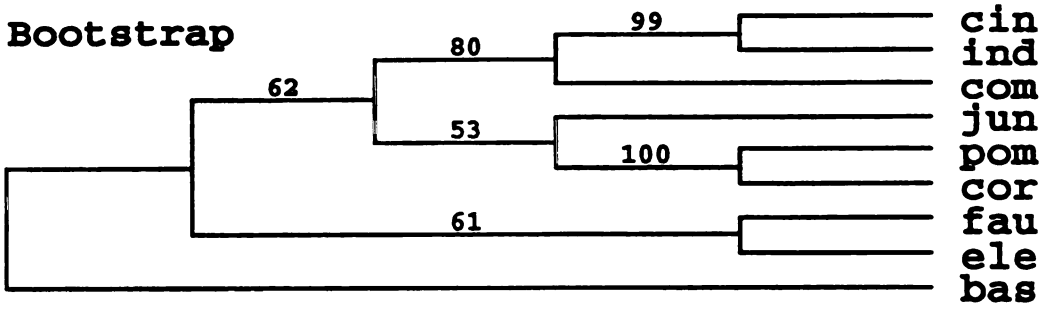
Figure 11

Figure 12. Phylogenetic analyses of *Rhagoletis* using the combined ITS1 and ITS2 sequences. a) This single most parsimonious tree (CI = 0.925 and RI = 0.917) was obtained after combining the phylogenetically informative characters from the PAUP analyses in Figures 10A and 11A (i.e. asterisks in Figures 8A and 9A) and then performing an exhaustive search. Gaps were treated as missing data. The number of character-state changes along each branch are indicated. b) 50% majority-rule bootstrap consensus tree of the combined ITS1 and ITS2 data set (as in part a) with the parameters set as in the legend for Figure 10B. BCLs are indicated. c) Bootstrap distance tree analysis on the combined ITS1 and ITS2 sequences as indicated in Figure 10C. However, only positions 1-340 of the ITS2 alignment (Figure 9A) were included in this analysis. The scale corresponds to 0.05 average percent nucleotide substitutions.

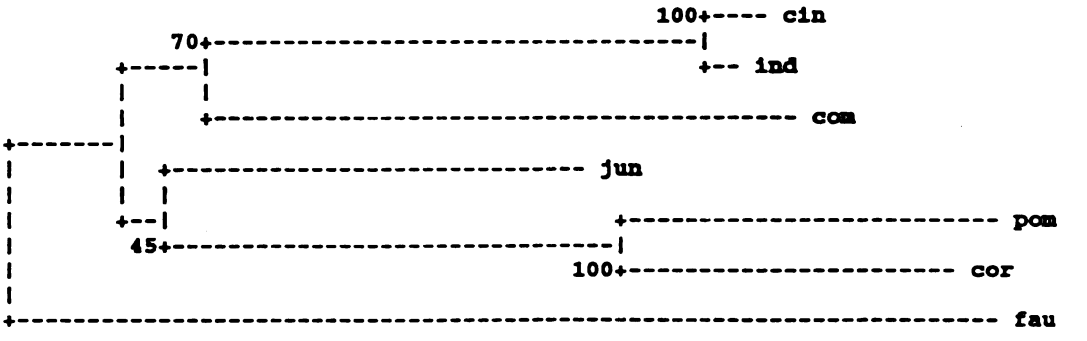
A



B



C



Scale: each - is about equal to 0.05 percent nucleotide substitutions

Figure 12

73% support for the *R. fausta* and *R. electromorpha* branch (Figure 11D). *R. completa* was grouped with the *cingulata-indifferens* pair 71% of the time, and the placement of *R. juniperina*, again, remained unresolved.

By combining informative characters from ITS1 (32) and ITS2 (11, only those found between positions 1 and 340 of ITS2; Figure 9A), I found one most parsimonious tree of length 53, CI = 0.925, HI = 0.075 and RI = 0.917 (Figure 12A). In this PAUP analysis of nine taxa, the absent *electromorpha* and *basiola* ITS1 sequences were treated as missing data (i.e., by dots). To check the effect of this treatment on the tree topology I repeated the PAUP analysis of the combined ITS sequence data, but with the ITS2 of *electromorpha* and *basiola* excluded (i.e., 7 taxa analyzed). One most parsimonious tree with the same topology was obtained, but with a shorter tree length (data not shown). Bootstrap analysis of the combined sequence data showed 99% support for the *R. cingulata* and *R. indifferens* branch, 100% support for the *R. pomonella* and *R. cornivora* branch, 61% support for the *R. fausta* and *R. electromorpha* branch (Figure 12B). *R. completta* was grouped with the *cingulata-indifferens* pair 80% of the time and *R. juniperina* was grouped with the *pomonella-cornivora* pair 53% of the time. Similarly, a bootstrap test of NJ distance tree from the combined ITS1 and ITS2 (again, only positions 1-340) sequences showed 100% support for the *R. cingulata* and *R. indifferens* branch, and 100% support for the *R. pomonella* and R. cornivora branch (Figure 12C; *electromorpha* and *basiola* sequences were excluded from the analysis because of their ITS1 absent in the alignment). *R. completta* was grouped with the *cingulata-indifferens* pair 70% of the time and *R. juniperina* was grouped with the *pomonella-cornivora* pair 45% of the time.

Discussion

ITS Sequence Divergence

Generally, the values for the average substitution rates are much less than the saturation level (around 40%) found between the *Drosophila melanogaster* species group

and *D. pseudoobscura* (or *D. virilis*; Schlotterer et al. 1994). This observation is consistent with the generally high transition to transversion ratios observed for the pairwise sequence comparisons (Table 8) when compared with those from the *Drosophila* study (Schlotterer et al. 1994). This may imply that the evolutionary history of the *Rhagoletis* species surveyed in this study is not as long as those of the *Drosophila* species. However, the observed differences in substitution rates and transition/transversion ratios may also reflect the fact that the rates of nucleotide changes are faster in *Drosophila* because they have many more generations per year than *Rhagoletis*. In general, gene sequences of closely related taxa will differ predominantly in transitions because, these occur at a far greater rate than do transversions (Quicke 1993). A decrease in the transition to transversion ratio is expected when the period of divergence increases (Brown et al. 1982). In the large subunit mitochondrial rDNA of Hawaiian *Drosophila*, transitions account for 90% of the number of total substitutions between taxa that had been separated for ≤1 Myr but account for only 40% of the differences between taxa separated for ≥10 Myr (DeSalle et. al 1987). For *cingulata-indifferens* and *pomonella-cornivora* species pairs, 100% of the total substitutions were transitions suggesting that the species diverged relatively recently (Table 8). However, nucleotide substitution rates and transition to transversion ratios are not always correlated well. For example, in the *completa/juniperina* pair of taxa, the transition to transversion ratios in the ITS2 were very low, while the corresponding substitution rates are not very high.

The significance of the high A-T content in the ITS sequences of *Rhagoletis* is not clear but may have some relation to a recent observation that *R. pomonella* rDNA clusters are located at the periphery of fibrillar centers in the nucleolus (Procunier and Smith 1993; also see Chapter II). The high A-T content found for the ITS sequences of *Rhagoletis* is not generally typical for eukaryotic organisms although high AT contents have been also observed in *Drosophila* (Tautz et al. 1988; Schlotterer at al. 1994) and in *Cicindela* beetles (Volger and DeSalle 1994). In several *Aedes* mosquito species, only ≤50% A-T content

were reported for the ITS2 (Wesson et al. 1992). The ITS sequences in vertebrates and plants have an A-T content as low as 20-30% and 30-50%, respectively (Torres et al. 1990; Yokota et al. 1989). The decreased AT content in higher animals has been proposed to be associated with the shift from cold-blooded to warm-blooded vertebrates (Bernardi et al. 1988). The low AT content in hematophagous mosquitoes may be linked to their feeding on warm-blooded animals (Wesson et al. 1992).

The Feasibility of Using the ITS for *Rhagoletis* Phylogenetic Analysis

The rDNA ITS sequences have become attractive markers for phylogenetic studies. Variation in these regions has been analyzed using restriction fragment length polymorphism (RFLP) (Hillis and Dixon 1991) and more recently, also using PCR and sequencing technologies. The rDNA array in animal species typically consists of several hundred tandemly repeated copies. The sequence diversity among those copies within a single individual (i.e., intra-individual variation) and within a population (i.e., inter-individual variation) can exist. With current techniques, it is impossible to determine the sequence for each of the hundreds repeating units and to obtain the total amount of genetic diversity within a population or even within a single individual. As a consequence the existence of undetected polymorphism among repeat units within a single individual or within a population can complicate the use of the rDNA in phylogenetic analysis. However, those tandem repeats are generally homogenized to produce a uniform sequence in all repeating units of a given species through several mechanisms, such as unequal crossing over and gene conversion, (i.e. concerted evolution; Dover 1982). Empirical evidence of the homogenization of tandem repeats through gene conversion has been provided by studying parthenogenetic lizards (Hillis et al. 1991). Probably because of this general assumption of concerted evolution, some studies on the rDNA do not pay much attention to intraindividual and intraspecific variation although Williams et al. (1988) warned that homogenization of the tandem repeats within a species may take a relatively

long period of evolutionary time. However, substantial variation within certain *Drosophila* species (Williams et al . 1987) have been noticed and a great amount of polymorphism in the ITS2 sequence has been reported in *Aedes* mosquitoes with highest intraindividual variation 1.52% and intraspecific variation 2.55% (Wesson et al. 1992). High intra-individual variation has also been described in a *Cicindela* beetle ITS1 sequence (Volger and DeSalle 1994). Those finding further support that homogenization of the rDNA repeats indeed requires long periods of evolutionary time and thus it becomes necessary to access the intraspecific variation in the ITS before it is used for phylogenetic analysis. Since events of unequal crossing over, gene conversion and interchromosomal reciprocal exchange are possible mechanisms responsible for homogenization of tandem repeats of rDNA the location of rDNA arrays on chromosomes.(i.e. multiple arrays on different chromosomal loci or different chromosomes) in a given species may affect the sequence uniforming process. In most *Anopheles* mosquitoes the rDNA is localized on the X chromosome while the rDNA arrays of *Aedes* can occur at least at two loci on different chromosomes (Park and Fallon 1990; Fritz et al. 1994; Kumar and Rai 1990; Wesson et al. 1992). This may partially account for the disparity in the degree of intraindividual and intraspecific ITS variation between *Aedes* and *Anopheles*, several species in the later genus shows little or no such variation in their ITS sequences (Porter and Collins 1991; Fritz et al. 1994). *D. melanogaster* is known to have rDNA loci on both the X and Y chromosomes and different rDNA sequences on the two chromosomes have been detected (Yagura et al. 1979), with the X-linked rDNA arrays probably under selective constraints (Williams et al. 1987). Using *in situ* hybridization, Procunier and Smith (1993) determined that the rDNA in *Rhagoletis pomonella* is located on two homologues of chromosome number one, which may be sex-linked with the X chromosome containing a larger block of rDNA. Therefore, high levels of intra-individual/intraspecific variation in *Rhagoletis* ITS sequences should not be surprising. The investigation of intraspecific polymorphism carried out in the previous chapter became necessary before the ITS

sequence is used in the phylogenetic analysis of the taxa covered in this chapter. Compared with the interspecific variation among the concerned taxa in this study (most of them representing different species groups) the intraspecific variation found in the previous chapter was significantly low, indicating that intraspecific variation can be regarded as negligible in phylogenetic analysis of taxa at the species group level, but not among sibling species of some species groups, such as the *cingulata* group.

Phylogenetic Implications from the ITS sequences

The lack of sufficient morphological characters between species groups of *Rhagoletis* has made it difficult to establish a reliable phylogenetic framework for this genus. Allozyme analyses of certain species in the genus *Rhagoletis* have provided insight into the phylogeny of this genus. However, a good consensus between the results from morphology and allozyme data is still far from being reached and some conflict appears to exist between the biochemical and molecular data (i.e. allozyme vs. mtDNA) with regard to the placement of certain species in the phylogeny of this genus. The results from my ITS study are in agreement with one or another of the earlier systematic studies on the genus *Rhagoletis* (Bush 1966; Berlocher and Bush 1982; Smith and Bush, unpublished data). There are also conflicts, with the previous studies, with regard to the placement of certain species. First, the ITS data support the view that *R. cornivora* belongs to the *pomonella* group. This placement is consistent with all the analyses performed in this study, which includes the PAUP and NJ analyses of both ITS1 and ITS2 as separate as well as combined regions. The BCL values for all the analyses were consistently high, usually about 100%. The closeness of *R. cornivora* and *R. pomonella* is in agreement with their affinity in their morphology. *Rhagoletis cornivora*, morphologically, resembles *R. pomonella* so much that it had been treated as a sympatric subspecies of *R. pomonella* until Bush (1966) recognized it as a distinct species. The unique karyotype and several features in wing pattern unambiguously place *R. cornivora* in the *pomonella* group (Bush 1966).

Although *R. pomonella/R. cornivora* hybridization are unlikely to occur in nature, a few hybrids between these two species were obtained in the laboratory and confirmed as hybrids with electrophorectic analysis (Smith et al., 1992). However, this placement is contradictory to the results of mtDNA analysis, which places *R. cornivora* far away from the *pomonella* group, and groups *R. cornivora* with *R. juniperina* or with the *tabellaria* group (Smith and Bush, unpublished data). Electrophoretic study (Berlocher and Bush 1982) indicates that the genetic distance between *R. cornivora* and the other members of the *pomonella* species group is about 9 times greater than the distance among the other members of this group themselves. Therefore, *R. cornivora* appears to have diverged much earlier than its sibling species during the evolution of this group. Although UPGMA clustering analysis of Nei standard distance (Berlocher and Bush 1982) places *R. cornivora* outside the *pomonella* group (Berlocher and Bush 1982), that analysis did not group *R. cornivora* with *R. juniperina* as in the mtDNA analysis. Although the ITS regions of other sibling species in the *pomonella* group, such as *R. mendax*, *R. zyphyria* and a *pomonella*-like species on the flowering dogwood in Florida, were not analyzed in this study, the species of this group are predicted to form a unambiguous clade if their ITS sequences are available because *R. cornivora* — the most genetically divergent based on allozyme and mtDNA data — is shown here to be closely related to *R pomonella*. Because of the important role of the *pomonella* group in the theory of sympatric speciation (*via* host shifting) and because this group also contains major economic pests, such as *R. pomonella* on apples and *R. mendax* on blueberries, accurate relationship among the members of this group is of particular interest both practically and theoretically. The nucleotide substitution rates between the *R. pomonella* and *R. cornivora* pairs of taxa were 2.99 in ITS1 and 1.59 in ITS2 while the highest rates among members of the *R. cingulata* species group were only 0.52 in ITS1 and 0.73 in ITS2. It is clear that the genetic distance in the *pomonella/cornivora* species pair is higher than that among sibling species in the *cingulata*

group. However, *R. pomonella* and the other *pomonella* group species are almost as closely related as *cingulata* group species.

