This is to certify that the

dissertation entitled

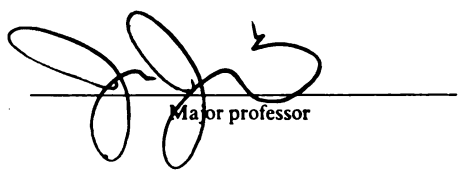
Documenting the history of speciation in Rhagoletis zephyria (Diptera: Tephritidae) using molecular markers

presented by

Vesna Gavrilovic

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Zoology



Major professor

Date 05/10/2001

DOCUMENTING THE HISTORY OF SPECIATION IN *Rhagoletis zephyria* (Diptera: Tephritidae) USING MOLECULAR MARKERS

By

Vesna Gavrilovic

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Department of Zoology and
Program in Ecology, Evolutionary Biology and Behavior

2001

ABSTRACT

DOCUMENTING THE HISTORY OF SPECIATION IN *Rhagoletis zephyria* (Diptera: Tephritidae) USING MOLECULAR MARKERS

By

Vesna Gavrilovic

Rhagoletis pomonella (apple maggot) and *R. zephyria* (snowberry fly) belong to the *R. pomonella* species group of the true fruit flies (Tephritidae) which are the center of a long-lasting debate about modes of speciation. Speciation in this group appears to have been occurring sympatrically via host-plant shifts, in the absence of the geographic barriers traditionally thought to be necessary for genetic divergence of populations and establishment of reproductive isolation. The distribution of *R. pomonella* and *R. zephyria* has been described as parapatric, while all the other species of the *pomonella* group are broadly sympatric. In addition, while most of the species in the group show only allele frequency differences for electrophoretic loci, there is a fixed difference at the *Had* locus (hydroxyacid dehydrogenase) in *R. zephyria* (fixed for Had^{111}) and *R. pomonella* (two other alleles – Had^{100} and Had^{125}). Therefore the relationship between these two species provides a contrast to other species pairs in the group. This project investigates whether *R. zephyria* has diverged by mechanisms similar to the ones that lead to divergence of other taxa in the *pomonella* group. To address this question, the phylogenetic position of *R. zephyria* within the group is studied using sequences of nuclear and mitochondrial genes. Patterns of genetic variation in both apple maggot and snowberry flies throughout

the geographic ranges of both species are characterized not only using gene sequence data, but also using amplified fragment length polymorphism (AFLP) patterns, in order to infer polarity and time of the host shift that presumably occurred causing speciation and infer genetic mechanisms involved in species divergence. *Rhagoletis pomonella* appears to represent a large and variable gene pool from which new species arise. The zone of overlap of native geographic ranges of *R. zephyria* and *R. pomonella* is shown to be much larger than previously described. Host shift that led to divergence of *R. zephyria* from *R. pomonella* and speciation appears to have occurred from ancestral hawthorn to snowberry. Speciation of *R. zephyria* appears to be very recent, as indicated by the unresolved phylogeographic relationships, incomplete lineage sorting and lack of fixed differences at all loci whose sequences were studied. The amount of genetic variation found in *R. zephyria* was large, indicating that speciation was not accompanied by bottlenecks. Little geographic structuring was revealed by analyses of anonymous nuclear loci, mitochondrial COI/COII gene and AFLP patterns. All analyses show that *R. zephyria* appears to be a native inhabitant of eastern North America. Some AFLP fragments were only amplified either in *R. zephyria* or *R. pomonella*. These fragments can potentially be used as diagnostic markers for distinguishing between these two species.

ACKNOWLEDGEMENTS

Numerous people have provided guidance, help and support in various ways throughout the years I was working on this dissertation. Without them the project would have been much weaker and the whole graduate school experience much less fun. Dr. Jim Smith, my advisor, provided knowledgeable and professional guidance, support and continuous understanding and a simple "Thanks" is just not enough to express how much I appreciate it. It is difficult to find adequate words of gratitude to Dr. Guy Bush for his continuous intellectual guidance and financial support. Thanks to Dr. Rich Lenski and Dr. Jim Hancock for the time and effort they invested in serving on my Graduate Committee.

Special thanks are due to present and past members of Smith and Bush or Bush and Smith lab, who were always extremely supportive and provided technical help, valuable discussions and suggestions: Dr. George Weiblen, Angela Roles, Joe Crossno, Dr. Vanderlei Martins, Matt Jaycox, Vince Borla, Chhaya Patel, Martha Smith Bressan Caldas, Jessica Wolf, Elizabeth Kruszewski and Danielle Christman.

Amplification and sequencing of the nuclear alleles would not be possible without Joe Roethele (University of Notre Dame) who constructed the genetic linkage map of *Rhagoletis pomonella*, designed primers for the nuclear loci, helped tremendously in cloning and sequencing, trained me to do sequencing myself, initiated valuable discussions and always had friendly words of understanding and support. I would like to thank Dr. Jeff Feder who kindly allowed me to use the primers designed in his laboratory and do a large part of the work there. Help and support of members of his lab - Uwe

Stolz, Dr. Bill Perry, Hattie Dambroski and Dr. Ken Filchak are highly appreciated. Thanks also to Nancy Perry for providing the warmth of home one winter when I worked at Notre Dame.

For direct or indirect help with collecting the samples throughout the country I would like to acknowledge Dr. Guy Bush, Dr. Jim Smith, Joe Crossno, Dr. Mark Wiedrlechner (Iowa State University), Mike Klaus (Washington Department of Agriculture), Dr. Diane Olsten (Utah State University), Charles Linn (Agriculture Experimental Station, Geneva, NY), Dr. Ron Prokopy (University of Massachusetts, Amherst) and Dr. Timothy Dickinson (Royal Ontario Museum, Toronto, ON). For help with sequencing and AFLP GeneScanning thanks go to Sue Soltzfus, Therese Best, Sherry Tjugum-Holland and Barbara Christian at the MSU DNA Sequencing Facility.

When data analyses seemed to be going to a dead end, Dr. Ron DeBry (University of Cincinnati) and Dr. Susan Masta (San Francisco State University) provided advice that helped me put things back on track. Their help is greatly appreciated.

Branislav Blagojevic was cheering for me from the very beginning, waiting impatiently for a long time on the other side of the country, making sure that I did many things in an unorthodox way, at least time-wise. Including the 3am phone calls which I will surely miss, but gladly trade for conversations in person.

It would be very difficult to get through all the challenges without the support of my family, even from across the ocean. Here I say thanks to Vladimir Gavrilovic, Ana and Dragan Gavrilovic and Jelena Todic.

Friends whose encouragement and sanity keeping are priceless include Sandra Bencic, Dr. Danny Rozen, Jelena and Goran Poprzen, Natasa Nestorovic, Igor Pastirk,

Sanela Lampa Pastirk, Eva Maria Muecke, Alexandra Burch, Hilschers – Klara, Roman and Ondrej, Silvia Fransen and Katarina Martic. And then also, thanks to Micaela Szykman, Russ Van Horn and Ed Siuda for all the kind words and support.

Encouragement from Maxygen, Inc. that came from Dr. Lori Giver, Dr. Steve delCardayre, Cristina Ivy, Troy Obrero, Walker Lutringer, Fred Gaume and many others contributed greatly to my motivation to finish the project and finish it quickly, for which I am very grateful to all of them and to Dr. Danny Rozen for helping me start that network.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO SYMBOLS AND ABBREVIATIONS

aap (ptx) – *Rhagoletis pomonella*, Waxahatchie, Texas

bas – *R. basiola*

bop (pga) - *R. pomonella*, Macon, Georgia

c – *R. cornivora*

cng – *R. cingulata*

com – *R. completa*

eL (electrom) – *R. electromorpha*

f – *R.* nr. *mendax* (flowering dogwood fly)

fa (nr. mendax a) - *R.* nr. *mendax* (flowering dogwood fly), adult

fp (nr.mendax p) - *R.* nr. *mendax* (flowering dogwood fly), pupa

fu – *R. fausta*

ind – *R. indiferens*

jbt – *R. juniperina*

m – *R. mendax*

mel – *R. mendax*, East Lansing, Michigan

mga – *R. mendax*, Georgia

mmi - *R. mendax*, East Lansing, Michigan

mnj – *R. mendax*, Rutgers, New Jersey

mns - *R. mendax*, Nova Scotia

mon - *R. mendax*, Ontario

mOtis – *R. mendax*, Otis Lake, Michigan

mxH – *R.* nr. *pomonella*, Mexico (highland)

mxhd – *R.* nr. *pomonella*, La Jolla, Mexico

mxL – *R.* nr. *pomonella*, Mexico (lowland)

NC (pmiNC) – *R.* nr. *pomonella*, Grant, Michigan

nr. mendax GA - *R.* nr. *mendax* (flowering dogwood fly), Georgia

nr. pom MX - *R.* nr. *pomonella*, Mexico

nwmx – *R. pomonella*, New Mexico

p - *R. pomonella*, Grant, Michigan

pco - – *R. pomonella*, Boulder, Colorado

pel - *R. pomonella*, East Lansing, Michigan

pga (bop) - *R. pomonella*, Macon, Georgia

pia - *R. pomonella*, Ames, Iowa

pil - *R. pomonella*, Riverwoods, Illinois

pma - *R. pomonella*, Amherst, Massachusetts

pmhbj - *R.* nr. *pomonella*, Louisiana (mayhaw fly)

pmi - *R. pomonella*, Grant, Michigan

pmiNC (NC) - *R. pomonella*, Grant, Michigan

pmn - *R. pomonella*, Staples, Minnesota

pne - *R. pomonella*, Nebraska

pns - *R. pomonella*, Kentville, Nova Scotia

pny - *R. pomonella*, Geneva, New York

pon - *R. pomonella*, Toronto, Ontario

ppa - *R. pomonella*, Biglerville, Pennsylvania

ptx (aap) - *R. pomonella*, Waxahatchie, Texas

put - *R. pomonella*, Wellsville, Utah

pwa - *R. pomonella*, St. Cloud Ranch, Washington

rpmx - *R.* nr. *pomonella*, Mexico City, Mexico

sparkga – *R.* nr. *mendax*, Georgia (sparkleberry fly)

str – *R. striatella*

sv – *R. suavis*

t – *R. tabellaria*, Washington

tzt – *R. tabellaria*, Ontario

zca – *R. zephyria*, Honeydew, California

zco – *R. zephyria*, Boulder, Colorado

zel – *R. zephyria*, East Lansing, Michigan

zid - – *R. zephyria*, Elmira, Idaho

zma – *R. zephyria*, Amherst, Massachusetts

zmio – *R. zephyria*, Mio, Michigan

zmn – *R. zephyria*, Hawby, Minnesota

zmt – *R. zephyria*, Swan Lake, Montana

zmtB – *R. zephyria*, Terry, Montana

znd – *R. zephyria*, Bismarck, North Dakota

zne – *R. zephyria*, Brady, Nebraska

zny – *R. zephyria*, Geneva, New York

zon – *R. zephyria*, Rice Lake, Ontario

zor – *R. zephyria*, Grants Pass, Oregon

zpa – *R. zephyria*, College Park, Pennsylvania

zsb – *R. zephyria*, Glen Haven (Sleeping Bear Dunes), Michigan

zsd – *R. zephyria*, Custer, South Dakota

zwa – *R. zephyria*, Dixie, Washington

zwi – *R. zephyria*, Waukesha, Wisconsin

zwy – *R. zephyria*, Moiser Gulch, Wyoming

In all trees, green branches represent *R. zephyria*, red *R. pomonella*, pink *R.* nr. *pomonella* (Mexico), blue *R. mendax* and light-blue *R.* nr. *mendax* (flowering dogwood fly). Squares next to *R. zephyria* labels and circles next to *R. pomonella* labels correspond to geographic regions from which the samples were taken: red – Northeast, orange – Great Lakes, green – Great Plains, blue – Rocky Mountains and Colorado Plateau, violet (purple) – Northwest, and gray – South.

CHAPTER 1

INTRODUCTION

Tephritid fruit flies in the genus *Rhagoletis* have been a model system for studying sympatric speciation via host-plant shifts with no geographic separation of populations. Some of the species in this genus (*suavis* species group, Bush and Smith 1998) conform to the traditional view that new species arise when genetic differences between populations accumulate as a result of geographic barriers to gene flow (Mayr 1963). However, species of the *pomonella* species group represent one of the examples of incipient sympatric speciation, not uncommon in phytophagous insects (Mitter et al. 1991, Guldemond and Mackenzie 1994, Emelianov et al. 1995) and some vertebrates (Schliewen et al. 1994, Sturmbauer 1998, Hatfield and Schluter 1999). *Rhagoletis pomonella* group species are also one of the best-understood examples where ecology has played an important role in speciation (Orr and Smith 1998).

SPECIATION IN THE ABSENCE OF GEOGRAPHIC BARRIERS

Species and speciation are the center of the long-standing debate focused on explaining how gene pools become split and isolated, giving rise to new species. Many biologists still hold the view that this is possible only if physical barriers to gene flow between populations exist, allowing accumulation of different mutations, different selection pressures and stochastic processes to shape gene complexes within these

populations in such a way that, if they come into secondary contact, they will no longer be able to produce viable and fertile offspring. If they accept the view that speciation can occur without physical barriers to gene flow, they often say it happens only rarely (e.g Barraclough and Vogler 2000).

Parapatric divergence in a classic sense may be initiated in a small region at the periphery of a widely distributed species. The new ecological "race" becomes adapted to a new "niche" and, once established, spreads to occupy the range of the new habitat or host (after Bush 1975). It is unlikely that a shift and adaptation to the new niche occurs simultaneously along the entire ecotone. Recently, mosaic distribution of populations throughout the species range has been referred as to parapetric distribution of populations (for example, Gavrilets et al 2000). Models of speciation under this definition of parapatry assume that geographic variation exists between populations of a widely distributed species and some gene flow occurs between neighboring populations. It is intuitive that the geographic ranges of most species are much larger than the dispersal distances of their individuals or gametes. Even though any particular mutation is a rare event, the number of possible mutations is almost unlimited and therefore different mutations are likely to accumulate in different populations of species throughout the range. Different habitat conditions, both biotic and abiotic, are likely to exist in different parts of the range, creating different selection regimes, which can contribute to divergence of the populations despite some gene flow. The main questions here are how fast reproductive isolation can develop and where does the first split occur (Gavrilets et al 2000). According to Mayr (1963), peripheral populations are more likely to diverge because they are usually smaller, experience different selection pressures than the central

populations and are less affected by gene flow. Brown (1957), on the other hand, in his concept of centrifugal speciation, argues that the central populations represent the origin of new species because they are often the source of new genetic variability. In a recently developed model, Gavrilets et al (2000) have shown that both time until speciation and location of the first split depend on mutation rate, amount of genetic change required for reproductive isolation and population size. One of the important implications of this model is that it does not require strong divergent selection for rapid (few hundred to few thousand generations) speciation. Experimental study of premating reproductive mechanisms in grasshoppers (Tregenza et al 2000) corroborated that the long periods of allopatry are not necessary for speciation.

At the opposite end to allopatric speciation in the geographic continuum is speciation in sympatry. Key elements of sympatric speciation are resource-based disruptive selection, assortative mating and habitat (host) fidelity (Johnson et al 1996, Futuyma 1997, Berlocher 1998, Schluter 1998). Habitat choice behavior is also attributed the status of a key component in generating conditions under which reproductive isolation between sympatric populations can evolve (Via et al 2000), because behavioral changes often initiate the use of a new environment, allowing selection to operate on morphological and physiological characters (Futuyma and Moreno 1988, Fry 1996). Earlier models (Felsenstein 1981) involved non-habitat-choice assortative mating and divergent selection. In these models, recombination easily separates alleles for assortative mating from alleles for performance in two habitats, and little or no divergence is possible. Felsenstein (1981) therefore concluded that sympatric speciation is only likely under highly restricted conditions. Models developed later (Rausher 1984) included loci

for habitat choice and demonstrated that habitat preference can develop easily even in absence of fitness differences across environments. Therefore characters that influence both specialized resource use (habitat preference) and assortative mating are particularly important in evolution of ecological specialization and speciation. The significance of ecological specialization in speciation has been widely recognized and documented for a number of species (Bush 1994, Johnson et al 1996, Schluter 1996).

Experimental corroboration of this model comes from work of Via on pea aphids. In several studies her group has shown that under experimental conditions two host races of pea aphids have poor performance and fitness on alternative hosts simply because of unwillingness to feed on that host, not because of physiological trade-off. Several other studies have shown that trade-offs are not essential for ecological specialization (Fry 1996, Kawecki 1996, Whitlock 1996). Caillaud and Via (2000) suggest that the physiological trade-off (toxicity of the alternative host) initially might have been the selective force favoring host plant acceptance, but it no longer plays that role. In host races of *Rhagoletis pomonella*, no evidence of larval feeding trade-off have been found (Prokopy et al 1988) – larvae of both races performed equally in hawthorn fruits, yet they are highly specialized for different temporal resources (Feder et al 1997, Feder and Filchak 1999, Filchak et al 1999).

Under the influence of natural selection, populations colonizing novel environments can diverge very rapidly from the ancestral populations, which is a critical early step in speciation (Coyne 1992, Travisano and Rainey 2000). Speciation by ecological specialization is most rapid when assortative mating in different environments is coupled with habitat preference (Johnson et al 1996). In the initial diversification phase

small amount of divergence occurs; this phase is followed by the quasi-equilibrium phase during which little divergence occurs and completion phase when gene flow between habitats stops quickly (in less than 1000 generations) and divergence is dramatic (Johnson et al 1996). Fast speciation in sympatry can also occur when traits for habitat choice, or traits correlated with them, are the direct target of disruptive selection (Kondrashov and Kondrashov 1999). Reproductive isolation can develop in the presence of limited gene flow via pleiotropy when divergent selection acts on multiple characters (Rice and Hostert 1993). When the effect of selection is larger than gene flow, some divergence will result; selection will then tend to further decrease gene flow, enabling divergence of traits that initially were not diverging (Rice and Salt 1990).

It is usually not easy to estimate how fast reproductive isolation evolves, especially in long-standing groups, since averaging rates over time obscures the biologically important short-term evolution (Hendry and Kinnison 1999). The most spectacular speciation rate has been estimated for cichlid fishes in lakes of Eastern Africa (Danley et al 2000, Wilson et al 2000). Several factors contribute to rapid diversification despite some gene flow in these lakes: philopatry, lack of dispersing larval stage, and varying water levels which affects extinction and reestablishment of populations (Danley et al 2000, Meyer 1993). A number of experimental studies with known timescales have shown that reproductive isolation can evolve very rapidly - examples of rapid divergence include limnetic and benthic forms of sticklebacks (Schluter 1996), evolution of soapberry bug populations on alternative hosts (Carroll et al 1997), morphological differentiation of *Anolis* lizards introduced to Bahaman Islands (Losos et al 1997), changes in life history traits and morphology in Neotropical guppies (Reznick et al 1997),

and morphological and genetic divergence between river and lake populations of salmon (Hendry et al 2000). In all these cases, diverging populations were adapting to local resources and genetic differences were small, especially in neutral markers, but isolation developed in as little as 13 generations (Hendry et al 2000). Accompanying morphological and ecological changes were large enough to prevent cross mating, which can further facilitate divergence and led to speciation in sympatry.

Ecological interactions, rather than ecological specialization, are the starting point in models of sympatric speciation with an evolutionary branching approach (Doebeli and Dieckmann 2000). These models are based on the theory of adaptive dynamics (Metz et al 1992, Geritz et al 1998). In these models, ecological interactions such as resource competition, predator-prey interactions or mutualism are the driving force of evolutionary change and branching. A number of empirical examples provide evidence for sympatric speciation under the evolutionary branching models. In sticklebacks (Schluter 1994, Nagel and Schluter 1998), *Anolis* lizards (Losos et al 1998) and Darwin's finches (Grant et al 1985, Schluter 1988) branching is attributed to resource competition; however, it is not clear that competition is an important factor in insect evolution. In orchid and orchid bees coevolution and cospeciation are initiated by mutualistic interactions (Kiester et al 1984), whereas speciation in sub-Antarctic weevils illustrates the significance of predator-prey interactions (Chown and Smith 1993).

Berlocher (1998) recognizes four stages in the process of sympatric speciation: host races, species isolated by host fidelity, species with prezygotic and/or postzygotic isolation unrelated to host fidelity, and totally isolated species. Taxa at all of the stages are found within the *Rhagoletis pomonella* species group. Host races of *R. pomonella*

mate on fruits of different host plants and display no other pre- or postzygotic reproductive isolation; however, gene flow between them is reduced by strong host fidelity and selection. Genetic differences between the host races are small; in the case of apple and hawthorn races of *R. pomonella* there are no unique allozyme alleles but frequency differences between races are maintained by host choice behavior and selection despite the 6% gene flow between them (Feder and Bush 1989). Species isolated by host fidelity in the *R. pomonella* species group (*R. pomonella* and "flowering dogwood fly", Berlocher 1998) have larger allele frequency differences but still no unique alleles have been identified. Hybridization between these species under laboratory conditions results in fully viable and fertile offspring (Smith 1986). In nature, host choice of these species is even more pronounced than in host races and this strong host fidelity prevents them from interbreeding (Berlocher et al 1993). Species with prezygotic and/or postzygotic reproductive isolation unrelated to host fidelity show non-fixed species-specific allozyme alleles (private alleles) and morphological autapomorphies (Feder and Bush 1989, Jenkins 1996). Low levels of gene flow are possible (Feder and Bush 1989), but the assortative mating occurs under laboratory conditions, even in the absence of host plants (Bierbaum and Bush 1990), and in nature where large host preference differences prevent interbreeding (Feder and Bush 1989). Hybrids have reduced viability but are fully fertile (Bierbaum and Bush 1990). Within the *R. pomonella* species group, this stage is illustrated by *R. mendax* (blueberry maggot) and *R. pomonella* species pair (Berlocher 1998). Totally isolated species, such as *R. cornivora* and *R. pomonella* are separated by large genetic distance and exhibit fixed allozyme differences at several loci (Berlocher et al 1993), as well as morphological polymorphisms (Jenkins 1996).

USE OF MOLECULAR MARKERS IN MAKING INFERENCES ABOUT SPECIATION

Information about speciation can be inferred from analyses of the geographic ranges of species, correlations between range overlaps and the degree of genetic divergences, and by a biogeographic analyses of allele or haplotype genealogies (Berlocher 1998). Phylogeography brings an historical perspective to population genetic phenomena such as population subdivision, genetic drift, gene flow and selection (Brown et al. 1996), and allows estimates to be made about the order and polarity of ancestor-descendant relationships and ecological niches (Berlocher 1998). Existing patterns of allele relationships depend on the time since speciation and the mode of speciation.

Deep gene trees with major lineages separated by relatively large mutational distances are usually observed when long-term, usually physical barriers to gene flow separate the populations (Avise 2000). The observed genetic distances may arise as a result of novel mutation accumulation in lineages after their geographic separation or as a result of lineage sorting from a polymorphic ancestral gene pool; these two cases are not mutually exclusive. However, the same results could occur following sympatric divergence. Deep splits between distinct mtDNA lineages have been observed in contiguous populations of black-backed jackal (Wayne et al 1990). The authors suggest that this finding may be the result of secondary admixture of populations that were separated in the past or retained distinct ancestral lineages despite high gene flow.

Recent sympatric speciation is reflected by alleles of ancestral species forming a paraphyletic group, with alleles of derived species arising from within the ancestral clade (Harrison 1991). Generally, if populations have been reproductively isolated for a

relatively short time, little differentiation is expected to be observed in neutral genes, leading to unresolved relationships between the taxa, since parental genes are still being shared among derived groups. The ancestral species is expected to be more geographically structured than the descendant species, since populations of the ancestral species have been adapting to local habitat conditions for a longer time (Brown et al 1996). For mitochondrial DNA, Neigel and Avise (1986) have shown by computer simulations that, following speciation, relationships among haplotypes go through phases of poly-, para- and monophyly and that the time to reaching monophyly depends on population sizes at and after speciation.

It should be noted that the general expectations about lineage sorting and coalescence are only applicable to selectively neutral alleles/haplotypes. Different types of natural selection will change the expectations in different ways. When directional selection drives the frequency of an advantageous allele, it eliminates the variation at linked sites by a selective sweep or hitchhiking effect (Stephan et al 1992). New variation is recovered very slowly by accumulation of new neutral mutations in copies of advantageous alleles. This shortens the coalescent time for the gene and linked loci, depending on the strength of linkage (Li 1997).

Selection against deleterious mutations (background selection, Charlesworth et al 1993) also reduces neutral variation at linked sites, reducing the coalescent time at that region as well. The effect of background selection is strongest when the mutation rate is high and the recombination rate is low; it decreases rapidly with the increase in recombination rate (Li 1997). Balancing selection, on the other hand, can slow down the

elimination (extinction) of alleles at a locus and linked sites, maintaining some lineages over time scales much longer than expected and across speciation events (Clark 1997).

Polymorphisms can also be misleading in the estimation of patterns and time of speciation when trans-species polymorphisms exist (Klein 1986). In such a case, the divergence of allelic lineages occurrs before the divergence of species, so that polymorphisms observed within species are older than the species itself, as illustrated by the highly polymorphic human MHC loci (Klein et al 1993).

The type of molecular markers chosen for phylogeographic studies thus largely influences the conclusions drawn from gene genealogies. Although the importance of the marker choice is not questioned, the majority of the studies published are based on only one type of molecular marker (for exceptions see Bernardi et al. 1993, Burton and Lee 1994, Hare and Avise 1998).

About 70% of all phylogeographic studies are based primarily or exclusively on analyses of mitochondrial DNA (Avise 2000), which is a maternally inherited, haploid marker. The mitochondrial genome is relatively small in size (16 kb in *Drosophila*) and present in thousands of copies in each individual. mtDNA does not undergo recombination, shows stable gene arrangement, ample polymorphism within species, and often evolves faster than typical single-copy nuclear DNA (Avise 1994). Isolation of haplotypes is relatively easy through PCR. Trees based on mitochondrial DNA sequences reflect only a matrilineal portion of the evolutionary history (Maddison 1995), which might be quite different from the history of nuclear genes.

Nuclear markers, depending on their nature (single-copy genes, repeated short sequences, size polymorphisms) and position in the genome (in regions presumably under

selection vs. neutral), can reveal different patterns of variation and evolution. Nuclear markers often used for studying genetic structure and relationships of populations and species include allozymes, RFLPs (restriction fragment length polymorphism, reviewed in Avise 1994), microsatelites (Weber and May 1989), VNTRs (variable number of tandem repeats or minisatellites, Jeffreys et al 1985) and AFLPs (amplified fragment length polymorphism, Voss et al 1995). These markers do not require DNA sequencing and allow estimates of population genetic parameters such as allele frequencies, numbers of alleles, heterozygosity, effective population size, N_em and $N_e\mu$ parameters, genetic distances among populations, and F or θ statistics (Wright 1965, Weir and Cockerham 1984).

The sequences of single-copy nuclear genes are more difficult to obtain and few studies have used them for estimation of intra- or interspecific gene trees and phylogeographic analyses (Hare and Avise 1998). In diploid organisms, if the individual is heterozygous at a locus under study, PCR amplifies both alleles and it is important to separate them and obtain both sequences. This is possible by cloning of the amplified PCR products through a vector. However, misincorporation of nucleotides during the previously performed PCR is possible (Keohavong and Thilly 1989). Some authors ignore possible misincorporations in their analyses (Palumbi and Baker 1994, Sota and Vogler 2001). Another approach to this problem involves sequencing multiple clones from a single individual so that the sequences can be compared and allelic variants distinguished (Bernardi et al 1993). Most phylogenetic information is recovered from gene trees obtained from loci in regions with low recombination frequencies. Recombination distorts the phylogenetic history of changes within and between lineages

by producing homoplasy, which can erroneously cluster the genotypes and violate the assumption that tree branches are non-reticulate (Avise 2000).

Estimates divergence times between species are based on the concept of a "molecular clock" (Hillis et al 1996). This concept is based on the simplifying assumptions that DNA sequences evolve at an approximately uniform rate over time in all evolutionary lineages (Zuckerkandl and Pauling 1965) and that the rate of neutral mutation equals the rate of evolution (Kimura 1968). For mtDNA it is estimated that 1-2% sequence divergence accumulates over a million years (Brown et al 1979). However, this clock has only been calibrated using insect and crustacean sequences (Brower 1994). It is known that variation in rate of mtDNA evolution exists not only among taxa but also among different genes (Simon et al 1994). Differences among taxa may stem from different efficiencies of their DNA repair systems (Britten 1986), different generation times (Laird et al 1969) or different metabolic rates (Martin and Palumbi 1993). More accurate estimates of the rate of evolution are obtained if multiple loci are used and averaged, since that reduces the variance (Takahata and Nei 1985). If the time since lineage separation can be estimated from fossil evidence as well, than the clock can be calibrated more accurately. Unfortunately, fossil data are not available for most taxa, so the molecular clock can only be used for rough estimates.

RHAGOLETIS POMONELLA GROUP TAXA, LIFE HISTORY, HOST USE AND GEOGRAPHY

The *Rhagoletis pomonella* species group consists of 4 taxonomically described species, at least two undescribed species (sparkleberry fly and flowering dogwood fly) and several taxa of uncertain taxonomic status (host races, geographic races) (Table 1.1). All *Rhagoletis* taxa are typical solitary parasites in that females search for host fruits for their offspring, and oviposit only on acceptable fruits (Price 1977). *R. pomonella* group species have specific hosts that belong to widely divergent plant families (Table 1.1). All taxa within the *R. pomonella* species group have almost identical life cycles. They are univoltine, with adults emerging over short periods of time in spring or summer. Emergence of each species is synchronized with the beginning of the fruiting season of their respective host. Adults feed on insect honeydew, bird feces, yeast and bacteria found on leaf surfaces of host or non-host plants (Bush 1992). Females become sexually mature 7 or 8 days after emergence, at which time they begin mating and ovipositing. Mating occurs almost exclusively on the host fruit (Prokopy et al 1971). Females are attracted to a suitable host fruit by chemical (Prokopy et al 1973, Frey et al 1998) and visual cues (Moericke et al 1975), where they lay their eggs into the fruit by inserting their ovipositor into the fruit. Oviposition is followed by marking of the fruit with pheromones. That deters other females from ovipositing into the same fruit (Prokopy 1972). This probably reduces competition and perhaps enhances resource partitioning. Each female typically lays one egg per fruit; however, multiple infestations have been observed, particularly in large fruit such as apples (Filchak, pers. comm.; Gavrilovic and Crossno, pers. observ.) and experimentally corroborated in *R. zephyria* (van Randen and

Roitberg 1996). Multiple ovipositions per single fruit were originally explained in parasitoids by mistakes in oviposition (van Lenteren 1981), but more recently were argued from a standpoint of adaptive superparasitism theory (van Alphen and Visser 1990). This theory states that when ecological or physiological conditions are unfavorable (e.g. no uninfested fruit, high egg load) and there is positive probability that an additional larva could survive feeding within the same fruit, females can increase their reproductive fitness component by accepting previously marked (i.e. infested) host fruit.

Larvae feed and pass through 3 instars within the fruit over a 2-4 week period (Bush 1966). Upon completion of larval development, which is synchronized with fruit falling to the ground, larvae leave the fruits and pupate 3-5 cm in the soil, enter diapause and overwinter in the pupal stage. The following summer adults emerge and complete the life cycle in 20-30 days. Host fidelity in all species of the genus *Rhagoletis* is very high and adults are known to disperse over large distances in search of favorable hosts when local host fruit is not available (Bush, 1966). Many of the taxa in the *pomonella* species group can be crossed in the laboratory (Bierbaum and Bush 1990); however, they rarely interbreed in nature (Feder and Bush 1989). This indicates that host preferences and host fidelity accompanied with mating on the fruits of the host plant serve as effective premating reproductive isolation mechanisms between taxa utilizing different host plants. Four species, two undescribed species close to *R. mendax* and two taxa of uncertain status close to *R. pomonella* belong to the *R. pomonella* species group (Table 1.1). The status of Mexican populations of *R. pomonella* is also uncertain – whether they represent geographic race of *R. pomonella* or a separate species remain under investigation (Bush, pers. communication; Feder et al, in prep.)

Rhagoletis pomonella is native to eastern North America, from Nova Scotia and Maine southward to central Florida and westward to northwestern Minnesota and eastern Texas (Figure 1.1). Populations present in the Pacific Northwest and western slopes of the Rocky Mountains are thought to be introduced (McPheron 1990). Ranges of all other taxa, except for *R. zephyria*, are fully contained within the much broader range of *R. pomonella* (see Bush and Smith 1998). *Rhagoletis zephyria* has been described as occurring throughout western North America and is parapatric in Minnesota and Manitoba with *R. pomonella* (Bush 1966). However, *R. zephyria* has recenty been reared from native snowberries in Ontario (Smith, pers. comm.), Michigan (Crossno, pers. comm.) and southeastern Wisconsin (Gavrilovic, pers. observation, see also Chapter 2), indicating that the zone of overlap is much broader (see Figure 1.1). Populations of *R. zephyria* also exist in eastern North America, where they are thought to be introduced from western North America with the host plant, *Symphoricarpos albus* var. *laevigatus*, that is grown as an ornamental shrub (Feder et al 1999).

Table 1.1. Taxa belonging to the *Rhagoletis pomonella* species group and their host plants.

TAXA	HOST PLANTS	FAMILY
Rhagoletis pomonella	hawthorn (*Crataegus* 15 spp.) apple (*Malus pumilla*)	Rosaceae
R. zephyria	snowberry (*Symphoricarpos* 3 spp.)	Caprifoliaceae
R. mendax	blueberry (*Vaccinium* 7 spp.) huckleberry (*Gaylusaccia* 3 spp.)	Ericaceae
R. cornivora	shrubby dogwoods (*Cornus* 2 spp.)	Cornaceae
R. nr. *mendax* (sparkleberry fly)	sparkleberry (*Vaccinium arboreum*)	Ericaceae
R. nr. *mendax* (flowering dogwood fly)	flowering dogwood (*Cornus florida*)	Cornaceae
R. nr. *pomonella* (plum fly)	plums (*Prunus* 2 spp.)	Rosaceae
R. nr. *pomonella* (mayhaw fly)	mayhaw (*Crataegus* 2 spp.)	Rosaceae

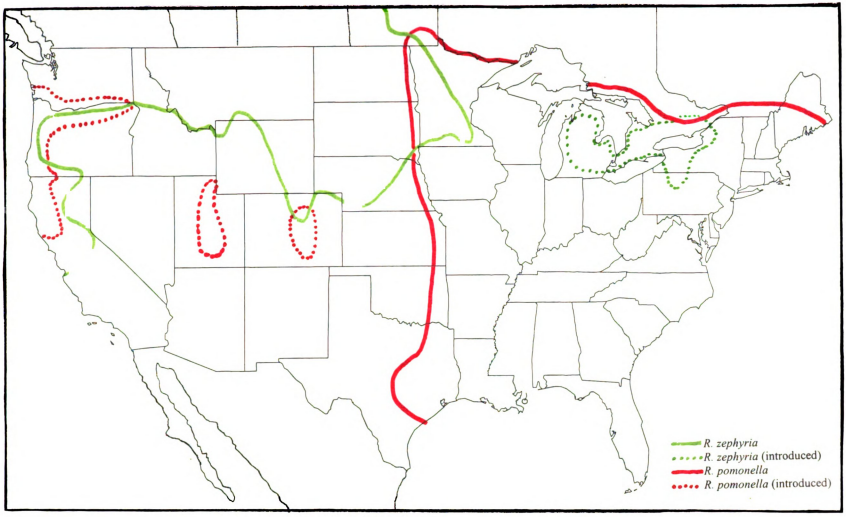


Figure 1.1. Geographic distribution of *Rhagoletis zephyria* and *R. pomonella*. Images in this dissertation are presented in color.