Rhagoletis completa, the walnut husk fly, represents the *suavis* group in which speciation is believed to be allopatric (Bush 1975), unlike that for the *R. pomonella* and *R. cingulata* species groups. The close relatives of this group has been difficult to determine based on morphological characters (Bush 1966). There are contradictory molecular data regarding the placement of this group in the genus *Rhagoletis*. This study, unlike the result from the allozyme study (Berlocher and Bush 1982) but in support of the mtDNA analysis (Smith and Bush, unpublished data), indicates that the *suavis* group is more closely related to the *cingulata* group than to the *pomonella* group. However, bootstrap values (BCL) supporting the closeness of these two groups are not consistent and vary from 36% (mtDNA analysis) to 88% (ITS2 PAUP bootstrap in this study). Analysis of the ITS sequences of other sibling species in the *suavis* group, such as *R. juglandis*, *R. boycei* and *R. zoqui*, may eventually provide more information on the relationship of this group with other *Rhagoletis* species groups.

Rhagoletis electromorpha is a representative of the *tabellaria* species group which conventionally also includes *R. juniperina* and two other species (Bush 1966). In this study, however, *R. juniperina* is far removed from the *tabellaria* group, which is congruent with allozyme and mtDNA analyses. Morphologically, *R. juniperina* is probably the most divergent among the other members of the *tabellaria* group. It is becoming clearer that *R. juniperina* may indeed not belong to the *tabellaria* group. However, to verify this point, the ITS of other members of the *tabellaria* group could be analyzed. Based on allozyme analysis, *R. juniperina* is more closely related to the *cingulata* group; but the ITS data indicates *R. juniperina* is either more closely related to the *pomonella* group (Figure 10) or its placement is unresolved (Figure 11). The closeness of *R. juniperina* to the *pomonella* group has also been noted in the mtDNA protein parsimony cladogram (Smith and Bush, unpublished data). Consequently, the ambiguous placement of *R. juniperina* is still in need

of a more extended ITS and mtDNA sequence analysis of this genus and some reanalysis of morphological and other biochemical data sets.

Morphologically, *R. fausta*, which is a serious cherry pest, did not show enough affinity with any North American *Rhagoletis* species groups and remained unplaced (Bush 1966). Allozyme data (Berlocher and Bush 1982) on the other hand consistently placed this species with the *suavis* group. The ITS sequence of *R. fausta* does not support this placement but finds it more closely related to *R. electromorpha*, which is a representative of the *tabellaria* group, than to the *completa*, which belongs to the *suavis* group. This interpretation is consistent with mtDNA analysis which has several different placements for *R. fausta*, none of which supports the placement of *R. fausta* with the *suavis* group either (Smith and Bush, unpublished data). However, the placement of *R. fausta* with *R. electromorhpa* in this study is very tentative because it is based on ITS2 sequence alone, showing low BCL. Therefore, the placement of *R. fausta* remains in doubt. To better resolve this question other members of the *suavis* group, such as *juglandis*, *suavis* and *boycei*, as well as several other species from the *tabellaria* group should be sampled and a comprehensive analysis of their ITS sequences performed.

Rhagoletis basiola and *R. striatella* have undergone extensive sequence divergence in the ITS regions. Except for the ITS2 of *R. basiola*, the ITS sequences for these two species, especially *R. striatella*, are very difficult to confidently align with other *Rhagoletis* species surveyed here. Their large genetic distance in the ITS sequences from other *Rhagoletis* species are also reflected in their possessing of several primitive morphological characters (Bush 1966) and in their divergent allozyme frequencies (Berlocher and Bush 1982). The striking difference in allozyme frequency between *R. striatella* and other *Rhagoletis* has even led to the placement of this species outside of the genus *Rhagoletis* (Berlocher and Bush 1982). In mtDNA COII nucleotide parsimony cladograms (Smith and Bush, unpublished data), *R. striatella* formed a polytomy with non-*Rhagoletis* outgroups such as *Zonosemata electa* and *Oedicarena latifrons*. *R. basiola* belongs to the *R. alternata*

species group which mainly consists of Eurasian species and are considered to be distinct genus by some authors (Bush 1966). It is not surprising that the monophyly of the genus *Rhagoletis* as currently defined has been questioned by several workers (Norrbom 1989; Foote et al. 1993; Smith and Bush, unpublished data; and Jenkins, personal communication). To further test the monophyly of the genus *Rhagoletis* using ITS sequences, some close relatives of *Rhagoletis*, such as *Z. electa* and *O. latifrons*, need to be analyzed and compared with the respective *Rhagoletis* species.

In summary, phylogenetic implications from the ITS sequence study are partially in agreement with some of the previous studies on *Rhagoletis* phylogeny, and partially contradictory to the previous results. The results from this study indicate that *R. cornivora* belongs to the *pomonella* group, further confirming the placement based on morphology and karyotype Bush (1966), but contradictory to the mtDNA study (Smith and Bush, unpublished data); *R. juniperina* is removed from the *tabellaria* group, in support of allozyme and mtDNA analyses, and this species may more closely relate to the *pomonella* group rather than the *cingulata* group, which is in agreement with the mtDNA (Smith and Bush, unpublished data), but not with allozyme analysis; close relatives of the *cingulata* group are more likely the members of the *suavis* group rather than those in the *pomonella* group, which is congruent with the mtDNA study, but not with the allozyme results; *R. fausta* may relate to the *tabellaria* group, which was not implied by previous studies; *R. basiola* and *R. striatella* are most divergent in the ITS sequences, correlating with their possession of several ancient morphological characters (Bush, 1966) and their basal branching in the analyses of mtDNA and allozyme frequency, further supporting the idea that *Rhagoletis*, as currently defined, may not be monophyletic. This is the first study of rDNA ITS sequences of the genus *Rhagoletis*. Their application in phylogenetic study of *Rhagoletis* has provided some insight into the relationship between certain taxa from different species groups of this genus. The study of ITS should be expanded in the genus *Rhagoletis* which should provide a new avenue for the understanding of the evolutionary

history of this genus and providing a reliable phylogenetic framework to test some important evolutionary theories such as sympatric speciation through host shifting.

BIBLIOGRAPHY FOR CHAPTER III

Bibliography for Chapter III

Berlocher, S. H., and G. L. Bush. 1982. An electrophoretic analysis of *Rhagoletis* (Diptera: Tephritidae) phylogeny. Systematic Zoology 31:136-155.

Berlocher, S. H., B. A. McPheron, J. L. Feder, and G. L. Bush. 1993. Genetic differentiation at allozyme loci in the *Rhagoletis pomonella* (Diptera: Tephritidae) species complex. Ann. Entomol. Soc. Am. **86**:716-727.

Berlocher, S. H., B. A. McPheron, J. L. Feder, and G. L. Bush. 1993. Genetic differentiation xx allozyme loci in the *Rhagoletis pomonella* (Diptera: Tephritidae) species complex. Ann. Ent. Soc. Amer. **86**:1-11.

Bernardi, G., D. Mouchiroud, C. Gautier, and G. Bernardi. 1988. Compositional patterns in vertebrate genomes: conservation and change in evolution. J. Mol. Evol. **28**:7-18.

Brown, W. M., E. M. Prager, A. Wang, and A. C. Wilson. 1982. Mitochondrial DNA sequences of primates: Tempo and mode of evolution. J. Mol. Evol. **18**:225-239.

Bush, G. L. 1966. The taxonomy, cytology, and evolution of the genus *Rhagoletis* in North America (Diptera: Tephritidae). Bull. Mus. Comp. Zool. **134**: 431-562.

Bush, G. L. 1969a. Mating behavior, host specificity, and the ecological significance of sibling species in frugivorous flies of the genus *Rhagoletis* (Diptera: tephritidae). Amer. Nat. **103**: 669-672.

Bush, G. L. 1969b. Sympatric host race formation and speciation in frugivorous flies of the genus *Rhagoletis* (Diptera: Tephritidae). Evolution **23**:237-251.

Bush, G. L. 1975. Sympatric speciation in phytophagous parasitic insects. Pp. 187-205 *in* P. W. Price, ed. Evolutionary strategies of parasitic insects and mites. Plenum Publishing Corp., New York.

Bush, G. L. 1993. Host race formation and sympatric speciation in *Rhagoletis* fruit flies (Diptera: Tephritidae). Psyche **99**: 335-357.

Bush, G. L. 1994. Sympatric speciation in animals: new wine in old bottles. Trends Ecol. Evol. **9**:285-288.

De Salle, 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.

Diehl, S. R., and G. L. Bush. 1984. An evolutionary and applied perspective of insect biotypes. Annu. Rev. Entomol. **29**:471-504.

Dover, G. 1982. Molecular drive: a cohesive mode of species evolution. Nature **299**:111-117.

Dover, G. A. 1989. Linkage disequilibrium and molecular drive in the rDNA gene family. Genetics **122**: 249-252.

Feder, J. L., and G. L. Bush. 1991. Genetic variation among apple and hawthorn host races of *Rhagoletis pomonella* (Diptera: Tephritidae) across an ecological transition zone. Entomol. Exp. Appl. **59**:249-265.

Fritz, G. N., J. Conn, A. Cockburn, and J. Seawright. 1994. Sequence analysis of the ribosomal DNA internal transcribed spacer 2 from populations of *Anopheles nuneztovari* (Diptera: Culicidae). Mol. Biol. Evol. **11**:406-416.

Hillis, D. M., and M. T. Dixon. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. Q. Rev. Biol. **66**:411-453.

Hillis, D. M., C. Moritz, C. A. Porter, and R. J. Baker. 1991. Evidence for biased gene conversion in concerted evolution of ribosomal DNA. Science **251**:308-310.

Jukes, T. H., and C. R. Cantor. 1969. Evolution of protein molecules. Pp. 21-132 *in* H. N. Munro, ed. Mammalian protein metabolism. Academic Press, New York.

Kimura, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**: 111-120.

Norrbom, A. L. 1989. The status of *Urophora acuticornis* and *U. sabroskyi* (Diptera: Tephritidae). Entomol. News **100**:59-66.

Park, Y. J., and A. M. Fallon. 1990. Mosquito ribosomal RNA genes: characterization of gene structure and evidence for changes in copy number during development. Insect Biochem. **20**:1-11.

Porter, C. H., and F. H. Collins. 1991. Species-diagnostic differences in a ribosomal DNA internal transcribed spacer from the sibling species *Anopheles freeborni* and *Anopheles hermsi* (Diptera: Culicidae). Am. J. Trop. Med. Hyg. **45**:271-279.

Procunier, W. S., and J. J. Smith. 1993. Localization of ribosomal DNA in *Rhagoletis pomonella* (Diptera: Tephritidae) by *in situ* hybridization. Insect Mol. Biol. 2:163-174.

Quicke, D. L. J. 1993. Principles and techniques of contemporary taxonomy. Chapman and Hall, New York.

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

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **12**:5463-5467.

Schlotterer, C., M. Hauser, A. von Haeseler, and D. Tautz. 1994. Comparative evolutionary analysis of rDNA ITS regions in *Drosophila*. Mol. Biol. Evol. **11**:513-522.

Swofford, D. L. 1993. PAUP: phylogenetic analysis using parsimony, version 3.1.1. University of Illinois, Champaign.

Tamura, K. 1992. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C content biases. Mol. Biol. Evol. **9**:678-687.

Tautz, D., J. M. Hancock, D. A. Webb, C. Tautz, and G. A. Dover. 1988. Complete sequence of the rDNA genes of *Drosophila melanogaster*. Mol. Biol. Evol. **5**: 366-376.

Torres, R. A., M. Ganal, and V. Hemleben. 1990. GC balance in the internal transcribed spacers ITS1 and ITS2 of nuclear ribosomal RNA genes. J. Mol. Evol. **8**:669-686.

Vogler, A. P., and R. DeSalle. 1994. Evolution and phylogenetic information content of the ITS-1 region in the tiger beetle *Cicindela dorsalis*. Mol. Biol. Evol. **11**:393-405.

Wesson, D. M., C. H. Porter, and F. H. Collins. 1992. Sequence and secondary structure comparisons of ITS rDNA in mosquitoes (Diptera: Culicidae). Mol. Phylogenet. Evol. **1**:253-269.

Williams, S. C., G. Furnier, E. Fuog, and C. Strobeck. 1987. Evolution of the ribosomal DNA spacers of *Drosophila melanogaster*: different patterns of variation on X and Y chromosomes. Genetics **116**:225-232.

Williams, S. C., R. W. DeBry, and J. L. Feder. 1988. A commentary on the use of ribosomal DNA in systematic studies. Syst. Zool. **37**:60-62.

Yagura, T., M. Yagura, and M. Muramatsu. 1979. *Drosophila melanogaster* has different ribosomal RNA sequences on X and Y chromosomes. J. Mol. Biol. **133**:533-547.

Yokota, Y., T. Kawata, Y. Iida, A. Kato, and S. Tanifuji. 1989. Nucleotide sequences of the 5.8S rRNA gene and internal transcribed spacer regions in carrot and broad bean ribosomal DNA. J. Mol. Evol. **29**:294-301.