Except for *R. cornivora*, species of the *pomonella* group are morphologically virtually indistinguishable. In the case of *R. pomonella* and *R. zephyria*, only males can be distinguished on the basis of small difference in the shape of surstyli (Bush 1966, Jenkins 1996). All *R. pomonella* group taxa show the same distinctive wing banding pattern, that resembles the legs of jumping spider when observed from above.

Genetic variation at various levels and evolutionary relationships within the *pomonella* group have been examined by allozymes (Berlocher and Bush 1982, Feder et al. 1988, McPheron et al. 1988, McPheron 1990, Berlocher et al. 1993, Berlocher 1995, Berlocher 2000) and mitochondrial DNA sequences (Smith and Bush 1997, Han and McPheron 1997, McPheron and Han 1997). Most taxa of the group show only allozyme allele frequency differences (Berlocher et al. 1993). Six diagnostic loci show constant host-related frequency differences between apple and haw races of *R. pomonella* (Feder et al 1988, McPheron et al 1988), as well as among other taxa within the group (Berlocher et al 1993). *Rhagoletis pomonella* and *R. mendax* possess private alleles at polymorphic loci aspartate-aminotransferase-1 (*Aat*-1^{100} in *R. pomonella*), diaphorase-2 (*Dia*-2^{100} in *R. pomonella*) and fumarase (*Fum*^{158} in *R. mendax*), with other alleles at these loci being shared among species. However, there is a nearly fixed allele difference between apple maggot and snowberry fly at the hydroxyacid dehydrogenase (*Had*) locus – *R. zephyria* is fixed for *Had*^{111}, while *R. pomonella* has two other alleles, *Had*^{100} and *Had*^{125} (Berlocher et al. 1993). It should be noted that Feder et al. (1999) found very low frequencies of *Had*^{100} in populations of *R. zephyria* from Michigan and central Minnesota, and *Had*^{111} in populations of *R. pomonella* from Washington, northern and central Minnesota. The presence of these rare alleles is explained by low-level

hybridization or occasional mistakes in oviposition. It should be noted again that *R. cornivora*, the most genetically distant of the species within the *R. pomonella* group, along with the fore-mentioned morphological differences, possesses 9 unique allozymes alleles and appears to have diverged first from the remainder of the group (Berlocher et al 1993, Jenkins 1996).

Although much work has been done in order to characterize the genetic variability of, and the relationships between, the species comprising the *Rhagoletis pomonella* group (Berlocher et al. 1993, Smith and Bush 1997, Berlocher 2000), many questions remain open, from the full resolution of phylogenetic relationships to the taxonomic status of currently unplaced populations, and to testing the sympatric hypothesis of speciation. The knowledge gained through extensive research and deployment of new molecular markers raises unresolved questions, keeping the field vibrant and exciting for years to come. This dissertation focuses on the placement of *R. zephyria* and its relationship to *R. pomonella* (the putative ancestor for all the taxa in the *R. pomonella* species group), in order to infer the mode of speciation that led to divergence of these two species. Speciation via host shifts has been proposed for most of the broadly sympatric *pomonella* group taxa and it is interesting to determine whether *R. zephyria* has diverged by a similar mechanism. *Rhagoletis zephyria* is the only species in the group with distribution extending far outside of the range of *R. pomonella*. It is also the only species in the *pomonella* group with fixed allele difference at one of the allozyme loci (*Had*). That provides an important contrast to all other species pairs in the group.

SPECIFIC OBJECTIVES AND HYPOTHESES

Population genetic studies based on allozymes indicate that *R. pomonella* is the most genetically diverse (with highest number of alleles at any of the allozyme loci and highest heterozygosity) and most highly structured (as inferred from F_{st} values) of the *pomonella* group species (Berlocher et al. 1993, Berlocher 1995, Berlocher 2000). This has led to a working hypothesis that the apple maggot fly represents a large and variable ancestral gene pool from which new species of the *pomonella* group arise. Specific hypotheses tested in this project are:

Hypothesis 1: *Rhagoletis pomonella* represents a large and variable gene pool, containing ancestral polymorphisms, from which new species arise.

Hypothesis 2: *Rhagoletis zephyria* has diverged from *R. pomonella* more recently than the blueberry maggot, *R. mendax*.

Hypothesis 3: The host shift that led to divergence of *R. zephyria* from *R. pomonella* and speciation occurred from ancestral hawthorn host to snowberry.

Hypothesis 4: *Rhagoletis zephyria* in Eastern North America has been introduced with the host plant (snowberry).

The extent of the range of overlap between *R. pomonella* and *R. zephyria* has important implications for inferring mechanisms and geographic location of divergence that led to speciation. Berlocher (1998) suggests that the range of overlap for these two species, based on the previously described range of *R. zephyria*, is only about 15% whereas mean range of overlap for all host-shifting species he examined was about 50%. The size of overlap for apple maggot and snowberry flies is close to the mean range of

overlap found for non-host-shifters (12.4%), where speciation can be explained by allopatric models (Berlocher 1998). To be able to address any hypotheses of speciation of *R. zephyria* and its origin in eastern North America, I needed to obtain reliable geographic distribution data (Chapter 2).

To test the hypotheses 1-3 I used a phylogeographic approach. *Rhagoletis pomonella* and *R. zephyria* populations representative of the species geographic ranges were surveyed for variation using molecular markers. In order to estimate levels of genetic variation in *R. pomonella* and *R. zephyria* and compare them to other species of the *pomonella* species group, as well as other species of *Rhagoletis*, I used DNA sequences from three anonymous single-copy nuclear loci from three different linkage groups in the *R. pomonella* genome (P220 from linkage group I, P2956 from linkage group II and P2480 from linkage group III; Figure 3.1; Roethele et al. 1997), as well as mitochondrial DNA (cytochrome oxidase subunits I and II with intergenic $tRNA^{Leu}$ coding region; COI/COII) (Chapter 3). The identities of the nuclear loci used here are unknown. They display significant linkage disequilibrium with allozymes which display host-specific frequency differences between host races of *R. pomonella* (Roethele et al. 1997). The degree of linkage to the allozyme loci (as determined by levels of recombination, Roethele et al 1997) is different for these markers. Each of these loci reflects the evolutionary history of the distinct part of the genome. This study thus relies on a representative portion of a genome. As mentioned earlier, relationships between taxa inferred from gene trees obtained using one gene can differ dramatically from those obtained using a different gene or a different type of marker. Therefore gene trees may not accurately correspond to species trees (Maddison 1995, Wang et al 1997). To

overcome this problem, I used not only sequences from different genes, but also amplified fragment length polymorphism (AFLP) genotyping (Voss et al 1995). This technique has recently been suggested for studying relationships between very closely related and rapidly evolving species (Albertson et al 1999), for studying population structure and differentiation and estimating population genetics parameters (Reineke et al 1998). It generates a large quantity of information (polymorphic loci) by screening the entire genome. I used two selective amplification primers to generate AFLP patterns for individual flies from populations representative of the geographic ranges of *R. pomonella* and *R. zephyria* and used these data to study phylogeographic relationships of the populations of both species (Chapter 4).

If *R. pomonella* indeed represents a large, ancestral gene pool, it can be expected that it will show a higher number of alleles, higher average nucleotide diversity, higher heterozygosity, and higher F_{st} values across its geographic range than *R. zephyria.* It can also be expected that *R. pomonella* is paraphyletic, while other species of the group are expected to show autapomorphic genetic differences, with alleles arising from *pomonella* alleles. The polarity and time of the presumed host shift can also be inferred from allele genealogies. If the data indicate that the divergence of *R. zephyria* from the ancestral gene pool of *R. pomonella* was recent, then the polarity of host shift can be inferred to be from ancestral hawthorns to snowberry. Cladograms and distance-based dendrograms obtained from sequences of different genes and comparison of structure of *R. zephyria* populations across the geographic range of the species and to the structure of *R. pomonella* populations in the zone of overlap, should provide information on the origin of *R. zephyria* in eastern North America, where one host plant species has been introduced.

CHAPTER 2

GEOGRAPHIC DISTRIBUTION OF *R. ZEPHYRIA*

INTRODUCTION

The *Rhagoletis pomonella* species group is one of the best-understood natural model systems in which ecology is proposed to have played an important role in species divergence (Orr and Smith 1998). Several closely related species in the *R. pomonella* species group appear to have speciated in sympatry, with shifts to new hosts accompanying all speciation events. *Rhagoletis zephyria* (snowberry fly) is a sister species to *R. pomonella* (apple maggot). The geographic ranges of the *R. pomonella* group taxa, except for *R. zephyria*, are completely contained within that of *R. pomonella* (Bush 1966, Foote et al 1993). Speciation of *R. zephyria*, which infests fruits of *Symphoricarpos* spp., is an interesting evolutionary question. Not only has *R. zephyria* different geographic distribution (overlapping with *R. pomonella* and mostly allopatric to other *pomonella* group taxa), but also this species is the only one of the *R. pomonella* group taxa that displays fixed genetic differences with *R. pomonella*.

The working hypothesis about the origin of *R. zephyria* is that it has diverged and speciated by a host shift. The shift must have had occurred in the zone of contact between the ancestral and derived host, therefore in the zone of overlap of *R. zephyria* and its ancestor. Another assumption is that the ancestral host is hawthorn, and that the ancestral population was *R. pomonella*-like. It has been proposed that *R. pomonella* (apple maggot fly) has the ability to shift from the original (ancestral) host to fruits of different host

plants and utilize them as a new resource, thus giving rise to new species (as implied in Bush 1969). Host fidelity is an important factor in the speciation process, since adults of all *pomonella* group taxa mate exclusively on or near their host fruits (Bush 1966). This establishes and maintains premating reproductive isolation between the newly diverged taxa. Although *R. pomonella* and *R. zephyria* can be crossed under laboratory conditions, in nature that happens only rarely (Feder and Bush 1989, McPheron 1990, Smith et al 1993, Feder et al 1999).

To begin addressing the questions of how, when and where *R. zephyria* diverged from an *R. pomonella*-like ancestor, it is important to establish the current distribution of the two species. The geographic distribution of *R. zephyria* has not been well characterized, and therefore it is not well known where and how large the zone of parapatry is or has historically been. In addition, the host distribution of *R. zephyria* is also uncertain. A precise map of the present geographic distribution of host and fly, as well as evidence suggesting past distributions, are two of several variables that should be known to document the history of speciation of any two taxa (White 1978). Many of the other variables are already well documented for the *R. pomonella* species group taxa. These include detailed morphological descriptions of taxa (Bush 1966, Foote et al 1993, Jenkins 1996), ecological data on preferred habitats and the life history of the organisms (Bush 1966, Foote et al 1993), behavioral information about possible ethological isolation or habitat selection (Prokopy et al 1971, 1972, Prokopy and Bush 1972, 1973), genetic descriptions based on cytological, biochemical and molecular data (Bush 1966, Berlocher and Bush 1982, Feder et al 1988, 1997, 1999, McPheron et al 1988, Berlocher et al 1993, Berlocher 1995, 2000, McPheron and Han 1997, Smith and Bush 1997), data

on morphological and genetic geographic variation (McPheron 1990, Feder et al 1999), experimental hybridization studies (Smith et al 1993) and information on the frequency of hybridization in nature where the ranges overlap (McPheron 1990, Feder et al 1999).

R. zephyria infests fruits of *Symphoricarpos* spp. (Caprifoliaceae). The primary host of *R. zephyria* in western North America is *S. albus* var. *laevigatus* (*S. rivularis*). It is a woody shrub, 1-2m tall, with dark-green leaves and small, pink, bell shaped flowers, which give rise to pearl-like white berries in clusters. The fruit remains on the plants into the late fall. The natural geographic distribution of this plant is limited to west of Rocky Mountains (Jones 1940, U.S. Department of Agriculture, Forest Service 2001 online), but following the Lewis and Clark expedition of 1804-05 this variety of snowberry was introduced into the eastern part of North America and has escaped from cultivation over a wide range (Bush 1966). *R. zephyria* infestations of the fruits of *S. albus* var. *laevigatus* have been observed in eastern North America and have commonly been assumed to have come along with the introduction of the host (Feder et al 1999).

On the east side of the Rocky Mountains and throughout the Great Plains, the most common species of snowberry, and the primary host of *R. zephyria*, is *S. occidentalis*, a small shrub up to 1m tall, with small pale pink flowers, which give rise to round, dull greenish-white spongy fruits in clusters; the fruits quickly turn brown in early fall. *Symphoricarpos occidentalis* has a very wide distribution, from British Columbia to Ontario and south to Washington, Utah, New Mexico and Oklahoma (U.S. Department of Agriculture, Forest Service 2001 online). It is most commonly found and is very abundant in the prairies of the northern Great Plains. Although records of *S. occidentalis* exist for Wisconsin, Illinois, Michigan, Ontario, Pennsylvania and New York (herbarium

data, Soper and Heimburger 1994, U.S. Department of Agriculture, Forest Service 2001 online), *R. zephyria* was not considered to be present in these states, except in association with the cultivated *S. albus* var. *laevigatus*.

The third potential host species for snowberry flies is *S. albus* var. *albus* although so far it has not been reported as a host of *R. zephyria*. *S. albus* var. *albus* is smaller in size than *S. albus* var. *laevigatus* (up to 1m) with smaller flowers and fruits which grow as solitary rather than in clusters. *Symphoricarpos albus* var. *albus* occurs in the eastern United States and Canada, from eastern Quebec south to North Carolina and extends west to the Rocky Mountains where it is only found on the eastern slopes (Jones 1940). This variety readily hybridizes and is interfertile with *S. albus* var. *leavigatus* introduced to the eastern North America (Smith, pers. comm.). This species of snowberry is only abundant in the northern part of its range (US Department of Agriculture, Forest Service online), but its presence throughout the entire region implies that the range of the three native potential host plants of *R. zephyria* is more or less continuous throughout North America.

The geographic range of *R. zephyria* has been described as primarily western, from southern British Columbia south to northern California, extending east to the Mississippi River valley in Minnesota (Foote et al 1993, Bush 1966). This implies that the zone of parapatry between apple maggot and snowberry flies is relatively narrow, since the native range of *R. pomonella* does not extend west of Minnesota (Figure 1.1, after Bush 1966). Infestations by *R. zephyria* were previously reported for fruits of *S. albus* var. *laevigatus* and *S. occidentalis* (Bush 1966). In order to determine precisely the geographic range of the snowberry flies, fruits from all three potential hosts were sampled throughout the ranges of these plants. The objectives were: i) to determine where

R. zephyria occurs; ii) whether all three potential hosts are infested and to what extent; iii) whether infestations of the native hosts are limited to the west of the Mississippi River valley; iv) where introduced hosts were infested; and v) how may the origin of *R. zephyria* in the eastern North America be explained.

Adults of *R. zephyria* were reared from fruits of all three host plants as a result of the extensive collection. The geographic range of *R. zephyria* extends south to South Dakota and Nebraska; infestations of native hosts in Wisconsin, Michigan, Ontario and New York are reported for the first time. The natural range of *R. zephyria* thus extends east of the Mississippi River valley through the Great Lakes area. These findings have important bearing on inferences about the speciation of *R. zephyria*, since they indicate that the zone of overlap with *R. pomonella* or *pomonella*-like ancestor is or has been much broader than previously thought.

MATERIALS AND METHODS

Field survey: Snowberry stands were located by searching herbarium records, visiting sites with the proper ecological characteristics but previously not known to be inhabited by potential host plants, or from personal contacts. Fruits of *S. albus* var. *laevigatus* , *S. albus* var. *albus* and *S. occidentalis* were hand-picked between the months of August and October each year from 1993 to 2000 from localities in Massachusetts, New York, Pennsylvania, Michigan, Wisconsin, Minnesota, North and South Dakota, Nebraska, Colorado, Wyoming, Montana, Idaho, Washington, Oregon and California and

the province of Ontario (Table 2.1). Identifications of plant species were made in the field. When uncertainty about the plant identity arose or when new sites were found, plants were pressed and brought to the laboratory for identification.

Insect rearing: Fruits were collected into Ziploc^{TM} bags containing a small amount of vermiculite to prevent early rotting. Bags were stored in coolers in the field at 10-20°C until returned to the laboratory. All fruits were placed into plastic trays with approx. 2cm layer of moist Grade 3 (medium) vermiculite and allowed to dry for 3-5 weeks at room temperature. During this time larvae completed their development, left the fruits and pupated in trays. Fruits were removed and in most instances counted. Vermiculite was sifted through a screen which retained the pupae, that were then counted and infestation rates calculated as the number of pupae recovered divided by the number of fruits. Pupae were placed in Petri plates (100x15mm) with a layer of approx. 1cm of moist vermiculite. The plates were then placed in a 4°C refrigerator for 6-7 months to simulate overwintering. Water was added to vermiculite monthly to keep the pupae hydrated. After 6-7 months Petri plates were transferred to a controlled environmental incubator (25°C, 14h light, 21°C, 10hr dark). Emerging adults were collected daily and kept alive for 5-7 days on diet of 50% (w/v) yeast and 25% (w/v) sucrose.

Flies reared from *Symphoricarpos* spp. were assumed to be *R. zephyria*. Host fruit remains one of the most reliable indications of species identification. Since the host fidelity in *R. pomonella* species group is very high and oviposition mistakes are rare (Bush 1966, Bierbaum and Bush 1990) it is generally accepted that fly species can be assigned based on fruits from which they were reared.

RESULTS

Host plants: Infestations of fruits were observed for all three host plants (*S. albus* var. *laevigatus*, *S. albus* var. *albus* and *S. occidentalis*) between 1993 and 2000 (Table 2.1). *Symphoricarpos albus* var. *laevigatus* was sampled at 34 localities in its native range west of the Rocky Mountains. It was found to be infested at 28 (82.4%) localities. Infestations were also observed in Washington, Oregon, California, Idaho and Montana. At three localities in California and three in Idaho fruits of *S. albus* var. *laevigatus* were not infested (Table 2.1). Stands of introduced *S. albus* var. *laevigatus* were infested at all localities east of the continental divide that were sampled. However, it should be noted that uninfested *S. albus* var. *laevigatus* have been observed in Door County, WI (Bush, pers. comm.) and at 17 localities in Nova Scotia (Crozier, pers. comm.).

S. occidentalis was sampled at 26 localities both west of the Rocky Mountains where its range overlaps with the range of *S. albus* var. *laevigatus* and in the Great Plains where it is the predominant snowberry species. In all cases it was infested (Table 2.1).

S. albus var. *albus* was sampled at eight localities and observed at two more where it was not sampled because permits to collect in National Parks and National Monuments were not available. It was infested at six localities (75%) in New York, Ontario, Michigan and Wisconsin. A low number of pupae were recovered from samples from Ontario (2-18, Table 2.1). At several other localities *S. albus* var. *albus* was not found to be infested, although *R. zephyria* was reared from the fruits of other hosts at relatively close localities (Table 2.1).

Localities: Northwest (Washington, Oregon and California). Native *S. albus* var. *laevigatus* was sampled at 11 localities in Washington, seven localities in Oregon and seven localities in Northern California; *S. occidentalis* was sampled at one locality in Washington. All *S. albus* var. *laevigatus* samples from Washington and Oregon were found to be infested; *R. zephyria* was reared from four samples from California. At three localities in California fruits were not infested with *R. zephyria.* Infestation rates ranged from 0.8% at Olympia, WA to 62.9% at Dixie, WA, which was the highest infestation rate observed for all samples (Table 2.1). *S. occidentalis* collected at Beacon Rock State Park in the Columbia River Gorge was also infested with *R. zephyria* (infestation rate 30.0%, Table 2.1).

Rocky Mountains and Colorado Plateau (Idaho, western Montana, Wyoming and Colorado). *S. albus* var. *laevigatus* was sampled at four localities in Idaho. A sample from one locality sampled in 1995 (Elmira) was infested with *R. zephyria*. At three localities sampled in 1997 no infestation was found (Table 2.1). In Montana, both *S. occidentalis* and *S. albus* var. *laevigatus* were sampled at localities west of the continental divide, where *S. albus* var. *laevigatus* is generally more common. Three *S. occidentalis* and five *S. albus* var. *laevigatus* populations were sampled. All the samples were found to be infested with *R. zephyria.* Most of the samples from Montana were taken in 1994 and 1995 when the fruit number was not determined. All three potential host species were sampled in Wyoming. *S. albus* var. *laevigatus* was sampled at a single locality at high altitude (6700', Glendo State Park). A large number of fruits was collected (2689, Table 2.1), enabling detection of a very low rate infestation (0.04% - only one pupa was recovered from the fruits). At lower altitudes, two samples of *S.*

occidentalis were found to be infested (1.0-3.2%, Table 2.1). *S. albus* var. *albus* was found growing at the Devil's Tower National Monument where it was not collected. Fruits did not look infested; however, this does not necessarily imply that they are not infested, but only the inability to detect infestation by observation. At another site (Bear Lodge campground), fruits of *S. albus* var. *albus* were relatively scarce and no pupae were found. In Colorado, a small stand of native *S. albus* var. *albus* was found at Estes Park. Fruits were sampled on several occasions (1996, 1997, 1999) but yielded no pupae. Cultivated *S. albus* var. *laevigatus* growing as an ornamental plant on campus of the University of Colorado in Boulder were sampled in early August of 1997, when an infestation rate of 0.8% was observed (Table 2.1). In early September 1999, very large berries were showing signs of heavier infestation; however, only 12 pupae were recovered from 973 fruits (1.2%, Table 2.1)

Great Plains (eastern Montana, North Dakota, South Dakota, Nebraska, Minnesota). *S. occidentalis* is a dominant host in eastern Montana. It is very common across the broad areas of North and South Dakota and Minnesota where large stands are frequently found and easily spotted along the roads. Three samples were taken from Montana, four from North Dakota, three from South Dakota, two from Nebraska and five from Minnesota. All samples yielded pupae; infestation rates ranged from 0.4% at Wasta, SD to 4.2% at Rapid City, SD. Several large stands of *S. occidentalis* were observed in Badlands National Park (Gavrilovic, pers. obs.), but no collections were made in the Park. A stand of *S. albus* var. *albus* was seen at Custer State Park, SD (Crossno, pers. comm.) where no collections were made, also due to the lack of necessary permits.

Great Lakes (Wisconsin, Michigan and Ontario). Herbarium records from the University of Wisconsin, Madison indicated that *S. occidentalis* grows in Wisconsin. However, most of the sites these records listed were converted to cornfields (Gavrilovic, pers. obs.). A stand of *S. occidentalis* found along the railroad tracks in Waukesha, about 15 mi west of Milwaukee, was heavily infested with *R. zephyria* (27.2%, Table 2.1). *S. occidentalis* was also found near Mio, MI and pupae were recovered from the fruits. Native stands of *S. albus* var. *albus* form the predominant ground cover at three sites in Ontario where it was sampled in 1996 and 1998. All three samples were infested; however, adults emerged only from samples collected at Glen Miller and Rice Lake (Table 2.1). *S. albus* var. *albus* was also observed in Pinery Provincial Park (Gavrilovic and Crossno, pers. obs.) but it was not collected. At Glen Arbor, MI (Sleeping Bear Dunes National Lakeshore), plants found were either *S. albus* var. *albus* or its hybrid with *S. albus* var. *laevigatus*. These fruits were also found to be infested (Table 2.1). Infestations of *S. albus* var. *laevigatus* introduced to Michigan as horticultural plant were previously known (Bush pers. comm., Smith pers. comm.). This host plant was sampled at five localities in Michigan and found to be infested at all of them (Table 2.1). The plant is common in Lansing/East Lansing area and heavily infested. Lower numbers of pupae were recovered from samples from Mio Dam and Harrisville (Table 2.1).

Northeast (Pennsylvania, New York and Massachusetts). Only introduced *S. albus* var. *laevigatus* was found in Pennsylvania and Massachusetts, infested with *R. zephyria* (Table 2.1), consistent with previous reports (Smith pers. comm., Bush pers. comm.). In New York, two samples of *S. albus* var. *laevigatus* were found to be infested. The plants at Wellesley Island were likely *S. albus* var. *albus*, however, the possibility that they were

hybrids with var. *laevigatus* cannot be excluded. This sample was also infested (Table 2.1), and was collected near the town of Alexandria Bay where infested *S. albus* var. *laevigatus* was observed (Smith, pers. comm.).

Table 2.1. Collection records for *R. zephyria*.

State/Province	Locality (County)	Host Species	Date	Collector	Fruits	Pupae(adults)	Infestation rate
Washington	Trout Lake (Klickitat)	*S.albus* var. *laevigatus*	8/30/1995	GLB/DB	ND	113 (76)	
	Olympia (Thurston)	*S.albus* var. *laevigatus*	9/01/1995	GLB/DB	ND	262 (63)	
	Olympia (Thurston)	*S.albus* var. *laevigatus*	9/2000	MK	740	6	0.8%
	Cle Elum (Kittitas)	*S.albus* var. *laevigatus*	9/06/1995	GLB/DB	ND	42 (22)	
	Dixie 1500' (Walla Walla)	*S.albus* var. *laevigatus*	8/31/1997	GLB/DB	948	596 (395)	62.9%
	St. Cloud Ranch (Skamania)	*S.albus* var. *laevigatus*	9/2000	MK/VG	325	58	17.8%
	Beacon Rock St. Park/ (Skamania)	*S.albus* var. *laevigatus*	9/2000	MK/VG	100	30	30.0%
	Yakima (Yakima)	*S.albus* var. *laevigatus*	9/2000	MK	200	3	1.5%
	Ellensburg area (Kittitas)	*S.albus* var. *laevigatus*	9/2000	MK	2000	~500	25.0%
	Sedro Wooley (Skagit)	*S.albus* var. *laevigatus*	9/2000	MK	550	20	3.6%
	Okanogan county	*S.albus* var. *laevigatus*	9/2000	MK	900	50	5.6%
	Chelan county	*S.albus* var. *laevigatus*	9/2000	MK	850	82	9.6%
Oregon	Wahkeena Falls SP (Multnomah)	*S.albus* var. *laevigatus*	8/10/1994	JJS	~250	16 (9)	6.4%
	Mulino (Josephine)	*S.albus* var. *laevigatus*	8/14/1994	JJS	ND	31 (7)	
	Grants Pass-3900' (Josephine)	*S.albus* var. *laevigatus*	8/28/1997	GLB/DB	2325	376 (308)	16.2%
	CamusValley-2000' (Douglas)	*S.albus* var. *laevigatus*	8/29/1997	GLB/DB	2203	617 (522)	28.0%
	Monroe (Benton)	*S.albus* var. *laevigatus*	8/29/1997	GLB/DB	1436	141 (78)	9.8%
	Philomath - 750' (Benton)	*S.albus* var. *laevigatus*	8/30/1997	GLB/DB	1594	424 (303)	26.6%
	Gaston (Washington)	*S.albus* var. *laevigatus*	8/30/1997	GLB/DB	2422	257 (213)	10.6%

Fruits – number of fruits collected; pupae – number of pupae recovered; adults – number of adults emerged. Collector initials: BAM – Bruce McPheron, BB – Branislav Blagojevic, DB – Dorie Bush, GLB – Guy Bush, HA – Hugh Atkinson, JEC – Joe Crossno, JJ – John Jenkins, JJS – Jim Smith, JNT – J Thomas, MK – Mike Klaus, RJP – Ron Prokopy, TG – Teri Genshorek, VG – Vesna Gavrilovic

Table 2.1 (cont.)

State/Province	Locality (County)	Host species	Date	Collector	Fruits	Pupae(adults)	Infestation rate
California	Berkeley (Contra Costa)	*S.albus* var. *laevigatus*	8/24/1997	GLB/DB	659	109	16.5%
	Ukiah – 1900' (Mendocino)	*S.albus* var. *laevigatus*	8/24/1997	GLB/DB	2007	0	0.0%
	Honeydew (Humboldt)	*S.albus* var. *laevigatus*	8/25/1997	GLB/DB	798	226	28.3%
	Capetown (Humboldt)	*S.albus* var. *laevigatus*	8/25/1997	GLB/DB	1211	0	0.0%
	Philipsville (Humboldt)	*S.albus* var. *laevigatus*	8/25/1997	GLB/DB	96	0	0.0%
	Chezan Rd (Humboldt)	*S.albus* var. *laevigatus*	8/27/1997	GLB/DB	521	19	3.6%
	Happy Camp (Siskiyou)	*S.albus* var. *laevigatus*	8/28/1997	GLB/DB	398	191	48.0%
Idaho	Elmira (Bonner)	*S.albus* var. *laevigatus*	8/24/1995	GLB/DB	ND	55 (19)	
	Nez Perce Res. (Nez Perce)	*S.albus* var. *laevigatus*	8/31/1997	GLB/DB	1703	0	0.0%
	Shotgun Creek	*S.albus* var. *laevigatus*	8/31/1997	GLB/DB	32	0	0.0%
	Bitterroot Wilderness (Idaho)	*S.albus* var. *laevigatus*	9/18/1994	JNT	300	0	0.0%
Montana	Terry (Prairie)	*S. occidentalis*	9/09/1994	GLB/DB	ND	50 (17)	
	Melstone (Musselshell)	*S. occidentalis*	9/09/1994	GLB/DB	ND	108 (54)	
	Avon (Powell)	*S. occidentalis*	9/10/1994	GLB/DB	ND	55 (21)	
	Kalispell (Flathead)	*S. occidentalis*	9/11/1994	GLB/DB	ND	59 (21)	
	Swan Lake (Flathead)	*S.albus* var. *laevigatus*	9/12/1994	GLB/DB	ND	61 (39)	
	Swan Lake (Flathead)	*S.albus* var. *laevigatus*	9/06/1995	GLB/DB	ND	82(60)	
	Libby (Lincoln)	*S.albus* var. *laevigatus*	8/24/1995	GLB/DB	ND	26 (19)	
	Clark River (Sanders)	*S.albus* var. *laevigatus*	9/07/1995	GLB/DB	ND	28 (14)	
	Perma (Sanders)	*S.albus* var. *laevigatus*	9/07/1995	GLB/DB	ND	61 (22)	
	Missoula elev. 2000' (Missoula)	*S. occidentalis*	9/01/1997	GLB/DB	2041	308 (187)	15.1%
	Missoula elev. 3500' (Missoula)	*S. occidentalis*	9/01/1997	GLB/DB	936	340 (175)	36.3%
Wyoming	Eagle Creek (Park)	*S. sp (?)*	9/04/1997	GLB/DB	2414	0	0.0%
	Bighorn (Sheridan)	*S. occidentalis*	9/05/1997	GLB/DB	560	18 (9)	3.2%
	Buffalo elev. 4300' (Johnson)	*S. occidentalis*	9/05/1997	GLB/DB	1761	18 (10)	1.0%
	Glendo Sp elev. 6700' (Converse)	*S.albus* var. *laevigatus*	9/06/1997	GLB/DB	2689	1 (1)	0.04%
	Devil's Tower (Crook)	*S. albus* var. *albus*	9/08/1999	VG			None collected
	Bear Lodge campground (Crook)	*S. albus* var. *albus*	9/08/1999	VG	160	0	0.0%

Table 2.1 (cont.)

State/Province	Locality (County)	Host species	Date	Collector	Fruits	Pupae(adults)	Infestation rate
Colorado	Boulder – Univ.Colorado (Larimer)	*S.albus* var. *laevigatus*	8/13/1997	GLB/DB	389	3 (3)	0.8%
	Boulder – Univ.Colorado (Larimer)	*S.albus* var. *laevigatus*	9/5/1999	VG/BB	973	12 (12)	1.2%
	Estes Park (Larimer)	*S. albus* var. *albus*	9/6/1999	VG/BB	365	0	0.0%
North Dakota	Tower City (Barnes)	*S. occidentalis*	9/7/1994	GLB/DB	ND	192 (111)	
	St. Anthony (Morton)	*S. occidentalis*	9/8/1994	GLB/DB	ND	206 (108)	
	Medora (Billings)	*S. occidentalis*	9/9/1994	GLB/DB	ND	130 (35)	
	Belfield (Stark)	*S. occidentalis*	9/8/2000	VG	985	15	1.5%
South Dakota	Rapid City (Pennington)	*S. occidentalis*	9/8/1999	VG	767	32	4.2%
	Custer (Pennington)	*S. occidentalis*	9/9/1999	VG	1613	15	0.9%
	Wasta (Pennington)	*S. occidentalis*	9/9/1999	VG	1497	6	0.4%
	Custer State Park (Custer)	*S. albus* var. *albus*	8/97, 9/99	JEC, VG			None collected
	Badlands NP (Jackson)	*S. occidentalis*	9/10/1999	VG			None collected
Nebraska	Brady (Lincoln)	*S. occidentalis*	9/07/1997	GLB/DB	1610	65 (11)	4.0%
	Sutherland (Lincoln)	*S. occidentalis*	9/04/1999	VG	683	3	0.4%
Minnesota	Staples (Todd)	*S. occidentalis*	9/05/1994	GLB/DB	ND	300 (180)	
	Park Rapids (Hubbard)	*S. occidentalis*	9/05/1994	GLB/DB	ND	157 (116)	
	Fertile (Polk)	*S. occidentalis*	9/06/1994	GLB/DB	ND	102 (33)	
	Syre (Norman)	*S. occidentalis*	9/06/1994	GLB/DB	ND	157 (91)	
	Hawley (Clay)	*S. occidentalis*	9/07/1994	GLB/DB	ND	192 (111)	
Wisconsin	Waukesha (Waukesha)	*S. occidentalis*	9/12/99	VG	578	157	27.2%
Michigan	Lansing Area (Ingham)	*S.albus* var. *laevigatus*	1993-98	JJS/JJ/VG	ND	~150	
	Glen Haven (Leelanau)	*S. albus* var. *albus* (?)	9/04/1994	JJS	ND	34 (7)	
	East Lansing Area (Ingham)	*S.albus* var. *laevigatus*	1994-00	JJS/VG	~6000	~200	~33.3%
	Mio Dam (Oscoda)	*S. albus* var. *albus*	8/30/1997	JEC	ND	78 (4)	
	Mio (Oscoda)	*S. occidentalis*	8/30/1997	JEC	ND	23 (6)	
	Mio Dam (Oscoda)	*S.albus* var. *laevigatus*	8/30/1997	JEC	ND	13 (1)	
	Harrisville (Alcona)	*S.albus* var. *laevigatus*	8/1997	JEC	ND	1 (0)	
	Springport (Alcona)	*S.albus* var. *laevigatus*	8/1997	JEC	ND	13	
Ontario	Glen Miller (Peterborough)	*S. albus* var. *albus*	8/1996	TG/HA	ND	3 (1)	
	Rice Lake (Peterborough)	*S. albus* var. *albus*	8/1996	TG/HA	ND	18 (8)	
	Rondeau Provincial Park (Kent)	*S. albus* var. *albus*	8/1998	VG	~500	2 (0)	0.4%

Table 2.1 (cont.)