CHAPTER IV

CONSTRUCTION OF AN *RHAGOLETIS POMONELLA* GENOMIC LIBRARY AND IDENTIFICATION OF THE COMPLETE rRNA REPEATING UNIT

Introduction

Higher eukaryotic nuclear rDNA is composed of tandemly repeated transcriptional units separated from each other by the non-transcribed intergenic spacer regions (IGS). Each transcriptional unit contains the genes for the 18S, 5.8S and 28S ribosomal RNA (rRNA) and the external and internal transcribed spacers (ETS and ITS, respectively; reviewed in Gerbi 1985; see Figure 13). The entire rDNA unit is transcribed by RNA polymerase I as a single precursor molecule, which is then processed to yield mature 18S, 5.8S and 28S rRNAs (Hadjiolov 1985; Sollner-Webb and Tower 1986). The transcriptional promoter of rDNA is located in the IGS/ETS boundary region and upstream in the IGS there are also multicopies of promoter-like sequences called transcription enhancers. Sequence comparison between species shows that the rDNA spacers including IGS and ETS diverge much faster than the gene coding regions (reviewed in Long and Dawid 1980). Further evidence that the transcription initiation region (i.e, IGS/ETS boundary) undergoes rapid evolutionary change comes from the fact that the RNA polymerase I of a particular species usually transcribes rDNA from that species only and fails to recognize the rDNA transcription initiation region of another species in *in vivo* and *in vitro* transcription assays among mammals, insects, and protozoans (for reviews, see Arnheim 1983; Dover and Flavell 1984). Therefore, sequence conservation of enhancer regions in the IGS and initiation recognition site in the ETS within a given species may be functionally important and the sequence divergence in these regions may be expected to be

Figure 13. Overall map of the *Drosophila melanogaster* rDNA repeat unit. Adapted from studies described by Tautz et al. (1987 and 1988). Abbreviations: IGS, intergenic spacer; ETS, external transcribed spacer; ITS, internal transcribed spacer; rRNA coding regions (18S, 5.8S and 28S); E, *Eco*RI; B, *Bgl*II; H, *Hin*dIII; Ha, *Hae*III. Only 1 *Hae*III is shown here. DNA probes used in this study: 1) from *D. melanogaster*, 11.7 kb *Eco*RI fragment from plasmid pDM238; and 2) from *D. virilis*, approx. 5 kb *Hin*dIII/*Hin*dIII fragment which includes IGS and ETS, DvH25. The forward orientation, direction of the primary transcript, is shown as reference for the primers used in the experiments using the polymerase chain reaction (PCR). The scale is indicated by the bar.

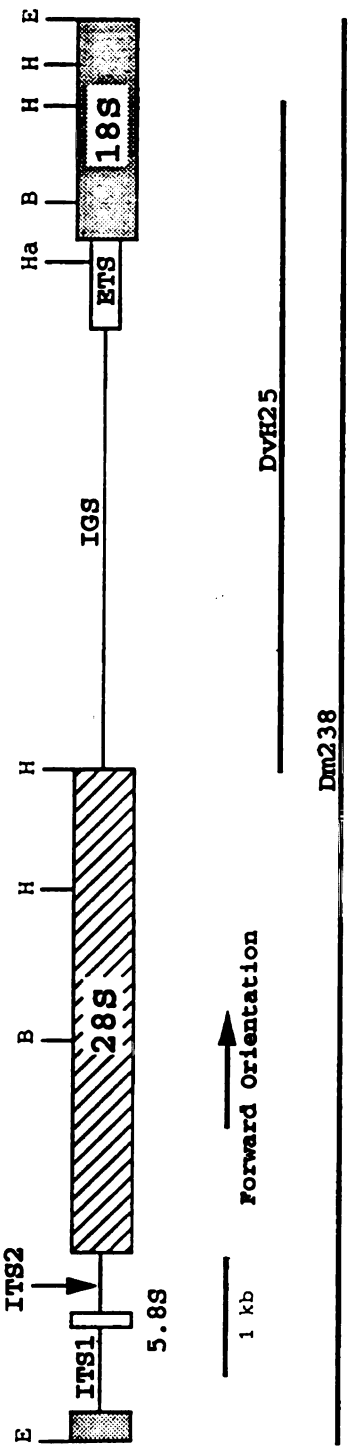

E
H
H
18S
B
Ha
ETS
IGS
H
H
B
28S
ITS2
5.8S
ITS1
E
1 kb
Forward Orientation
DvH25
Dm238

Figure 13

species specific. Furthermore, Williams et al. (1985) found that individual flies of *Drosophila mercatorum* from different geographic regions have characteristic length of the IGS, which range from about 4kb to more than 6.5kb. Those length variations may serve as markers for determining the geographic origin of individuals. In light of these findings, the IGS/ETS is a promising region for phylogenetic analysis of closely related species, in particular those species (or host races) that are morphologically indistinguishable, such as those often encountered in the genus *Rhagoletis*.

Although an increasing body of literature on rDNA has been accumulated for *Drosophila*, *Aedes*, *Glossina* and *Xenopus* (Furlong et al. 1983; Gerbi 1985; Cross and Dover 1987a; Cross and Dover 1987b; Gale and Crampton 1989), the rDNA of agriculturally important insects in the genus *Rhagoletis* has not been characterized. To evaluate the potential use of *Rhagoletis* rDNA in studies on speciation and phylogenetic reconstruction, I carried out a preliminary characterization on the organization of *R. pomonella* rDNA. I first constructed a genomic DNA library from *R. pomonella* , then took advantage of the highly conserved 18S and 28S genes using the complete repeating rDNA unit of *Drosophila melanogaster* (provided by Dr. G. Dover, see Figure 13) as a probe to identify *R. pomonella* rDNA clones. In this study, for the first time, I report several positive rDNA clones for *R. pomonella*, which together cover all of the rDNA coding regions and the various spacers. In addition, the *R. pomonella* ITS1 and ITS2 regions were sequenced and compared with those from *D. melanogaster*. The *R. pomonella* rDNA clones from this study will form the basis for a more detailed study of the sequence organization of the rDNA repeating unit for the genus *Rhagoletis* in the future. This is of particular interest since the IGS and ETS are potentially useful regions for phylogenetic study of closely related species or even host races in the genus *Rhagoletis*. Phylogeny inferred from the IGS/ETS will allow us to compare the phylogeny obtained from the ITS regions.

Materials and Methods

Biological Material

Rhagoletis pomonella larvae were collected in late summer from the fruit of *Malus pumula* (domestic apple; Door Co., WI) by Jeff Feder and allowed to pupate in the laboratory. Upon emergence the following spring flies were stored at -70°C since 1984 for rDNA analysis.

Genomic DNA Isolation and *Sau*3AI Partial Digestion

R. pomonella genomic library construction is briefly outlined in Figure 14. Total genomic DNA was isolated from 50 flies as described by Procunier and Smith (1993). Twenty µg of the genomic DNA was partially digested with 0.29 U of *Sau*3AI (GIBCO-BRL) in 500 µl (reaction volume) containing 10 mM Tris-HCl, pH 7.5, 100 mM NaCl and 7 mM $MgCl_2$ at 37°C for 30 min. The reaction products were extracted twice with an equal volume of phenol/chloroform/isoamyl alcohol (PCI; 25:24:1) and once with an equal volume of chloroform/isoamyl alcohol (CI; 24:1). The partially digested DNA was precipitated with NH_4OAc/ethanol and then dissolved in TE buffer (10 mM Tris-HCl, pH 7.5, and 1 mM EDTA). The *Sau*3AI digestion conditions were optimized to generate genomic DNA fragments ranging in size between 6 and 16 kb. About 10 µg of the *Sau*3AI-treated genomic DNA was used in a partial fill-in reaction of the first two nucleotides of the *Sau*3AI site (Figure 15A). The 25 µl final reaction volume contained 50 mM Tris-HCl, pH 7.2, 10 mM $MgSO_4$, 0.1 mM dithiothreitol, 12.5 µg acetylated BSA (Promega), 1 mM dATP, 1 mM dGTP and 3 U of the Klenow enzyme (GIBCO-BRL). The partial fill-in reaction was carried out at 37°C for 30 min. The reaction products were extracted with PCI and CI as before, precipitated with ethanol and then dissolved in 2 µl of TE buffer.

Figure 14. Schematic outline of the procedure for generating a *Rhagoletis pomonella* genomic DNA library and screening for the rDNA repeat unit.

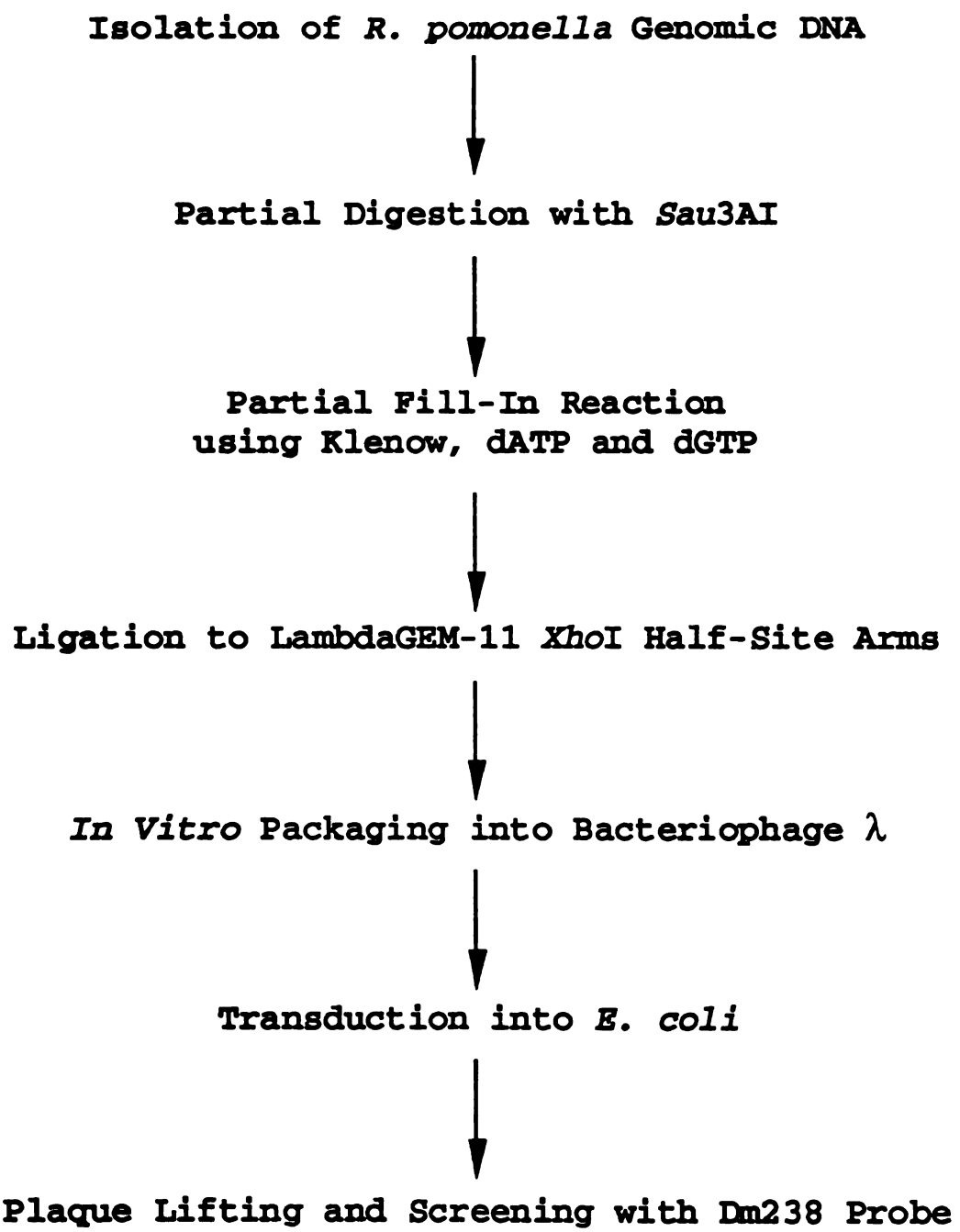

Isolation of R. pomonella Genomic DNA
Partial Digestion with Sau3AI
Partial Fill-In Reaction
using Klenow, dATP and dGTP
Ligation to LambdaGEM-11 XhoI Half-Site Arms
In Vitro Packaging into Bacteriophage λ
Transduction into E. coli
Plaque Lifting and Screening with Dm238 Probe

Figure 14

Figure 15. Detailed outline of several specific steps in Figure 14. Panel a represents the *Sau*3AI partial-digestion reaction of *Rhagoletis pomonella* genomic DNA and the subsequent partial fill-in reaction with dATP and dGTP. Panel b presents details of the LambdaGEM-11 *Xho*I half-site arms purchased from Promega and the cloning product. The two *Sal*I restriction sites are about 100 bp away from the *Xho*I half-sites. Bold letters indicate the nucleotides that were added during the partial fill-in reactions. The figure is not drawn to scale.

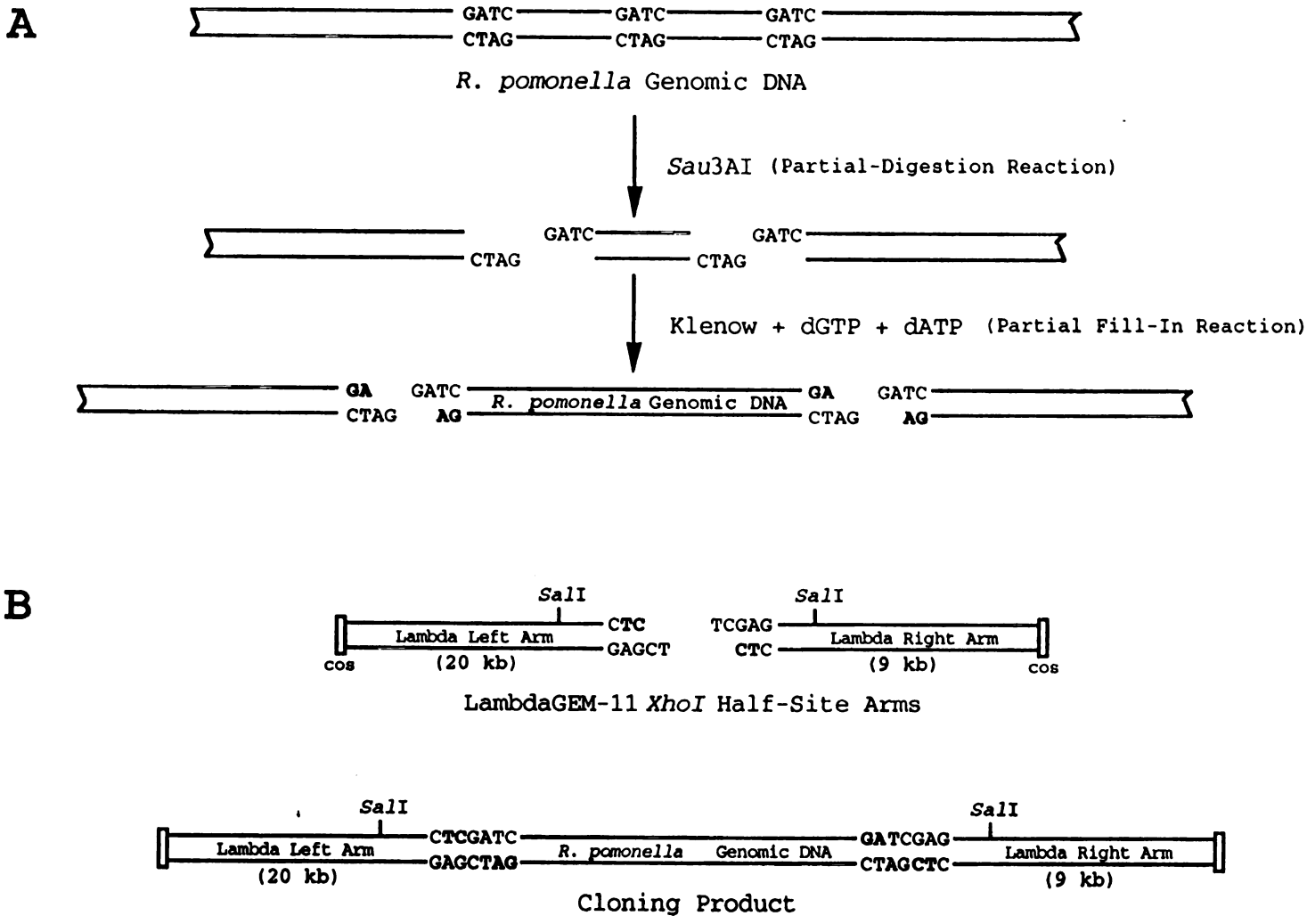

A
GATC GATC GATC
CTAG CTAG CTAG
R. pomonella Genomic DNA
Sau3AI (Partial-Digestion Reaction)
GATC GATC
CTAG CTAG
Klenow + dGTP + dATP (Partial Fill-In Reaction)
GA GATC GA GATC
CTAG R. pomonella Genomic DNA CTAG
AG AG
B
SalI SalI
CTC TCGAG
Lambda Left Arm Lambda Right Arm
GAGCT CTC
cos (20 kb) (9 kb) cos
LambdaGEM-11 XhoI Half-Site Arms
SalI SalI
CTCGATC GATCGAG
Lambda Left Arm R. pomonella Genomic DNA Lambda Right Arm
GAGCTAG CTAGCTC
(20 kb) (9 kb)
Cloning Product