State/Province	Locality (County)	Host species	Date	Collector	Fruits	Pupae(adults)	Infestation rate
Pennsylvania	State College (Centre)	*S.albus* var. *laevigatus*	09/1994	BAM	ND	ND	
New York	Goat Island (Niagara)	*S.albus* var. *laevigatus*	8/17/1997	JJS	ND	81 (41)	
	Geneva (Ontario)	*S.albus* var. *laevigatus*	8/18/1997	JJS	ND	137 (86)	
	Wellesley Island (Niagara)	*S.albus* var. *albus (?)*	8/22/1997	JJS	ND	6 (3)	
Massachusetts	Amherst (Hampshire)	*S.albus* var. *laevigatus*	9/1998	RJP	ND	52 (36)	
	Amherst (Hampshire)	*S.albus* var. *laevigatus*	9/4/1999	GLB	ND	52 (0)	

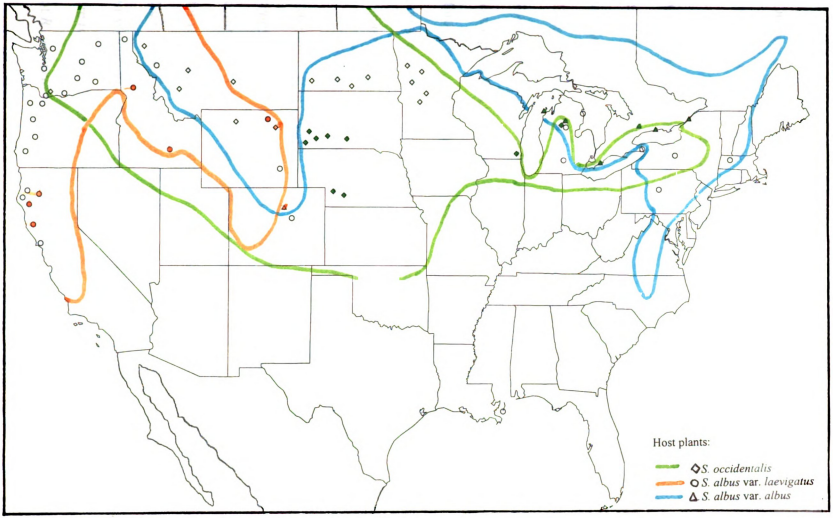


Figure 2.1. Collection sites of *R. zephyria*. Green symbols represent new findings of *R. zephyria* on native hosts. Orange symbols represent non-infested snowberry samples. (Images in this dissertation are presented in color)

DISCUSSION

Native snowberry species in the western part of North America (Washington, Oregon, California, Idaho, Montana, Wyoming) and in the Great Plains (North Dakota and Minnesota) were previously known to be host plants for *R. zephyria* (Bush 1966, Foote et al 1993). The natural range of the fly has been reported to extend east into Minnesota and no further than the Mississippi River Valley. Findings of *R. zehyria* in eastern North America have been attributed to human introduction, as the host, *S. albus* var. *laevigatus* was introduced and cultivated as an ornamental plant (Feder et al 1999). Thus, the range of *R. zephyria*, including all introduced and native populations, was thought to be discontinuous (Bush and Smith 1998). The main reason for this was the assumption that the suitable host plants were not found in the Great Lakes area and further east, into the northeast United States. *Symphoricarpos albus* var. *albus*, the native snowberry species in the eastern North America, has never been reported infested with *R. zephyria*.

In this field survey, stands of native snowberry species were located in a broad geographic area covering the regions previously known as parts of the range of *R. zephyria* and those where the fly has not been found although the hosts were known to commonly occur (such as South Dakota and Nebraska), or where the only host available for flies was thought to be the introduced *S. albus* var. *laevigatus* (Wisconsin, Michigan, New York, Ontario). Here the observations of *R. zephyria* infesting native *S. occidentalis* in Wisconsin and Michigan and *S. albus* var. *albus* in Michigan, Ontario and New York are reported. These are the first reports of *R. zephyria* in Wisconsin and the first reports of *R. zephyria* infesting native hosts east of Minnesota. The rearing of *R. zephyria* pupae

and adults from native hosts (*S. occidentalis*) in South Dakota and Nebraska is also reported, and the findings of *R. zephyria* in Wyoming are extended into the eastern part of the state. By these findings the known natural geographic range of *R. zephyria* is extended.

Another species of the *R. pomonella* group, *R. mendax* (blueberry maggot), has recently been found to infest native host plants in the Great Lakes area (Smith et al, in review), a finding that questions the previously accepted view that the pest was introduced with the cultivated host plants from the regions where it commonly occurs. Similarly, this report of fly populations infesting native hosts raises a question of the origin of *R. zephyria* in the eastern North America.

S. occidentalis is a native inhabitant of the Great Lakes area. This species is typically found in prairies, which became established in this region after the Pleistocene glaciers began receding (Thompson and Smith 1970, Catling and Catling 1993). Herbarium records indicate that *S. occidentalis* was common in Wisconsin, Illinois, Michigan, Ohio and New York as recently as 60 years ago, and distribution data from the US Forest Service list this species from British Columbia to Ontario and as far south as New Mexico and Oklahoma. Stands of *S. occidentalis* in Wisconsin and Michigan may represent prairie remnants from the period after the Pleistocene glaciers began receding and pupae and adults of *R. zephyria* were reared from both localities (Waukesha, WI and near Mio, MI). Catling and Catling (1993) suggest that prairie remnants exist around Lake Ontario as well, with *S. albus* var. *albus* being the predominant ground cover at some areas, whereas Soper and Heimburger (1994) report *S. occidentalis* in Ontario, at areas that correspond to Catling and Catling's (1993) prairie remnants. Adults of *R.*

zephyria were reared from *S. albus* var. *albus* from two such sites (Glen Miller and Rice Lake, ON). It is interesting to note that herbarium records from the University of Michigan indicate that *S. occidentalis* was growing on the streets of Grand Rapids, MI and in the Black River valley near Port Huron, MI about 100 years ago. An attempt was made to locate native stands of either *S. ocidentalis* or *S. albus* var. *albus* in Ohio, but information from the Department of Natural Resources indicated that the snowberries were extirpated from Ohio relatively recently (Cusick, pers. comm.). All this suggests that *S. occidentalis* was much more common in the Great Lakes area in the recent past, indicating that the zone of overlap of *R. zephyria* and *R. pomonella* may have been much wider than previously thought. This observation is important with respect to whether populations of *R. zephyria* in eastern North America have been established by importation of flies with the host plants or by spreading to the introduced host from local native plants.

S. occidentalis appears to be the preferred host of *R. zephyria*. All the samples of *S. occidentalis* were found to be infested (Table 2.1), whereas some samples of native *S. albus* var. *laevugatus* were not (three samples from California and three from Idaho yielded no pupae). Introduced *S. albus* var. *laevigatus* in Door County, WI, outside of the native range of any snowberry species or variety, has never showed signs of infestations (Bush, pers. comm.) Cultivated *S. albus* var. *laevigatus* in Nova Scotia is also not known to be infested with *R. zephyria* (Crozier, pers. comm.). The same variety has been widely introduced into Great Britain from the Pacific coast of North America, but no *R. zephyria* associated with it has been observed (Gilbert 1995). Therefore it appears that infestations of introduced *S. albus* var. *laevigatus* occur only where the native host is present as well.

This supports the hypothesis that *R. zephyria* may be a native inhabitant of eastern North America.

AFLP analysis showed that *R. zephyria* reared from native host plants in New York, Ontario, Michigan and Wisconsin possess AFLP fragments characteristic of *R. zephyria.* Neighbor-joining analysis of AFLP patterns placed these individuals within the *R. zephyria* cluster (Chapter 4). While mitochondrial and nuclear DNA sequences of flies from these populations are very similar to other *R. zephyria* (Chapter 3), a lack of geographic structuring does not allow me to distinguish between different scenarios about the origin of snowberry flies in the east. However, the amount of variation observed suggests that bottlenecks did not play a role in population divergence, which would be expected if the populations in eastern North America were introduced and founded from a small number of individuals. The possibility that these populations were established by range expansion of *R. zephyria* west of the Mississippi River valley and colonization of available hosts in the region cannot be eliminated. However, under such a scenario it would be expected to see isolation by distance, which the data at this point do not support.

The results of this field survey and rearing of *R. zephyria* from a previously unrecorded host, *S. albus* var. *albus*, raise a question as to whether other species of *Symphoricarpos* may serve as hosts for snowberry flies. It is important to note that where *S. albus* var. *albus* was co-occurring or growing relatively close to *S. albus* var. *laevigatus* or *S. occidentalis*, no infestations of *S. albus* var. *albus* were observed, indicating that the other hosts are preferentially chosen. However, it may also indicate that short dispersal distances and the patchiness of the suitable habitats may limit further

range expansion, even though the hosts may be present, as in the case of geometrid moths *Itame andersoni* feeding on patchy distributed host plant, *Dryas drummondii* (Doak 2000). The relatively isolated, although physically close, stands of *S. albus* var. *albus* in Colorado and Wyoming were not found to be infested (Table 2.1). The same was true for stands of different species or varieties of *Symphoricarpos* observed or sampled in Utah, Nevada and Arizona, where no signs of infestations were obvious and no flies were reared from any of the samples (Bush, pers. comm., Smith, pers. comm.). This may indicate that more or less geographically continuous presence of host plants is necessary for dispersal of flies and range expansion.

The described findings of *R. zephyria* infesting native hosts in the Great Lakes region may be similar to the observation of Smith et al (in review) of blueberry maggot (*R. mendax*) infesting native *Vaccinium corymbosum* and *V. angustifolium* in the same region. Similar to infestations of *S. occidentalis*, *Vaccinium stamineum* is reported to be infested with *R. mendax* throughout its natural geographic range in eastern North America, with *V. corumbosum* and *V. angustifolium* serving as "range extenders". Even so, the latter two hosts are not infested everywhere they grow, presumably because abiotic factors render the habitat unsuitable for *R. mendax* (Smith et al, in review). The same may be true for *R. zephyria* whose preferred host appears to be *S. occidentalis*, with *S. albus* var. *laevigatus* and *S. albus* var. *albus* serving as "range extenders" respectively to the west and east.

CHAPTER 3

PHYLOGEOGRAPHIC RELATIONSHIPS OF *R. POMONELLA* AND *R. ZEPHYRIA*

INTRODUCTION

The history of species divergence can be inferred from DNA sequence polymorphism patterns at randomly selected genes (Kliman et al 2000) by comparing the patterns between species, estimating DNA sequence divergence, inferring ancestral states and reconstructing speciation events from the phylogenies. White (1978) stressed the importance of describing genetic variation across the entire geographic ranges of two sister taxa for documenting the history of their speciation. Previous DNA sequence studies of phylogenetic relationships of taxa in the *R. pomonella* species group and species of the genus *Rhagoletis* as a whole have sampled only a small number of individuals (usually one) from each species (Smith and Bush 1997, McPheron and Han 1997, Han and McPheron 1997). This study represents the first effort to make inferences about speciation of *R. zephyria* using a phylogeographic approach.

In this study I characterized genetic variation across the geographic ranges of *Rhagoletis zephyria* and *R. pomonella* by analyzing DNA sequence divergence in mitochondrial DNA and three anonymous nuclear loci. The general objective was to examine biogeographic patterns within and between species at the DNA sequence level and make inferences not only with respect to where the divergence of *R. zephyria* from *R. pomonella* may have occurred, but also to the mode of speciation and subsequent species divergence. Within this general framework, I addressed several specific questions: i) what is the phylogenetic position of *R. zephyria* within the *R. pomonella* species group;

ii) how closely related are *R. zephyria* and *R. pomonella*; iii) is there geographic structure within *R. zephyria* and/or *R. pomonella* and iv) are there any fixed differences between the two species?

The phylogenetic position of *R. zephyria* within the *R. pomonella* group is uncertain. Phylogenetic studies based on allozymes (Berlocher et al 1993, Berlocher 2000) have placed *R. zephyria* at the base of the closely related *R. pomonella* group species (with *R. cornivora* at the base of the entire group). However, the parsimony tree in which *R. zephyria* was a sister to one of the *R. pomonella* clades was not a significantly worse fit (Berlocher 2000). Studies based on mitochondrial DNA sequences (16S – McPheron and Han 1997, Han and McPheron 1997; COI/COII – Smith and Bush 1997) support the latter relationship.

Studies of genetic variation in multiple populations of *R. pomonella* so far have all been based on allozymes (McPheron 1990, Berlocher 2000). Several populations of *R. zephyria* were analyzed for allozyme and mitochondrial RFLP variation by Feder et al (1999); their study focused on hybridization between *R. pomonella* and *R. zephyria*. No large-scale geographic studies to date in either species have been based on DNA sequence data.

R. pomonella and *R. zephyria* display a fixed difference at one of the six diagnostic allozyme loci (*Had*) identified by Feder et al (1988) and McPheron et al (1988). *R. zephyria* is fixed for Had^{111} while *R. pomonella* is polymorphic for two other alleles, Had^{100} and Had^{125}. Also, *R. pomonella* and *R. zephyria* display a fixed difference in the shape of male claspers (surstyli) (Westcott 1989), and the lengths of female ovipositors (however, the variation in the latter trait is large and ovipositor lengths

overlap, Bush 1966). No fixed differences in mitochondrial COII and 16S DNA sequences have been found (Smith and Bush 1997, McPheron and Han 1997).

As mentioned above (Chapter 1), relationships between taxa inferred from gene trees obtained using one gene can differ dramatically from those obtained using a different gene or a different type of marker. Therefore gene trees may not accurately correspond to species trees (Maddison 1995, Wang et al 1997). Maddison (1997) remarks that "phylogeny has a variance as well, represented by the diversity of trees of different genes". The use of multiple loci for estimating relationships between taxa permits one to distinguish the forces that act on many genes (and ideally the entire genome) from those affecting individual loci (Hudson et al 1987).

In this study I used DNA sequences of three anonymous single-copy nuclear genes and a well-characterized mitochondrial COI/COII gene in order to minimize the limitations of the single-gene approach. The nuclear loci represent different linkage groups in the *R. pomonella* genome (Figure 3.1; Roethele et al 1997). While the identities and function of the nuclear loci used here are unknown, they are each in significant linkage disequilibrium with the allozyme loci that display host-specific frequency differences between host races of *R. pomonella* (Roethele et al. 1997). The degree of linkage to the allozyme loci as determined by levels of recombination (Roethele et al 1997) is different for each of these markers and ranges between 1.8 and 7.5cM.

The working hypothesis in this study is that the speciation of *R. zephyria* occurred by a host shift of an *R. pomonella*-like ancestor from hawthorns to snowberries in a zone of parapatry or sympatry in the upper Mississippi River valley of Minnesota. In studies based on allozyme data (Berlocher and Bush 1982, Berlocher 2000) *R. pomonella* has

been shown to be the most genetically variable species of the group, with a large and variable gene pool containing ancestral polymorphisms. In at least one case it has been documented that *R. pomonella* has shifted from one host (hawthorns) to another (apples) thus forming a new host race. This has led to a hypothesis that *R. pomonella* may be the ancestor of other taxa in the group, which speciate by host shifts. It is further hypothesized that the descendent *R. zephyria* spread west, and possibly east, on snowberries. If there were no snowberries east of the Mississippi River valley prior to this host shift, *R. zephyria* found in eastern North America today must have been introduced with the infested host after the Lewis and Clark expedition (see Chapter 2). One alternative hypothesis is that the host shift occurred at the Great Plains/Eastern Woodland boundary, with subsequent dispersal of snowberry flies both east and west, assuming that snowberries existed in eastern North America prior to the host shift (see Chapter 2).

MATERIALS AND METHODS

Sample: Infested fruits of *R. zephyria* and *R. pomonella* host plants (snowberries, hawthorns and apples) were collected during late summer and fall of 1994, 1995, 1997, 1998, 1999 and 2000 throughout the geographic ranges of both fly species (Table 3.1) and brought to the laboratory, where the files were reared as described in Chapter 2. Pupae of *R. pomonella* from Nova Scotia were obtained from R. Smith in 95% ethanol. Adults were used for DNA extractions for all *R. zephyria* and most *R. pomonella*

samples; DNA was extracted from pupae of *R. pomonella* from Nova Scotia (ethanol preserved), Massachusetts, New York and Washington State (samples collected in fall 2000) and Ontario (no adults emerged during two seasons). Individuals belonging to other taxa included in the analyses were either reared in the laboratory of J. Feder at the University of Notre Dame or obtained from S. Berlocher (University of Illinois at Urbana-Champaign) or J. Smith. These flies were subjected to the same procedures as *R. zephyria* and *R. pomonella* samples (see below) and sequences from them were obtained in collaboration with J. Roethele (University of Notre Dame).

DNA extraction from *R. zephyria* and *R. pomonella*: Total genomic DNA was isolated from individual flies using a protocol modified from Han and McPheron (1997). Individual flies were homogenized in buffer (10 mM Tris, pH 8, 60 mM NaCl, 150 mM sucrose, 10 mM EDTA, 0.5% (w/v) SDS, 0.1 mg/mL proteinase K) and crude homogenates were incubated at 55°C for 30 minutes. Buffer containing 300 mM Tris, 150 mM sucrose, 100 mM EDTA and 0.75% (w/v) SDS was added to precipitate protein. After incubation on ice for 10 minutes an equal volume of phenol was added and protein and cell debris were removed by centrifugation (14,000 rpm, 5 min). Residual protein contaminants were removed by subsequent phenol:chloroform:isoamyl alcohol (25:24:1, v/v/v) and chloroform:isoamyl alcohol (24:1, v/v) extraction followed by ethanol precipitation at -20°C overnight. DNA was resuspended in TE buffer (10mM Tris, 1mM EDTA) containing 100μg/mL RNase A and stored at 4°C.

Table 3.1. Populations of *Rhagoletis zephyria* and *R. pomonella* included in the phylogeographic analyses

Species/ State	Locality	Label
R. zephyria		
Massachusetts	Amherst	ZMA
New York	Geneva	ZNY
Pennsylvania	Penn State campus	ZPA
Ontario	Rice Lake	ZON
Michigan	Mio	ZMIO
Michigan	Sleeping Bear	ZSB
Michigan	East Lansing	ZEL
Wisconsin	Waukesha	ZWI
Minnesota	Hawby	ZMN
North Dakota	Bismarck	ZND
South Dakota	Custer SP	ZSD
Nebraska	Brady	ZNE
Colorado	Boulder	ZCO
Wyoming	Moiser Gulch	ZWY
Montana	Swan Lake	ZMT
Montana	Billings	ZMTB
Idaho	Elmira	ZID
Washington	Dixie	ZWA
Oregon	Grants Pass	ZOR
California	Honeydew	ZCA
R. pomonella		
Nova Scotia	Kentville	PNS
Massachusetts	Amherst	PMA
New York	Geneva	PNY
Pennsylvania	Biglerville	PPA
Ontario	Toronto	PON
Michigan	E. Lansing	PEL
Michigan	Grant	PMI
Georgia	Macon	PGA
Illinois	Riverwoods	PIL
Minnesota	Staples	PMN
Iowa	Ames	PIA
Nebraska	(I80E,exit285)	PNE
Colorado	Boulder	PCO
Texas	Waxahatchie	PTX
New Mexico	(unknown)	PNM
Utah	Wellsville	PUT
Washington	St. Cloud	PWA
Mexico	Mexico City	PMX
Mexico	La Jolla	MXHD

Choice of markers: Primer pair sequences for 10 anonymous nuclear loci located in different linkage groups (Figure 3.1) within the genome, and showing different levels of recombination with allozymes which show host-specific frequency differences in host races of *R. pomonella*, were provided by J. Roethele and J. Feder (University of Notre Dame). These loci (P181 and P220 from linkage group I; P114 and P2956 from linkage group II; P7, P2156 and P2480 from linkage group III; P100 and P661 from linkage group IV and P454 from linkage group V) were screened for variation and level of phylogenetic signal. Based on the results of a preliminary study, P220, P2956 and P2480 were chosen for the phylogeographic study. P181, P2156 and P454 showed very little variation; the phylogenetic signal from P661 was complicated by high recombination rates within this gene; P114 provided less resolution than P2956; P7 was very difficult to amplify and clone, and P100 did not amplify in *R. zephyria*.

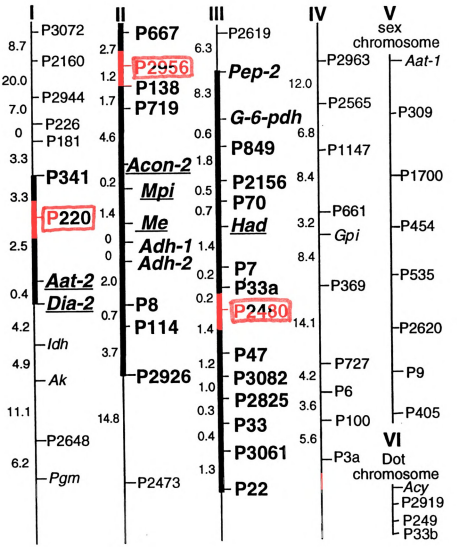


Figure 3.1. Genetic linkage map of *Rhagoletis pomonella*. Map constructed by Roethele et al. (1997) based on genetic recombination data obtained using allozyme and cDNA markers. Underlined allozymes show allele frequency differences between the apple and hawthorn host races of *R. pomonella*. cDNA markers are labeled with a "P", followed by a number; bold lines and marker labels indicate loci in gametic disequilibrium with an adjacent allozyme locus. The loci used in this study are framed in red.

PCR amplification, cloning and sequencing for *R. zephyria* and *R. pomonella*: Primers used for PCR amplifications of nuclear alleles were designed by J. Roethele and are listed in Table 3.2. Reactions were performed in 50μL or 200μL volumes using 1xGibcoBRL PCR buffer, 6.25 mM $MgCl_2$, 1.25 mM of each of dATP, dCTP, dGTP and dTTP, 10 pM of each primer, 5 units of Gibco Taq polymerase and 1μL of template DNA (not quantified). The PCR cycling regime was: 95°C, 5 min; 35 cycles of 94°C 2 min, 52°C 1 min 30 sec, 72°C 2 min; 72°C 7 min. PCR products were initially gel-purified by centrifugation through an Amicon Ultrafree-DA filter (Millipore, Bedford, MA) followed by isopropanol precipitation for 2 hr at 4°C and cloned into the pCR2.1 Dual Promoter plasmid using a commercially available TA cloning kit (Invitrogen, Carlsbad, CA). After overnight ligation at 14°C, chemically competent EF1α *E. coli* cells were transformed and grown on LB medium plates containing 50 mg/mL ampicillin and 1.6 mg X-gal. PCR products were also directly cloned without gel-purification, which yielded higher numbers of putatively positive clones, so this technique was adopted for all subsequent clonings. Putative positive clones were checked for the presence of inserted DNA by direct PCR amplification from bacterial colonies, using the same primers and same cycling regime as in the original PCR. Four positive clones from each individual fly were purified (Wizard*SV* Miniprep, Promega, Madison, WI) and sequenced. For about 30% of the number of flies, sequencing was performed in both directions by cycle sequencing with dye termination using universal primers M13 and M7 and the ALF automated sequencer (Amersham-Pharmacia) at the University of Notre Dame. The rest of the sample was sequenced unidirectionally, using universal primer M13 and ALF automated sequencer at the University of Notre Dame, or specific T7 primers (Table 3.2) with the

BigDye termination (Applied Biosystems, Foster City, CA) and ABI377 sequencer at Iowa State University or the University of Minnesota. Some sequences were obtained by sequencing purified PCR products from screening reactions (Qiaquick, Qiagen, Valencia, CA). Several clones were sequenced at both University of Notre Dame and Iowa State University to check for discrepancies. The results were consistent regardless of the primers (universal or specific), template (purified plasmids or direct screening PCR products) and sequencing method.

The mitochondrial COI/COII coding gene region was PCR amplified as a single 1kB fragment using the protocol of Smith and Bush (1997) and primers George (CI-J-2792, Bogdanowicz et al 1993) and Eva (TK-N-3722, Bogdanowicz et al 1993) (Table 3.2), with the modification that DNA was not quantified prior to amplification. PCR products were purified (Qiaquick, Qiagen, Valencia, CA) and sequenced directly in both directions using BigDye termination (Applied Biosystems, Foster City, CA) and ABI377 sequencer at Iowa State University.

Table 3.2. Oligonucleotide sequences used for PCR amplifications and sequencing

Primer	Sequence
220T3	5' CTG AAG TGG AAG ATG AAG AG 3'
220T7	5' TTC GCG TAG TTA CAT ATT TAC 3'
2956T3	5' CTG CGT TGC TGT TTT TGC 3'
2956T7	5' CGC TAT TTA TTC CTG AAC ATA TTT TC 3'
2480T3	5' GCC AAA GGG TAA TTT GTT TGA TAG 3'
2480T7	5' TGC GAT TTG ACT TAT CTT AAT GG 3'
George	5' ATA CCT CGA CGT TAT TCA GA 3'
Eva	5' GAG ACC ATT ACT TGC TTT CAG TCA TCT 3'

Data analyses: Editing of nucleotide sequences was performed by using Sequencher 3.1 (Gene Codes Corp. Ann Arbor, MI) or Mac Vector 6.0 (Oxford Molecular Ltd., Oxford, UK). Sequences were aligned by eye using Se-Al v1.0 (Rambaut 1996 online) and compared to unpublished sequences of cDNA for each nuclear locus (courtesy of J. Roethele) or published COI/COII sequences (Smith and Bush 1997). Alignments (Appendices A-D) were exported as Nexus files for analyses of polymorphism and phylogenetic relationships. GenBank accession numbers for the sequences are being obtained at the time of the submission of this dissertation.

Measures of DNA polymorphism (number of haplotypes, haplotype diversity h (estimate of heterozygosity), nucleotide diversity π and θ, and number of polymorphic sites) were obtained using the program DnaSP ver. 3.50 (Rozas and Rozas 1999). This program was also used to calculate values of Tajima's D (Tajima 1989) as well as to perform the Hudson-Kreitman-Aguadé test (HKA, Hudson et al 1987) of neutrality (Kimura 1983). The average number of nucleotide substitutions per site between *R. pomonella* and *R. zephyria* (D_{xy}) and the number of net nucleotide substitutions per site between these two species (D_a) were also estimated using DnaSP, as well as the parameters that estimate gene flow (Nm).

Phylogenetic analyses were performed in PAUP* 4b (Swofford 1999) version 6. For nuclear loci, analyses were performed on a PowerMacintosh G3, with 512MB RAM memory and 400MB allocated to PAUP. For COI/COII, analyses were performed on a PowerMacintosh 8600/200, with 64MB RAM and 48MB allocated to PAUP. For each individual dataset, neighbor-joining and maximum parsimony analyses were done. Neighbor-joining analysis was based on pairwise Jukes-Cantor distances (Jukes and

Cantor, 1969), with *R. cingulata* as outgroup for P220, *R. basiola* as outgroup for P2956, *R. tabellaria* as outgroup for P2480 and *R. suavis* as outgroup for COI/COII. Identical alleles (haplotypes) were then removed from the analysis in order to decrease the number of possible rearrangements. Insertion/deletion events (indels) were hypothesized based on the aligned sequences. For P220 and P2480 indels were scored as present or absent and appended to the datasets (Sültman et al 1995). For P2956, the majority of the indels were overlapping so it was not possible to score them separately. For COI/COII there was only one 5-bp deletion in one sequence and this indel was also not scored. Gaps were then treated as missing data. In parsimony analyses, MP trees were first obtained using the heuristic search option with simple sequence addition and TBR branch swapping. The search was limited to saving 5,000 trees due to the large number of most parsimonious trees. Multiple searches were than performed to search for the islands of shorter trees as recommended by Maddison (1991) and using the method of Masta (2000) with 1000 replicate searches, each limited to saving 50 trees one step longer than the shortest trees found. In all cases these searches found trees that were shorter than the ones obtained by the initial simple sequence addition search. MPRs from these searches were then used as starting trees for TBR branch swapping.

Images in this dissertation are presented in color. In all dendrograms and phylograms presented here, green branches indicate *R. zephyria*, red *R. pomonella*, blue *R. mendax* and light-blue *R.* nr. *mendax* (flowering dogwood fly). Squares to the right of taxon label indicate *R. zephyria,* circles *R. pomonella*. Different colors of these symbols correspond to geographic regions: red – Northeast, orange – Great Lakes, green – Great Plains, blue – Rocky mountains, violet – Northwest and gray – South.

To assess the confidence estimates on the groups contained in the MP trees, Felsenstein's (1985) bootstrap approach was used with 500 replicates for both neighbor-joining and MP trees. For parsimony bootstrapping, the fast stepwise sequence addition option in PAUP* 4b6 was used. Branches with bootstrap confidence limits above 90 are considered to be well supported, between 70 and 90 moderately well supported and between 50 and 70 weakly supported.

RESULTS

Polymorphism. A summary of the number of sequences obtained at each locus is presented in Table 3.3. DNA sequence variation is presented in Table 3.4. Two main sampling strategies are usually found in phylogenetic studies based on DNA sequences (Funk 1999). In one strategy, a small number of individuals (or only one) is sampled from a large number of taxa and the phylogenies obtained from the sequence data are usually well resolved and supported (Funk et al. 1995). An alternate strategy consists of sampling several individuals from a number of geographic units, usually from the same species, focusing on population genetic processes within species (Avise et al. 1987). Both of these strategies have their limitations. When a small number of individuals per taxon is sampled, intraspecific variation may not be detected; if sample sizes are large but limited to one species inferences are also limited to that species. In this study, populations of *R. zephyria* and *R. pomonella* were sampled across their geographic ranges and a small number of alleles from each population was obtained for three nuclear and one mitochondrial gene. Therefore the sampling for the phylogeographic analysis of the

divergence between *R. zephyria* and *R. pomonella* conforms to the "large number of taxa, small number of individuals per taxon" strategy, since the populations of the two species were the main objects (or taxa) of the study, with the goal to obtain information that can provide insight on the process of speciation and divergence of *R. zephyria*.

The uneven number of sequences obtained from each taxon at each locus resulted from unsuccessful amplifications or clonings or sequences that were too ambiguous at multiple sites so that the calls could not be easily made. At P220 and P2956 haplotype diversity in *R. zephyria* is slightly higher than in *R. pomonella*, whereas at P2480 and COI/COII it is higher in *R. pomonella*. Summed over all nuclear loci, the weighted average value of θ/bp was higher in *R. zephyria* (0.0652) than in *R. pomonella* (0.0513). For mitochondrial COI/COII *R. pomonella* displays higher nucleotide polymorphism than *R. zephyria*. The percentage of polymorphic sites was similar in the two species at P2956 and P2480 (7.7 and 20.5% in *R. pomonella*, 7.9 and 19.8% in *R. zephyria*). At P220, 21.9% of sites were polymorphic in *R. zephyria*, compared to 13.9% in *R. pomonella*. At COI/COII, a low level of polymorphism was observed – only 3% of sites were polymorphic in *R. pomonella* and 2.3% in *R. zephyria*. No fixed differences were observed between *R. pomonella* and *R. zephyria* at any of the loci.

Table 3.3. Summary of the sample and loci analyzed in *R. zephyria*, *R. pomonella* and other taxa of the genus *Rhagoletis* in this phylogeographic study.

Locus	n_z	n_p	n_o	Link. gr.	Allozyme	Length
P220	67	46	32	I	*Aat-2*, *Dia-2*	425
P2956	13	36	96	II	*Acon-2*, *Mpi*, *Me*	521
P2480	70	34	11	III	*Had*	439
COI/COII	25	23	32	mt	n/a	1074

n_z – number of *R. zephyria* sequences. n_p – number of *R. pomonella* sequences. n_o – number of sequences obtained from other taxa. Link. gr. – linkage group (chromosome) where the locus is located in the *R. pomonella* genome (Roethele et al 1997). Allozyme – allozyme locus to which the locus under study is linked. Length – number of base pairs aligned.

Table 3.4. Summary of the polymorphism at three anonymous nuclear loci and mitochondrial COI/COII locus in *R. pomonella*, *R. zephyria* and other taxa of the genus *Rhagoletis*.

Locus	Length		N_{al}	h	N_{poly}	π	θ
P220	425	pom	30	0.926	57	0.031	0.059
		zep	48	0.950	93	0.028	0.073
		other	28	0.992	85	0.046	0.056
P2956	521	pom	22	0.943	40	0.020	0.021
		zep	11	0.962	41	0.018	0.032
		other	92	0.999	175	0.048	0.087
P2480	439	pom	32	0.996	90	0.070	0.080
		zep	51	0.984	87	0.057	0.097
		other	11	1.000	81	0.065	0.077
COI/COII	1074	pom	13	0.874	32	0.009	0.020
		zep	11	0.800	25	0.006	0.014
		other	22	0.970	139	0.041	0.059

pom – *R. pomonella*. zep – *R. zephyria*. Other – other taxa. N_{al} – number of different alleles (haplotypes). N_{poly} - number of polymorphic sites. h – haplotype diversity. π - nucleotide diversity. θ – nucleotide diversity, estimate of $4N_e\mu$.

Tests of selective neutrality. Natural selection acting on different loci can change expectations about lineage sorting and coalescence, so it is important to estimate whether selection is acting on the loci under study. Selection can be detected using Tajima's parameter D based on the comparison of two different estimates of θ which are strongly correlated (Tajima 1989). If the loci are neutral, both estimates are expected to be the same and D is expected to be 0. A values of D significantly different from 0 indicates that natural selection acts on the locus. Using this method, selection was detected at P220 and P2956 in *R. zephyria* and COI/COII in *R. pomonella* and *R. zephyria* (Table 3.5). An alternative method (Fu and Li, 1993) is based on weakly correlated estimates of θ. The test statistic G was significant at P220 and COI/COII in *R. pomonella* and *R. zephyria* and at P2480 in *R. zephyria* (Table 3.5). All estimates of D and G had negative values, indicating an excess of low frequency polymorphisms and slightly deleterious alleles (Li 1997).

Table 3.5. Detection of natural selection at the four loci in *R. pomonella, R. zephyria* and all other *Rhagoletis* taxa (pooled) using Tajima's (1989) and Fu and Li's (1993) tests.

Locus	Species	Tajima's D	Fu & Li's G
P220	pom	-1.648	-3.018*
	zep	-2.122*	-4.704**
	other	-0.674	-0.221
P2956	pom	-1.086	-1.966
	zep	-1.930*	-2.153
	other	-1.504	-1.731
P2480	pom	-1.062	-1.673
	zep	-1.404	-3.377**
	other	-0.722	-0.367
COI/COII	pom	-2.172**	-2.973*
	zep	-2.219**	-3.108**
	other	-1.190	-0.480

pom – *R. pomonella*, zep – *R. zephyria*, other – other *Rhagoletis* species. * - P<0.05, ** - P<0.01

Neutral mutation hypothesis predicts that the levels of within- and between-species DNA variation should be positively correlated (Kimura and Ohta 1971). This prediction can be tested by the Hudson-Kreitman-Aguadé (HKA) test which estimates whether the levels of observed polymorphism and divergence are consistent across loci (Hudson et al 1987). HKA test showed that levels of polymorphism and divergence across the three regions were consistent within and between species and none of the loci caused significant departure from neutrality in *R. pomonella* and *R. zephyria* (χ^2 non-significant, Table 3.6). Pairwise comparison of P2956 and COI/COII was not possible because none of the *R. pomonella* sequences at P2956 were obtained from the individuals from which sequences of COI/COII were obtained.

Table 3.6. HKA (Hudson-Kreitman-Aguadé) test for silent-site differences in four regions between *R. pomonella* and *R. zephyria*.

Region	seg		pwd		χ^2
	obs	exp	obs	exp	
P220	38	38.20	16.27	16.07	
vs. P2956	25	24.80	10.23	10.43	0.001
P220	33	33.89	9.64	8.75	
vs. P2480	65	64.11	15.67	16.56	0.018
P220	37	38.09	9.31	8.22	
vs. COII/COII	21	19.91	3.21	4.30	0.075
P2480	29	28.33	23.81	24.47	
vs. P2956	8	8.67	8.15	7.49	0.012
P2480	66	66.95	15.21	14.26	
vs. COI/COII	24	23.05	3.97	4.91	0.035

seg – number of segregating sites, pwd – average pairwise number of differences

Divergence between *R. pomonella* and *R. zephyria*. The average numbers of nucleotide substitutions per site between *R. pomonella* and *R. zephyria* (D_{xy}) for each locus are given in Table 3.7. The highest number of substitutions per site was observed at P2480, whereas the lowest was in COI/COII. The number of net nucleotide substitutions per site (D_a) between the two species was highest at P220 (Table 3.7) and lowest at COI/COII. Using these parameters and the Lynch and Crease (1990) method in DnaSP, gene flow between the species was estimated (Nm, Table 3.7). Values of Nm at P2480 and COI/COII were high (Table 3.7). At these two loci natural selection was detected using Fu and Li's test (Table 3.5). The estimate of Nm was the lowest at P220 for which Fu and Li's G was also significant in both *R. pomonella* and *R. zephyria* (Table 3.5).

Table 3.7. Summary of divergence parameters between *R. pomonella* and *R. zephyria*.

Locus	D_{xy}	D_a	Nm
P220	0.0526	0.0233	0.31
P2956	0.0205	0.0056	0.66
P2480	0.0682	0.0039	3.97
COI/COII	0.0075	0.0002	15.24

Phylogenetic analyses. Neighbor-joining analyses based on Jukes-Cantor distances and parsimony analyses were performed for each locus. The neighbor-joining trees (Figures 3.2, 3.5, 3.8 and 3.11) illustrate the relative genetic distances between the analyzed sequences. If the substitution rates at each locus were uniform (constant molecular clock within and across loci), branch lengths of neighbor-joining trees would be proportional to time since divergence. The constancy of the molecular clock at each locus so far has not been demonstrated in *Rhagoletis*; at different loci branch lengths differ considerably (Figures 3.2, 3.5, 3.8 and 3.11), which obscures estimates of times since divergence. Generally, the amount of change in the *R. pomonella* species group was found to be small (especially at COI/COII and P2956) which suggests recent divergence of the taxa (Figures 3.2, 3.5, 3.8, 3.11).

A summary of the parsimony analyses is given in Table 3.8. At all loci the number of informative characters used for parsimony analyses was low, and in some cases (P220) lower than the number of individual sequences analyzed. This complicated the analyses and resulted in thousands of equally parsimonious trees. Except for P2480, consensus trees were poorly resolved and bootstrap support for clades was low (Figures

3.4, 3.7, 3.10 and 3.13). Consistency indices were also low, while retention indices were relatively high (Table 3.8).

Table 3.8. Summary of the parsimony analyses of the DNA sequences at three anonymous nuclear loci and mitochondrial COI/COII locus.

Locus	n	length	n_i	n_{MPR}	TL	CI	RI	RC	HI
P220	145	425	105	264500[a]	262	0.523	0.896	0.468	0.477
P2956	145	521	148	285400[a]	309	0.628	0.937	0.588	0.372
P2480	115	439	139	1920	582	0.308	0.698	0.214	0.692
COI/COII	80	1074	141	89200[a]	300	0.613	0.780	0.483	0.387

n – number of sequences, n_i – number of informative characters, n_{MPR} – number of equally parsimonious trees obtained, TL – tree length, CI – consistency idex, RI – retention index, RC – rescaled consistency index, HI – homoplasy index; [a] indicates that the search was terminated when the number of saved MPRs exhausted the RAM before swapping on all trees.

Maximum parsimony analysis of P220 using 1000 replicate island searches yielded 26500 equally parsimonious trees of length 262. Additional TBR swapping on these trees yielded only trees of the same length (262) and exhausted the RAM after saving 264,500 trees. A random MPR (most parsimonious reconstructions) is shown in Figure 3.3. Strict consensus of the 264,500 trees is shown in Figure 3.4. Only 10 nodes within the *R. pomonella* species group were supported and overall bootstrap support was low.

Figure 3.2. P220 neighbor-joining tree based on Jukes-Cantor distances. Bootstrap values greater than 50% are shown. Green branches represent *R. zephyria*, red *R. pomonella*, blue *R. mendax* and light-blue *R.* nr. *mendax* (flowering dogwood fly). Squares next to the taxon name symbolize *R. zephyria*, and circles symbolize *R. pomonella*. Colors of the symbols represent geographic regions: red – Northeast, orange – Great Lakes, green – Great Plains, blue – Rocky Mountains and Colorado Plateau, violet – Northwest and gray – South. For taxon labels please refer to Key to symbols and abbreviations (p. xi).

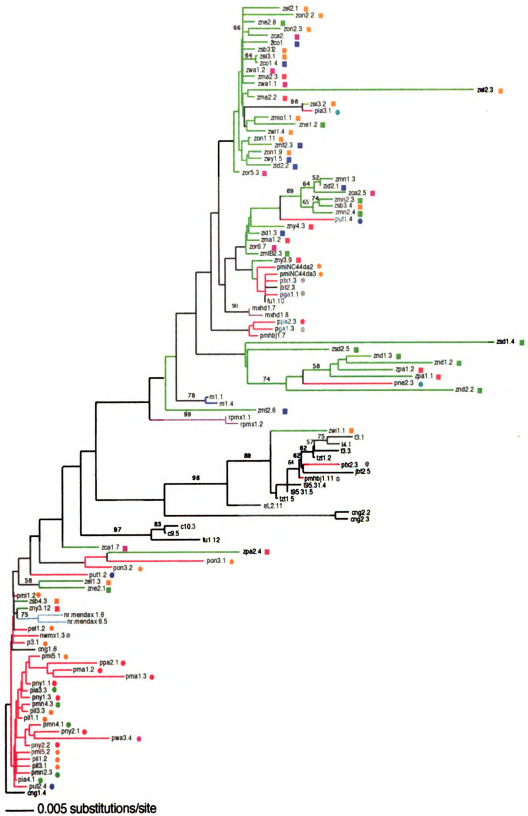


Figure 3.2.

Figure 3.3. P220 – a random most parsimonius tree (of the 264500 equal MPRs). Green branches represent *R. zephyria*, red *R. pomonella*, blue *R. mendax* and light-blue *R.* nr. *mendax* (flowering dogwood fly). Squares next to the taxon name symbolize *R. zephyria*, and circles symbolize *R. pomonella*. Colors of the symbols represent geographic regions: red – Northeast, orange – Great Lakes, green – Great Plains, blue – Rocky Mountains and Colorado Plateau, violet – Northwest and gray – South. For taxon labels please refer to Key to symbols and abbreviations (p. xi)

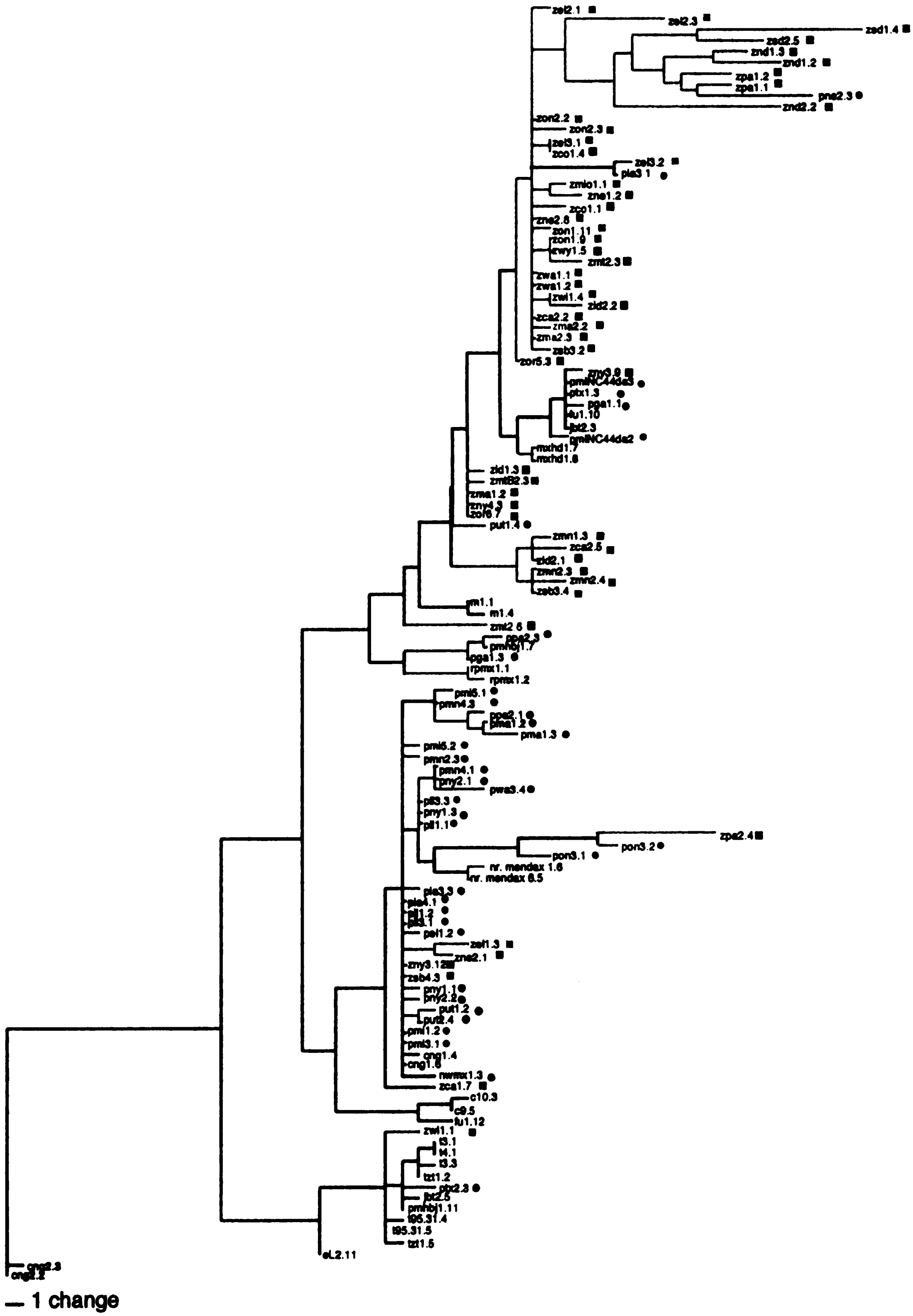

zel2.1
zel2.3
zed1.4
zed2.5
znd1.3
znd1.2
zpa1.2
zpa1.1
pna2.3
znd2.2
zon2.2
zon2.3
zel3.1
zco1.4
zel3.2
zmlo1.1
zne1.2
zco1.1
zne2.6
zon1.11
zon1.9
zwy1.5
zmt2.3
zwa1.1
zwa1.2
zwi1.4
zld2.2
zca2.2
zma2.2
zma2.3
zsb3.2
zor5.3
pmINC44da3
ptx1.3
pga1.1
fu1.10
pmINC44da2
mxhd1.7
mxhd1.6
zld1.3
zmlB2.3
zma1.2
put1.4
zmn1.3
zca2.5
zld2.1
zmn2.3
zmn2.4
zsb3.4
m1.1
m1.4
zml2.6
pga1.3
rpmx1.1
rpmx1.2
pmn4.3
pma1.3
pml6.2
pmn2.3
pmn4.1
pny2.1
pwa3.4
pll3.3
pny1.3
pll1.1
zpa2.4
pon3.2
pon3.1
nr. mendax 1.6
nr. mendax 6.5
pel1.2
zel1.3
zne2.1
zsb4.3
put1.2
pml1.2
pml3.1
cng1.4
cng1.6
nwmx1.3
zca1.7
c10.3
c9.5
fu1.12
zwl1.1
t3.1
t4.1
t3.3
tzt1.2
pbt2.3
pmnbj1.11
t95.31.4
t95.31.5
tzt1.5
eL2.11
cng2.3
1 change

Figure 3.3.

Figure 3.4. P220 - strict consensus of 264500 most parsimonious trees of length 262. CI=0.523, RI=0.896, RC=0.468. Bootstrap values greater than 50% are shown above branches. Green branches represent *R. zephyria*, red *R. pomonella*, blue *R. mendax* and light-blue *R.* nr. *mendax* (flowering dogwood fly). Squares next to the taxon name symbolize *R. zephyria*, and circles symbolize *R. pomonella*. Colors of the symbols represent geographic regions: red – Northeast, orange – Great Lakes, green – Great Plains, blue – Rocky Mountains and Colorado Plateau, violet – Northwest and gray – South. For taxon labels please refer to Key to symbols and abbreviations (p. xi).

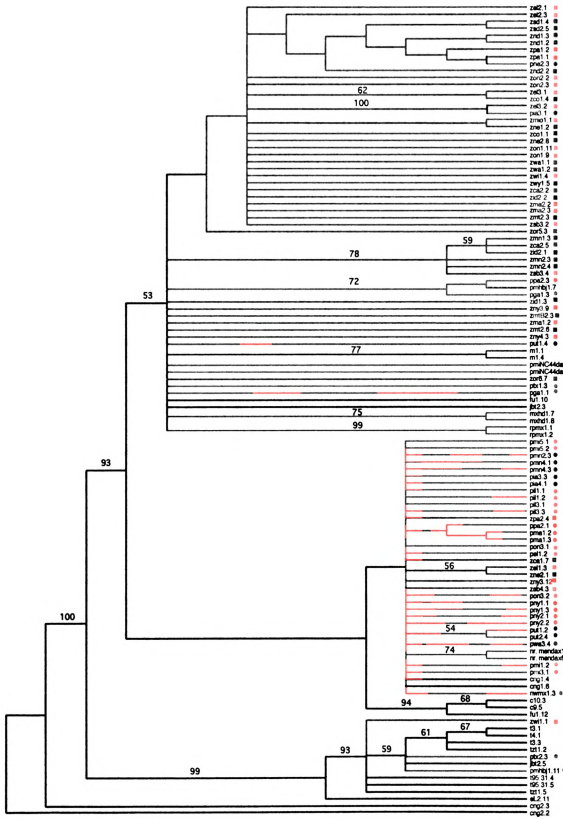


Figure 3.4.

P2956 was the only locus for which majority of the sequences obtained were from the taxa outside of the *R. pomonella* group (Table 3.3, Figures 3.5, 3.6 and 3.7). Both neighbor-joining and parsimony analyses at this locus suggest that *R. zephyria* groups within the *R. pomonella* species group (Figures 3.5, 3.6 and 3.7), with the majority of the groups weakly to moderately supported. Replicate island parsimony searches yielded 4550 trees of length 309, which were used as starting trees for TBR branch swapping. The search was aborted after saving 285,400 trees and exhausting the RAM. A random MPR (most parsimonious reconstruction) is shown in Figure 3. 6. A strict consensus of the MPRs at P2956 (Figure 3.7) retains the *R. pomonella* species group clade with poorly resolved relationships between the taxa and *R. zephyria* forming an unresolved group within this clade, with the exception of a single *R. zephyria* from New York. Bootstrap support for the clades was low to moderate (Figure 3.7).

Replicate island searches for the shortest parsimony trees at P2480 yielded 650 trees of length 582. After additional TBR swapping on these trees, 1920 trees of the same length were found. A random tree representing the 1920 trees is shown in Figure 3.9, and a strict consensus of these trees is shown in Figure 3.10. Although the resolution at this locus was good (better than at other loci), consistency index and retention index had low values (Table 3.8) and a low number of nodes was bootstrap supported (Figure 3.10).

Figure 3.5. P2956 – neighbor-joining tree (Jukes-Cantor distances). Bootstrap values greater than 50% are shown. Green branches represent *R. zephyria*, red *R. pomonella*, blue *R. mendax* and light-blue *R.* nr. *mendax* (flowering dogwood fly). Squares next to the taxon name symbolize *R. zephyria*, and circles symbolize *R. pomonella*. Colors of the symbols represent geographic regions: red – Northeast, orange – Great Lakes, green – Great Plains, blue – Rocky Mountains and Colorado Plateau, violet – Northwest and gray – South. For taxon labels please refer to Key to symbols and abbreviations (p. xi).

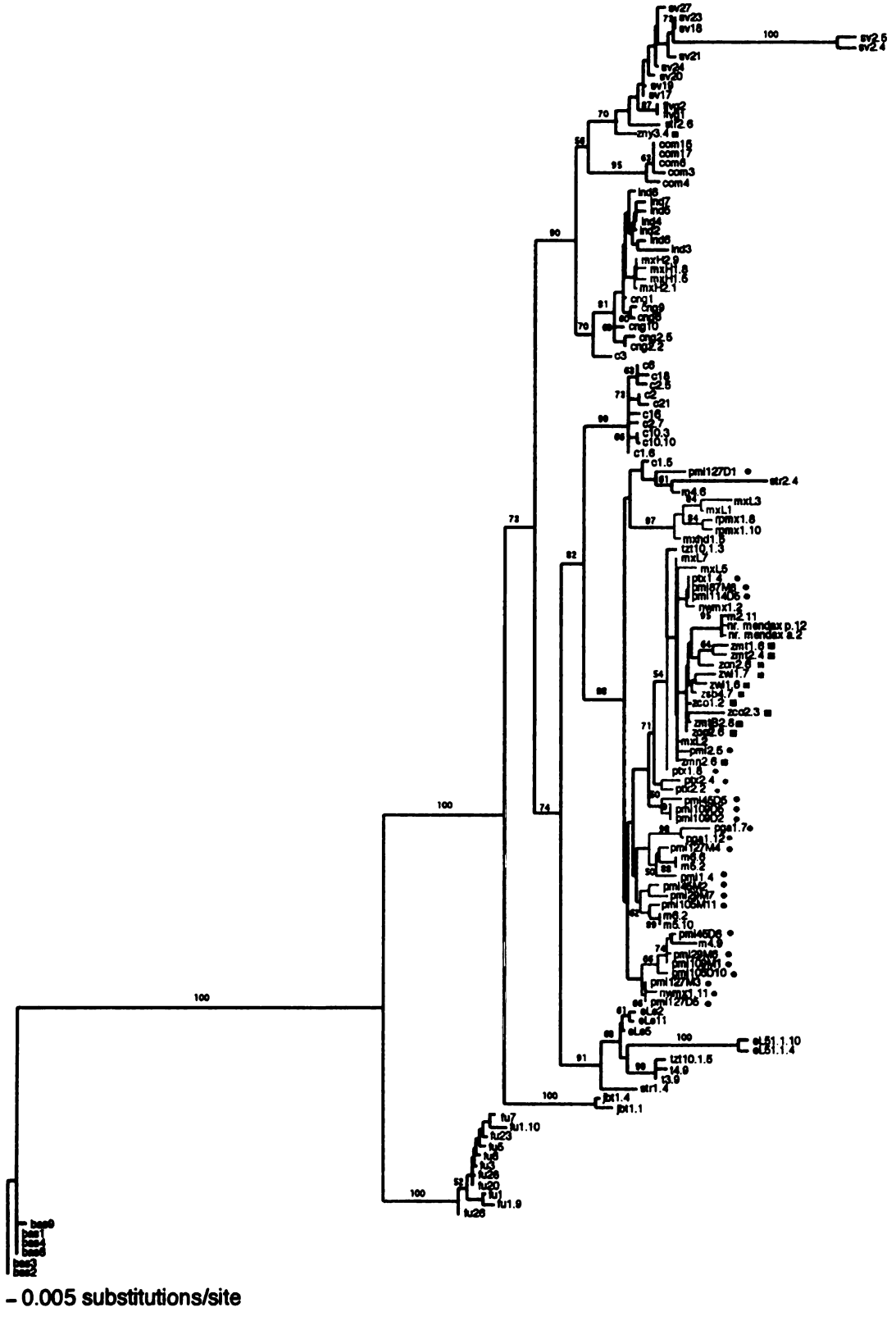

sv27
sv23
sv18
sv21
sv24
sv20
sv19
sv17
sv2.5
sv2.4
str2.6
zmy3.4
com16
com17
com6
com3
com4
ind6
ind7
ind5
ind4
ind2
ind8
ind3
mxH2.9
mxH1.8
mxH1.6
mxH2.1
cng1
cng9
cng8
cng10
cng2.5
cng2.2
c3
c6
c16
c2.6
c2
c21
c16.7
c10.3
c10.10
c1.6
c1.5
pmi127D1
str2.4
mxL3
mxL1
pmx1.8
pmx1.10
mxhd1.5
tzl10.1.3
mxL7
mxL5
pmi127M8
pmi114D5
pmx1.3
m2.11
nr. mendax p.12
nr. mendax a.2
zmi1.6
zmi2.4
zzn2.6
zwi1.7
zwi1.6
zsb3.7
zco1.2
zco2.3
zmi82.6
zzs2.6
mxL2
pmi2.5
zmn2.6
ppc1.8
ppc2.4
ppc2.2
pmi45D5
pmi108B8
pmi108D2
ppa1.7
ppa1.12
pmi127M4
m6.9
m6.2
pmi1.4
pmi45M2
pmi28M7
pmi108M11
m9.2
m5.10
pmi45D8
m4.8
pmi28M8
pmi108M1
pmi108D10
pmi127M3
pmx1.11
pmi127D6
eLe2
eLe11
eLe5
tzl1.1.10
eLt1.1.4
tzl10.1.5
t4.9
t3.9
str1.4
jbt1.4
jbt1.1
fu7
fu1.10
fu23
fu6
fu8
fu3
fu28
fu20
fu1
fu1.9
fu26
– 0.005 substitutions/site

Figure 3.5.

Figure 3.6. P2956 – a random most parsimonious tree (of 285400 equal MPRs). Green branches represent *R. zephyria*, red *R. pomonella*, blue *R. mendax* and light-blue *R.* nr. *mendax* (flowering dogwood fly). Squares next to the taxon name symbolize *R. zephyria*, and circles symbolize *R. pomonella*. Colors of the symbols represent geographic regions: red – Northeast, orange – Great Lakes, green – Great Plains, blue – Rocky Mountains and Colorado Plateau, violet – Northwest and gray – South. For taxon labels please refer to Key to symbols and abbreviations (p. xi).

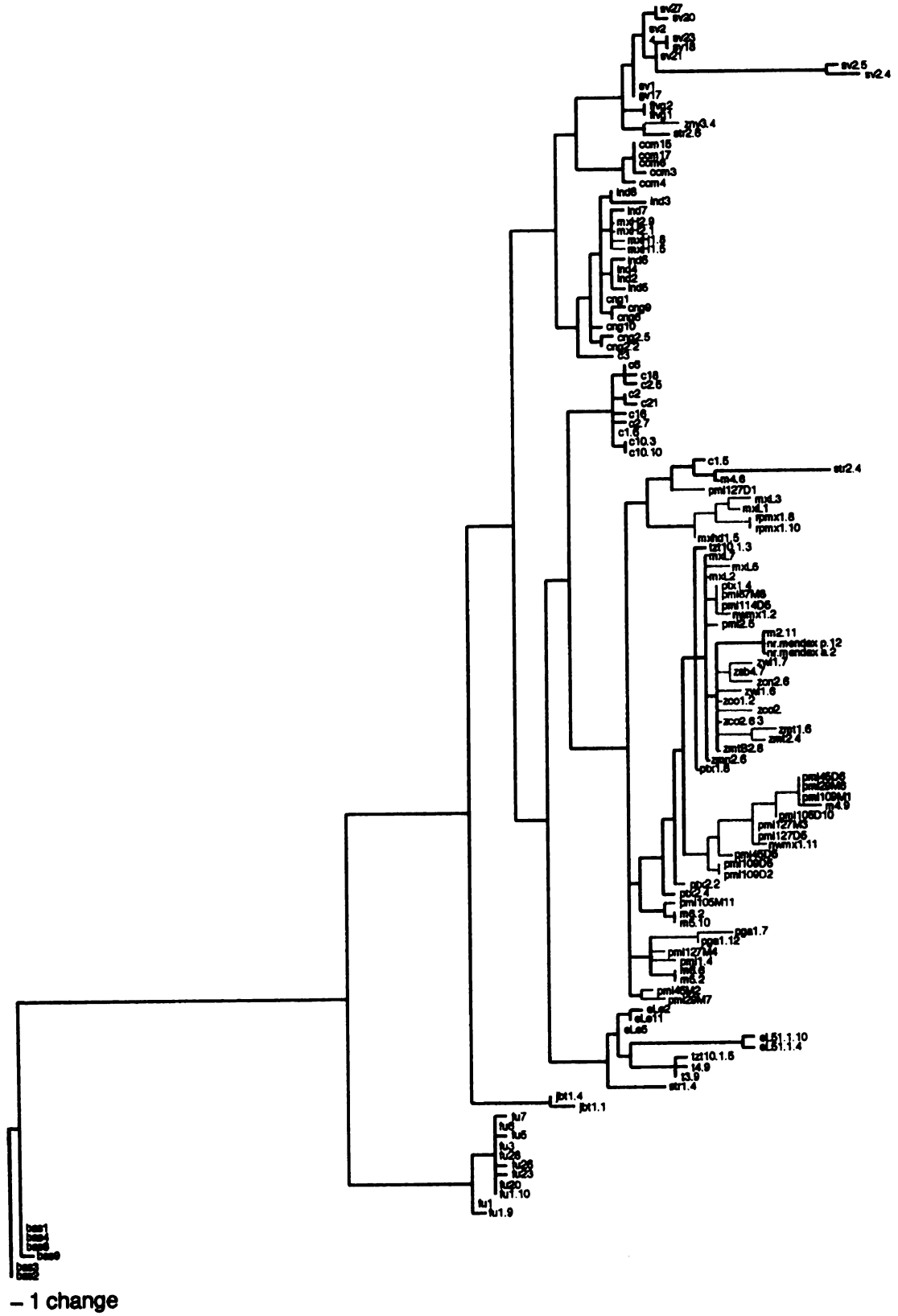


Figure 3.6.

Figure 3.7. P2956 - strict consensus of 285400 most parsimonious trees of length 309. CI=0.628, RI=0.937, RC=0.588. Bootstrap values greater than 50% are shown above branches. Green branches represent *R. zephyria*, red *R. pomonella*, blue *R. mendax* and light-blue *R.* nr. *mendax* (flowering dogwood fly). Squares next to the taxon name symbolize *R. zephyria*, and circles symbolize *R. pomonella*. Colors of the symbols represent geographic regions: red – Northeast, orange – Great Lakes, green – Great Plains, blue – Rocky Mountains and Colorado Plateau, violet – Northwest and gray – South. For taxon labels please refer to Key to symbols and abbreviations (p. xi).

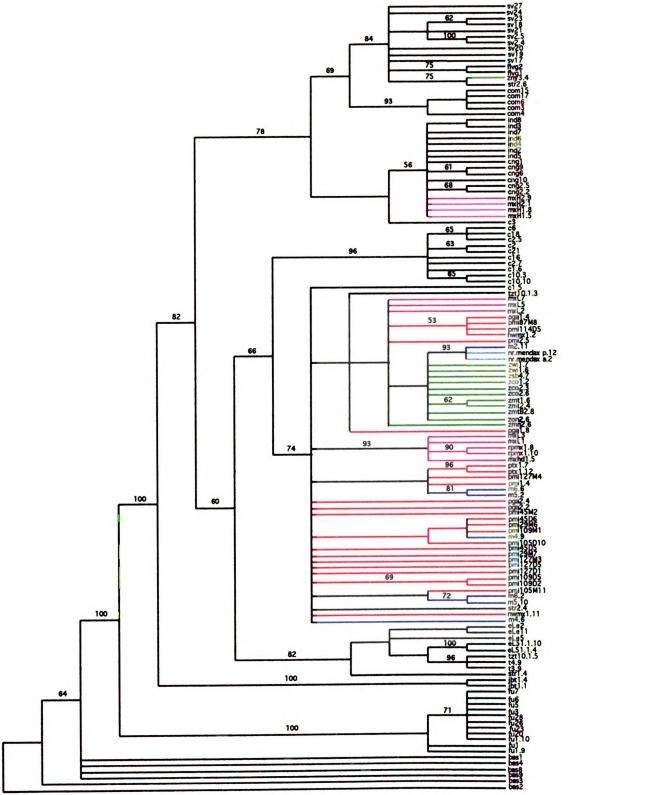


Figure 3.7.

Figure 3.8. P2480 – neighbor-joining tree (Jukes-Cantor distances). Bootstrap values greater than 50% are shown. Green branches represent *R. zephyria*, red *R. pomonella*, blue *R. mendax* and light-blue *R.* nr. *mendax* (flowering dogwood fly). Squares next to the taxon name symbolize *R. zephyria*, and circles symbolize *R. pomonella*. Colors of the symbols represent geographic regions: red – Northeast, orange – Great Lakes, green – Great Plains, blue – Rocky Mountains and Colorado Plateau, violet – Northwest and gray – South. For taxon labels please refer to Key to symbols and abbreviations (p. xi).


zor5.3
znd1.3
zmio1.2
zeb3.6
zwy2.4
zwy1.1
zoo2.2
zwy1.3
pmn2.2
pna1.2
pmn2.4
pmn1.1
pmn2.1
pna1.1
pna2.3
zny3.2
zna2.5
zma2.6
zwa95.6.4
zny3.3
zny4.2
zny4.5
zel1.1
zma1.
zel2.3
pel1.3
zoo1.3
zoo1.1
znd2.3
znd2.1
zwa95.1.3
zwa95.1.5
zel2.1
zon1.3
zor6.8
zor6.9
zmi95.10.2
zwa95.6.6
zmiB2.2
zta1.1
zid1.1
zeb3.3
zna2.2
zoo1.2
pmn1.2
pna2.2
znd1.3
zoo1.4
zwa2.3
zor2.4
zel1.3
znd2.2
zoo2.1
zon1.2
zwy1.1
zmio1.3
pma1.1
zpa1.1
pel1.1
zmiB1.3
zmn2.1
zca2.1
zmio2.6
zmio2.5
pny1.2
pny2.1
pny2.2
zwi3.2
zor2.2
pmi1.5
ppa2.6
zmt1.5
zmt1.7
zmtB1.2
nr. mendax.a3
zmt1.4
pny1.1
pors3.2
pwa3.4
pta3.1
pta3.3
pta4.3
ppa2.5
pt2.6
nr. mendax.p1
nr. mendax.p6
mt1.8
pmi2.5
zne2.3
zne2.1
zwi3.3
nr. mendax.a5
znd1.1
put1.6
put1.5
pmi5.1
pta4.1
pt2.5
mt1.11
C10.3
C10.4
c9.7
t4.2
t4.1
0.005 substitutions/site

Figure 3.8.

Figure 3.9. P2480 – a random most parsimonious tree (of 1920 MPRs). Green branches represent *R. zephyria*, red *R. pomonella*, blue *R. mendax* and light-blue *R.* nr. *mendax* (flowering dogwood fly). Squares next to the taxon name symbolize *R. zephyria*, and circles symbolize *R. pomonella*. Colors of the symbols represent geographic regions: red – Northeast, orange – Great Lakes, green – Great Plains, blue – Rocky Mountains and Colorado Plateau, violet – Northwest and gray – South. For taxon labels please refer to Key to symbols and abbreviations (p. xi).

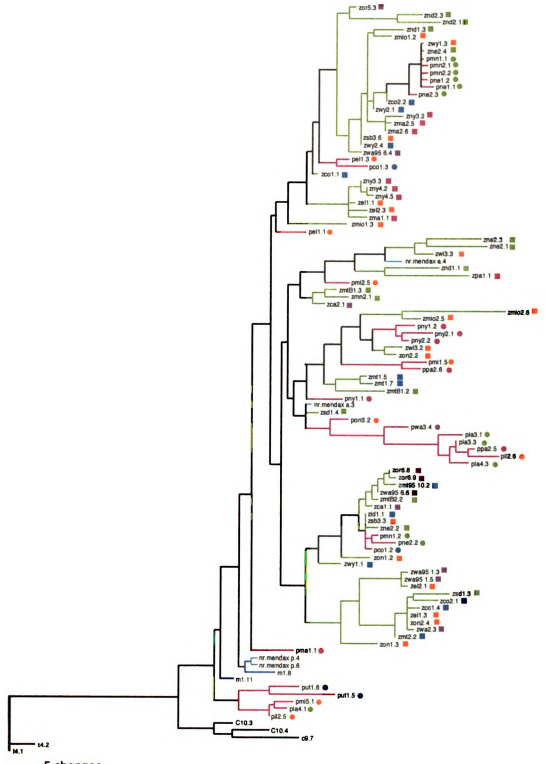
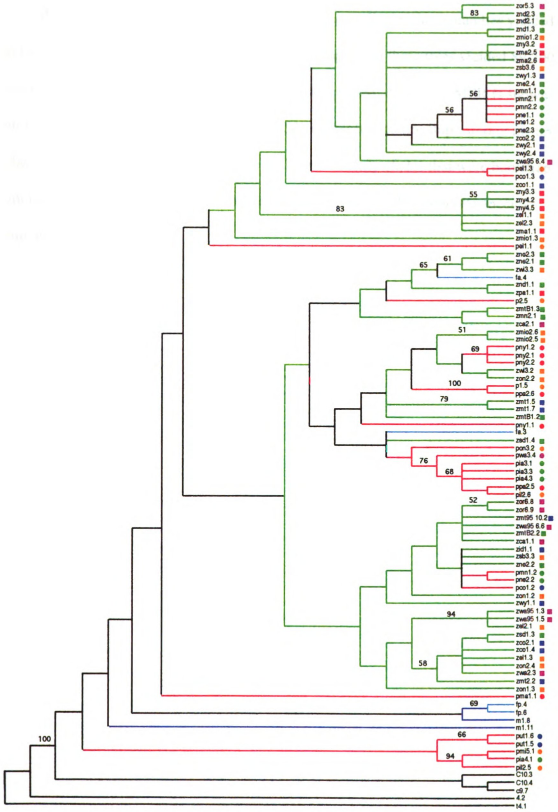

zor5.3
znd2.3
znd2.1
znd1.3
zmio1.2
zwy1.3
zwa2.4
pmn1.1
pmn2.1.9
pmn2.2
pne1.2
pne1.1
pne2.3
zoo2.2
zwy2.1
zmy3.2
zma2.5
zma2.6
zsb3.6
zwy2.4
zwa95 6.4
pei1.3
zco1.3
zco1.1
zmy3.3
zmy4.2
zmy4.5
zei1.1
zei2.3
zma1.1
zmio1.3
pei1.1
zwa2.3
zwa2.1
zwt3.3
nr.mendax a.4
znd1.1
zpa1.1
pmi2.5
zwt01.3
zmi2.1
zco2.1
zmio2.5
pny1.2
pny2.1
pny2.2
zwt3.2
zor2.2
pmi1.5
ppa2.6
zmt1.5
zmt1.7
zmt01.2
pny1.1
nr.mendax a.3
zsd1.4
pon3.2
pwa3.4
pla3.1
pla3.3
ppa2.5
pil2.6
pla4.3
zor6.8
zor5.9
zmt95 10.2
zwa95 6.6
zwt02.2
zca1.1
pil1.1
zsb3.3
zna2.2
pmn1.2
zna2.2
poo1.2
zon1.2
zwy1.1
zwa95 1.3
zwa95 1.5
zei2.1
zod1.3
zoo2.1
zco1.4
zei1.3
zor2.4
zwa2.3
zna2.2
zon1.3
pma1.1
nr.mendax p.4
nr.mendax p.6
ms1.8
mi1.11
pul1.6
pul1.5
pm5.1
pla4.1
pil2.5
C10.3
C10.4
c9.7
t4.2
t4.1
5 changes

Figure 3.9.