Figure 15

Cloning into Bacteriophage λ Half-Site Arms and Packaging

The partial fill-in reaction products were cloned into LambdaGEM-11 *Xho*I half-site arms (Promega; Cat No. B1960; Figure 15B) using T4 DNA ligase (GIBCO-BRL) according to the manufacturer's (Promega) instructions. Control cloning reactions without insert and with a test insert provided by the manufacturer were also performed. The ligation reactions were packaged using the Packagene *in vitro* packaging system (Promega; Cat. No. K315) according to the manufacturer's instructions. Packaged λ phage titers were checked on Luria-Bertaini (LB) plates using *Esherichia coli* LE392 (Promega), according to the manufacturer's instructions, and were found to range between 1.2×10^4 - 2.8×10^5 pfu/ml. As expected, controls without inserts showed no plaques on LB plates.

DNA Probe Generation and Genomic Library Screening

The 11.7 kb Dm238 *Eco*RI insert (Figure 13) from plasmid pDM238 (provided by Dr. G. Dover) was prepared as described in Procunier and Smith (1993). Purified Dm238 *Eco*RI insert (200 ng) was labeled with digoxygenin-dUTP (Boehringer Mannheim; Cat. No. 1093088) by random-priming using a commercially available kit (Boehringer Mannheim) according to the manufacturer's specifications. In addition to the Dm238 EcoRI fragment, the approximately 5 kb *Hin*dIII/*Hin*dIII fragment of *D. virilis* rDNA repeat unit, DvH25 (Tautz et al. 1987), was also used to generate digoxygenin-labelled probes. Labelled probes were precipitated from an 80 µl reaction mixture by adding 10 µl of 3M NaOAc, pH 5.0, 1 µl of 20 mg tRNA per ml (Boehringer Mannheim; Cat. No. 109495) and 250 µl 100% ethanol. After centrifugation, the pellets were washed with 70% ethanol, air dried and dissolved in 100 µl TE containing 0.1% SDS. The probes (about 1-2 ng/µl) were stored at 4°C until use.

Bacteriophage λ plaques were immobilized onto Gene Screen nylon membranes (DuPont Inc.; Cat No. NEF-983) as described elsewhere (Sambrook et al. 1989). Membranes were prehybridized in buffer A (2X SSC [0.15 M NaCl, 20 mM sodium

citrate, pH 7.0], 0.5% [w/v] casein [Sigma; Cat. No. C5890], 0.1% [w/v] N-lauroyl sarcosine and 0.2% [w/v] SDS) at 68°C for at least 2 hr. Hybridization was performed overnight at 68°C with fresh buffer A containing digoxygenin-labelled Dm238 probe (approx. 3 ng labelled Dm238 per ml buffer A). Filters were washed twice at room temperature with 2X SSC/0.1% SDS for 10 min each and once at 68°C with 0.2X SSC/0.1% SDS for 15 min. Hybridized DNA was detected with an anti-dioxygenin alkaline phosphatase conjugate (50 mU/ml) and LumiPhos substrate of the Genius system (Boehringer Mannheim) using the protocol outlined in the Boehringer Mannheim Technical Bulletin for LumiPhos 530 (900264R3/10M; Jan 1991) substituting casein for the blocking reagent. Membranes were then exposed to X-OMAT film (Kodak) for 1 min to 2 hr for visualization.

Characterization of Recombinant Phage by Gel Electrophoresis and Southern Hybridization

After identifying several positive plaques, the λ bacteriophages corresponding to those plaques were purified and the phage DNA was isolated using protocols described by Sambrook et al. (1989). Isolated phage DNA was digested with various restriction enzymes and the products analyzed by agarose gel electrophoresis according to common procedures described by Ausubel et al. (1987). After electrophoresis, the gel was stained with 0.5 μg/ml ethidium bromide and visualized using a transilluminator. The gel was also photographed using a red filter with Polaroid positive/negative films (type 55). DNA in the gel was then transferred by capillary action to Gene Screen nylon membranes (DuPont Inc.; Cat No. NEF-983) and Southern hybridization was performed using the labelled Dm238 and DvH25 probes (see earlier section), as described elsewhere (Sambrook et al. 1989).

PCR Amplification and DNA Sequencing

The ITS1 region was amplified using primers 1406F 5'CCTTTGTACACACCGCCCGT (matching the 3' end of 18S) and 35R-GB14

5'AGCTRGCTGCGTTCTTCATCGA (matching the 5' end of 5.8S). The ITS2 region was amplified using primers 108F 5'GAACATCGACHHKTYGAACGCA (matching the 3' end of 5.8S) and 52R 5'GTTAGTTTCTTTTCCTCCSCT (matching the 5' end of 28S). The letters F and R refer to the forward and reverse orientation of the primers, respectively, with respect to complete rDNA repeat unit (Figure 13). Amplification by the polymerase chain reaction (PCR) was carried out in 25 µl (final volume) containing 10 mM Tris-HCl, pH 8.5, 50 mM KCl, 3 mM MgCl2, 375 µM of each dNTP, 0.1-0.4 µM primer 1406F for ITS1 (or 108F for ITS2) and 0.1-0.4 µM primer 35R-GB14 for ITS1 (or 52R for ITS2), and 1.25-2.50 units of Ampli Taq DNA polymerase (Perkin Elmer Cetus). Amplification parameters were 92°C for 3 min 10 sec; 35 cycles each with 92°C for 15 sec, 65°C for 15 sec and 72°C for 1 min 10 sec; and 72°C for 6 min 10 sec. Amplified DNA was subjected to electrophoresis on 0.8-1.0% agarose gels and visualized with ethidium bromide. Bands containing the DNA were excised from the gel and the DNA purified using the Prep-A-Gene DNA purification matrix (Bio-Rad), according to the manufacturer's instructions.

The TA Cloning kit (Invitrogen) was used in a one-step cloning strategy for the direct insertion of the purified PCR products into a plasmid vector, followed by transformation into competent cells. Plasmid vector and competent cells were supplied by the manufacturer. Plasmid DNA was purified from individual clones using the Magic-Prep DNA purification kit (Promega), following the manufacturer's instructions. DNA sequencing was performed according to the chain-termination method of Sanger et al. (1977), and using the Sequenase Version 2.0 DNA sequencing kit (USB) and ^{35}S-dATP (Amersham). The same primers as in the amplification reactions (see earlier sections) were used to determine the DNA sequence in both directions. Once a certain stretch of DNA was sequenced other primers were employed to complete the sequencing of the ITS1 and ITS2 regions; primers 35R-GB27 5'ACC(CT)AAACATTTTCAAGT(CT)GCG and 108F-GB25 5'A(AT)(AG)(AG)AATC(AT)(CT)AGTATTCCC were used for the ITS1 and ITS2 regions, respectively.

Computer Analyses

A number of programs, including FETCH, GAP, MAPSORT, SEQED, and STRINGSEARCH, in the GCG package of the University of Wisconsin Genetics Computer Group (UWGCG package, version 8.0) were used in sequence comparison of ITS1 and ITS2 between *R. pomonella* and *D. melanogaster*. The sequence of *D. melanogaster* rDNA repeating unit, from clone pDm238, was found using Stringsearch under the locus name DROGRAB and accession numbers M21017 and M29800. Bases 1 to 7232 and 7206 to 12026 were described in Tautz et al. (1988) and Tautz et al. (1987), respectively. In the *D. melanogaster* DROGRAB locus, the actual region between the 28S and 18S subunits of rDNA, which includes the IGS and ETS regions, subtends between bases 7206 to 11729.

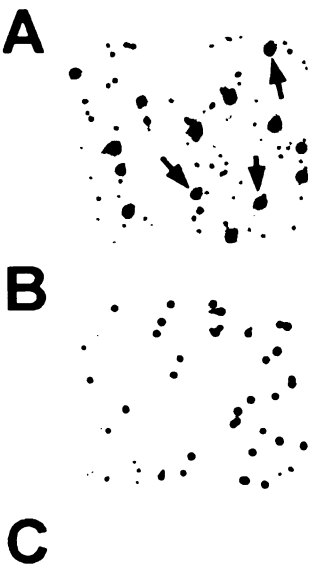
Results and Discussion

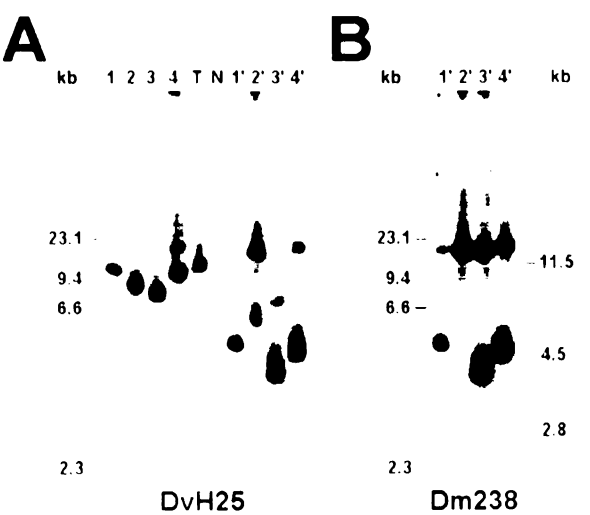
Genomic DNA Library Analysis

A high number of positive recombinant bacteriophage plaques were identified after plaque-lifting and probing membranes with Dm238 (Figure 16A shows only one of the several membranes). After collecting the well-separated plaques and repeating the transduction and screening protocols, almost all of the plaques on the plates turned out to be positive recombinants when probed with Dm238 (Figure 16B). The well-separated plaques from the second screening were picked and named RpA, RpB, RpC, and RpD (Figure 16B only shows the filter for clone RpD). Several plaques from the first screening with relatively low or no signals when probed with Dm238 were also selected (Figure 16C). Some of these latter plagues did indeed have genomic inserts, but non-rDNA , and were used as negative controls in subsequent experiments (see below). Surprisingly, plaques containing the control test insert (provided by Promega as negative control) also showed positive signals with Dm238 (data not shown). Upon correspondence with researchers at Promega, it was confirmed that their test insert contained significant amounts

Figure 16. Screening of the *Rhagoletis pomonella* genomic DNA library with Dm238 probe. In panel A arrows indicate several positive recombinant plaques (positive for nuclear rDNA). Smaller spots represent non-rDNA recombinant and/or nonrecombinant (background) plaques. One such spot was used as a negative control. On the LB plates all plaques appeared of equal size. Panel b represents clone RpD after the second screening protocol. Panel c represents the negative control plaque from the first screening in Panel a.

Figure 17. Southern blot analysis of the restriction digestion products of DNA purified from *R. pomonella* rDNA clones. Panels a and b represent blots hybridized with DvH25 and Dm238, respectively. Lanes 1 (1'), 2 (2'), 3 (3') and 4 (4') correspond to DNA from clones RpA, RpB, RpC and RpD, respectively. DNA in lanes 1, 2, 3 and 4 from Panel A was digested with *SalI*. DNA in lanes 1', 2', 3' and 4' from Panels A and B was digested with *BglII*. T and N stand for test insert (from Promega) and negative control (see Material and Methods, and Results sections), respectively.





Figures 16(top) and 17(bottom)

of rDNA from mice. To further characterize the *R. pomonella* rDNA clones, purified DNA from the positive recombinant plaques was digested with *Sal*I. This releases the complete DNA insert flanked on each side with about 100 bp of λ DNA (see Figure 15B for the *Sal*I sites). The choice for using *Sal*I restriction enzyme to release the insert was made in part because a search, using MAPSORT in GCG, for *Sal*I site(s) in the complete *Drosophila* rDNA repeat unit showed no such site; consequently, I reasoned that *Rhagoletis* rDNA repeat units may also lack *Sal*I site(s). Separation of the digested products by gel electrophoresis and Southern hybridization using DvH25 probe (also provided by Dr. G. Dover; Figure 13) showed one major band from clones RpA, RpB, RpC and RpD with the fragment size varying between 8 and 12 kb among the various clones (Figure 17A lanes 1-4). Another minor band was found in clones RpB, RpC and RpD in around the compression region of the gel (about 23 kb). This latter band is most likely due to incomplete digestion with *Sal*I because the *Sau*3AI partial digestion products of the *R. pomonella* genomic DNA ranged in size between 6 and 15 kb and, consequently, any *Sal*I site in the *R. pomonella* rDNA repeat should only lead to shorter DNA fragments and not as large as 20-23 kb. When the same membranes were stripped and probed with Dm238 (longer than DvH25) only the same *Sal*I fragments were observed as using DvH25 and no additional fragments were detected indicating that indeed SalI released the inserts as a whole fragment in the above positive clones (Figure 17A only shows results using DvH25). The size of the insert in clone RpD was comparable to the size of the complete *D. melanogaster* rDNA repeat unit, which is about 11.7 kb (Coen et al. 1982). Further digestion of the clones with *Bgl*II and hybridization of membranes with DvH25 and Dm238 probes (Figure 17; Panels A and B) showed DNA fragments comparable in size to the 4.6 kb *Bgl*II/*Bgl*II fragment of *D. melanogaster* rDNA which contains ITS1 and ITS2. In all cases, the negative control which contained a DNA insert (when gels were visualized by ethidium bromide; data not shown) did not hybridize with either of the mentioned two probes (Figure 17; Panels A and B).