Figure 3.10. P2480 strict consensus of 1920 most parsimonious trees of length 582. CI=0.308, RI=0.698, RC=0.214. Bootstrap values greater than 50% are shown above branches. Green branches represent *R. zephyria*, red *R. pomonella*, blue *R. mendax* and light-blue *R.* nr. *mendax* (flowering dogwood fly). Squares next to the taxon name symbolize *R. zephyria*, and circles symbolize *R. pomonella*. Colors of the symbols represent geographic regions: red – Northeast, orange – Great Lakes, green – Great Plains, blue – Rocky Mountains and Colorado Plateau, violet – Northwest and gray – South. For taxon labels please refer to Key to symbols and abbreviations (p. xi).



zor5.3
znd2.3
znd2.1
znd1.3
zmio1.2
zny3.2
zma2.5
zma2.6
zsb3.6
zwy1.3
zne2.4
pmn1.1
pmn2.1
pmn2.2
pne1.1
pne1.2
pne2.3
zco2.2
zwy2.1
zwy2.4
zwa95 6.4
pel1.3
pco1.3
zco1.1
zny3.3
zny4.2
zny4.5
zel1.1
zel2.3
zma1.1
zmio1.3
pel1.1
zne2.3
zne2.1
zwi3.3
fa.4
znd1.1
zpa1.1
p2.5
zmtB1.3
zmn2.1
zca2.1
zmio2.6
zmio2.5
pny1.2
pny2.1
pny2.2
zwi3.2
zon2.2
p1.5
ppa2.6
zmt1.5
zmt1.7
zmtB1.2
pny1.1
fa.3
zsd1.4
pon3.2
pwa3.4
pia3.1
pia3.3
pia4.3
ppa2.5
pil2.6
zor6.8
zor6.9
zmt95 10.2
zwa95 6.6
zmtB2.2
zca1.1
zid1.1
zsb3.3
zne2.2
pmn1.2
pne2.2
pco1.2
zon1.2
zwy1.1
zwa95 1.3
zwa95 1.5
zel2.1
zsd1.3
zco2.1
zco1.4
zel1.3
zon2.4
zwa2.3
zmt2.2
zon1.3
pma1.1
fp.4
fp.6
m1.8
m1.11
put1.6
put1.5
pmi5.1
pia4.1
pil2.5
C10.3
C10.4
c9.7
4.2
t4.1
83
56
56
55
83
65
61
51
69
100
79
76
68
52
94
58
69
66
94
100

Figure 3.10.

Parsimony analysis of the COI/COII data using replicate island searches yielded 280 trees of length 300. Additional TBR branch swapping on these trees yielded 89,200 trees of the same length (Table 3.8) before exhausting RAM. A random MPR is shown in Figure 3.12. A strict consensus of the MPRs did not resolve relationships among the taxa of the *R. pomonella* species group (Figure 3.13) and the only clades that were bootstrap supported were the clades consisting of the taxa outside of the *R. pomonella* species group and *R. cornivora*.

Figure 3.11. COI/COII – neighbor-joining tree (Jukes-Cantor distances). Bootstrap values greater than 50% are shown. Green branches represent *R. zephyria*, red *R. pomonella*, blue *R. mendax* and light-blue *R.* nr. *mendax* (flowering dogwood fly). Squares next to the taxon name symbolize *R. zephyria*, and circles symbolize *R. pomonella*. Colors of the symbols represent geographic regions: red – Northeast, orange – Great Lakes, green – Great Plains, blue – Rocky Mountains and Colorado Plateau, violet – Northwest and gray – South. For taxon labels please refer to Key to symbols and abbreviations (p. xi).

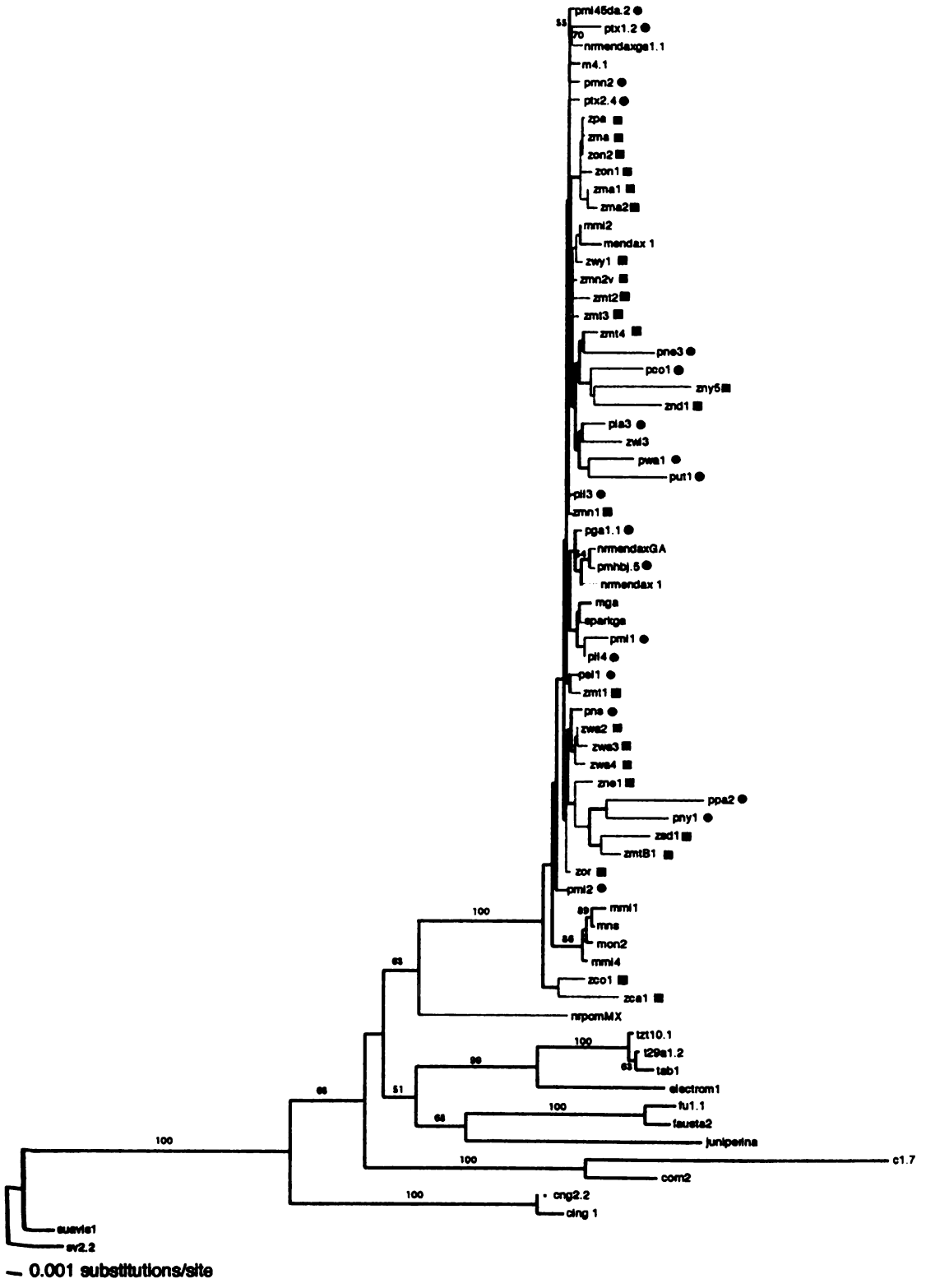

pml45da.2
ptx1.2
nrmendaxga1.1
m4.1
pmn2
ptx2.4
zpe
zma
zon2
zon1
zma1
zma2
mml2
mendax 1
zwy1
zmn2v
zmt2
zmt3
zmt4
pne3
pco1
zny5
znd1
pla3
zwl3
pwa1
put1
pll3
zmn1
pga1.1
nrmendaxGA
pmhbj.5
nrmendax 1
mga
spartga
pml1
pll4
pel1
zmt1
pne
zwa2
zwa3
zwa4
zne1
ppa2
pny1
zad1
zmtB1
zor
pml2
mml1
mns
mon2
mml4
zco1
zca1
nrpomMX
tzt10.1
t29a1.2
tab1
electrom1
fu1.1
tausta2
juniperina
c1.7
com2
cng2.2
cing 1
suavis1
sv2.2
0.001 substitutions/site

Figure 3.11.

Figure 3.12. COI/COII – a random most parsimonious tree (of 89200 MPRs). Green branches represent *R. zephyria*, red *R. pomonella*, blue *R. mendax* and light-blue *R.* nr. *mendax* (flowering dogwood fly). Squares next to the taxon name symbolize *R. zephyria*, and circles symbolize *R. pomonella*. Colors of the symbols represent geographic regions: red – Northeast, orange – Great Lakes, green – Great Plains, blue – Rocky Mountains and Colorado Plateau, violet – Northwest and gray – South. For taxon labels please refer to Key to symbols and abbreviations (p. xi).

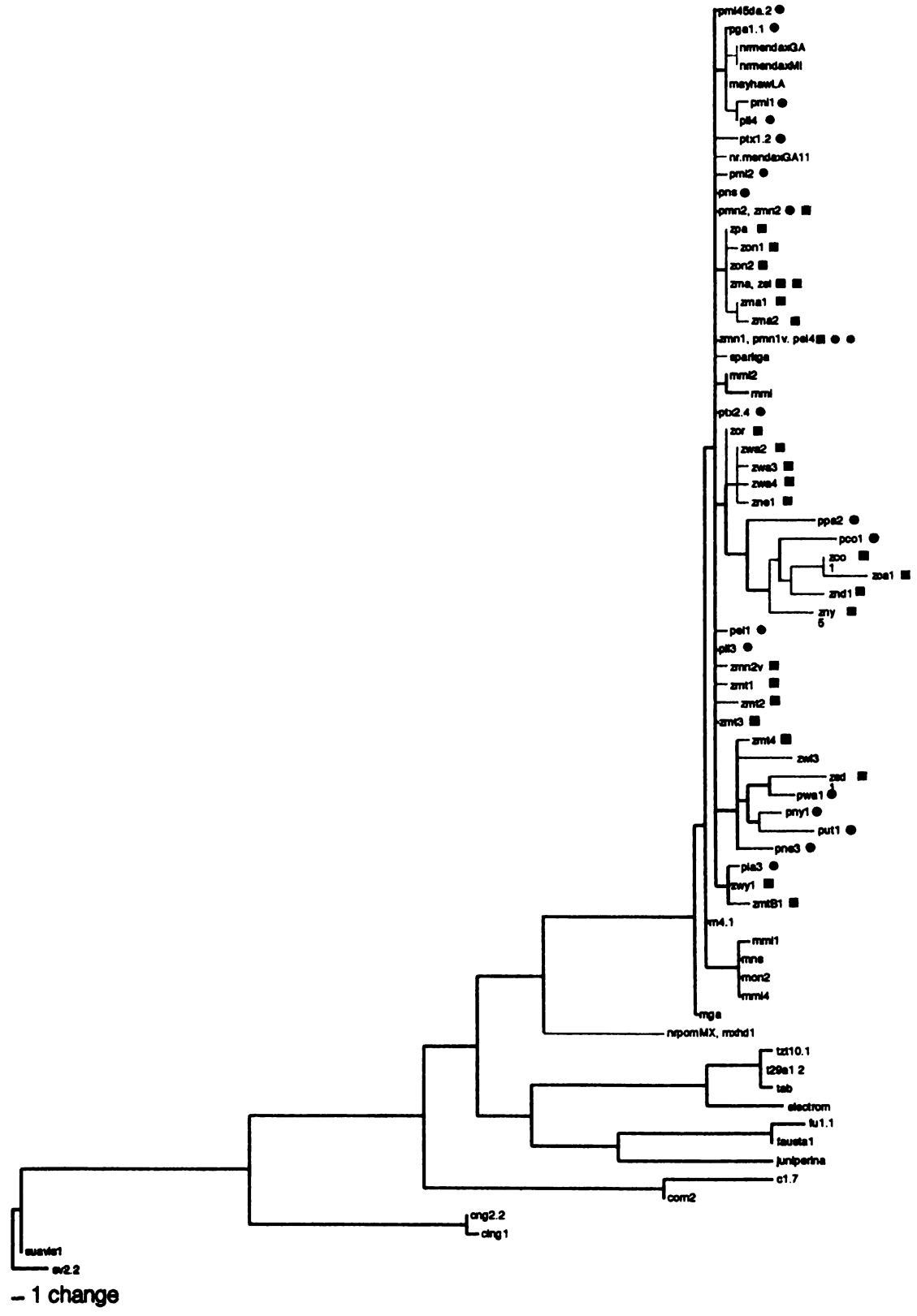

pmi45da.2
pge1.1
nrmendaxGA
nrmendaxMI
mayhawLA
pmi1
pti4
ptx1.2
nr.mendaxGA11
pmi2
pns
pmn2, zmn2
zpa
zon1
zon2
zma, zat
zma1
zma2
zmn1, pmn1v, pel4
sparliga
mmi2
mmi
ptx2.4
zor
zwa2
zwa3
zwa4
zwa1
ppa2
pco1
zpo
1
zpa1
znd1
zny
5
pel1
pti3
zmn2v
zmt1
zmt2
zmt3
zmt4
zwt3
zad
pwa1
pny1
put1
pna3
pla3
zwy1
zmtB1
m4.1
mmi1
mna
mon2
mmi4
mga
nrpomMX, mdhd1
tzt10.1
t29a1 2
tab
electrom
fu1.1
fausta1
juniperina
c1.7
com2
cng2.2
cng1
eu.vir1
ev2.2
– 1 change

Figure 3.12.

Figure 3.13. COI/COII - strict consensus of 89200 most parsimonious trees of length 300. CI=0.613, RI=0.780, RC=0.483. Bootstrap values greater than 50% are shown above branches. Green branches represent *R. zephyria*, red *R. pomonella*, blue *R. mendax* and light-blue *R.* nr. *mendax* (flowering dogwood fly). Squares next to the taxon name symbolize *R. zephyria*, and circles symbolize *R. pomonella*. Colors of the symbols represent geographic regions: red – Northeast, orange – Great Lakes, green – Great Plains, blue – Rocky Mountains and Colorado Plateau, violet – Northwest and gray – South. For taxon labels please refer to Key to symbols and abbreviations (p. xi).

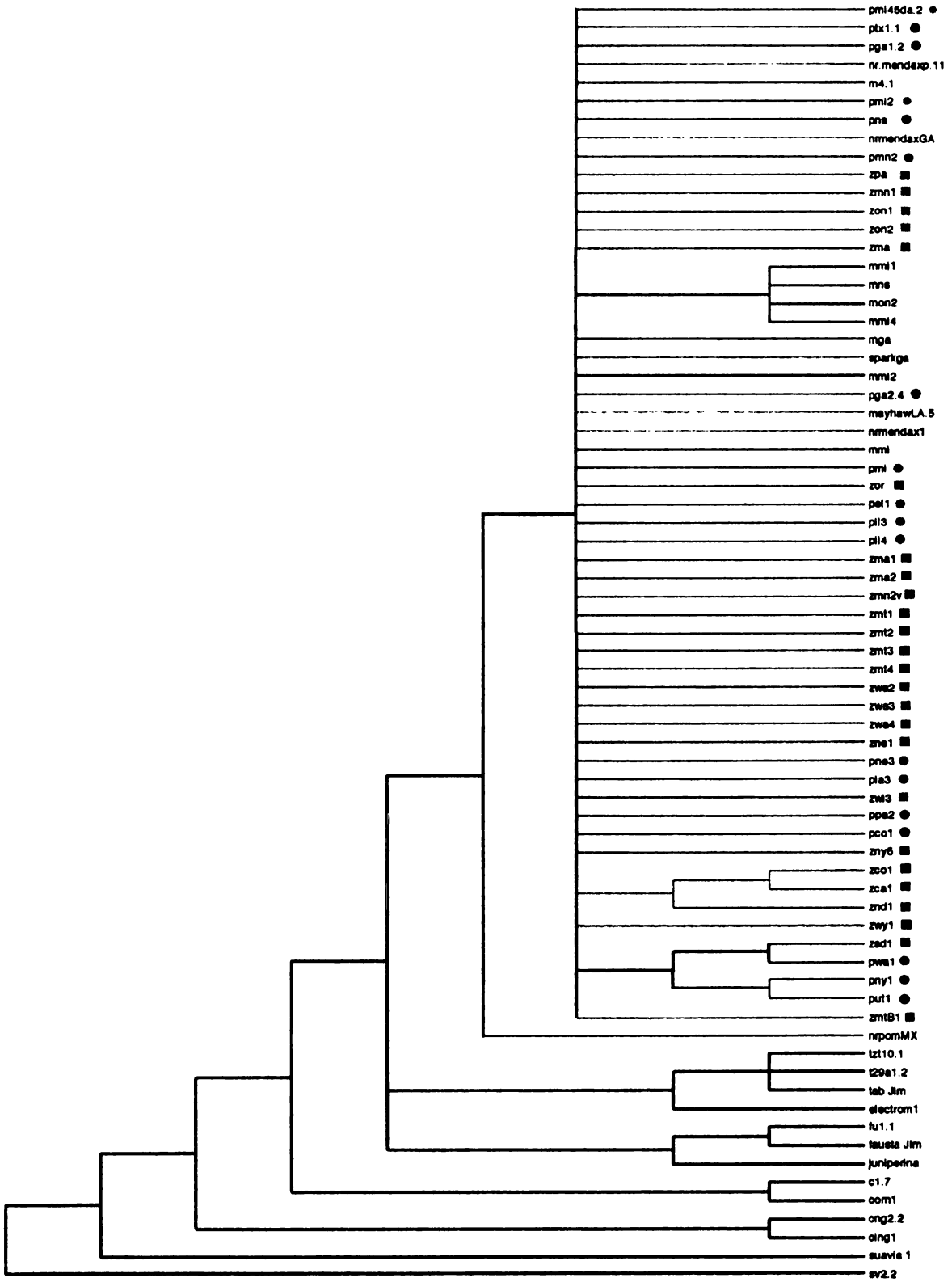

pml45da.2
ptx1.1
pga1.2
nr.mendaxp.11
m4.1
pml2
pns
nrmendaxGA
pmn2
zpa
zmn1
zon1
zon2
zma
mml1
mns
mon2
mml4
mga
sparkga
mml2
pga2.4
mayhawLA.5
nrmendax1
mml
pml
zor
pel1
pll3
pll4
zma1
zma2
zmn2v
zmt1
zmt2
zmt3
zmt4
zwe2
zwe3
zwe4
zne1
pne3
pla3
zwl3
ppa2
pco1
zny6
zco1
zca1
znd1
zwy1
zsd1
pwa1
pny1
put1
zmtB1
nrpomMX
tzt10.1
t29a1.2
tab Jim
electrom1
fu1.1
fausta Jim
juniperina
c1.7
com1
cng2.2
cing1
suavis 1
sv2.2

Figure 3.13.

DISCUSSION

Phylogenetic position of *R. zephyria*. At all four loci analyzed here, both *R. pomonella* and *R. zephyria* appear to be polyphyletic (Figures 3.2-2.13). Results obtained from the different loci support the hypothesis that *R. zephyria* is a sister taxon to *R. pomonella* and that it has diverged from the common ancestor more recently than *R. mendax*. At all loci sequences of *R. zephyria* and *R. pomonella* are placed within the same clades, and at P2956, P2480 and COI/COII alleles of *R. mendax* appear to have diverged from the common ancestor with the present-day *R. pomonella* before alleles of *R. zephyria* (Figures 3.7, 3.10 and 3.12). This is consistent with findings of Bush and Smith (1997) and McPheron and Han (1997) based on mitochondrial DNA sequences. Allozyme data (Berlocher et al 1993) suggested a different pattern and placement of *R. zephyria* at the base of the closely related taxa of the *R. pomonella* species group (without the more diverged *R. cornivora*). Recent analysis based on allozyme data agreed with the previous one on basal placement of *R. zephyria*; however, this analysis included only two populations of *R. zephyria* and it was suggested that the alternative placement (within the group as a sister to *R. pomonella*) was not significantly poorer (Berlocher 2000).

At P220 most of the *R. zephyria* sequences were included in the same clade (Figure 3.3) which also included *R. pomonella* from Michigan, Texas and Georgia, *R.* nr. *pomonella* from La Jolla, MX, *R. fausta* and *R. juniperina*. Within this clade two smaller clades of *R. zephyria* were observed. Within the larger of these two clades two *R. pomonella* sequences were found, both from the Great Plains region. Other sequences of *R. zephyria* or *R. pomonella* do not fall into the clades which reflect geographic regions.

A large *R. zephyria* clade is retained in the strict consensus with the same *R. pomonella* sequences from Great Plains within the clade. A clade formed by *R. tabellaria* and *R. electromorpha*, sister species from the *R. tabellaria* species group, was also retained in the strict consensus tree (Figure 3.4). The remainder of the tree was poorly resolved (Figure 3.4).

At P2956, species of the *R. pomonella* group formed a clade within which a single *R. striatella* allele sequence was placed (Figure 3.6). *R. zephyria* formed a sister clade to one of the *R. pomonella* clades, except for a single individual from New York (zny 3.4), which was a sister taxon to another *R. striatella* and placed in a clade with *R. suavis* and *R. flavigenualis* which are distantly related to the *R. pomonella* species group. The clade of *R. pomonella* species group (excluding some of the *R. cornivora* sequences which were placed as a sister group to *R. suavis* and *R. flavigenualis*, Figure 3.7) was weakly supported (bootstrap = 66). Moderate support was observed for the clade consisting of *R. pomonella* group taxa without *R. cornivora* (bootstrap = 74). Clades consisting of the taxa outside of the *R. pomonella* species group were weakly to moderately supported (Figure 3.7).

The relationships among the *R. pomonella* species group taxa suggested by the maximum parsimony analysis of the P2480 were consistent with the relationships suggested by allozyme data (Berlocher et al 1993). *R. cornivora* formed a sister clade to the other *R. pomonella* group species (Figure 3.10). At the base of the more closely related *R. pomonella group* taxa (without *R. cornivora*) was the clade formed by five *R. pomonella* sequences. *R. mendax* and *R.* nr. *mendax* branched off from the remainder of the taxa before branching of *R. zephyria* and *R. pomonella*. Sequences of *R. zephyria* and

R. pomonella (except for five in the basal clade) were placed in three distinct clades, none of which corresponded to species. One of the three clades consisted mostly of *R. zephyria* from the populations west of the Mississippi River Valley, with three *R. pomonella* sequences (from Minnesota, Nebraska and Colorado) within this clade. Another large clade consisted of *R. pomonella* and *R. zephyria* from both sides of the Mississippi, with some structuring of *R. zephyria* into a predominantly western and predominantly eastern group. *R. pomonella* sequences were placed with predominantly eastern *R. zephyria* (except one *R. pomonella* from Michigan found in the "western" *R. zephyria* clade, Figure 3.10). The third large clade contained mostly sequences of *R. zephyria* from the Northeast and Great Lakes, with *R. pomonella* predominately from the Great Plains.

At COI/COII branch lengths within the *R. pomonella* species group were short, which is consistent with the low levels of nucleotide polymorphism (Table 3.4) and low sequence divergence between *R. pomonella* and *R. zephyria* (Table 3.7). A unique haplotype was found in *R. zephyria* populations ranging from Massachusetts to Michigan. Of all the eastern *R. zephyria*, only haplotypes found in Wisconsin and New York populations was placed into the clade consisting of western *R. zephyria* and both eastern and western *R. pomonella* (Figure 3.12). All mtDNA haplotypes observed in western *R. zephyria* were either shared with or more closely related to *R. pomonella*.

The inability to better resolve the relationships among taxa in this study at three out of four loci may stem from the very low number of informative characters and high number of taxa (sequences) analyzed. High number of taxa also deflates consistency indices, regardless of any change in information content (Kitching et al 1998). It is interesting to note that the only well-resolved locus in this study was P2480, which is

tightly linked to *Had*, the only allozyme locus that displays fixed difference between *R. pomonella* and *R. zephyria*.

Some of the discrepancies between the gene trees presented here may be caused by different forms of selection acting on the different genes. At P220, alleles from the same heterozygous individuals were found in divergent clades, suggesting that balancing selection may have played a role in shaping the phylogenetic pattern observed at this locus. Estimates of Tajima's D and Fu and Li's G at all loci were negative (Table 3.5). The more robust Fu and Li's statistic detected significant selection at P220 and COI/COII in both *R. pomonella* and *R. zephyria* and at P2480 in *R. zephyria* (Table 3.5). This finding was expected for the mitochondrial COI/COII gene region which codes for the two subunits of cytochrome oxidase and has been shown to be conserved within the *R. pomonella* species group (Smith and Bush 1997). Low nucleotide diversity in both species and high similarity of mitochondrial COI/COII haplotypes indicate that this gene region remains highly conserved under strong selection pressure that does not eliminate only the mutations that maintain the functionality of the gene. Negative values of Tajima's D or Fu and Li's G reflect an excess of low frequency polymorphisms, which can result from recent selection so that the variation is removed, or by a recent population size expansion (Tajima 1989, Kliman et al 2000). The HKA test, on the other hand, did not indicate a significant contribution of any of the loci to deviations from the neutral mutation hypothesis (Table 3.6). Presence of selection also influences the estimates of gene flow between *R. pomonella* and *R. zephyria* (Table 3.7). The indirect methods of estimating gene flow (average number of migrants per generation, Nm) are based on the parameters of divergence between the two species (average number of nucleotide

substitutions and number of net substitutions per site at each locus, Table 3.7) and the assumption that the alleles at the loci under study are selectively neutral. If alleles are not selectively neutral, and if selection favors the same allele everywhere (directional selection), the rate of gene flow tends to be overestimated (Futuyma 1997), which can explain the high Nm estimated for P2480 and COI/COII (Table 3.7). For these two loci selection was detected by Fu and Li's test (Table 3.5). Rate of gene flow can also be overestimated if speciation has been recent and the two sister species have not had enough time to differentiate. The high Nm at these two loci seems to support the hypothesis of a recent speciation event. If selection favors different alleles in the two species, the rate of gene flow tends to be underestimated. Selection was detected at P220 by Fu and Li's test (Table 3.5), however, Nm estimated for this locus was smaller (0.31) than for the neutral P2956 (0.66).

Inferences about speciation from phylogeography. Overall phylogeographic patterns observed seem to be consistent with recent speciation. If the speciation has occurred within the past 10,000 years, unresolved relationships and shared ancestral genes are expected to be found (Masta 2000). Host fidelity in *Rhagoletis* is high (Bush 1966) and hybridization between *R. pomonella* and *R. zephyria* occurs only rarely (Feder et al 1999). The phylogenetic patterns observed at all the loci, lack of fixed differences, sharing of alleles and haplotypes between the two species and the placement of alleles derived from geographically very distant populations can be explained by the recent speciation and retention of ancestral polymorphism. Alternatively, these patterns could be explained by interspecific gene flow. Other types of data are consistent with the recent speciation hypothesis as well – *R. pomonella* and *R. zephyria* share alleles at all allozyme

loci except for *Had* (Berlocher and Bush 1982, Berlocher et al 1993, Feder et al 1999) and are morphologically identical except for the shape of male genitalia (Bush 1966). Low levels of genetic divergence are often found between ecologically differentiated species which speciated by habitat shifts (sticklebacks – Schluter 1996, cichlid fishes – Albertson et al 1999, Danley et al 2000, damselflies – Brown et al 2000) and the results reported here are consistent with those findings. One possible alternative explanation for sharing alleles and haplotypes between species would be if the recurrent mutations occur in the two species independently (Kliman et al 2000). However, it would be difficult to explain the low levels of sequence divergence observed at all unlinked loci by this phenomenon.

Generally, little geographic structuring was observed at any of the loci. When the sequences of *R. pomonella* were found in predominantly *R. zephyria* clades (such as at P220 and P2480), they tended to be from the populations of *R. pomonella* sampled in the Great Plains (Minnesota, Nebraska, Iowa), indicating that the host shift from hawthorns to snowberries and speciation of *R. zephyria* occurred in the Plains, in the zone where *Crataegus* sp. and *Symphoricarpos* sp. were in contact when the glaciers receded after the last glaciation period.

At P220 and P2480, some of the *R. pomonella* sequences were placed at the base of the *R. pomonella* species group, supporting the hypothesis that the common ancestor may have been *R. pomonella*-like. Based on the working hypothesis that the common ancestor of *R. pomonella* and *R. zephyria* was *R. pomonella*-like and infested hawthorn fruits (host of the present-day *R. pomonella*), the expectation was that *R. pomonella*

would show a higher number of alleles, higher average nucleotide diversity and higher heterozygosity across its geographic range than *R. zephyria*.

Both species have high levels of polymorphism estimated by nucleotide diversities (π and θ) and haplotype diversity (h, also an indirect estimate of heterozygosity) (Table 3.4). The levels of polymorphism observed in *R. pomonella* and *R. zephyria* were much higher than in *Drosophila simulans*, *D. mauritiana* and *D. sechellia* (Kliman et al 2000). Only at COI/COII was the amount of variation low, as expected for *Rhagoletis* mitochondrial genes (Smith and Bush 1997, Smith and Bush 2000). Over all loci studied, *R. zephyria* had higher weighted average of θ/bp (0.0652) than *R. pomonella* (0.0513), contrary to the expectation that *R. pomonella* should have higher diversity. However, at some loci (P2480) sequences of *R. pomonella* were not obtained from the entire geographic range, and generally numbers of sequences obtained were uneven at all loci, which may have contributed to the unexpected finding that *R. zephyria* appears to be more genetically variable than *R. pomonella*. Even so, it appears that the speciation of *R. zephyria* was not accompanied by bottleneck.

It should be noted that high genetic diversity does not necessarily negate population bottleneck at speciation – whether the bottleneck can be reflected by the present genetic structure depends on how recent the speciation event is and how long the bottleneck lasted relative to the effective population size (Eyre-Walkers 1998). Given the lack of fixed differences between *R. pomonella* and *R. zephyria* at the loci studied here and the similarity of their sequences reflected by the small numbers of phylogenetically informative characters, shared alleles and haplotypes between individuals from different species, relatively short branch lengths in neighbor-joining analyses and unresolved

phylogenetic relationships (Table 3.8, Figures 3.2-3.13, Smith and Bush 1997, McPheron and Han 1997), the scenario in which the speciation occurred relatively recently seems plausible. Therefore, it is unlikely that if the bottleneck played a role in speciation it would not be reflected by the present structure.

It is difficult to infer the direction of population expansion from the phylogenies at different loci (Figures 3.2-3.13), since little geographic structuring is revealed. However, the better resolved P2480 MPRs suggest some structuring of *R. zephyria*, supporting the hypothesis that dispersal to the west preceded the expansion to the east (Figure 3.10). Of the three clades where *R. zephyria* is placed in the strict consensus of 1920 most parsimonious trees, the basal one consists mostly of alleles from populations west of the Mississippi, indicating that these populations may have had more time to adapt to local conditions and accumulate new mutations than the populations in the east.

Incomplete lineage sorting and the presence of both eastern and western alleles of *R. zephyria* in the same clades at P2480 and P220 seems to support the hypothesis that *R. zephyria* was not introduced to the Eastern North America, consistent with findings of infestations of native hosts east of the Mississippi River and infestations of introduced hosts only when the native hosts have been present (Chapter 2). If introduction had occurred, populations in the east should have gone through a bottleneck, which would be reflected in the MPRs by shorter branch lengths, similarity of the sequences in the east and their placement together. This was not observed (Figures 3.2-3.13). However, a hypothesis that multiple introductions of *R. zephyria* from different areas in the west into the east cannot be ruled out.

One possible way to obtain more information about the processes involved in divergence of the two closely related species may be to assess genetic variation at many more loci and more individuals, which I did using the amplified fragment length polymorphism (AFLP) DNA fingerprinting (Chapter 4).

CHAPTER 4

DIVERGENCE OF *R. POMONELLA* AND *R. ZEPHYRIA* AS REVEALED BY DIFFERENCES IN AFLP PATTERNS

INTRODUCTION

Attempts to resolve phylogenetic relationships among taxa of the *R. pomonella* species group and/or to determine phylogenetic position of *R. zephyria* so far have used allozymes (Berlocher et al 1993, Feder et al 1999, Berlocher 2000), morphological characters (Jenkins 1996), and sequences of mtDNA genes (Smith and Bush 1997, McPheron and Han 1997) or nuclear loci (Chapter 3, Feder et al in prep). These studies have shown that *R. pomonella* appears to be the most variable species of the group, which has led to a working hypothesis that *R. pomonella* represents a large ancestral gene pool from which new species have arisen. The lack of observed fixed differences between the *pomonella* group species, except for the Had^{111} allozyme allele in *R. zephyria*, and similarity between mtDNA haplotypes (Smith and Bush 1997, McPheron and Han 1997) indicates their recent divergence, with *R. pomonella* giving rise to new taxa faster than alleles can become fixed. Since mitochondrial COI/COII is highly conserved and reflects only a matrilineal fraction of the evolutionary history (Maddison 1995), nuclear gene phylogenies for 3 different loci were constructed (Chapter 3). These phylogenies all suggested different relationships between taxa and revealed little or no clear geographic structuring of *R. pomonella* and *R. zephyria* populations. In addition, the results only hinted to the origin of *R. zephyria* - where the host shift occurred, what the forces driving divergence and speciation could have been, and whether the snowberry fly is a native

inhabitant of Eastern North America. Furthermore, attempts to identify diagnostic markers that would be useful in determining whether specimens collected (trapped) in the field as adults belong to one or another species have so far not been very successful.

The problem in the phylogeographic approach may be that an insufficient number of sequences (individuals) or insufficient number of markers was used. It is not very likely that the sampling strategy used for phylogeographic analyses provided insufficient number of sequences, given the large number obtained from *R. pomonella* and *R. zephyria* (Chapter 3). Kliman et al (2000) argue that repetitive sampling within species tends to reveal true genealogical history, implying that the number of individuals can be dramatically reduced if multiple loci are sampled. For rapidly evolving clades, Moran and Kornfield (1993) suggest that only a method that assays variation at a large number of loci could overcome the effects of incomplete allele sorting in the time since species divergence.

Recently, amplified fragment length polymorphism (AFLP, Vos et al 1995) genotyping has been suggested as a method of choice for studying relationships between closely related species (Albertson et al 1999), population structure and differentiation and estimating population genetic parameters (Reineke et al 1998, Cardoso et al 2000). It is widely used in characterizing variation in microbial (Rademaker 2000) and plant isolates (Myashita et al 1999, Cho et al. 1996), as well as in mapping plant genomes (Voorips et al 1997, Brigneti et al. 1997, Vuylsteke et al 1999, Castiglioni et al 1999). It has less frequently been applied to animal systems, probably because microsatellites are well established for many species and have been extensively used as a method for studying genetic variation between populations (for example, Estoup et al 1996). However, a

number of recent studies indicate that AFLP fingerprinting is becoming more popular - it has been employed in assessing parentage in birds (Quisteau et al 1999), detecting hybridization (Prowell et al unpublished, Nijman et al 1999) and for establishing linkage maps (Liu et al 1999).

By screening loci distributed throughout the entire genome, AFLP can be used to generate a larger quantity of information than other markers such as allozymes, RAPDs, RFLPs or microsatellites (Mueller and Wolfenberger 1999). Characters produced by AFLP are highly replicable and provide good resolution of phylogenetic structure (Albertson et al 1999). A change in a restriction recognition site does not alter the size of the fragment but removes it from the profile. The absence of a particular fragment can also be the result of an insertion or deletion event between the cut sites, in which case the size of the fragment is altered; it can also result from the failure to amplify some of the fragments from unknown reasons. Therefore, one disadvantage of the AFLP technique is that the characters behave as dominant markers, rather than co-dominant produced by traditional RFLP. However, its advantages include that it does not require prior knowledge of DNA sequences, detects variation over the entire genome and is highly reproducible because it uses stringent reaction conditions (Vos et al 1995).

I used AFLP fingerprinting technique to compare intra- and interspecific variation at sites distributed randomly throughout *Rhagoletis* genome among 55 *R. pomonella* individuals from 13 populations and 59 *R. zephyria* individuals collected from 16 populations. These populations were chosen to represent the zone of overlap between the two species and to cover the entire geographic range of *R. zephyria*. This study was aimed at providing information about geographic structuring within *R. pomonella* and *R.*

zephyria and identifying relationships between their populations. This information is crucial if we are to better understand the mechanisms of speciation of *R. zephyria*, infer the location of the presumed host shift from ancestral hawthorns to snowberries and explain the present-day geographic distribution of *R. zephyria*. In addition, potentially useful characters that can be used to differentiate *R. pomonella* and *R. zephyria* genotypes by a PCR-based technique were identified, eliminating the need for the fresh material required for allozyme-based methods.

MATERIALS AND METHODS

Sample: Individual flies representing the zone of overlap of geographic ranges of *R. pomonella* and *R. zephyria* and covering the entire range of *R. zephyria* were used for the AFLP genotyping. I sampled at least 5 flies from each population (Figure 4.1). Not all the samples amplified successfully (Table 4.1), so for calculation of genetic distances and analyses of geographic structure samples were pooled according to geographic regions as follows (Table 4.1): Northeast (Massachusetts, New York and Pennsylvania), Great Lakes (Ontario, Michigan, Illinois and Wisconsin), Great Plains (Minnesota, Iowa, Nebraska, North Dakota, South Dakota and Montana "B"), Rocky Mountains and Colorado Plateau (Colorado, Montana, Utah and Idaho) and Pacific Northwest (Washington, Oregon and California). *Rhagoletis mendax* individuals from four populations (East Lansing, MI; Otis Lake, MI; Jasper Pulaski, IN and Rutgers, NJ) were also included in the neighbor-joining analysis (Table 4.1) in an attempt to infer the

polarity and order of host shifts, while two individuals of *R.* nr. *pomonella* sampled near Mexico City, Mexico, were used as an outgroup.

Table 4.1 Populations of *R. zephyria* and *R. pomonella* sampled for the AFLP analysis.

Species/ State	Locality	Label	n
R. zephyria			
Massachusetts	Amherst	ZMA	3
New York	Geneva	ZNY	2
Pennsylvania	PSU campus	ZPA	3
Michigan	East Lansing	ZEL	4
Wisconsin	Waukesha	ZWI	5
Minnesota	Hawby	ZMN	5
North Dakota	Bismarck	ZND	5
South Dakota	Custer SP	ZSD	5
Colorado	Boulder	ZCO	3
Montana	Swan Lake	ZMT	3
Montana	Melstone	ZMTB	2
Idaho	Elmira	ZID	4
Washington	Dixie	ZWA	5
Oregon	Grants Pass	ZOR	7
California	Honeydew	ZCA	3
R. pomonella			
Massachusetts	Amherst	PMA	5
New York	Geneva	PNY	2
Pennsylvania	Biglerville	PPA	3
Ontario	Toronto	PON	3
Michigan	E. Lansing	PMI	4
Illinois	Riverwoods	PIL	4
Minnesota	Staples	PMN	4
Iowa	Ames	PIA	5
Nebraska	(I80E,exit285)	PNE	3
Colorado	Boulder	PCO	5
Utah	Wellsville	PUT	4
Washington	St. Cloud	PWA	4
Mexico	Mexico City	PMX	2

n - number of individual fingerprints used for the analysis

Figure 4.1 Collecting sites for the populations sampled for AFLP analysis.

DNA extraction: DNA was isolated from the individual flies following the protocol of Han and McPheron (1997, see also Chapter 3). For AFLP genotyping, the concentration of DNA in each sample was determined using TKO100 fluorimeter (Hoeffer Scientific Instruments, San Francisco, CA), since equal amounts need to be used for each individual reaction.

AFLP genotyping: AFLP analysis was performed using a modified protocol developed by PE Applied Biosystems (Foster City, CA) after Vos et al (1995) (Figure 4.2). Restriction and ligation of adaptors was carried out in a single 11μL reaction for each individual DNA sample. Approximately 400ng of total genomic DNA was digested with two restriction enzymes: 1U of four-base recognition *Mse*I and 5U of six-base recognition *Eco*RI for 2.5 hours at 37°C in presence of 1U of T4 DNA ligase and adaptor pairs, short double-stranded sequences with one asymmetric end complementary to either *Eco*RI or *Mse*I end. Ligation of adaptor sequences modifies the restriction recognition sequence so that restriction does not occur after the ligation. The restriction-ligation samples were diluted to 200μL in TE buffer (20mM Tris-HCl, 0.1mM EDTA, pH 8.0) to obtain the appropriate concentration for subsequent amplification reaction. Preselective amplification reduces the number of fragments 16-fold by using the primers that match recognition site, adaptor sequence and 3'C for *Mse*I primer and 3'A for *Eco*RI primer (Figure 4.2). Amplifications were performed in 20μL total reaction volume using 15μL of Amplification Core Mix (PE Applied Biosystems), 1μL of preselective primer mix and 4μL of diluted restriction-ligation products. The cycling regime for preselective amplification was 72°C 2min, 20 cycles of 94°C 1sec, 56°C 30sec, 72°C 2 min, followed by an extension step of 30 min at 60°C, with all ramp times set to 0.01 on the PE9600

thermal cycler. PCR products from preselective amplification were diluted to 400μL with TE buffer to obtain the appropriate concentration for selective amplification. Selective amplification further reduces the number of fragments using the primers that match restriction enzyme recognition site, adaptor sequence and three adjacent nucleotides (Figure 4.2). *Eco*RI primers are labeled with fluorescent dye which enables the visualization of fragments after the separation of amplified products on a DNA sequencer. Each fragment for each primer combination, visualized as a band on the denaturing gel, corresponds to a single AFLP locus. One *Mse*I and one *Eco*RI primer should be chosen according to the PE Applied Biosystems protocol; after initial screening for polymorphism using 35 primer combinations, I chose the E-AGG/M-CTT and E-ACT/M-CAG primer combinations. Selective amplification reactions were performed in 20μL volumes, using 15μL of Amplification Core Mix (PE Applied Biosystems), 1μL of each primer and 3μL of diluted preselective amplification reaction products. Cycling parameters for selective amplification are given in Table 4.2. Amplification products were separated on 6% denaturing polyacrylamide gels for 3.5 hours using an ABI Prism 377 DNA sequencer with GS-500 ROX-labeled size standard. A subset of samples was independently prepared twice and electrophoresed at both MSU and Iowa State Sequencing Facilities to check for any possible discrepancies. The results were consistent, so the rest of the sample was run at MSU. Sizes of fragments were determined automatically by GeneScan 2.0.2 software (PE Applied Biosystems), by comparison to the size standard.

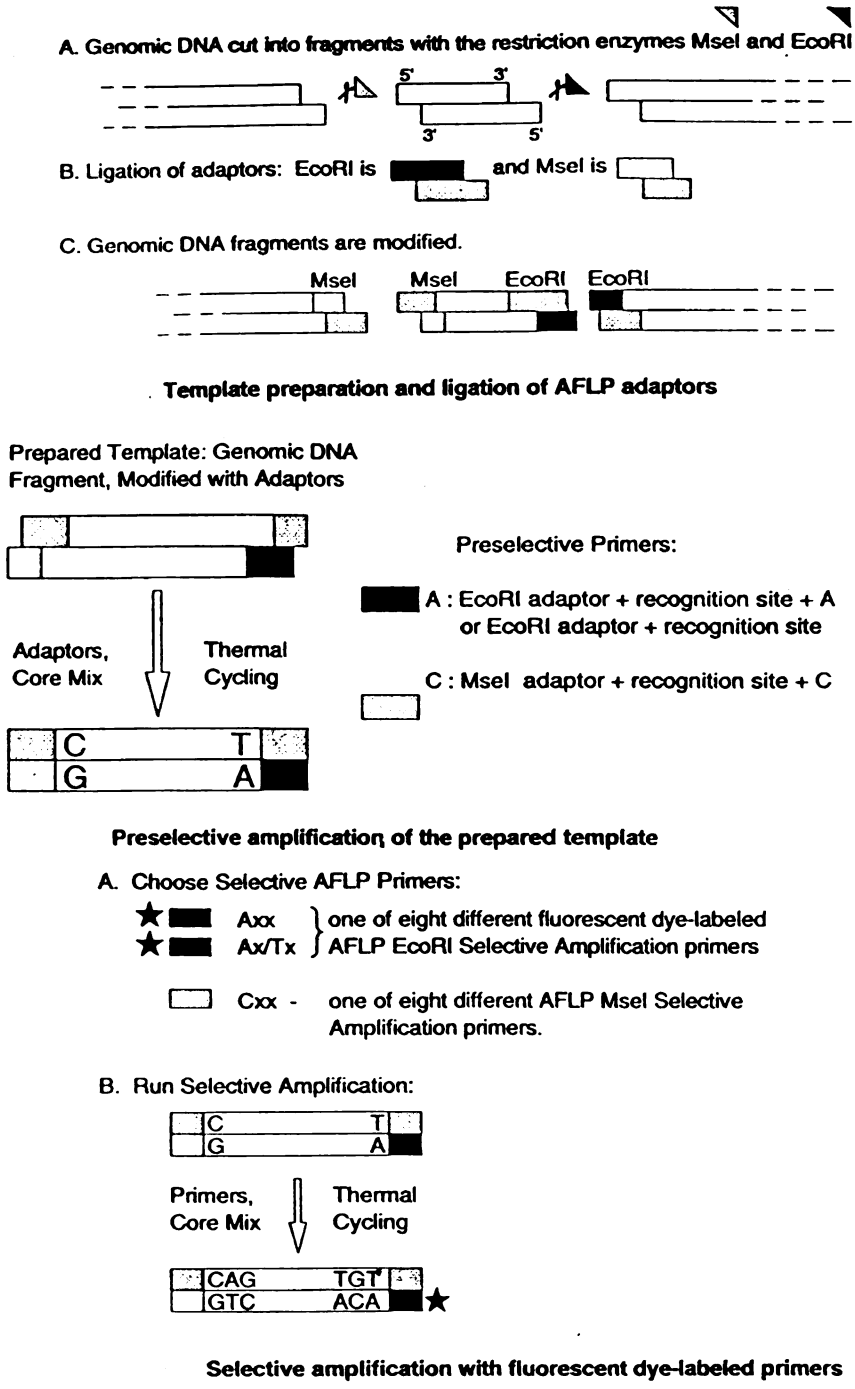


Figure 4.2. The principle of the AFLP DNA fingerprinting technique. Total genomic DNA is cut into fragments using two different restriction enzymes. At the same step, fragments are modified by ligation of adaptors to prevent reannealing. Number of fragments is subsequently reduced by two PCR amplifications (preselective and selective) with increasing specificity of primers. In selective amplification, one of the primers is labeled with fluorescent dye, which enables visualization of fragments after their resolution on a denaturing gel.

Table 4.2. Cycling parameters used for selective amplification

Hold	Cycle	Hold	# of cycles
94°C 2 min	65°C 30 sec	72°C 2 min	1
94°C 1 sec	64°C 30 sec	72°C 2 min	1
94°C 1 sec	63°C 30 sec	72°C 2 min	1
94°C 1 sec	62°C 30 sec	72°C 2 min	1
94°C 1 sec	61°C 30 sec	72°C 2 min	1
94°C 1 sec	60°C 30 sec	72°C 2 min	1
94°C 1 sec	59°C 30 sec	72°C 2 min	1
94°C 1 sec	58°C 30 sec	72°C 2 min	1
94°C 1 sec	57°C 30 sec	72°C 2 min	1
94°C 1 sec	56°C 30 sec	72°C 2 min	23
60°C 30 min			
4°C forever			

Data analysis: AFLP fragments of in the size range of 200-450bp for primer combination E-AGG/M-CTT and 150-450bp for primer combination E-ACT/M-CAG were scored as present (1) or absent (0) for each individual. GeneScan automatically assigns bands detected in each lane to a particular molecular weight. Because sizing by the automated sequencer has a certain standard deviation, genotype patterns were manually compared against each other and fragments were inferred to be the same if they differed by less than 1bp. Data were entered into a spreadsheet and subsequently examined to pool the fragments that differed by 1bp if no individuals displayed both smaller and larger fragment. A frequency distribution of the number of fragments obtained from each individual was plotted to check for deviations from normal. Five percent of the samples with unusually high and 5% with unusually low numbers of fragments were eliminated from the analysis. A high number of fragments has been shown in some cases to be the result of incomplete digestion after which all digested and partially digested fragments are amplified (PE Applied Biosystems 1996 protocol). A low number of fragments resulted from unsuccessful amplifications with one of the primer combinations. The dataset was then analyzed by neighbor-joining using PAUP*4b (Swofford 1999) version 6 with mean character difference as the distance measure. The relative strength of neighbor-joining clusters was evaluated by bootstrapping (Felsenstein 1985) based on 500 replicates. Characters supporting each cluster in the neighbor-joining tree were determined by MacClade 4.0 (Maddison and Maddison 2000). Percentages of polymorphic loci, heterozygosities, F_{st} values and genetic distances between geographic regions were estimated using TFPGA (Tools For Population Genetic Analyses, Miller 1997). Percentage of polymorphic loci was calculated using both 99% and 95% criteria.

Since AFLP bands represent the dominant genotype at a locus, heterozygosities were estimated based on a Taylor expansion of the estimate of frequencies of recessive alleles (Lynch and Milligan 1994). F_{st} was calculated using Weir and Cockerham's (1984) method with jackknifing over loci to obtain variance estimates and bootstrapping with 1000 replicates to generate 95% confidence intervals. Genetic distances were calculated as Nei's (1972, 1978) distance measures. Approximate geographic distances were estimated between the geographic centers of the populations sampled within the regions and used in TFPGA to perform a Mantel test (Mantel 1967) with 10000 random permutations to determine whether there is a correlation between unbiased Nei's (1978) genetic and geographic distances.

RESULTS

In a preliminary study I screened a total of 35 AFLP primer combinations for amplification of polymorphic fragments in four *R. zephyria* and four *R. pomonella* individuals, each from a different population. The majority of loci were polymorphic with the polymorphism being shared between the two species. However, two primer pair combinations (E-AGG/M-CTT and E-ACT/M-CAG) were identified that produced bands unique to *R. pomonella* or *R. zephyria*. These primer pairs were used to generate AFLP fingerprints for the entire sample. Fragments smaller than 200bp for primer combination E-AGG/M-CTT and smaller than 150bp for primer combination E-ACT/M-CAG were excluded from the analysis. The distribution of these fragments within each individual

genotype and between genotypes was too complex to score and interpret unambiguously. Primer combination E-AGG/M-CTT amplified a total of 112 fragments between 200 and 450bp, with the average of 19 per individual; E-ACT/M-CAG amplified 143 fragments between 150 and 450bp, with the average of 24 fragments per individual. Of these 255 AFLP loci, 239 (93.7%) were polymorphic across both species.

A frequency distribution of number of fragments amplified for each individual of *R. zephyria*, *R. pomonella* and *R. mendax* is shown in Figure 4.3. Ninety percent of the fingerprints obtained consisted of 29-62 fragments. Therefore all individuals for which less than 29 or more than 62 fragments were amplified were eliminated from the analysis.

Average heterozygosities and the percentage of polymorphic loci within regions for both species are given in Table 4.3. Levels of genetic variation, described by both heterozygosities and % polymorphic loci were somewhat lower in *R. pomonella* than in *R. zephyria* (Table 4.3). However, *R. pomonella* populations appear to be more genetically structured than *R. zephyria* – the value of parameter θ (which corresponds to F_{st}) in *R. pomonella* is 0.106 (with a 95% confidence interval of 0.072-0.142), while in *R. zephyria* it is 0.0615 (with a 95% confidence interval of 0.044-0.080). However, these values were not significantly different ($t=2.7510$, $df=110$, $P>0.05$).

The neighbor-joining tree (Figure 4.4) shows that, in general, each of the three species (*R. pomonella*, *R. zephyria* and *R. mendax*) clusters coherently. *Rhagoletis mendax* appears as a sister group to *R. pomonella*, with *R. zephyria* forming a sister taxon to them. The only *R. zephyria* that falls into the cluster with *R. pomonella* is a single individual from South Dakota (ZSD16). Two individuals of *R. zephyria* (one from Idaho – ZID5, and one from Washington – ZWA8), as well as one blueberry maggot fly from

New Jersey (MNJ7), were placed outside of the main clusters in the neighbor-joining analysis (Figure 4.4).

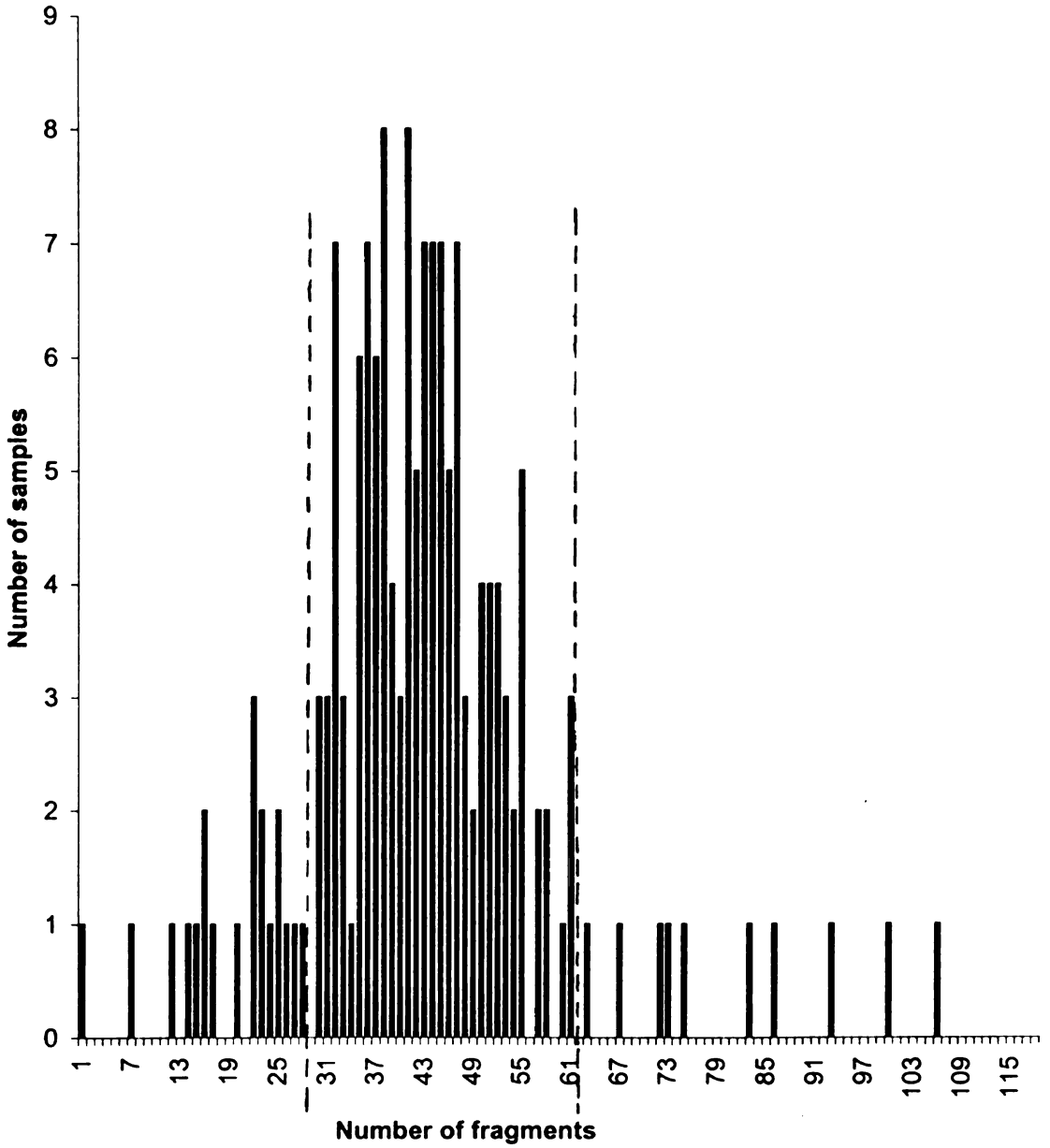


Figure 4.3 Frequency distribution of number of fragments amplified from each individual. Samples with numbers of fragments between the dash lines were used in the analysis.

Table 4.3 Estimated average heterozygosities and percentage of polymorphic loci in *R. zephyria* and *R. pomonella*.

R. zephyria	n	H	$poly_{99}$	$poly_{95}$	*R.pomonella*	n	H	$poly_{99}$	$poly_{95}$
Northeast	8	0.066	25.098	25.098	Northeast	10	0.109	48.235	28.235
Great Lakes	9	0.108	45.098	45.098	Great Lakes	18	0.073	44.706	26.667
Gr. Plains	17	0.110	51.765	33.333	Gr. Plains	12	0.087	41.961	26.274
Rocky Mountains	10	0.110	45.098	23.529	Rocky Mountains	9	0.086	32.157	32.157
Northwest	15	0.102	47.843	28.626	Northwest	4	0.080	25.490	25.490
total	59	0.111	61.176	31.765	total	53	0.101	57.255	31.327

n – sample size, H – average heterozygosity, $poly_{99}$ - % of polymorphic loci by the 99% criterion, $poly_{95}$ - % of polymorphic loci by the 95% criterion.

Figure 4.4. Neighbor-joining tree based on the analysis of 255 AFLP loci. Green branches indicate *R. zephyria*, red *R. pomonella* and blue *R. mendax*. Colored squares next to *R. zephyria* and circles next to *R. pomonella* sample labels indicate geographic regions: red – Northeast, orange – Great Lakes, green – Great Plains, blue – Rocky Mountains and purple – Northwest. For taxon labels please refer to Key to symbols and abbreviations (p. xi). (Images in this dissertation are presented in color)

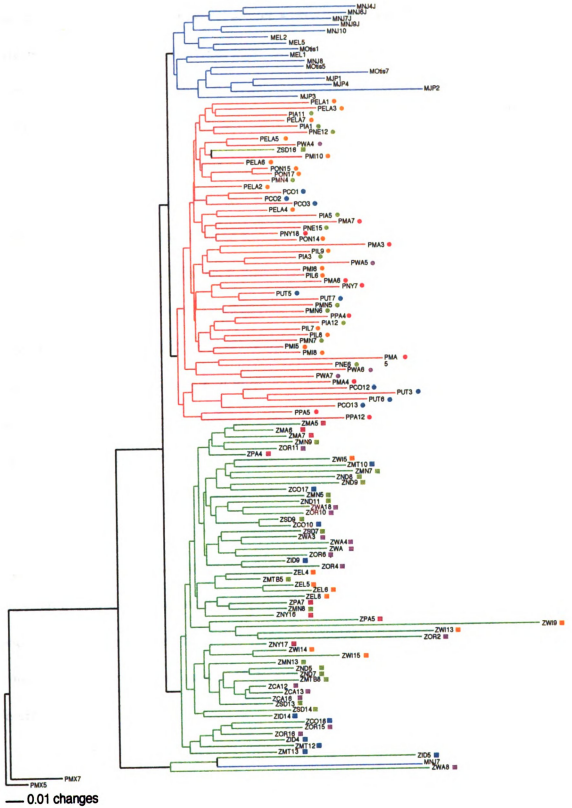

MNJ4J
MNJ5J
MNJ7J
MNJ8J
MNJ10
MEL2
MEL5
MOte1
MEL1
MNJ6
MOte5
MOte7
MJP1
MJP4
MJP2
MJP3
PELA1
PELA3
PIA11
PELA7
PIA1
PNE12
PELA5
PWA4
ZSD16
PMI10
PELA6
PCN15
PCN17
PMN14
PELA2
PCO1
PCO2
PCO3
PELA4
PIA5
PMA7
PNE15
PNY18
PCN14
PMA3
PIL9
PIA3
PWA5
PIN8
PIL6
PMA6
PNY7
PUT5
PUT7
PMN5
PMN9
PPA4
PIA12
PIL7
PIL5
PMN7
PMI5
PMI8
PMA
5
PNE8
PWA8
PWA7
PMA4
PCO12
PUT3
PUT6
PCO13
PPA5
PPA12
ZMA5
ZMA6
ZMA7
ZMN9
ZOR11
ZPA4
ZWI5
ZMT10
ZMN7
ZND5
ZND9
ZCO17
ZMN5
ZND11
ZCA10
ZOR10
ZSD8
ZCO10
ZND7
ZWA3
ZWA4
ZWA
ZOR6
ZID9
ZOR4
ZEL4
ZMT B5
ZEL5
ZEL6
ZEL5
ZPA7
ZMN8
ZNY16
ZPA5
ZWI9
ZWI3
ZOR2
ZNY17
ZWI4
ZWI15
ZMN13
ZND5
ZND7
ZMT B6
ZCA12
ZCA13
ZCA15
ZSD11
ZSD14
ZID14
ZCO18
ZOR15
ZOR16
ZID4
ZWI12
ZMT13
ZID5
MNJ7
ZWA8
PMX5
PMX7
0.01 changes

Figure 4.4.

Like in *R. pomonella* and *R. mendax*, groups within the *R. zephyria* cluster do not correspond to regions, and genotypes from the same population are scattered throughout the tree. Somewhat exceptional is the cluster of individuals from western populations at the base of the large *R. zephyria* cluster and 3 snowberry flies from Massachusetts and California each. In these cases genotypes from the same population cluster together, but within groups that consist of individuals from different regions (plains, northwest and northeast with Massachusetts flies and mainly plains with California flies).

Several characters define each cluster, and some of them are potentially useful for discrimination between species. The fragment of 273bp amplified by the E-AGG/M-CTT primer pair is present in all *R. zephyria* except the two individuals falling outside of the main clusters and is found in no *R. pomonella* except one individual from Washington. The same primer combination amplified a fragment of 316bp in all snowberry flies but one from Idaho and one from South Dakota, both of which are not placed within the main *R. zephyria* cluster. The 316bp fragment was amplified in only three *R. pomonella* and two *R. mendax*. In 41 of the 59 snowberry flies primer pair E-ACT/M-CAG amplified a fragment of 262bp, with frequency increasing from east to west. This fragment was amplified in two apple maggots from Nebraska, two blueberry maggots from Michigan (one from East Lansing, one from Otis Lake) and one from New Jersey (this individual groups with two *R. zephyria* outside of the main clusters in the neighbor-joining tree, Figure 4.4).

The primer combination E-ACT/M-CAG amplified a fragment of 150bp in most *R. zephyria* and *R. mendax*, but appears to be lost in the *R. pomonella* lineage since it amplified in only six *R. pomonella* from different populations. Fragment of 224bp

amplified by this primer pair was found in about 60% of snowberry flies, its frequency being highest in the Plains, Northwest and Northeast, and lower in the Great Lakes and Rocky Mountains populations. The same fragment was found only in one apple maggot from Iowa. Similarly, fragment of 244bp was amplified in about 60% of *R. zephyria*, with highest frequency in the Great Lakes and lowest in the Rocky Mountains. This fragment was present in only three *R. pomonella*, one from Pennsylvania and two from Washington. One of these apple maggots from Washington also had the 273bp fragment amplified by E-AGG/M-CTT primers and present only in *R. zephyria*. The 305bp fragment observed in all *R. pomonella* from Colorado and Utah (Rocky Mountains and Colorado Plateau ecoregion) was rare in apple maggots from other regions and found in snowberry flies from Wisconsin and North Dakota (four and one individuals, respectively).

R. pomonella/R. mendax cluster is defined by fragments of 266 and 277bp amplified by the E-AGG/M-CTT combination and 202bp amplified by the E-ACT/M-CAG primer pair. All these characters are potentially useful for species identification.

Mantel test indicated that there is no significant correlation between genetic and geographic distances. Coefficient of correlation between the two matrices in *R. zephyria* was 0.369 and in *R. pomonella* 0.402, both of them were not significant, which was consistent with the results of neighbor-joining analysis where no geographic structuring was observed (Figure 4.4)

Table 4.4 Genètic distances (below diagonal) and geographic distances (in miles, above diagonal) between *R. zephyria* and *R. pomonella* from different regions based on AFLP data.

	z-NE	z-GL	z-GP	z-RM	z-NW	p-NE	p-GL	p-GP	p-RM	p- NW
z-NE	0	400	1100	1600	2100	50	350	800	1400	1900
z-GL	0.0209	0	700	1000	1800	350	120	350	100	1400
z-GP	0.0139	0.0073	0	300	1000	1150	520	300	450	900
z-RM	0.0137	0.0087	0.0012	0	480	1650	1400	700	230	500
z-NW	0.0154	0.0097	0.0021	0.0011	0	2150	1900	1200	650	100
p-NE	0.0346	0.0349	0.0267	0.0256	0.0328	0	400	800	1400	1900
p-GL	0.0355	0.0395	0.0287	0.0272	0.0338	0.0100	0	420	1120	1500
p-GP	0.0273	0.0316	0.0207	0.0193	0.0246	0.0056	0.048	0	600	1100
p-RM	0.0478	0.0538	0.0420	0.0430	0.0500	0.0187	0.0161	0.0159	0	700
p-NW	0.0408	0.0490	0.0382	0.0401	0.0432	0.0131	0.0158	0.0117	0.0191	0

z- *R. zephyria*, p – *R. pomonella*; NE – Northeast, GL – Great Lakes, GP – Great Plains, RM – Rocky Mountains and Colorado Plateau, NW – Northwest.

DISCUSSION

The results obtained by the AFLP analysis of *Rhagoletis pomonella* and *R. zephyria* demonstrated that this technique is powerful and promising for studying genetic variation and relationships between the closely related species of *Rhagoletis*. The high number of loci sampled at random and as a cross-section of the entire genome reflects much higher variation and provides better resolution than sequences from any of the individual nuclear or mitochondrial loci that have been examined thus far (Chapter 3).

Neighbor-joining analysis based on the AFLP loci revealed three well-defined groups with almost perfect correlation of clusters with the three species (*R. zephyria*, *R. pomonella* and *R. mendax*). The only exception was the placement of a single snowberry fly from South Dakota within the *R. pomonella* cluster (Figure 4.4). This result is consistent with the recent study of divergence of the *pomonella* group species based on the allele frequencies at 29 allozyme loci (Berlocher 2000), where clustering of populations corresponded to species. In that study *R. zephyria* was placed at the base of the neighbor-joining tree, as a sister taxon to the cluster containing *R. pomonella*, *R. mendax* and *R.* nr. *mendax*. AFLP analysis similarly places *R. zephyria* basally to the *pomonella/mendax* group (Figure 4.4).

Branch lengths in the neighbor-joining tree indicate that the majority of genetic variation is distributed within populations and species, suggesting that speciation occurred without large population bottlenecks. This result is consistent with high allelic and nucleotide diversity found within populations and species in studies of nuclear and

mitochondrial DNA sequence variation (Chapter 3, Feder et al in prep) and with hypothesis that speciation occurred in sympatry.

Several characters define each cluster corresponding to species, although the groups reflected by the neighbor-joining tree in Figure 4.4 do not have bootstrap support; AFLP analysis thus for the first time provides a reliable PCR-based method for distinguishing between the closely related species of the *pomonella* group. The primer combinations used in this study amplified several fragments only in *R. zephyria* and very rarely in *R. pomonella* or *R. mendax*. These fragments can theoretically be used as diagnostic markers. Fragment of 273bp was successfully amplified by the primer combination E-AGG/M-CTT in all *R. zephyria* except in two individuals falling outside of the main cluster. In *R. pomonella* this fragment was only amplified in one individual from Washington. This individual also possessed the fragment of 244bp amplified by the primer combination E-ACT/M-CTT, present in 33 of the 59 *R. zephyria* and only two other *R. pomonella*, one of which was from the same Washington population. This suggests that very low-level hybridization with introgression of *zephyria* alleles into *pomonella* may be occurring between apple maggot and snowberry flies in the West. *R. pomonella* was introduced into the Northwest within the past 100 years (AliNiazee and Brunner 1986, McPheron 1990); allozyme alleles characteristic for *R. zephyria* (Had^{111}) have been found in low frequencies in flies reared from hawthorns and apples in Washington (McPheron 1990) and Minnesota (Feder et al 1999), suggesting male-mediated gene flow in the direction from native into the introduced taxon's gene pool. In both these studies (McPheron 1990, Feder et al 1999) multilocus genotypes, however, did not indicate the presence of F_1 hybrids, suggesting extensive backcrossing to *R.*

pomonella. The placement of the potential hybrid individuals observed in this study within the pomonella cluster (Figure 4.4) is consistent with this interpretation. Two snowberry flies, one from Idaho and one from Washington, placed outside the main *zephyria* cluster, both lack fragments characteristic for *R. zephyria*. The possibility that this is an artifact produced by failure to amplify these fragments from these flies cannot be ruled out. If there is some biological reason for the absence of *zephyria*-characteristic fragments in these flies, their placement at the base of the neighbor-joining tree may further indicate that the possible hybridization between apple maggot and snowberry flies is rather rare. In cladistic analyses hybrid taxa tend to be placed as basal, causing extensive homoplasy and reducing the resolution of relationships between other taxa (McDade 1992).

To detect population structure in *R. pomonella* group large sample sizes are needed (Berlocher 1995, 2000), as suggested by the allozyme studies (McPheron 1990, Berlocher 2000). In studies of larger geographic scale (McPheron 1990, Feder et al 1999) F_{st} values observed between the populations in native, eastern part of the *R. pomonella* range, were low (0.024-0.032). Feder et al (1999) report the values of parameter θ of 0.069 in *R. zephyria* and 0.073 in *R. pomonella*. Based on the analysis of 255 AFLP loci and populations grouped in five regions, θ in *R. zephyria* was 0.0615, while it was higher in *R. pomonella* (0.1060). This finding is consistent with the hypothesis that *R. pomonella*, as an ancestral species, is more geographically structured than the species that have diverged from it. Contrary to the expectation based on the allozyme data (Berlocher et al 1993, Berlocher 2000, Feder et al 1999) that it should also have higher average heterozygosity, AFLP data indicated somewhat higher value of this parameter in *R.*

zephyria (0.111 vs. 0.101 in *R. pomonella*). Lower average heterozygosity in Northeastern *R. zephyria* is consistent with the hypothesis that snowberry fly was introduced to this region from the west when the host plant, *Symphoricarpos albus* var. *laevigatus*, was introduced as ornamental shrub. However, the possibility cannot be ruled out that it reflects recent range expansion by colonization followed by genetic bottleneck in this part of the species range.

Rhagoletis zephyria appears to have differentiated more from the present-day *R. pomonella* than *R. mendax* (Figure 4.3). This result is consistent with morphological (there is a slight difference in shape of male surstyli between apple maggot and snowberry flies; Bush 1966, Jenkins 1996) and allozyme data (Had^{111} is fixed in *R. zephyria* and found only in very low frequencies in western *R. pomonella* - McPheron 1990; there are no fixed allele differences between *R. pomonella* and *R. mendax*). It may suggest that the differences accumulated after the host shift and speciation in zone of parapatry followed by the rapid range expansion to the west (Berlocher 1998), where the flies filled new ecological niches and encountered different selection regimes. Grouping of several flies from Northwest and Rocky Mountains at the base of the neighbor-joining tree seems to indicate that these populations diverged most. Individuals from the Plains region are present in every smaller cluster within *R. zephyria* and grouped with genotypes from all other regions, which may indicate that speciation occurred somewhere in the zone of parapatry in the Plains, as the hawthorns colonizing prairies after the last glaciation came into contact with snowberries, and highly genetically variable subpopulations spread both east and west. Very low genetic distances between the Great

Lakes and all other regions except Northeast support the hypothesis that *R. zephyria* is native to the Great Lakes area.

The only snowberry fly placed within the *R. pomonella* cluster (Figure 4.3) lacks some of the fragments characteristic for *R. zephyria* and absent from *R. pomonella*. However, it is difficult to assume that it shares *pomonella* genotype. This genotype may indicate past hybridization with low frequency introgression of genes from *R. pomonella* into the *R. zephyria* genome, or incomplete lineage sorting with retention of the ancestral polymorphisms in very low frequencies in *R. zephyria*, or mistake in oviposition by *R. pomonella*.