ITS Sequence Analysis

The ITS1 and ITS1 fragments, PCR amplified from one of the positive recombinant phage DNA (RpD) using primers matching the highly conserved regions of the 18S, 5.8S, and 28S genes are shown in Figure 18. There is one major DNA band of approx. 1 kb for ITS1 of *R. pomonella* and a similar size DNA fragment was also amplified for the ITS1 region of *Drosophila* (from Dm238 DNA) as a positive control (Figure 18A). Similarly, one major band of approx. 680 bp was detected for the ITS2 region of the same clone of *R. pomonella* and one band of approx. 550 bp from *Drosophila* (Figure 18B). The major PCR bands from both *R. pomonella* and *D. melanogaster* were purified (using Prep-A-Gene), cloned into the pCR II vector (from Invitrogen) and then sequenced. The sequence comparison of ITS for the two species were carried out using the GAP sequence alignment program in the GCG package and presented in Figure 19. The length of the ITS1 and ITS2 sequences from *R. pomonella* clone RpD were 680 and 475 bp, respectively. The identities of the ITS sequences were verified by sequencing about 100 nucleotides for the adjacent highly conserved coding regions (18S, 5.8S and 28S) since these regions are almost identical between *Rhagoletis* and *Drosophila* (Tautz et al. 1988). *R. pomonella* ITS1 and ITS2 sequences were 46 bp shorter and 90 bp longer than the respective *D. melanogaster* ITS sequences (Tautz et al. 1988). The A-T content of the *R. pomonella* ITS1 and ITS2 sequences were 80.1 and 81.7%, respectively; slightly higher than those of the *D. melanogaster* (73.1 and 80.0%, respectively; Tautz et al. 1988; Schlotterer et al. 1994). According to the ITS1 and ITS2 sequence alignments shown in Figure 19, there is substantial divergence between *R. pomonella* and *D. melanogaster,* including several insertion/deletion differences. However, a number of highly conserved regions are also observed between the two along the length of the spacers (for example, the underlined sequence in Figure 19B). The latter point is especially obvious for ITS2, indicating that some of these regions may be of structural/functional significance.

Figure 18. Products of the PCR amplification reaction of the ITS1 and ITS2 regions of *Rhagoletis pomonella*. Panels a and b correspond to ITS1 and ITS2, respectively. Lanes 1 and 2 in both panels represent products of PCR amplification reactions using Dm238 DNA and purified recombinant bacteriphage λ DNA from clone RpD, respectively. Lanes with M represent the 123 bp DNA ladder as molecular size markers.

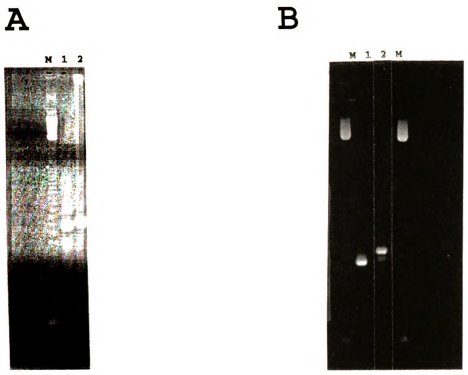


Figure 18

Figure 19. Nucleotide sequences of the ITS1 and ITS2 from *Rhagoletis pomonella* and pairwise comparison with their counterparts from *Drosophila melanogaster*. The GAP sequence alignment program in the GCG package was used to perform this pairwise comparison. GAP penalty parameters (factors) were set at: gap weight = 1.00; gap length weight = 0.10. Bars represent nucleotide identities and gaps are represented by dots. Panels a and b are for ITS1 and ITS2, respectively. The underlined region in Panel b represents part of a constrained structural element conserved between *Rhagoletis* and *Drosophila*, and discussed in Chapter II. R. pom = *R. pomonella* and D. mel = *D. melanogaster*.

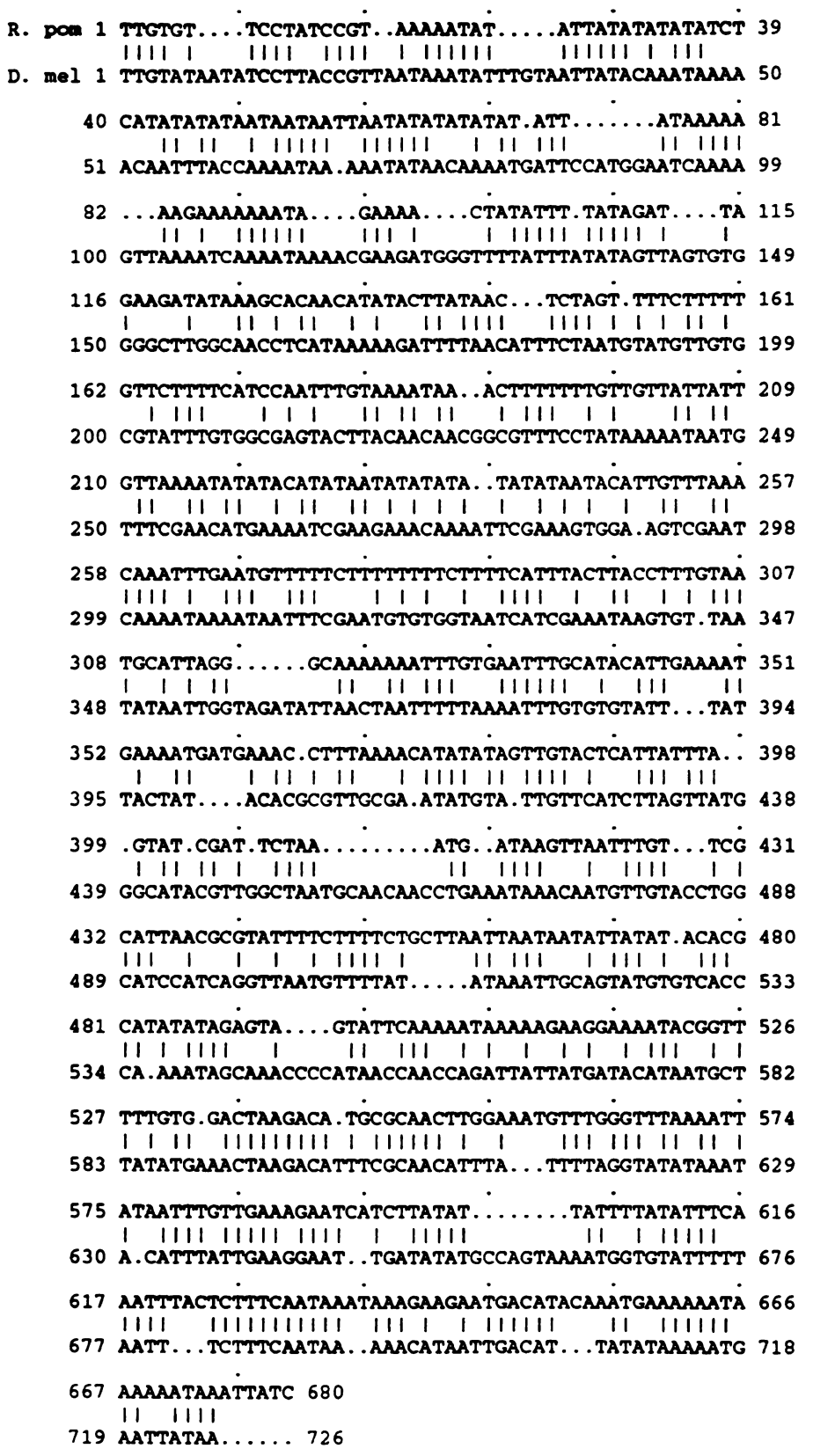

R. pom 1 TTGTGT....TCCTATCCGT..AAAAATAT.....ATTATATATATATCT 39
D. mel 1 TTGTATAATATCCTTACCGTTAATAAATATTTGTAATTATACAAATAAAA 50
40 CATATATATAATAATAATTAATATATATATAT.ATT.......ATAAAAA 81
51 ACAATTTACCAAAATAA.AAATATAACAAAATGATTCCATGGAATCAAAA 99
82 ...AAGAAAAAAATA....GAAAA....CTATATTT.TATAGAT....TA 115
100 GTTAAAATCAAAATAAAACGAAGATGGGTTTTATTTATATAGTTAGTGTG 149
116 GAAGATATAAAGCACAACATATACTTATAAC...TCTAGT.TTTCTTTTT 161
150 GGGCTTGGCAACCTCATAAAAAGATTTTAACATTTCTAATGTATGTTGTG 199
162 GTTCTTTTCATCCAATTTGTAAAATAA..ACTTTTTTTGTTGTTATTATT 209
200 CGTATTTGTGGCGAGTACTTACAACAACGGCGTTTCCTATAAAAATAATG 249
210 GTTAAAATATATACATATAATATATATA..TATATAATACATTGTTTAAA 257
250 TTTCGAACATGAAAATCGAAGAAACAAAATTCGAAAGTGGA.AGTCGAAT 298
258 CAAATTTGAATGTTTTTCTTTTTTTTCTTTTCATTTACTTACCTTTGTAA 307
299 CAAAATAAAATAATTTCGAATGTGTGGTAATCATCGAAATAAGTGT.TAA 347
308 TGCATTAGG......GCAAAAAAATTTGTGAATTTGCATACATTGAAAAT 351
348 TATAATTGGTAGATATTAACTAATTTTTAAAATTTGTGTGTATT...TAT 394
352 GAAAATGATGAAAC.CTTTAAAACATATATAGTTGTACTCATTATTTA.. 398
395 TACTAT....ACACGCGTTGCGA.ATATGTA.TTGTTCATCTTAGTTATG 438
399 .GTAT.CGAT.TCTAA.........ATG..ATAAGTTAATTTGT...TCG 431
439 GGCATACGTTGGCTAATGCAACAACCTGAAATAAACAATGTTGTACCTGG 488
432 CATTAACGCGTATTTTCTTTTCTGCTTAATTAATAATATTATAT.ACACG 480
489 CATCCATCAGGTTAATGTTTTAT.....ATAAATTGCAGTATGTGTCACC 533
481 CATATATAGAGTA....GTATTCAAAAATAAAAAGAAGGAAAATACGGTT 526
534 CA.AAATAGCAAACCCCATAACCAACCAGATTATTATGATACATAATGCT 582
527 TTTGTG.GACTAAGACA.TGCGCAACTTGGAAATGTTTGGGTTTAAAATT 574
583 TATATGAAACTAAGACATTTCGCAACATTTA...TTTTAGGTATATAAAT 629
575 ATAATTTGTTGAAAGAATCATCTTATAT........TATTTTATATTTCA 616
630 A.CATTTATTGAAGGAAT..TGATATATGCCAGTAAAATGGTGTATTTTT 676
617 AATTTACTCTTTCAATAAATAAAGAAGAATGACATACAAATGAAAAAATA 666
677 AATT...TCTTTCAATAA..AAACATAATTGACAT...TATATAAAAATG 718
667 AAAAATAAATTATC 680
719 AATTATAA...... 726