Data presented here do not seem to support the hypothesis of multiple speciation events of *R. zephyria* from *R. pomonella* or *pomonella*-like ancestor. The observed patterns do not allow distinguishing precisely between several scenarios of where the speciation of snowberry flies has occurred and how the range has expanded. The ability to further resolve relationships between populations of *R. zephyria* and *R. pomonella* should provide better insight into the processes involved in their divergence. Lack of geographic structuring and bootstrap support for the groups in neighbor-joining tree indicates that even with this large number of markers we may still not have sufficient data. To improve resolution and observe geographic structuring within species we may need to sample even more AFLP loci using more primer combinations and obtain AFLP fingerprints from more individuals from each population. Resolution, as well as the support for particular groups, is expected to increase with increased number of loci scored (Albertson et al 1999). It is also possible that resolution would improve with decreasing the number of loci included in the analyses (Prowell et al in prep), since the

large number of similar loci may mask differentiation between populations and species. This is reflected by the high levels of homoplasy, i.e. fragments present in some individuals and some populations without any regular pattern.

CONCLUSIONS

Hypothesis 1: *Rhagoletis pomonella* represents a large and variable gene pool, containing ancestral polymorphisms, from which new species arise.

Rhagoletis pomonella showed larger haplotype diversity than *R. zephyria* at one of the three anonymous nuclear loci analyzed here (P2480) and at the mitochondrial locus COI/COII. At all loci, nucleotide diversity π was higher in *R. pomonella* than in *R. zephyria*. Surprisingly, however, the nucleotide diversity parameter θ, as well as weighted θ/bp, was at all nuclear loci higher in *R. zephyria,* which may reflect the influence of natural selection on polymorphism patterns in *R. zephyria*. Phylogenetic analyses of two nuclear loci (P220 and P2480) placed some of the *R. pomonella* sequences at the base of the tree, indicating that alleles of *R. pomonella* may be more similar to ancestral.

Hypothesis 2: *Rhagoletis zephyria* has diverged from *R. pomonella* more recently than blueberry maggot, *R. mendax*.

Phylogenetic analyses of anonymous nuclear loci and mitochondrial COI/COII have indicated that *R. mendax* branched off from the ancestor of present-day *R. pomonella* and *R. zephyria* before *R. zephyria*. Analysis of 255 AFLP loci, however, showed that *R. zephyria* is more distant from *R. pomonella* than is *R. mendax*, contrary to results of the analyses of individual loci and this hypothesis. New models of sympatric speciation suggest that if species begin diverging in sympatry, the ranges of those that speciated the latest should still be sympatric, whereas older species may be less sympatric

because of range changes over time (Berlocher, pers. comm.). The geographic range of *R. zephyria* extends far outside of the range of *R. pomonella*, although the zone of overlap is much larger than previously thought. The range of *R. mendax* is completely contained within the range of *R. pomonella*, which may indicate that *R. zephyria* speciated before *R. mendax*. Therefore, it is difficult to infer the timeframe and order of speciation events in the *R. pomonella* species group. Nevertheless, polyphyly of *R. zephyria* and *R. pomonella*, unresolved phylogenetic relationships at anonymous nuclear loci and mitochondrial COI/COII analyzed in this study, shared alleles and haplotypes between species, and lack of fixed differences indicate that speciation may have been recent and could have happened in the past 10,000 years, after the glaciers of the last glaciation receded and host plants (*Crategus* sp. and *Symphoricarpos* sp.) came into contact.

Hypothesis 3: The host shift that led to divergence of *R. zephyria* from *R. pomonella* and speciation occurred from ancestral hawthorn host to snowberry.

If the ancestor of the present-day *R. pomonella* and *R. zephyria* was *R. pomonella*-like, as suggested by the analyses of DNA sequences at different loci, then the host shift occurred from the hawthorn host of the ancestor to snowberry. Placement of the *R. pomonella* alleles sampled from populations from the Great Plains within the clades consisting mainly of *R.* zephyria alleles seems to support the hypothesis that the host shift may have occurred in the Plains. Little geographic structuring was revealed by the analyses of DNA sequences and AFLP fingerprinting patterns, which makes inferences about the direction of population expansion difficult. However, analyses of P2480 (neighbor-joining and parsimony analysis) and AFLP loci (neighbor-joining) indicate that

western populations of *R. zephyria* appear to be more divergent and placed as basal to the rest of *R. zephyria*, suggesting that these populations may have been more isolated, perhaps because they dispersed first.

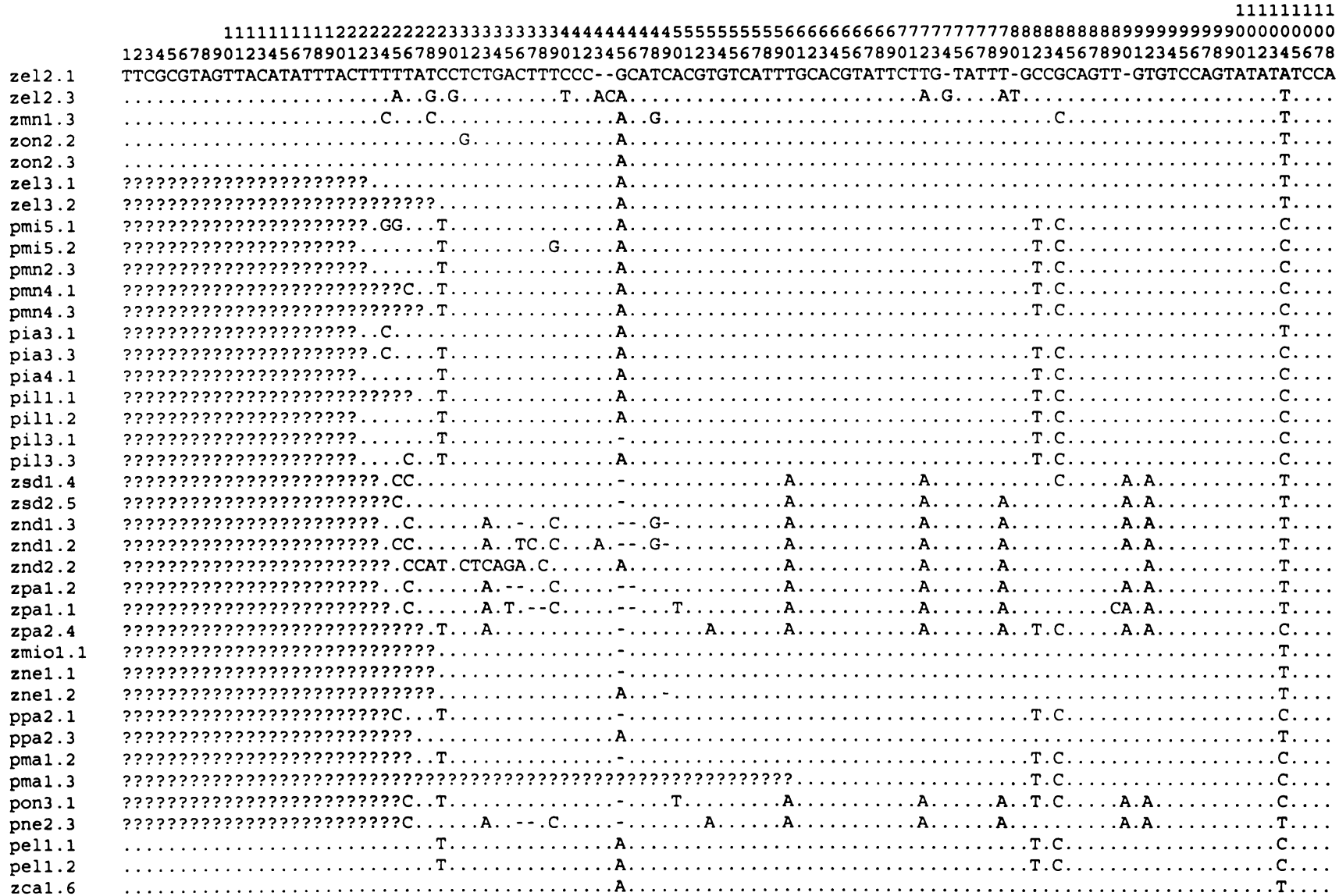
Hypothesis 4: *Rhagoletis zephyria* in Eastern North America has been introduced with the host plant (snowberry).

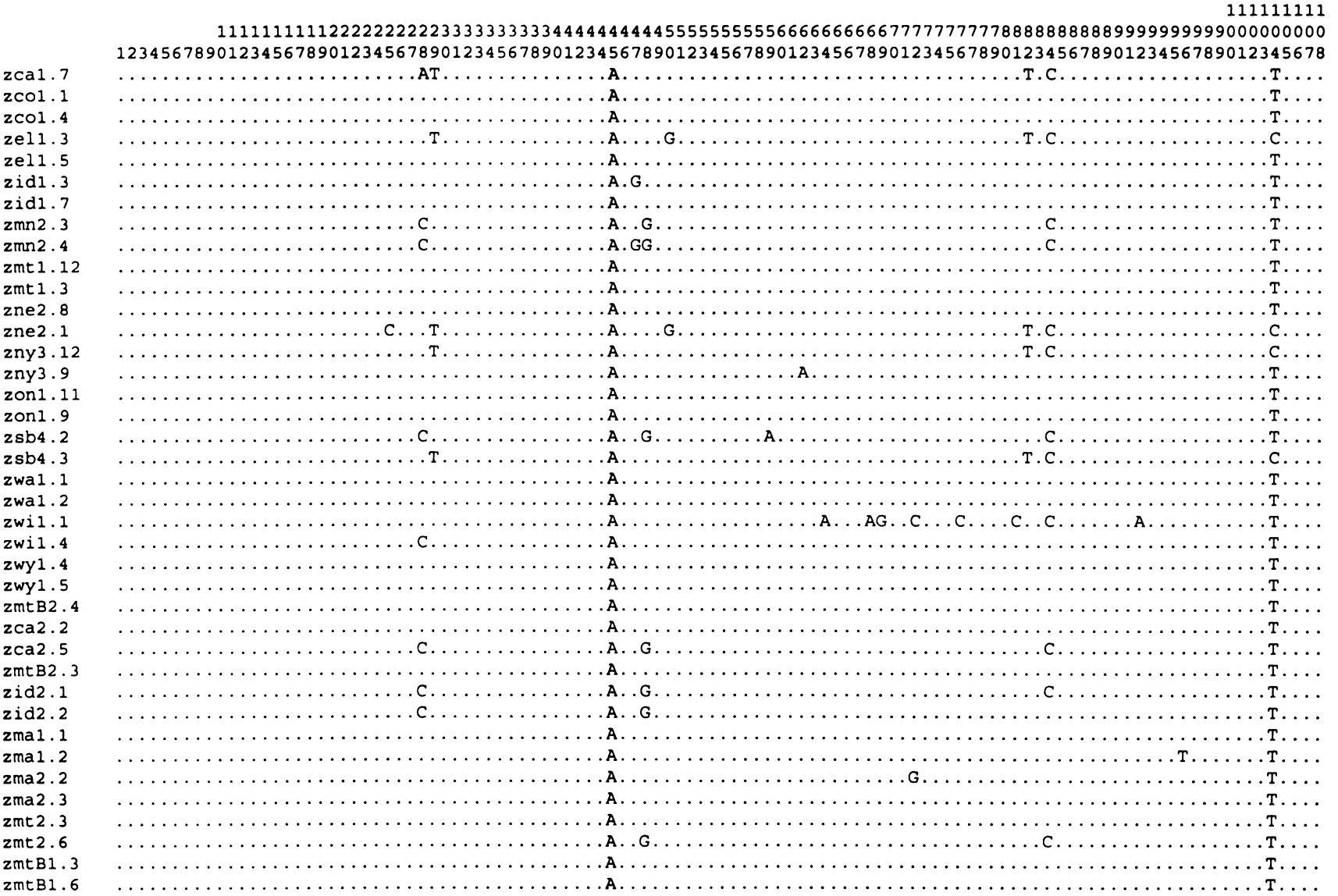
Evidence presented in this study supports the alternative to this hypothesis – that, in fact, *R. zephyria* is a native inhabitant of eastern North America. Native hosts have been found infested with *R. zephyria* in the Great Lakes region and in the Northeast, and introduced *Symphoricarpos albus* var. *laevigatus* has been found infested only when the native hosts were also available in relatively close proximity. Eastern populations of *R. zephyria* showed no evidence of population bottleneck at any of the nuclear loci, mitochondrial COI/COII or AFLPs, which might be expected if they were introduced with the host plant and founded from a small number of individuals. Alleles of the eastern *R. zephyria* were placed into same clades as the alleles of western *R. zephyria* without much geographic structuring. If *R. zephyria* were introduced to eastern North America, then it would be expected that alleles from eastern populations would group together and lose to particular western populations from which they were introduced, which was not the case in this study.

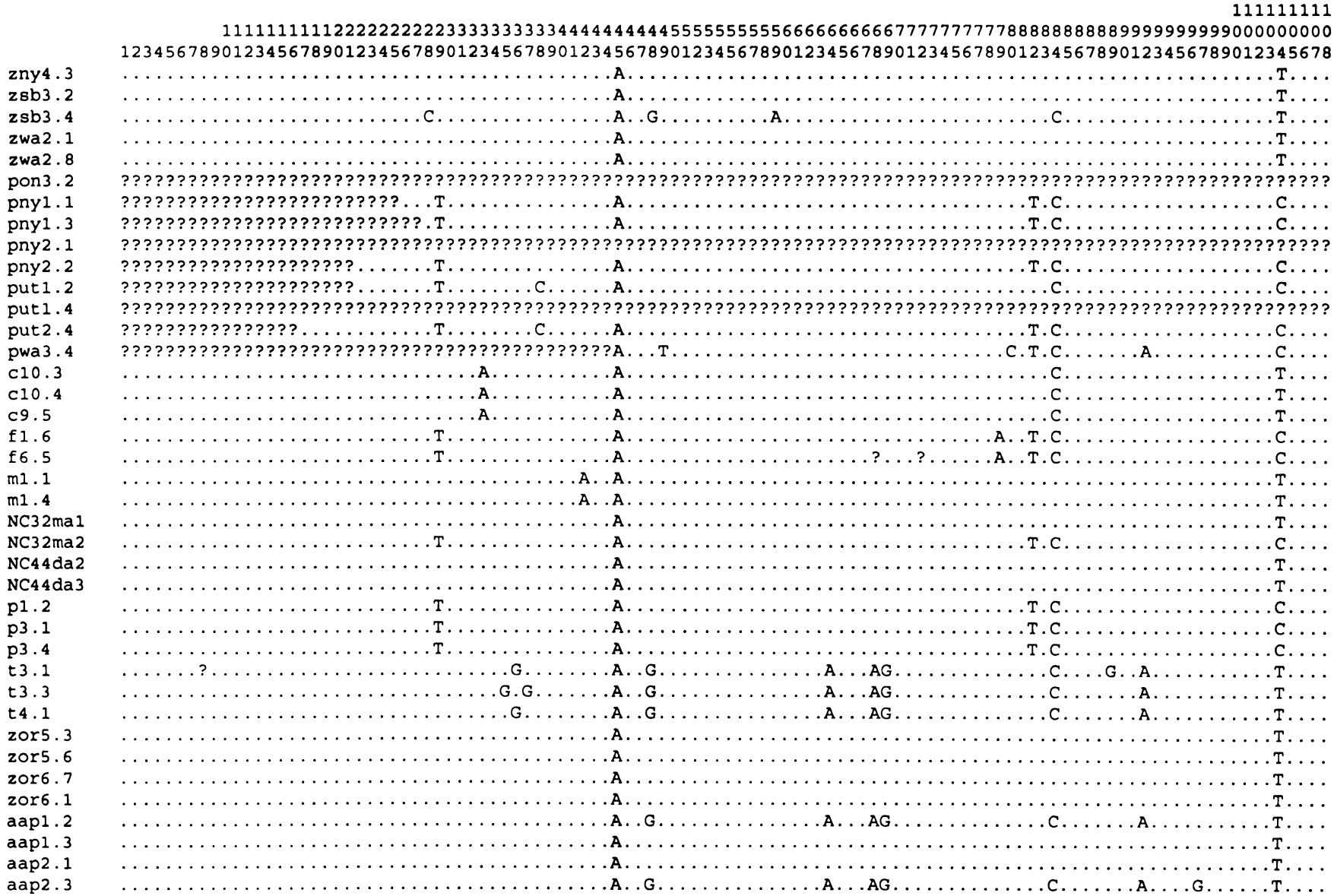
APPENDICES

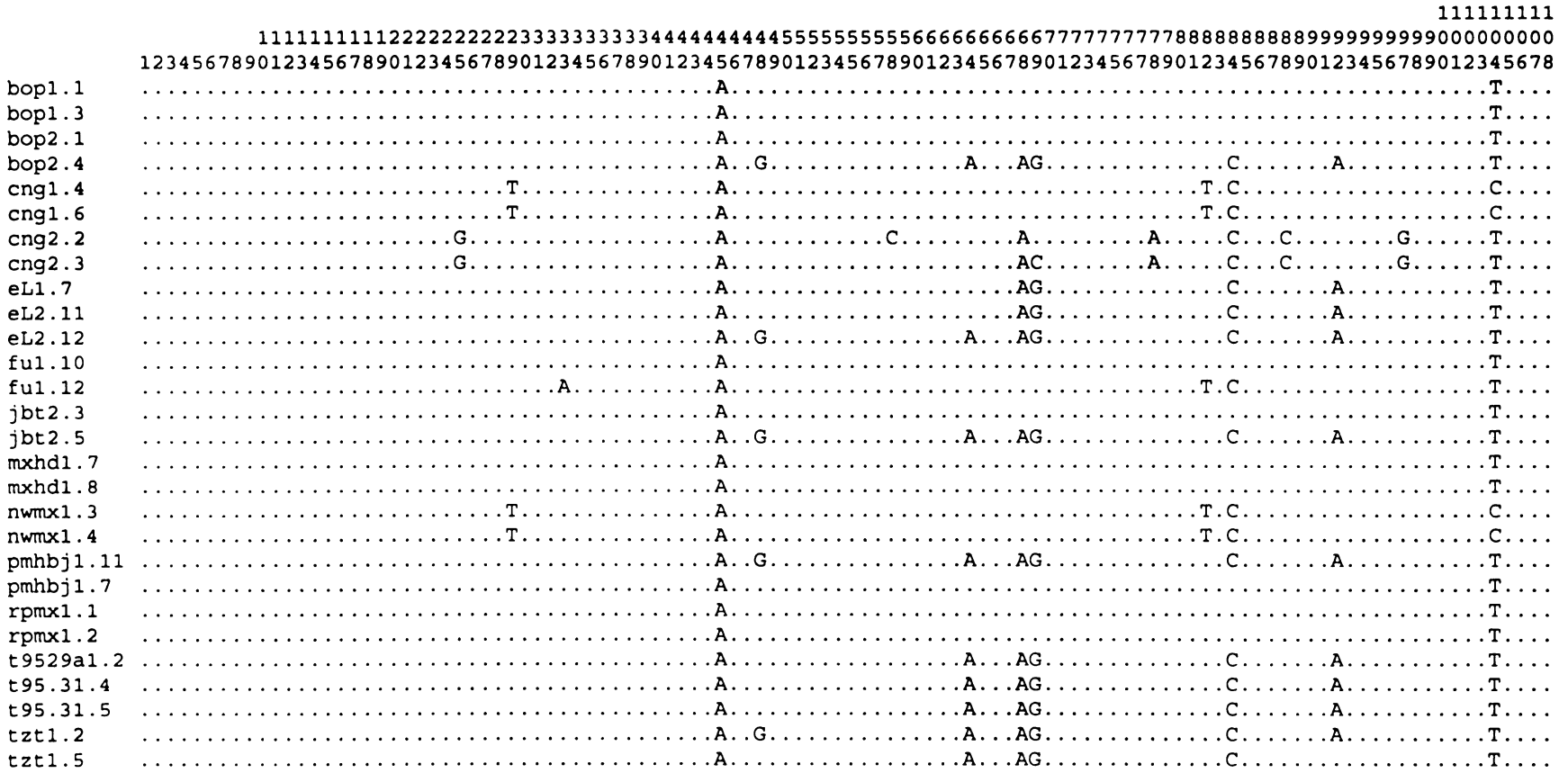
APPENDIX A

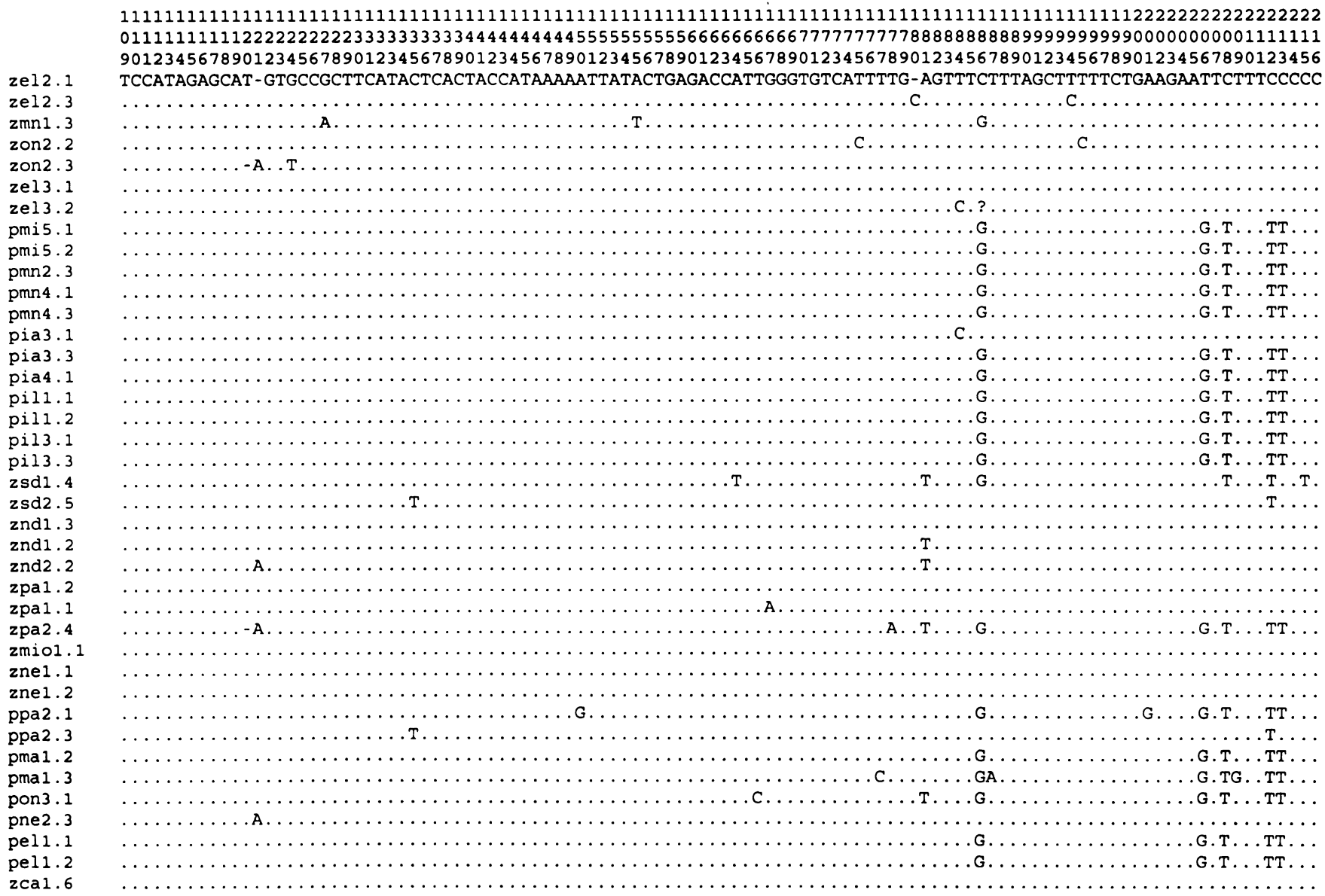
P220 sequence alignment

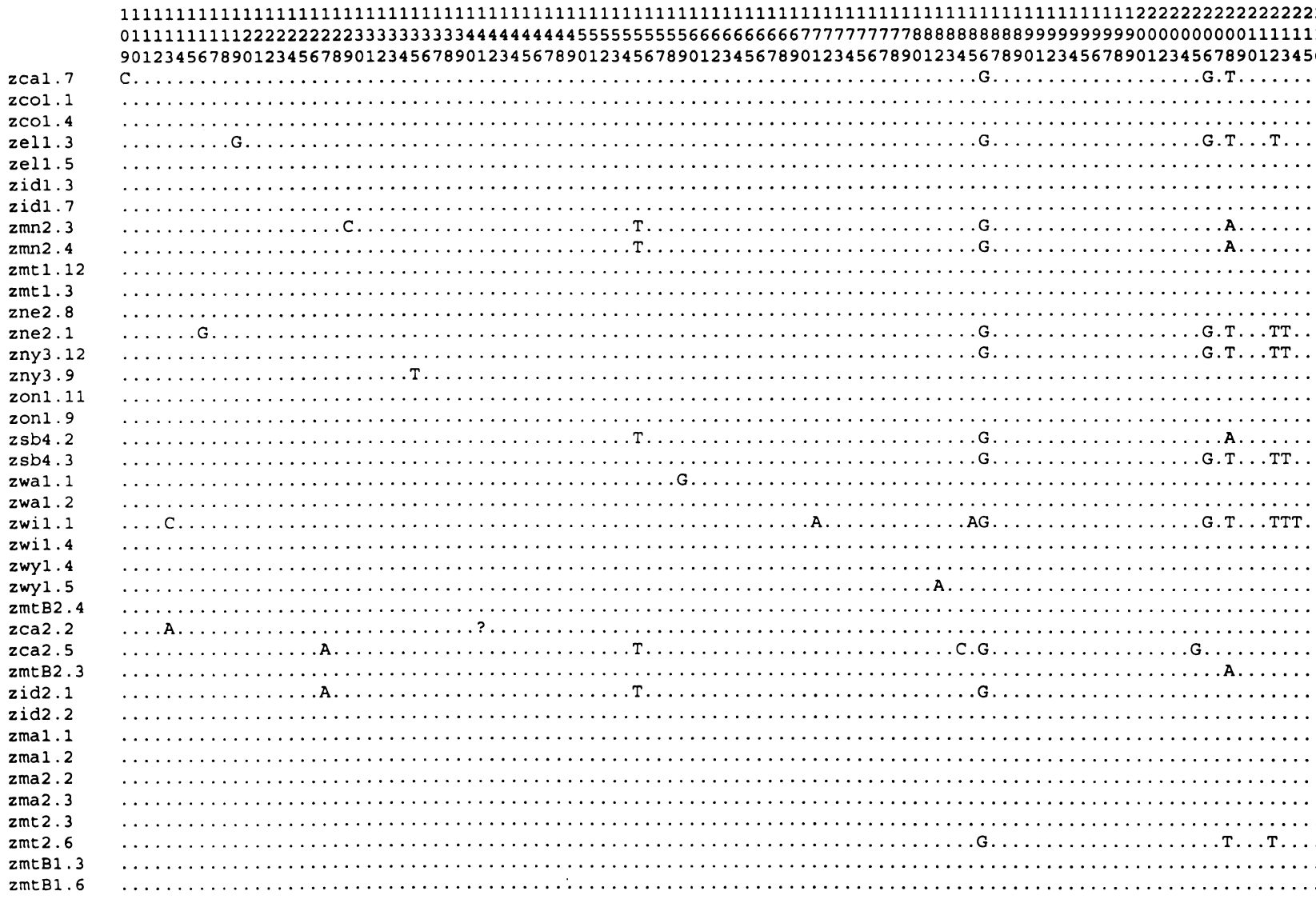

zel2.1 TTCGCGTAGTTACATATTTACTTTTTATCCTCTGACTTTCCC--GCATCACGTGTCATTTGCACGTATTCTTG-TATTT-GCCGCAGTT-GTGTCCAGTATATATCCA
zel2.3
zmn1.3
zon2.2
zon2.3
zel3.1
zel3.2
pmi5.1
pmi5.2
pmn2.3
pmn4.1
pmn4.3
pia3.1
pia3.3
pia4.1
pil1.1
pil1.2
pil3.1
pil3.3
zsd1.4
zsd2.5
znd1.3
znd1.2
znd2.2
zpa1.2
zpa1.1
zpa2.4
zmio1.1
zne1.1
zne1.2
ppa2.1
ppa2.3
pma1.2
pma1.3
pon3.1
pne2.3
pel1.1
pel1.2
zca1.6

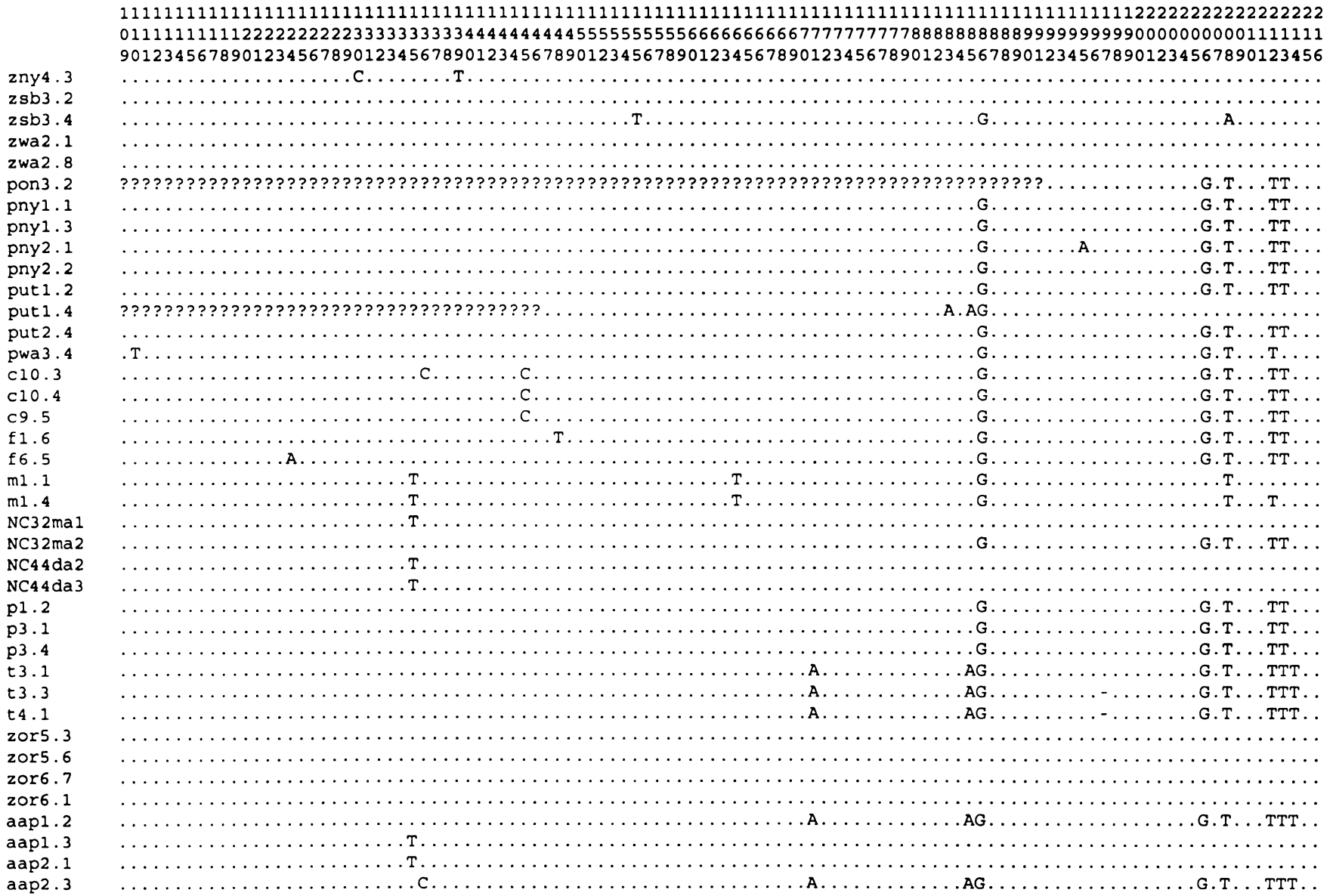

111111111
111111111122222222223333333333444444444455555555556666666666777777777788888888899999999990000000000
12345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678
zca1.7 .....AT.....A.....T.C.....T....
zco1.1 .....A.....T....
zco1.4 .....A.....T....
zel1.3 .....T.....A...G.....T.C.....C....
zel1.5 .....A.....T....
zid1.3 .....A.G.....T....
zid1.7 .....A.....T....
zmn2.3 .....C.....A..G.....C.....T....
zmn2.4 .....C.....A.GG.....C.....T....
zmt1.12 .....A.....T....
zmt1.3 .....A.....T....
zne2.8 .....A.....T....
zne2.1 .....C..T.....A...G.....T.C.....C....
zny3.12 .....T.....A.....T.C.....C....
zny3.9 .....A.....A.....T....
zon1.11 .....A.....T....
zon1.9 .....A.....T....
zsb4.2 .....C.....A..G.....A.....C.....T....
zsb4.3 .....T.....A.....T.C.....C....
zwa1.1 .....A.....T....
zwa1.2 .....A.....T....
zwi1.1 .....A.....A...AG..C...C....C..C.....A.....T....
zwi1.4 .....C.....A.....T....
zwy1.4 .....A.....T....
zwy1.5 .....A.....T....
zmtB2.4 .....A.....T....
zca2.2 .....A.....T....
zca2.5 .....C.....A..G.....C.....T....
zmtB2.3 .....A.....T....
zid2.1 .....C.....A..G.....C.....T....
zid2.2 .....C.....A..G.....T....
zma1.1 .....A.....T....
zma1.2 .....A.....T.....T....
zma2.2 .....A.....G.....T....
zma2.3 .....A.....T....
zmt2.3 .....A.....T....
zmt2.6 .....A..G.....C.....T....
zmtB1.3 .....A.....T....
zmtB1.6 .....A.....T....