Figure 19A

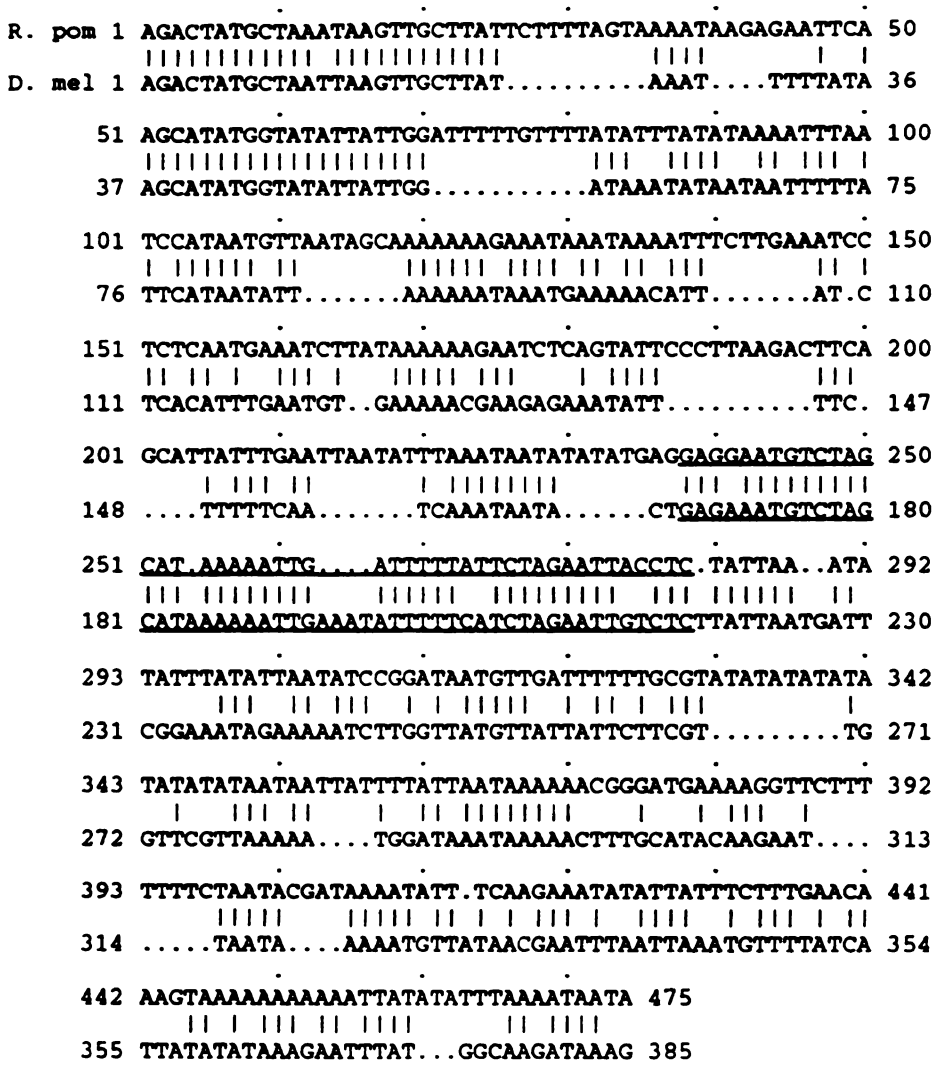
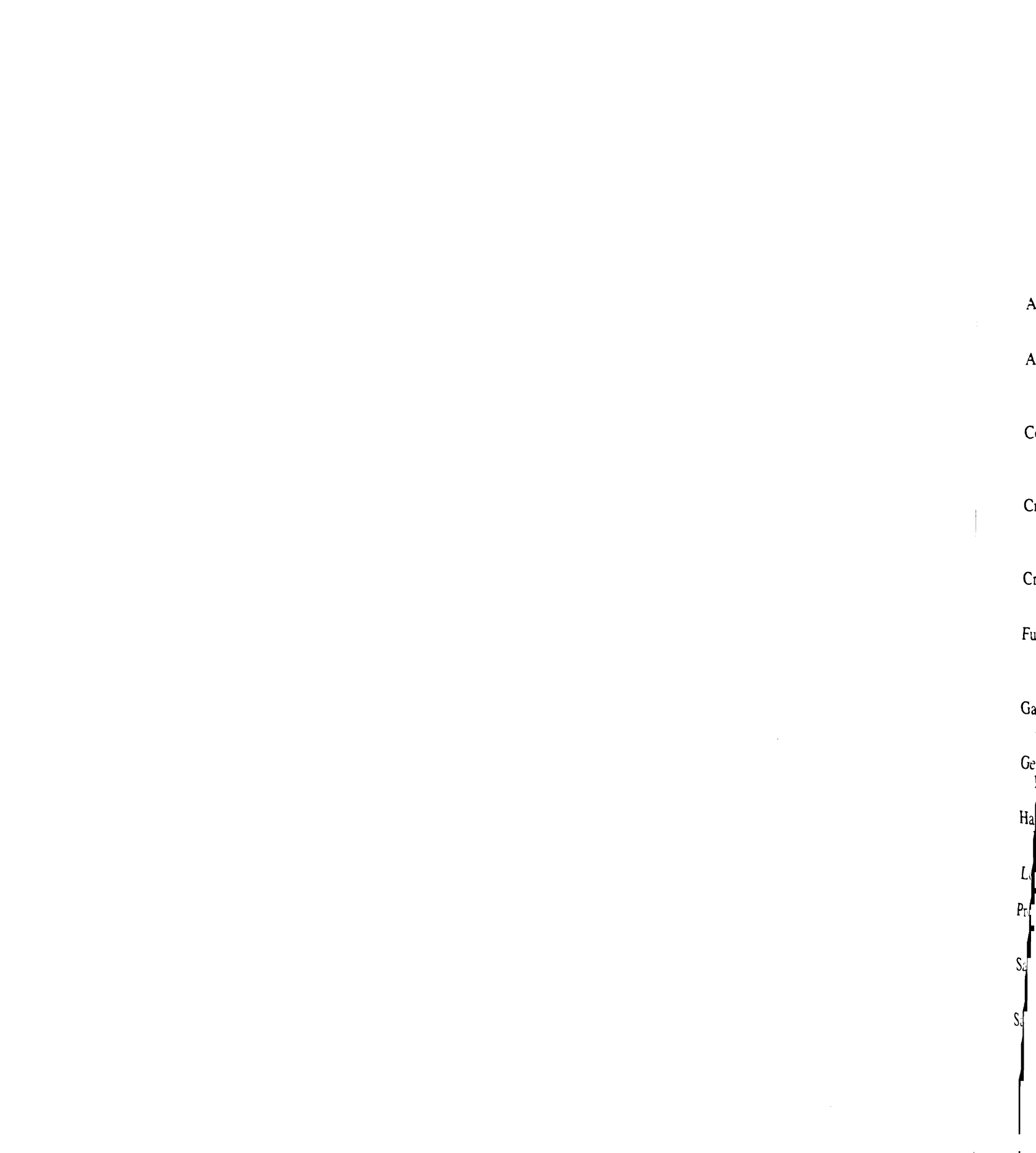

R. pom 1 AGACTATGCTAAATAAGTTGCTTATTCTTTTAGTAAAATAAGAGAATTCA 50
D. mel 1 AGACTATGCTAATTAAGTTGCTTAT.....AAAT....TTTTATA 36
51 AGCATATGGTATATTATTGGATTTTTGTTTTATATTTATATAAAATTTAA 100
37 AGCATATGGTATATTATTGG.....ATAAATATAATAATTTTTA 75
101 TCCATAATGTTAATAGCAAAAAAAGAAATAAATAAAATTTCTTGAAATCC 150
76 TTCATAATATT.....AAAAAATAAATGAAAAACATT.....AT.C 110
151 TCTCAATGAAATCTTATAAAAAAGAATCTCAGTATTCCCTTAAGACTTCA 200
111 TCACATTTGAATGT..GAAAAACGAAGAGAAATATT.....TTC. 147
201 GCATTATTTGAATTAATATTTAAATAATATATATGAGGAGGAATGTCTAG 250
148 ....TTTTTCAA.....TCAAATAATA.....CTGAGAAATGTCTAG 180
251 CAT.AAAAATTG....ATTTTTATTCTAGAATTACCTC.TATTAA..ATA 292
181 CATAAAAAATTGAAATATTTTTCATCTAGAATTGTCTCTTATTAATGATT 230
293 TATTTATATTAATATCCGGATAATGTTGATTTTTTGCGTATATATATATA 342
231 CGGAAATAGAAAAATCTTGGTTATGTTATTATTCTTCGT.....TG 271
343 TATATATAATAATTATTTTATTAATAAAAAACGGGATGAAAAGGTTCTTT 392
272 GTTCGTTAAAAA....TGGATAAATAAAAACTTTGCATACAAGAAT.... 313
393 TTTTCTAATACGATAAAATATT.TCAAGAAATATATTATTTCTTTGAACA 441
314 .....TAATA....AAAATGTTATAACGAATTTAATTAAATGTTTTATCA 354
442 AAGTAAAAAAAAAAATTATATATTTAAAATAATA 475
355 TTATATATAAAGAATTTAT...GGCAAGATAAAG 385

Figure 19B

Future Direction

The construction of a genomic library for *R. pomonella* has been successful. The primary characterization of the positive clones indicate that the inserts are indeed rDNA fragments. The ITS sequences from one of the clones further confirmed the presence of rDNA in that positive clone. The large insert size of the clones suggests that it is very possible that the complete repeating rDNA unit has been obtained, at least in some of the clones if not all. However, since this is only a preliminary investigation of the rDNA of *R. pomonella*, more detailed characterization needs to be done in order to determine which clone has which part of the rDNA. Accordingly, the desirable clones may be further subcloned into plasmid vectors, allowing one to sequence the IGS/ETS boundary region, together with the whole ETS , with a primer matching the 5' end of 18S. The sequence information obtained for the transcription initiation region, together with other available sequence information such as those of *Drosophila* and *Glossina* for the same region, will help to design the primers for PCR amplifying the ETS region of *Rhagoletis* flies and further conduct phylogenetic analysis of closely related species or host races in the genus *Rhagoletis*.

BIBLIOGRAPHY FOR CHAPTER IV



Bibliography for Chapter IV

Arnheim, N. 1983. Concerted evolution of multigene families. Pp. 38-61 *in* M. Nei and R. K. Koehn, eds. Evolution of genes and proteins. Sinauer, Sunderland, Mass.

Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl. 1987. Current protocols in molecular biology. John Wiley and Sons, New York.

Coen, E., T. Strachan, and G. Dover. 1982. Dynamics of concerted evolution of ribosomal DNA and histone gene families in the *melanogaster* species subgroup of *Drosophila*. J. Mol. Biol. **158**:17-35.

Cross, N. C. P., and G. A. Dover. 1987a. A novel arrangement of sequence elements surrounding the rDNA promoter and its spacer duplications in tsetse species. J. Mol. Biol. **195**:63-74.

Cross, N. C. P., and G. A. Dover. 1987b. Tsetse fly rDNA: an analysis of structure and sequence. Nucleic Acids Res. **15**:15-30.

Furlong, J. C., and B. E. H. Maden. 1983. Patterns of major divergence between internal transcribed spacers of ribosomal DNA in *Xenopus borealis* and *Xenopus laevis*, and of minimal divergence within ribosomal coding regions. EMBO J. **2**:443-448.

Gale, K., and J. Crampton. 1989. The ribosomal genes of the mosquito, *Aedes aegypti*. Eur. J. Biochem. **185**:311-317.

Gerbi, S. A. 1985. Evolution of ribosomal DNA. Pp. 419-517 *in* R. J. MacIntyre, ed. Molecular evolutionary genetics. Plenum Press, New York.

Hadjiolov, A. A. 1985. The nucleolus and ribosome biogenesis. Springer Verlag, New York.

Long, E. O., and I. B. Dawid. 1980. Annu. Rev. Biochem. **49**: 727-764.

Procunier, W. S., and J. J. Smith. 1993. Localization of ribosomal DNA in *Rhagoletis pomonella* (Diptera: Tephritidae) by *in situ* hybridization. Insect Mol. Biol. 2:163-174.

Sambrook, J., E. Fritsch, and T. Maniatis. 1989. Molecular cloning: a laboratory manual, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **12**:5463-5467.

Schlotterer, C., M. Hauser, A. von Haeseler, and D. Tautz. 1994. Comparative evolutionary analysis of rDNA ITS regions in *Drosophila*. Mol. Biol. Evol. **11**:513-522.

Sollner-Webb, B., and J. Tower. 1986. Transcription of cloned eukaryotic ribosomal RNA genes. Annu. Rev. Biochem. **55**:801-830.

Tautz, D., C. Tautz, D. Webb, and G. A. Dover. 1987. Evolutionary divergence of promoters and spacers in the rDNA family of four *Drosophila* species: implications for molecular coevolution in multigene families. J. Mol. Biol. **195**:525-542.

Tautz, D., J. M. Hancock, D. A. Webb, C. Tautz, and G. A. Dover. 1988. Complete sequence of the rDNA genes of *Drosophila melanogaster*. Mol. Biol. Evol. **5**: 366-376.

Dover, G. A., and R. B. Flavell. 1984. Molecular coevolution: DNA divergence and the maintenance of function. Cell **38**:622-623.

Williams, S. M., R. DeSalle, and C. Strobeck. 1985. Homogenization of geographical variants at the nontranscribed spacer of rDNA in *Drosophila mercatorum*. Mol. Biol. Evol. **2**:338-346.

CHAPTER V

PARTIAL CHARACTERIZATION OF THE EXTERNAL TRANSCRIBED SPACER REGIONS OF *RHAGOLETIS CINGULATA*

Introduction

Insect sibling species are difficult to distinguish morphologically. This is even more often the case with the genus *Rhagoletis* since many species have the ability to rapidly shift hosts from native wild plants to introduced cultivated crops. As a consequence, serious pest species with little or no morphological difference can arise in a relatively short period of time under sympatric conditions. Taxonomist and systematists are often challenged to seek alternative markers rather than morphological characters to detect and differentiate such species and to infer their phylogenetic relationship. Accurate identification of such sibling species is essential for appropriate management of these pest species. Furthermore, the correct assessment of genetic differences among the sibling species could be useful in our understanding of their evolutionary history, which in itself is necessary for testing various evolutionary theories regarding speciation.

In the past few years, as a candidate of alternatives to morphological method, nuclear rDNA spacers, especially the ITS sequences, have been applied to differentiate morphologically-indistinguishable species (Porter and Collins 1991) or subspecies (Volger and DeSalle 1994) and infer phylogenetic relationship among closely related species (Pleyte et al. 1992; Wesson et al. 1992; Schlotterer et al. 1994). In chapter II, I carried out a similar study on the *R. cingulata* species group, in which there are few morphological differences in the four North American species. However, except for a few potentially useful molecular markers to differentiate some of the species, the ITS sequences seem not

to display sufficient interspecific variations to confidently infer a phylogenetic relationship for this group. Moreover, between the two cherry-infesting species (*R. cingulata* and *R. indifferens*) only one potential molecular marker (position 371 in ITS2) existed in the whole ITS regions. Since several other species groups in *Rhagoletis*, such as the *pomonella* and *tabellaria* groups, also consist of a number of morphologically hard-to-distinguish sibling species (Chapter I), alternative methods need to be devised to distinguish them accurately.

In order to find a suitable gene marker for distinguishing sibling species and/or host races in *Rhagoletis* and for inferring the phylogenetic relationship of closely related species in this group, I explored the same non-coding regions of rDNA as in Chapter IV, namely the ETS and IGS (see Figure 13 in Chapter IV), but using a different approach. Because the IGS/ETS region is flanked by highly conserved rDNA coding regions, I took advantage of published sequences of distantly related organisms to design primers for PCR amplification of this highly variable region in *Rhagoletis*. Previous characterization of the IGS/ETS region in eukaryotic organisms were exclusively performed by the long drawn out procedures involving genomic library construction and screening. For the purpose of identifying species and phylogenetic analysis, which usually requires the estimation of intra- and inter-individual variations based on a large sample size of the same species, a simpler technique using PCR to obtain the DNA fragment(s) of interest must be developed.

The objectives of this study were 1) to design suitable primers, according to published sequences from other organism, which matches the 3' end of 28S and 5' end of 18S of *Rhagoletis* rDNA (i.e., flanking the IGS/ETS region); 2) to optimize conditions for amplifying the rDNA spacers of interest in *Rhagoletis*; 3) to clone the PCR products; 4) partially sequence the clones from both ends and verify that the appropriate region was obtained by comparing with published sequences of other organisms; and 5) to assess the value of the ETS region for phylogenetic reconstruction and species recognition.

I obtained two types of clones from PCR amplification products of *R. cingulata* flies on different host plants; approx. 2.6kb fragment from a fly on sour cherry and 3.9kb fragment from a fly on wild black cherry. The two types of clones were partially sequenced from each end and both types of clones appear to have the ends matching the 3' end of 28S and 5' end of 18S, indicating that both types of clones may contain the IGS and ETS of *R. cingulata.* However, the sequences of the two type clones are surprisingly diverged, even at some regions within 3' end of the 28S gene. Whether the observed divergence is related to their different host origins remains to be determined.

Materials and Methods

Biological Material

R. cingulata larvae were collected from two different field infested host plants, *Prunus cerasus* (sour cherry; from Hart, MI) and *Prunus serotina* (black cherry; from Roselake MI) during 1988-90. Larvae were allowed to pupate in fine moist vermiculite. Pupae were sifted from the vermiculite and subsequently stored at 4°C. After at least 5 months, pupae were removed from the cold and held at 22°C, under a 15:9 hr light:dark cycle to terminate diapause. Within 2-7 days after emergence most adult flies were frozen at -70°C for subsequent genomic DNA isolation. Specimens from each collection were pinned for species identification.

DNA Isolation and Amplification

Total genomic DNA was isolated from a single female *R. cingulata* fly reared from black cherries and a pool of 30 *R. cingulata* flies reared from sour cherries respectively as described by Procunier and Smith (1993). In addition, pDm238 (provided by Dr. G. A. Dover), which contains the complete *D. melanogaster* rDNA repeat unit, was used as a positive control for the PCR amplification reactions and also as a test for the accuracy of the sequencing procedures in this study. The region between the 28S and 18S subunits of

rDNA, which includes the IGS and ETS sequences, was amplified using the following primers:

GB16 5'GTTTAGACCGTCGTGAGACAGGTTA (matching the 3' end of 28S),

GB4 5'AGACATGCATGGCTTAATCTTTGAG (matching the 5' end of 18S), or

GB4' 5'ACAAGCATATAACTACTGGCAGGAT (matching the 5' end of 18S).

The sequences for the primers GB4 and GB16 were chosen after aligning the 18S and 28S regions of the several different organisms (see section below; Figures. 2 and 3a) and searching for highly conserved regions in the 5' end of 18S and 3' end of 28S genes, respectively. Amplification by the polymerase chain reaction (PCR) was carried out in 25 µl (final volume) containing 33.5 mM Tris-HCl, pH 8.8, 8.3 mM NH_4SO_4, 3.35 mM $MgCl_2$, 3.35 µM EDTA, 15.0 µM mercaptoethanol, 10% DMSO (Fluka), 160 µg/ml BSA (Boehringer Mannheim), 1.25 mM of each dNTP, 0.1-0.4 µM primer GB16 and 0.1-0.4 µM primer GB4, and 2.0 units of Ampli Taq DNA polymerase (Perkin Elmer Cetus). Amplification parameters were 95°C for 7 min; 35 cycles each with 94°C for 1 min, 60°C for 1 min and 65°C for 8 min; and 65°C for 15 min. Amplified DNA was subjected to electrophoresis on a 0.8% agarose gel and visualized with ethidium bromide. Bands containing the DNA were excised from the gel and the DNA purified using the Prep-A-Gene DNA purification matrix (Bio-Rad), according to the manufacturer's instructions.

Cloning and Sequencing

The TA Cloning kit (Invitrogen) was used in a one-step cloning strategy for the direct insertion of the purified PCR products (see earlier sections) into a plasmid vector, followed by transformation into competent cells. Plasmid vector and competent cells were supplied by the manufacturer. Clones were randomly selected and stored as frozen stock at -80°C. Plasmid DNA was purified from individual clones using the Magic-Prep DNA purification kit (Promega), following the manufacturer's instructions. DNA sequencing was performed according to the chain-termination method of Sanger et al. (1977), and

using the Sequenase Version 2.0 DNA sequencing kit (USB) and ^{35}S-dATP (Amersham). Vector primers NB7 5'AATACGACTCACTATAG and NB11 5'GTCATAGCTGTTTCCTG were employed to determine the DNA sequence.

Computer Software and Analyses

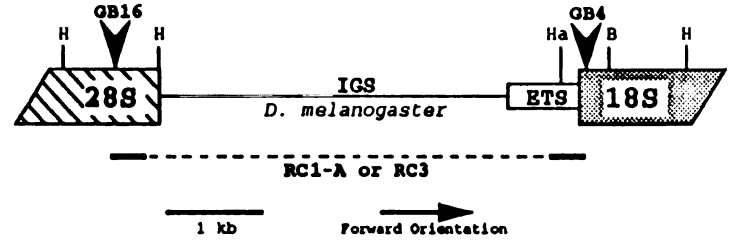
Several programs, such as FETCH, LINEUP, PILEUP, MAPSORT, SEQED and STRINGSEARCH, in the GCG software package of the University of Wisconsin Genetics Computer Group (UWGCG package, version 8.0) were used in this study. The 28S, 18S, IGS and ETS regions of rDNA from *Aedes albopictus* (mosquito; mqsrnagn.gb_in), *Daphnia pulex* (water flea; daprrnaspa:gb_in), *Drosophila melanogaster* (fruit fly; pdm238.gb_in), *Glossina morsitans* (tsetse fly; gmedna1.gb_in); *Lytechinus variegatus* (sea urchin); *Rattus norvegicus* (rat; rnrgm4a.em_ro), *Saccharomyces cerevisiae* (yeast); *Tenebrio molitor* (beetle; tmrn18s.gb_in); and *Xenopus laevis* (frog; xlrn01.gb_ov) were found using STRINGSEARCH. Alignments were performed using first LINEUP and then PILEUP, with gap weight set at 1.00 and gap weight length set at 0.10. Restriction enzyme sites on the sequences were located using MAPSORT.

Results and Discussion

At the start of this study, two PCR primers, named GB4 and GB16, were designed according to the highly conserved 18S (5'end) and 28S (3' end) regions of several distantly related organisms such as frog and water flea (Figure 20; Materials and Methods). Because the negative control with GB4 alone gave nonspecific PCR products with λ DNA (data not shown), another primer, named GB4', adjacent to GB4 was also designed and used concurrently. With the genomic DNA of pooled *R. cingulata* from sour cherry, a major band of approx. 2.6 kb and a slow migrating minor band were observed after amplification with GB4' and GB16 (Figure 21). Control amplification reactions using each primer separately showed faint low molecular weight bands (Figure 21; lanes 1 and 2) the origin

Figure 20. Designing the appropriate primers for PCR amplification of the IGS/ETS regions of *Rhagoletis*. Panel a shows a map of the region in question. Refer to legend of Figure 13 (Chapter IV) for notations and abbreviations. The partially sequenced regions of the two *R. cingulata* clones (RC1-A and RC3) are indicated by the solid line; regions not sequenced yet are denoted by the dashed line. Panel b shows the alignment of a 3' end region of the 28S gene from various organisms, including the two *R. cingulata* clones from this study. Panel c shows the alignment of the 5' end of the 18S gene from various organisms, including the two *R. cingulata* clones from this study. Nucleotide positions are numbered according to the start of coding region of the respective genes of *Drosophila melanogaster*. The PCR primers are underlined.

A



B

3' of 28S

```
+3458     +3467      +3477      +3487
.         .          .          .
GTTTAGACCG TCGTGAGACA GGTTAGTTTT ACCCTACT    Drosophila melanogaster (fruit fly)
---------- ---------- ---------- --------    Aedes albopictus        (mosquito)
---------- ---------- ---------- --------    Xenopus laevis          (frog)
---------- ---------- ---------- --------    Daphnia pulex           (water flea)
---------- ---------- ---------- --------    Rattus norvegicus       (rat)
---------- ---------- ---------- --------    Rhagoletis cingulata    (RC1-A from black cherry)
---------- ---------- ---------- --------    Rhagoletis cingulata    (RC3 from sour cherry)
```

GB16 Primer

C

5' end of 18S

```
+1        +10         +20         +30         +40         +50         +60
.         .           .           .           .           .           .
ATTCTGGTTG ATCCTGCCAG TAGTTATATG CTTGTCTCAA AGATTAAGCC ATGCATGTCT AA    Drosophila melanogaster (fruit fly)
---------- ---------- ---------- ---------- ---------- ---------- --    Glossina morsitans      (tsetse fly)
TCC------- ---------- ----C----- ---------- ---------- ---------- C-    Lytechinus variegatus   (sea urchin)
TAC------- ---------- ----C----- ---------- ---------- ---------- TC    Tenebrio molitor        (beetle)
TAC------- ---------- ---.C----- ---------- ---------- -----C--G- --    Xenopus laevis          (frog)
TAC------- ---------- ---.C----- ---------- ---------- ---------- --    Daphnia pulex           (water flea)
TA-------- ---------- ----C----- ---------- ---------- ---------- --    Saccharomyces cerevisiae(yeast)
TAC------- ---------- ----C----- ---------- ---------- ----------       Rhagoletis cingulata    (RC1-A)
TA-------- ---------- ---------- -----                                  Rhagoletis cingulata    (RC3)
```

GB4' Primer

GB4 Primer

Figure 20

Figure 21. Products of PCR amplification of the IGS/ETS region of *Rhagoletis cingulata* from sour cherry. Lanes 1 and 2 represent products from amplification using only GB4' and GB16 primers, respectively. Lane 3 represents the major PCR product (approx. 2.6 kb band) obtained after amplification using GB4' and GB16 together. Lanes M and M' represent the *Hin*dIII-*Pst*I double digest of λ DNA and 123 bp DNA ladder as molecular size markers.

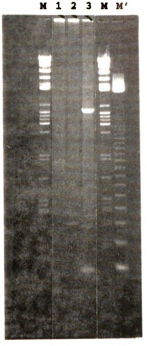


Figure 21

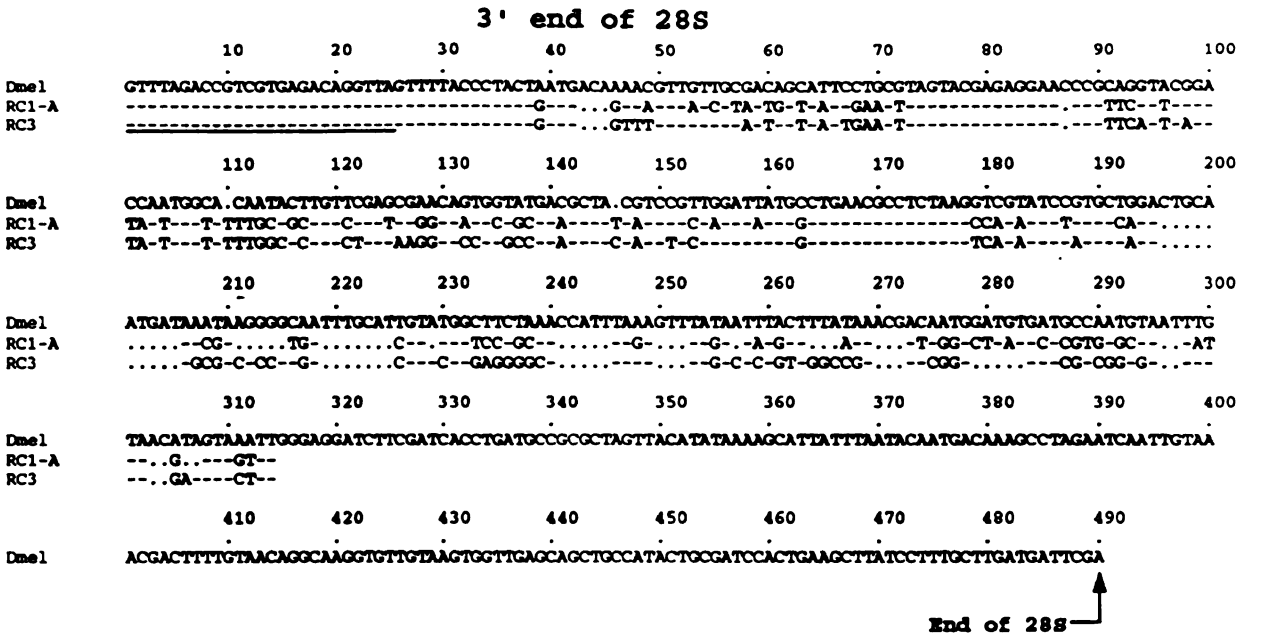
of which may be the result of the amplification condition used in this study. Using GB4 and GB16, one major band of approx. 3.9 kb was obtained for the single fly DNA preparation (*R. cingulata* from black cherry); in addition, using the same two primers, a major band of approximately 5 kb was obtained from pDm238 (data not shown).

After preparative scale amplifications, purification of the major DNA bands and cloning, seven clones from the single fly DNA preparation, named RC1-A through RC1-G, two clones from the pooled *R. cingulata* DNA preparation, named RC2 and RC3, and one clone from Pdm238, named Dm12, were randomly selected. The seven clones from the single fly DNA preparation were partially sequenced from both ends using the vector primers (see Materials and Methods). More than 200 nts were sequenced from both ends of each clone. Two of the seven clones (RC1-A and -B) had identical sequences (only the sequence of RC1-A is presented in Figure 22). Each of the other five clones, however, had identical sequences at both of their ends (the sequence matched the 5' end of 18S corresponding with GB4; data not shown). The latter results were surprising and may be explained as follows: 1) they may be an artifact generated as a result of the amplification step or 2) they may reflect the presence of high numbers of genetic inversions within the rDNA sequences, in particular within regions including the 5' end of 18S. To my knowledge, there appears to be no precedent for such high numbers of genetic inversions within rDNA sequences in other organisms; however, this possibility cannot be ruled out at this time. More detailed characterization and investigation was necessary to resolve this perplexing set of results and deemed, at the present time, to be beyond the scope of this study.

One of the two clones from the pooled DNA preparation (RC3) and the Dm12 from PDm 238 were also partially sequenced from both ends. The Dm12 sequences were identical with published data for Pdm238 (Tautz et al. 1987; 1988). The sequences from RC1-A and RC3 were aligned, using PILEUP, with the relevant *D. melanogaster* regions (Figure 22; Panels A and B). Several patches of well conserved regions with

Figure 22. Alignment of the sequenced regions from the two *Rhagoletis cingulata* clones RC1-A and RC3 with homologous regions from *Drosophila melanogaster*. Alignments were performed using PILEUP with gap weight = 1.00 and gap weight length = 0.10. Panel a shows the sequences downstream from the GB16 primer (underlined sequence) position (i.e., the 3' end of the 28S. Panel b shows the sequences upstream from the GB4 and GB4' primer (underlined sequences) positions (i.e., the 5' end of the 18S). Conserved nucleotides and gaps are denoted by dashes and dots, respectively. Unidentified nucleotides are denoted by a question mark. Arrows denote the start and end nucleotides of the 18S and 28S coding regions, respectively.

A


3' end of 28S
Dmel
RC1-A
RC3
End of 28S

B

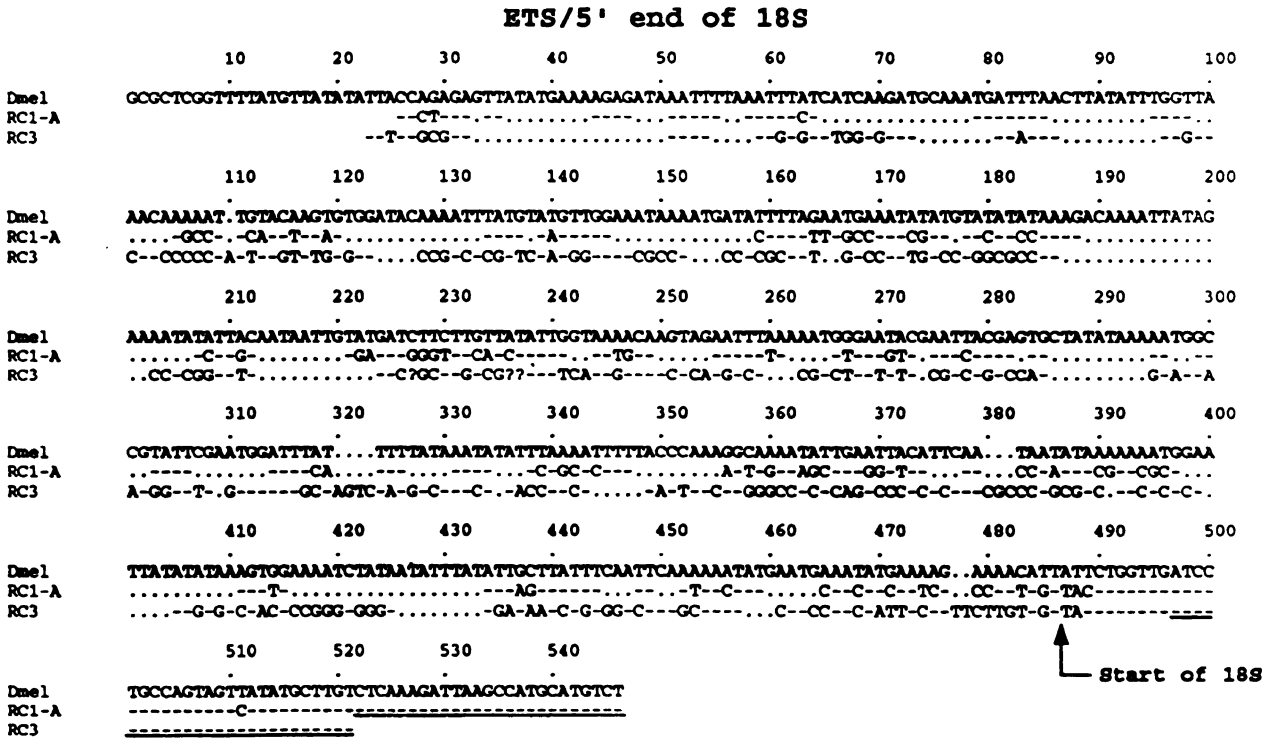

ETS/5' end of 18S
Dmel
RC1-A
RC3
Start of 18S

Figure 22

D. melanogaster were observed immediately downstream of the GB16 primer, such as positions 26-42, 71-90 and 160-178 (Figure 22A). Actually these three patches were even conserved among other organisms such as *Aedes albopictus*, *Xenopus laevis*, *Daphnia pulex* and *Rattus norvegicus* (Figure 20A; and other data not shown). Other well conserved patches were observed immediately upstream of the GB4 and GB4' primers (positions 490-521 for RC1-A and 489-496 for RC3; Figure 22B). The T to C transition detected at position 511 was also observed among other organisms (Figure 20C position 25). Therefore, the conservation observed at both ends of the inserts from clones RC1-A and RC3, clearly indicates that at least one of the two clones, if not both, contains the complete *R. cingulata* IGS and ETS regions.

Surprisingly, however, beyond the highly conserved regions mentioned above, the partially sequenced regions of the two *R. cingulata* clones appear to be considerably diverged — even the 3' end of the 28S gene (Figure 22; Panels A and B). This was unexpected given that these sequences are from the same species although from different host plants. Compared with their sequence divergence in the ITS regions (about 0.2 percent nucleotide substitutions), the level of genetic variation between the two different host-associated flies of *R. cingulata* seemed to be excessively high in those partially sequenced regions. However, I detected considerable variation in the 3' end of the 28S (downstream of GB16) among other organisms whose sequences were from the GCG database (e.g., *Drosophila, Aedes, Xenopus, Daphnia, Rattus*; data not shown). Indeed, the presence of variable regions at the 3' end of the 28S gene (also called expansion segments) has been reported for humans and *Drosophila* (Gonzalez et al. 1990; Linares et al. 1991).

Expansion segments are absent in prokaryotic rRNA genes but are known to vary considerably in length and sequence between different eukaryotic organisms (Linares et al. 1991). This fact has prompted the use of expansion segments for the resolution of phylogenetic relationships between closely related species (Gonzalez et al. 1990). It is also

noteworthy that the two *R. cingulata* clones were different in length; RC1-A and RC3 were about 3.9kb and 2.6kb long, respectively. Length variation in the IGS is probably due to different number of subrepeats in IGS. Large amount of length variation in several IGS clones from *D. melanogaster* have been reported and the variations arose from varying numbers of an internal repeat of about 250 bp long (Coen et al. 1982; Simeone et al. 1982). Similar length differences were also noted for *D. mercatorum*; individual flies from different geographic regions have characteristic length for the IGS which range from 4.0 kb to 6.5 kb (Williams et al. 1985). Williams et al. (1985) proposed that those length difference patterns may serve as markers for determining the geographic origin of individuals. In addition, a Y chromosome-linked length variant in the IGS is also present in *D. mercatorum* (Williams et al. 1985). Therefore, it is not yet clear whether the heterogeneity in sequence and length between these two *R. cingulata* clones is an artifact or a reflection of inherent variation between host-associated populations.

The *R. pomonella* clones described in Chapter IV may help resolve this puzzle since some of the clones, as mentioned in Chapter IV, very likely have the complete rDNA repeating unit of *R. pomonella* . Upon subcloning the DNA fragments which contain the appropriate *R. pomonella* IGS-ETS region, a direct comparison of this region with homologous regions from the RC1-A and/or RC3 could be undertaken. Therefore, in the future, the 3' end of 28S as well as the IGS/ETS boundary region could be useful regions for the resolution of the phylogenetic relationships between the host-associated populations of *R. cingulata.*

BIBLIOGRAPHY FOR CHAPTER V

Bibliography for Chapter V

Coen, E., T. Strachan, and G. Dover. 1982. Dynamics of concerted evolution of ribosomal DNA and histone gene families in the *melanogaster* species subgroup of *Drosophila*. J. Mol. Biol. **158**:17-35.

Gonzalez, I. L., C. Chambers, J. L. Gorski, D. Stambolian, R. D. Schmickel, and J. E. Sylvester. 1990. Sequence and structure correlation of human ribosomal transcribed spacers. J. Mol. Biol. **212**:27-35.

Linares, A. R., 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**:381-390.

Pleyte, K. A., S. D. Duncan, and R. B. Philips. 1992. Evolutionary relationships of the salmonid fish genus Salvelinus inferred from DNA sequences of the first internal transcribed spacer (ITS1) of ribosomal DNA. Mol. Phylogenet. Evol. **1**:223-230.

Porter, C. H., and F. H. Collins. 1991. Species-diagnostic differences in a ribosomal DNA internal transcribed spacer from the sibling species *Anopheles freeborni* and *Anopheles hermsi* (Diptera: Culicidae). Am. J. Trop. Med. Hyg. **45**:271-279.

Procunier, W. S., and J. J. Smith. 1993. Localization of ribosomal DNA in *Rhagoletis pomonella* (Diptera: Tephritidae) by *in situ* hybridization. Insect Mol. Biol. **2**:163-174.

Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **12**:5463-5467.

Schlotterer, C., M. Hauser, A. von Haeseler, and D. Tautz. 1994. Comparative evolutionary analysis of rDNA ITS regions in *Drosophila*. Mol. Biol. Evol. **11**:513-522.

Simeone, A., A. de Falco, G. Mancino, and E. Boncinelli. 1982. Sequence organization of the ribosomal spacer of D. melanogaster. Nucleic Acids Res. **10**:8263-8272.

Tautz, D., C. Tautz, D. Webb, and G. A. Dover. 1987. Evolutionary divergence of promoters and spacers in the rDNA family of four *Drosophila* species: implications for molecular coevolution in multigene families. J. Mol. Biol. **195**:525-542.

Tautz, D., J. M. Hancock, D. A. Webb, C. Tautz, and G. A. Dover. 1988. Complete sequences of the rRNA genes of Drosophila melanogaster. Mol. Biol. Evol. **5**:366-376.

Vogler, A. P., and R. DeSalle. 1994. Evolution and phylogenetic information content of the ITS-1 region in the tiger beetle *Cicindela dorsalis*. Mol. Biol. Evol. **11**:393-405.

Wesson, D. M., C. H. Porter, and F. H. Collins. 1992. Sequence and secondary structure comparisons of ITS rDNA in mosquitoes (Diptera: Culicidae). Mol. Phylogenet. Evol. 1:253-269.

Williams, S. M., R. DeSalle, and C. Strobeck. 1985. Homogenization of geographical variants at the nontranscribed spacer of rDNA in *Drosophila mercatorum*. Mol. Biol. Evol. 2:338-346.

CHAPTER VI

CONCLUSIONS

In this dissertation, I have presented the complete ITS sequences of the 4 members of the *cingulata* species group as well as that of *R. pomonella*. The *Rhagoletis* ITS sequences are highly A-T rich like those from *Drosophila* and beetles. I found low levels of interspecific ITS variation in the *cingulata* species group, implying that ITS sequences are of limited application in phylogenetic analysis of host-associated populations and/or closely related sibling species. Between the two cherry infesting species (*R. cingulata* and *R. indifferens*) there is only one potential nucleotide position that could differentiate them. However, a few molecular markers have been described, which can be potentially useful for distinguishing between the two olive infesting species themselves (*R. osmanthi* and *R. chionanthi*) and differentiate them from the two cherry flies species. The high sequence divergence found between the *cingulata* group and *R. pomonella* enabled me to use the ITS regions for analyzing the phylogenetic relationship of taxa from different *Rhagoletis* species groups. The overall computer generated secondary-structure models of the ITS sequences were presented for the *cingulata* species group and *R. pomonella*. Several highly conserved secondary-structural elements were determined by FOLDRNA and comparative analysis. One such element (I2R4) seems to be conserved even among two distant families, Tephritidae and Drosophilidae, which diverged in the Cretaceous period between 65 to 130 million years ago implying an important functional role for this structural-element in the processing of precursor rRNA. The conserved ITS secondary-structural elements have been helpful guide for increasing the accuracy of aligning regions based on homology

rather than similarity which might be a consequence either of common ancestry or of chance.

In addition, I have sequenced the ITS regions of 7 other *Rhagoletis* species, which include *cornivora, completa, juniperina, fausta, electromorpha, basiola*, and *striatella*. Phylogenetic implications from this study are in partial agreement with some of the previous studies on *Rhagoletis* phylogeny which use other methods, and partially contradictory to these previous results. This study indicates that *R. cornivora* belongs to the *pomonella* group, supporting the placement based on morphology and karyotype similarities but is not in agreement with the mtDNA analysis; *R. juniperina* was removed from the *tabellaria* group, in support of allozyme and mtDNA analyses, and this species may be more closely related to the *pomonella* group rather than the *cingulata* group. This result is in agreement with the mtDNA data but not with the allozyme analysis. With respect to the relatives of the *cingulata* group members of the *suavis* group rather than those in the *pomonella* group appear closest, which is congruent with the mtDNA study, but not with the allozyme results; *R. fausta* may be related with the *tabellaria* group, which had been left unplaced in previous studies; *R. basiola* and *R. striatella* are most divergent in the ITS sequences, correlating with their possession of several ancient morphological characters and their basal branching in the analyses of mtDNA and allozyme data and further supporting the idea that *Rhagoletis* may not be monophyletic. The application of ITS sequences in the phylogenetic study of *Rhagoletis* has provided some insight into the relationship between certain taxa from different species groups of this genus.

This is the first study of rDNA ITS sequences of the genus *Rhagoletis*. The study of the ITS should be expanded in the genus *Rhagoletis* which could open a new avenue for the understanding of the evolutionary history of this genus and providing a reliable phylogenetic framework to test some important evolutionary theories of speciation through host shifting. In addition, the mode of ITS evolutionary divergence should be useful in future investigations of the structure, function and processing of precursor rRNA. Future

studies on the ITS of other *Rhagoletis* species will allow one to elucidate the degree of functional and structural constraints on the ITS sequences in *Rhagoletis*.

Besides the studies on the ITS regions, I have also conducted preliminary investigations on other noncoding regions of *Rhagoletis* rDNA. The construction of a genomic library for *R. pomonella* has been successful. The primary characterization of the positive rDNA clones were substantiated by sequencing the ITS regions in these clones. The large insert size of the clones suggests that it is very possible that the complete repeating rDNA unit has been obtained, at lease in some of the clones if not all. However, since this is only a preliminary investigation of the rDNA of *R. pomonella*, more detailed characterization needs to be done in order to determine which clone has which part of the rDNA. Accordingly, the desirable clones may be further subcloned into plasmid vectors, allowing one to sequence the IGS/ETS boundary region, together with the whole ETS , with a primer matching the 5' end of 18S. The sequence information obtained for the transcription initiation region, together with other available sequence information such as those of *Drosophila* and *Glossina* for the same region, will help in the design of primers for PCR amplification of the ETS region of *Rhagoletis* flies, and further aid in the phylogenetic analysis of closely related species and host races in the genus *Rhagoletis*.

I also attempted to PCR amplify the IGS/ETS region from two *R. cingulata* flies of different host plants. Two types of clones were obtained; one with an approx. 2.6kb insert from the fly on sour cherry and the other with an approx. 3.9kb insert from the fly on wild black cherry. The two types of clones were partially sequenced from each end and both types of clones appear to have the ends matching the 3' end of 28S and 5' end of 18S, indicating that both types of clones may contain the IGS and ETS of *R. cingulata*. However, the sequences of the two type clones are surprisingly diverged, even at some regions within 3' end of the 28S gene. Whether the observed divergence is related to their different host origins remains to be determined.