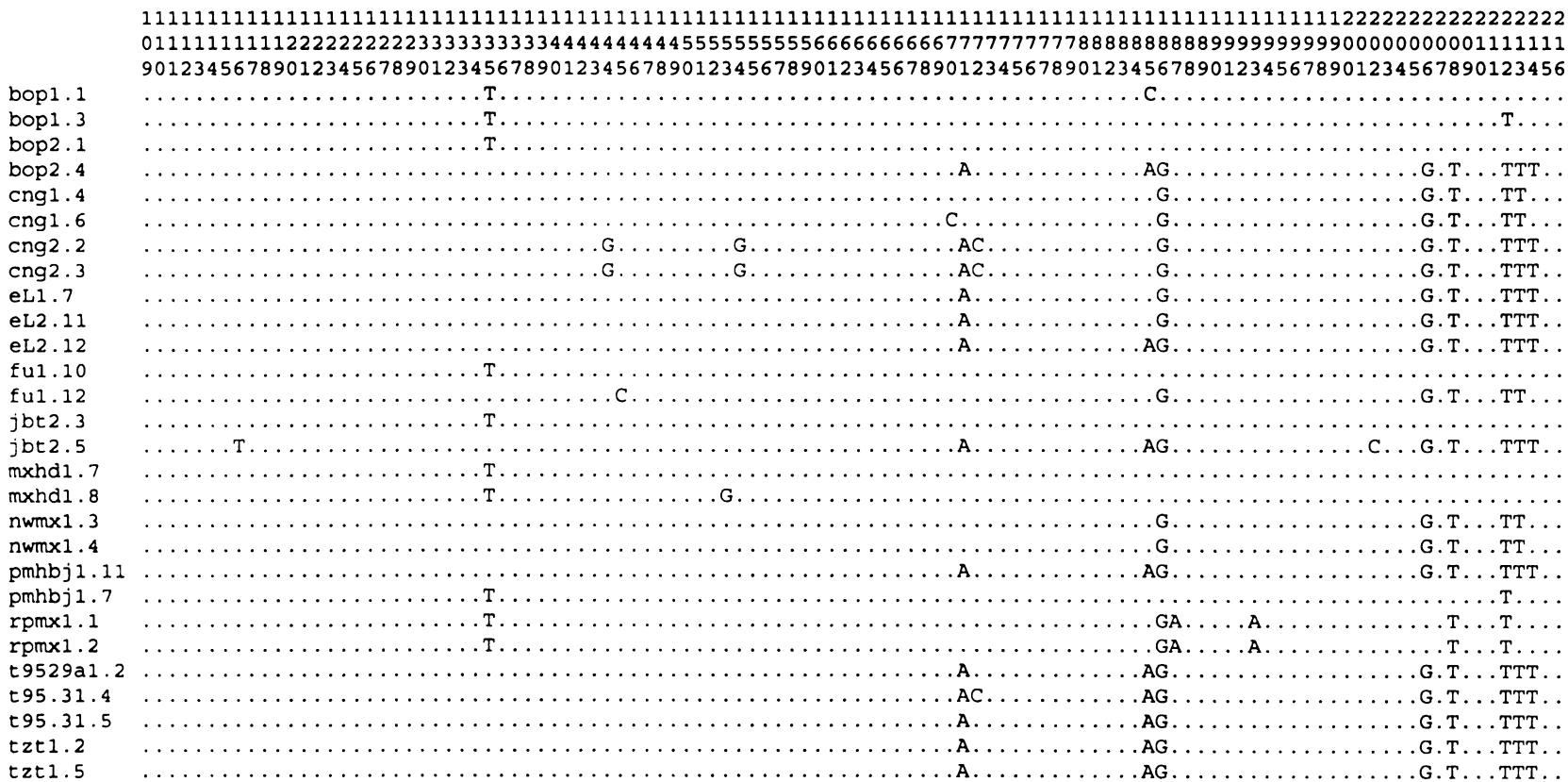

111111111
111111111122222222223333333333444444444455555555556666666666777777777788888888899999999990000000000
12345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678
zny4.3
zsb3.2
zsb3.4
zwa2.1
zwa2.8
pon3.2
pny1.1
pny1.3
pny2.1
pny2.2
put1.2
put1.4
put2.4
pwa3.4
c10.3
c10.4
c9.5
f1.6
f6.5
m1.1
m1.4
NC32ma1
NC32ma2
NC44da2
NC44da3
p1.2
p3.1
p3.4
t3.1
t3.3
t4.1
zor5.3
zor5.6
zor6.7
zor6.1
aap1.2
aap1.3
aap2.1
aap2.3

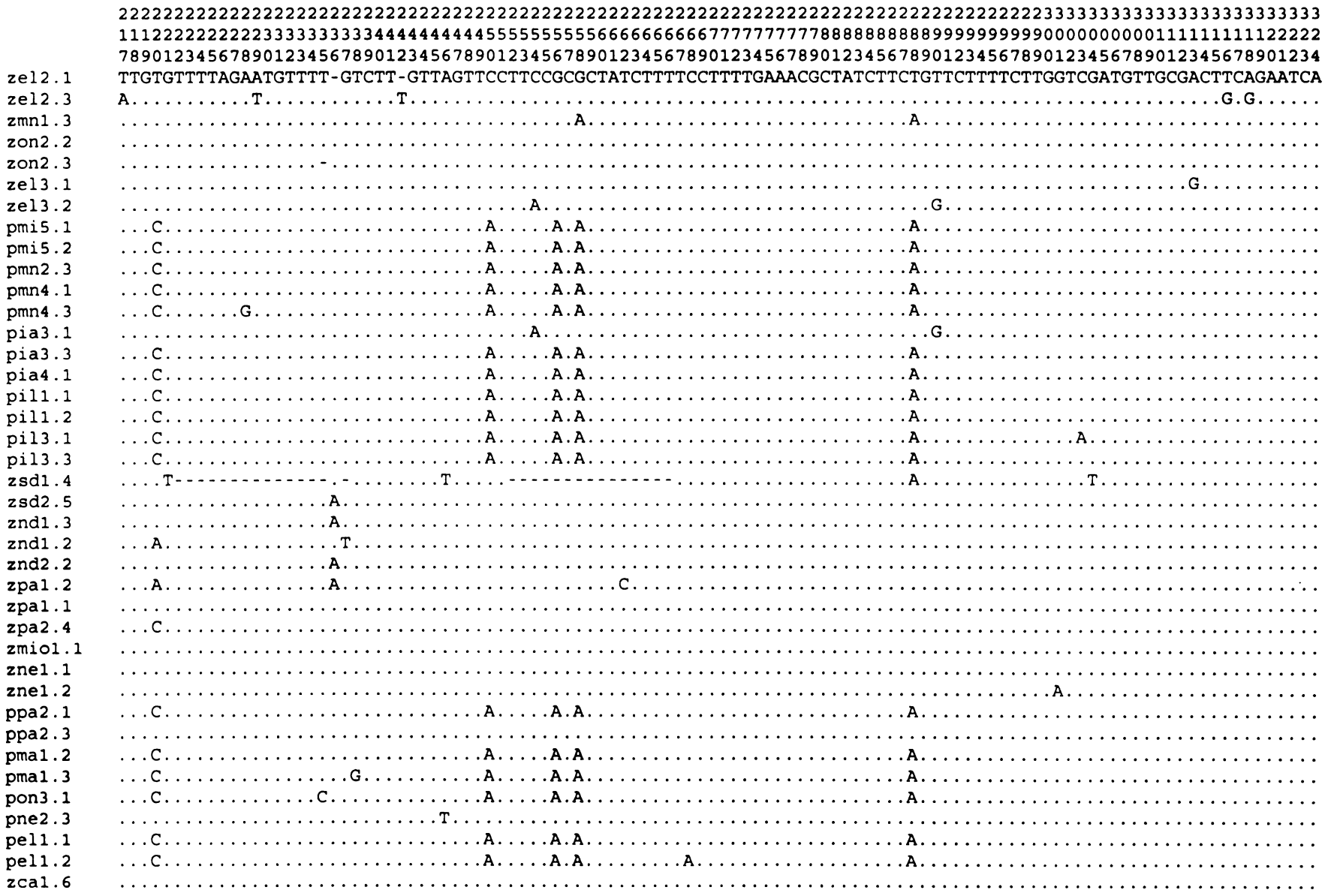

111111111
111111111122222222223333333333444444444455555555556666666666777777777788888888899999999990000000000
12345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678
bop1.1 A T
bop1.3 A T
bop2.1 A T
bop2.4 A G A AG C A T
cng1.4 T A T C C
cng1.6 T A T C C
cng2.2 G A C A A C C G T
cng2.3 G A AC A C C G T
eL1.7 A AG C A T
eL2.11 A AG C A T
eL2.12 A G A AG C A T
fu1.10 A T
fu1.12 A A T C T
jbt2.3 A T
jbt2.5 A G A AG C A T
mxhd1.7 A T
mxhd1.8 A T
nwmx1.3 T A T C C
nwmx1.4 T A T C C
pmhbj1.11 A G A AG C A T
pmhbj1.7 A T
rpmx1.1 A T
rpmx1.2 A T
t9529a1.2 A A AG C A T
t95.31.4 A A AG C A T
t95.31.5 A A AG C A T
tzt1.2 A G A AG C A T
tzt1.5 A A AG C T

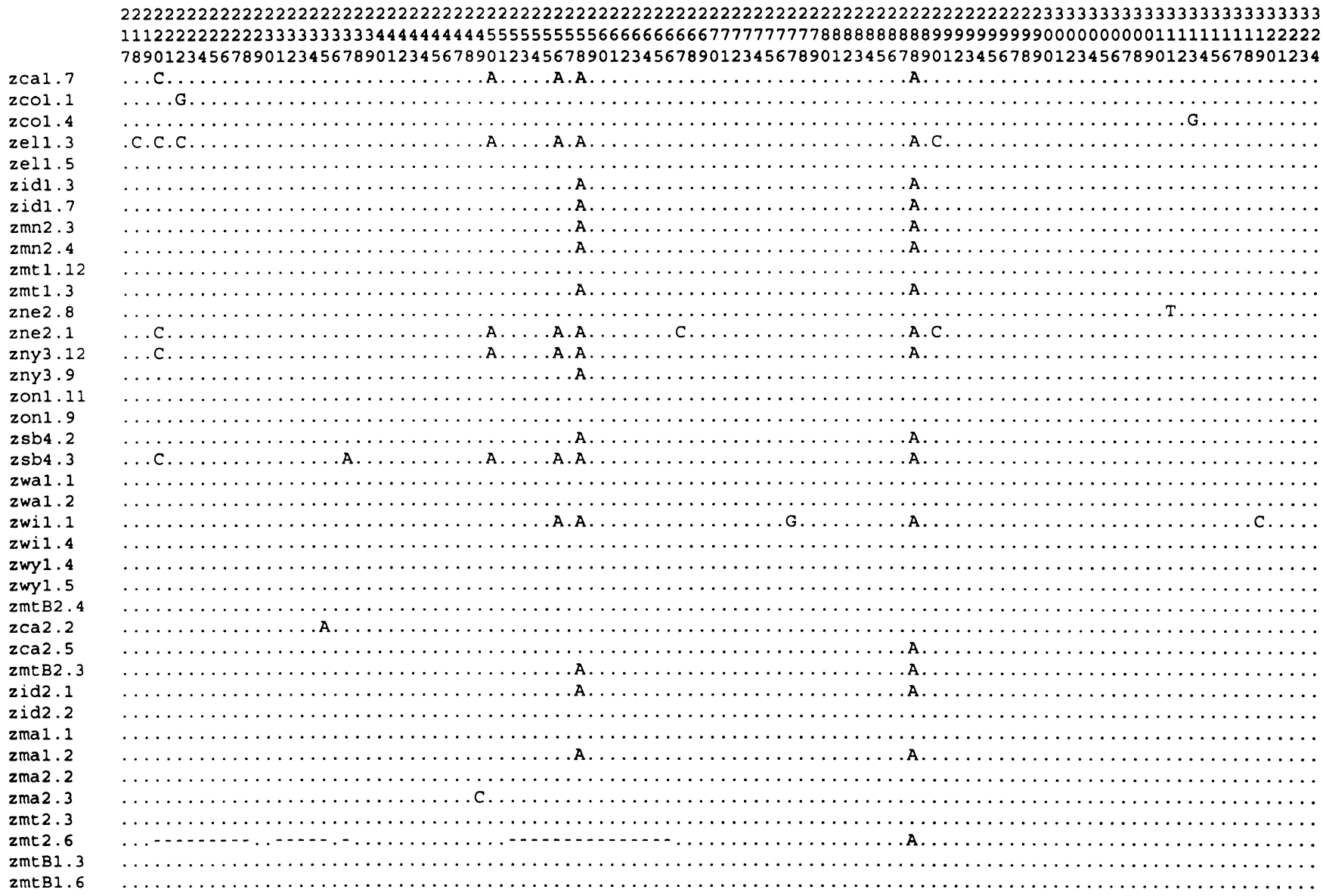

zel2.1 TCCATAGAGCAT-GTGCCGCTTCATACTCACTACCATAAAAATTATACTGAGACCATTGGGTGTCATTTTG-AGTTTCTTTAGCTTTTTCTGAAGAATTCTTTCCCCC
zel2.3
zmn1.3
zon2.2
zon2.3
zel3.1
zel3.2
pmi5.1
pmi5.2
pmn2.3
pmn4.1
pmn4.3
pia3.1
pia3.3
pia4.1
pil1.1
pil1.2
pil3.1
pil3.3
zsd1.4
zsd2.5
znd1.3
znd1.2
znd2.2
zpa1.2
zpa1.1
zpa2.4
zmio1.1
zne1.1
zne1.2
ppa2.1
ppa2.3
pma1.2
pma1.3
pon3.1
pne2.3
pel1.1
pel1.2
zca1.6

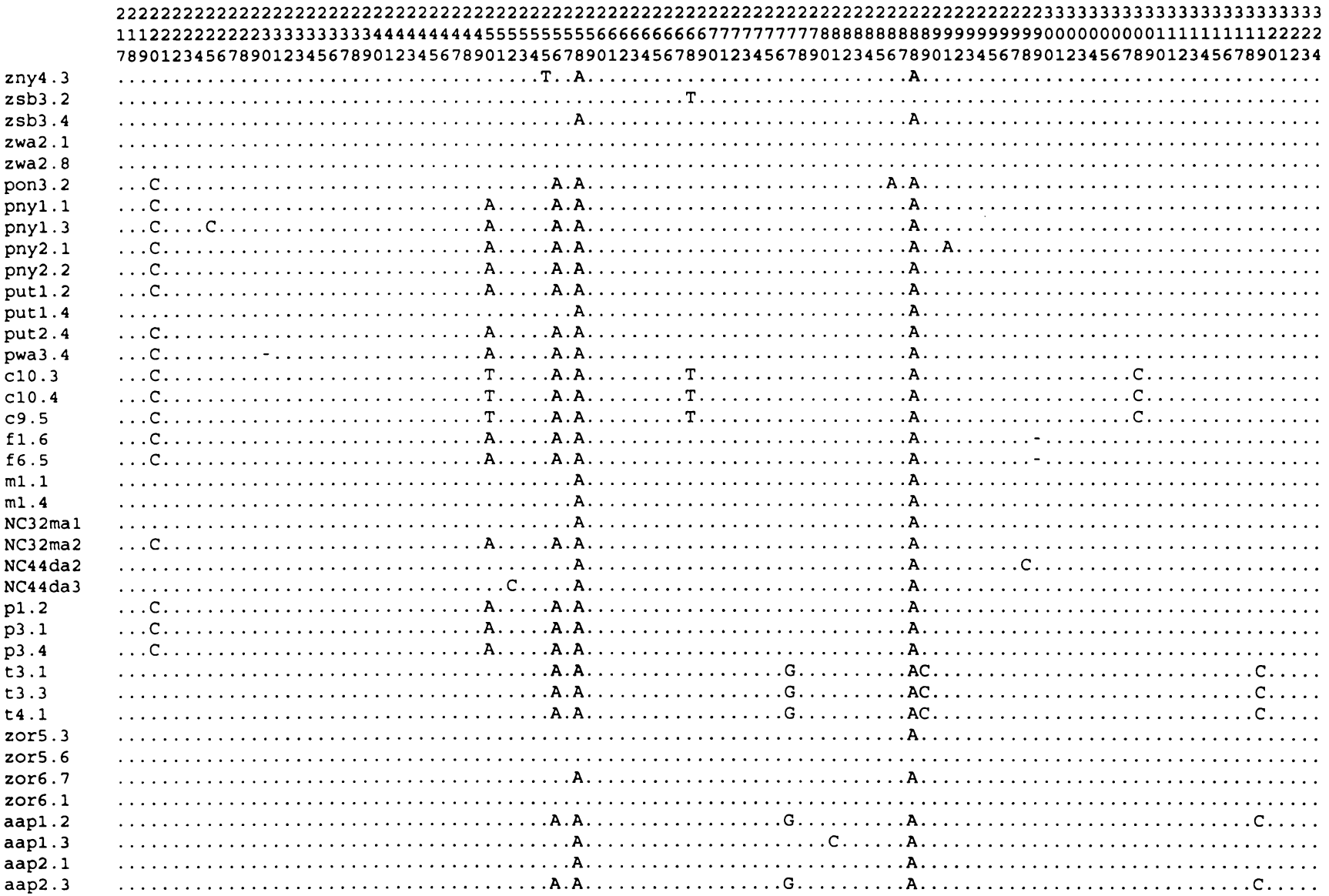

zca1.7
zco1.1
zco1.4
zel1.3
zel1.5
zid1.3
zid1.7
zmn2.3
zmn2.4
zmt1.12
zmt1.3
zne2.8
zne2.1
zny3.12
zny3.9
zon1.11
zon1.9
zsb4.2
zsb4.3
zwa1.1
zwa1.2
zwi1.1
zwi1.4
zwy1.4
zwy1.5
zmtB2.4
zca2.2
zca2.5
zmtB2.3
zid2.1
zid2.2
zma1.1
zma1.2
zma2.2
zma2.3
zmt2.3
zmt2.6
zmtB1.3
zmtB1.6

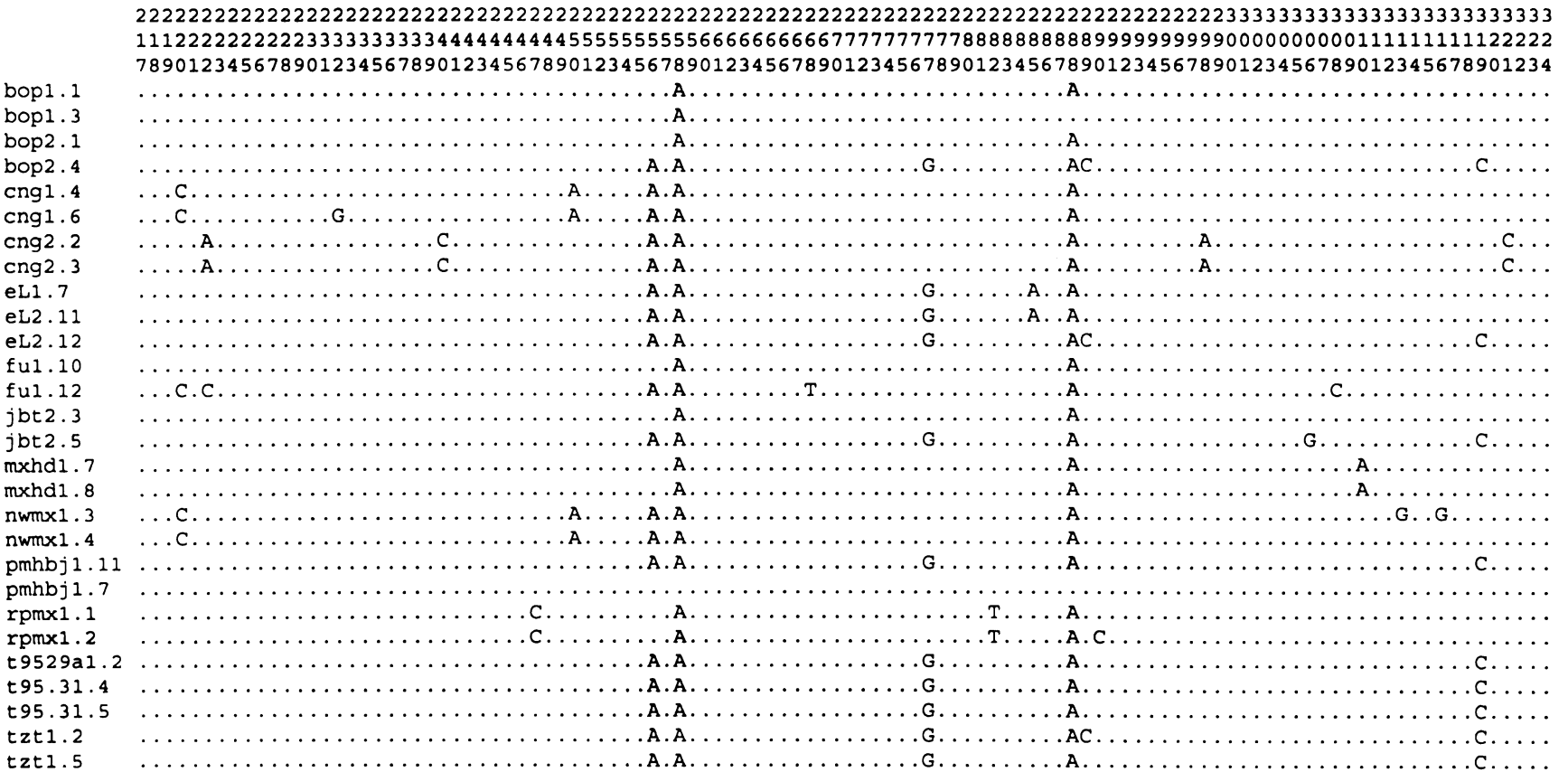

zny4.3
zsb3.2
zsb3.4
zwa2.1
zwa2.8
pon3.2
pny1.1
pny1.3
pny2.1
pny2.2
put1.2
put1.4
put2.4
pwa3.4
c10.3
c10.4
c9.5
f1.6
f6.5
m1.1
m1.4
NC32ma1
NC32ma2
NC44da2
NC44da3
p1.2
p3.1
p3.4
t3.1
t3.3
t4.1
zor5.3
zor5.6
zor6.7
zor6.1
aap1.2
aap1.3
aap2.1
aap2.3

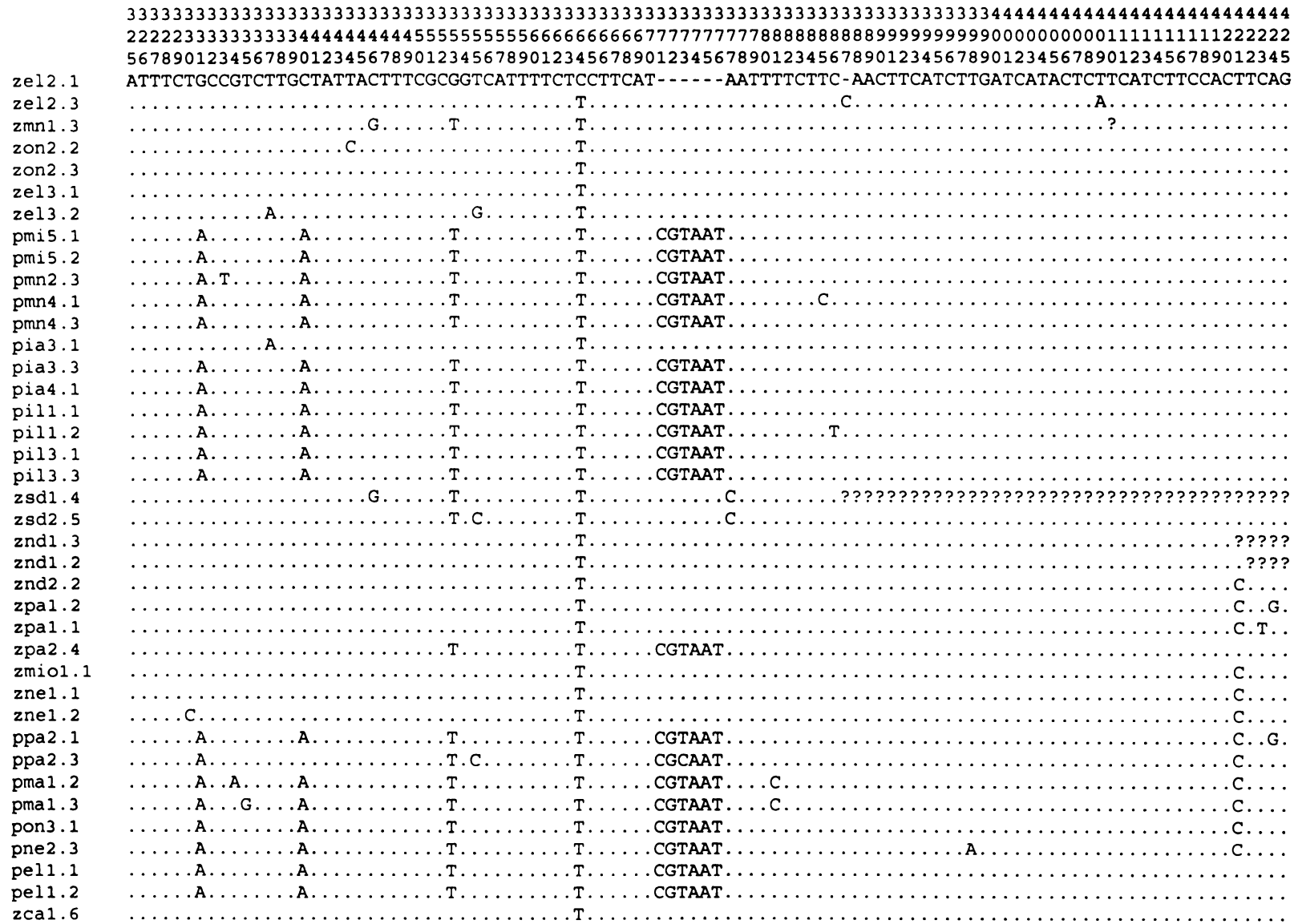

111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111222222222222222222
011111111112222222222333333333344444444445555555555666666666677777777778888888888999999999900000000001111111
901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456
bop1.1 ..........T..........C.....
bop1.3 ..........T..........T....
bop2.1 ....T..........
bop2.4 ..........A.....AG.....G.T...TTT..
cng1.4 ..........G.....G.T...TT...
cng1.6 ..........C.....G.....G.T...TT...
cng2.2 .....G.....G.....AC.....G.....G.T...TTT..
cng2.3 .....G.....G.....AC.....G.....G.T...TTT..
eL1.7 ..........A.....G.....G.T...TTT..
eL2.11 ..........A.....G.....G.T...TTT..
eL2.12 ..........A.....AG.....G.T...TTT..
fu1.10 ....T..........
fu1.12 .....C..........G.....G.T...TT...
jbt2.3 ....T..........
jbt2.5 .....T..........A.....AG.....C...G.T...TTT..
mxhd1.7 ....T..........
mxhd1.8 ..........T.....G..........
nwmx1.3 ..........G.....G.T...TT...
nwmx1.4 ..........G.....G.T...TT...
pmhbj1.11 ..........A.....AG.....G.T...TTT..
pmhbj1.7 ..........T..........T....
rpmx1.1 ..........T..........GA.....A.....T...T....
rpmx1.2 ..........T..........GA.....A.....T...T....
t9529a1.2 ..........A.....AG.....G.T...TTT..
t95.31.4 ..........AC.....AG.....G.T...TTT..
t95.31.5 ..........A.....AG.....G.T...TTT..
tzt1.2 ..........A.....AG.....G.T...TTT..
tzt1.5 ..........A.....AG.....G.T...TTT..

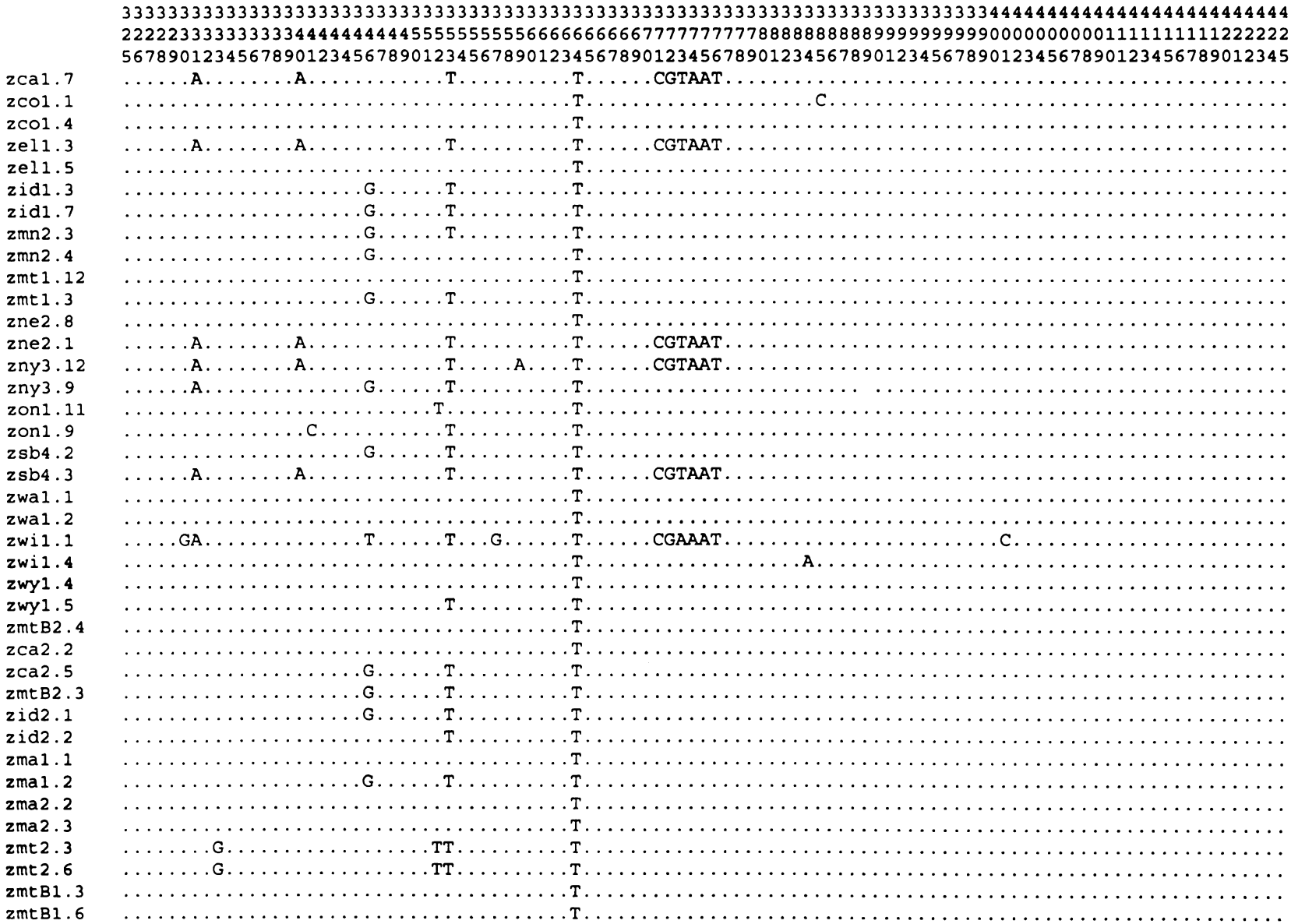

zel2.1 TTGTGTTTTAGAATGTTTT-GTCTT-GTTAGTTCCTTCCGCGCTATCTTTTCCTTTTGAAACGCTATCTTCTGTTCTTTTCTTGGTCGATGTTGCGACTTCAGAATCA
zel2.3 A.....T.....T.....G.G.....
zmn1.3 .....A.....A.....
zon2.2 .....
zon2.3 .....-.....
zel3.1 .....G.....
zel3.2 .....A.....G.....
pmi5.1 ...C.....A.....A.A.....A.....
pmi5.2 ...C.....A.....A.A.....A.....
pmn2.3 ...C.....A.....A.A.....A.....
pmn4.1 ...C.....A.....A.A.....A.....
pmn4.3 ...C.....G.....A.....A.A.....A.....
pia3.1 .....A.....G.....
pia3.3 ...C.....A.....A.A.....A.....
pia4.1 ...C.....A.....A.A.....A.....
pil1.1 ...C.....A.....A.A.....A.....
pil1.2 ...C.....A.....A.A.....A.....
pil3.1 ...C.....A.....A.A.....A.....A.....
pil3.3 ...C.....A.....A.A.....A.....
zsd1.4 ....T----------.....T.....----------.....A.....T.....
zsd2.5 .....A.....
znd1.3 .....A.....
znd1.2 ...A.....T.....
znd2.2 .....A.....
zpa1.2 ...A.....A.....C.....
zpa1.1 .....
zpa2.4 ...C.....
zmio1.1 .....
zne1.1 .....
zne1.2 .....A.....
ppa2.1 ...C.....A.....A.A.....A.....
ppa2.3 .....
pma1.2 ...C.....A.....A.A.....A.....
pma1.3 ...C.....G.....A.....A.A.....A.....
pon3.1 ...C.....C.....A.....A.A.....A.....
pne2.3 .....T.....
pel1.1 ...C.....A.....A.A.....A.....
pel1.2 ...C.....A.....A.A.....A.....A.....
zca1.6 .....

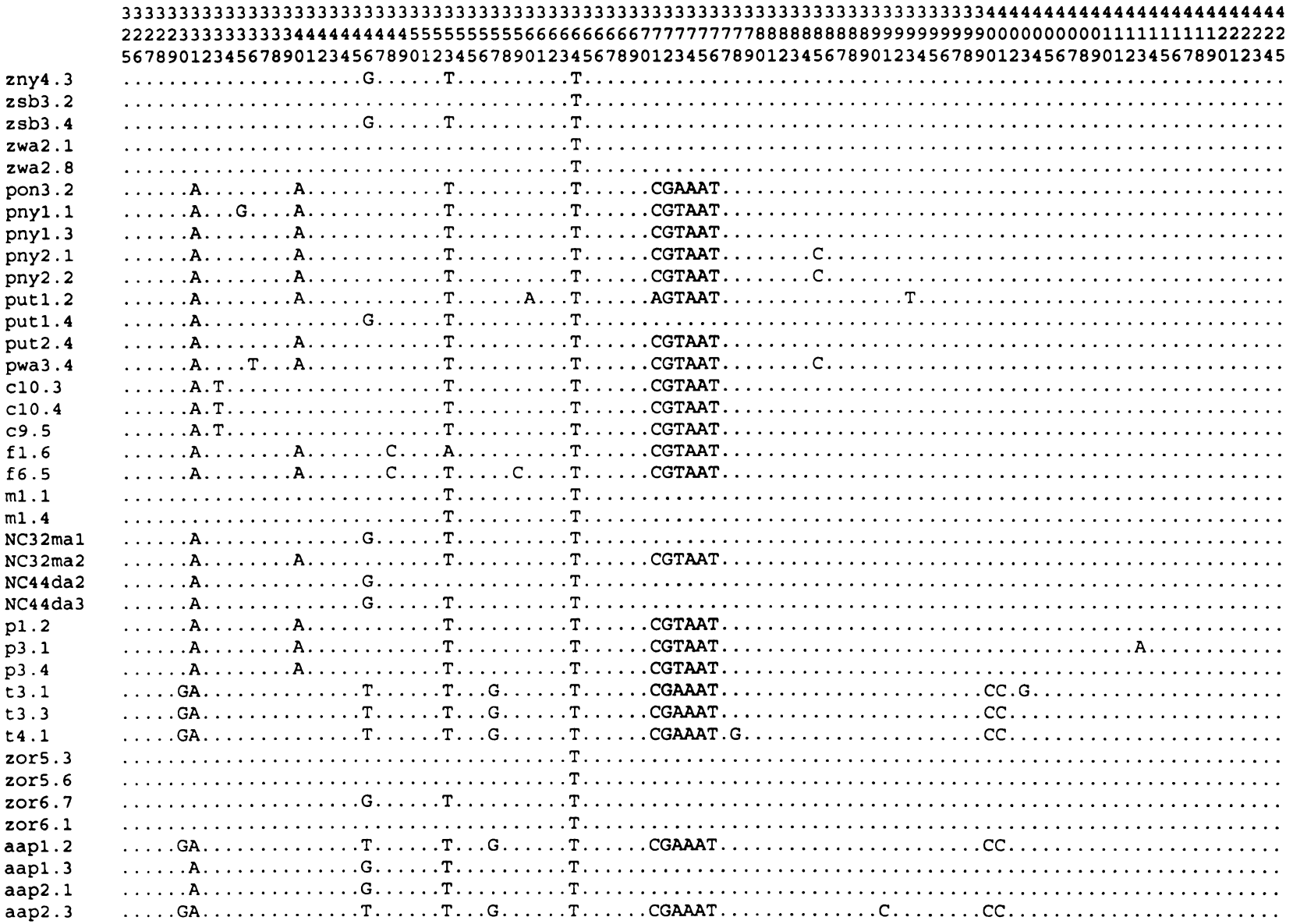

zca1.7
zco1.1
zco1.4
zel1.3
zel1.5
zid1.3
zid1.7
zmn2.3
zmn2.4
zmt1.12
zmt1.3
zne2.8
zne2.1
zny3.12
zny3.9
zon1.11
zon1.9
zsb4.2
zsb4.3
zwa1.1
zwa1.2
zwi1.1
zwi1.4
zwy1.4
zwy1.5
zmtB2.4
zca2.2
zca2.5
zmtB2.3
zid2.1
zid2.2
zma1.1
zma1.2
zma2.2
zma2.3
zmt2.3
zmt2.6
zmtB1.3
zmtB1.6

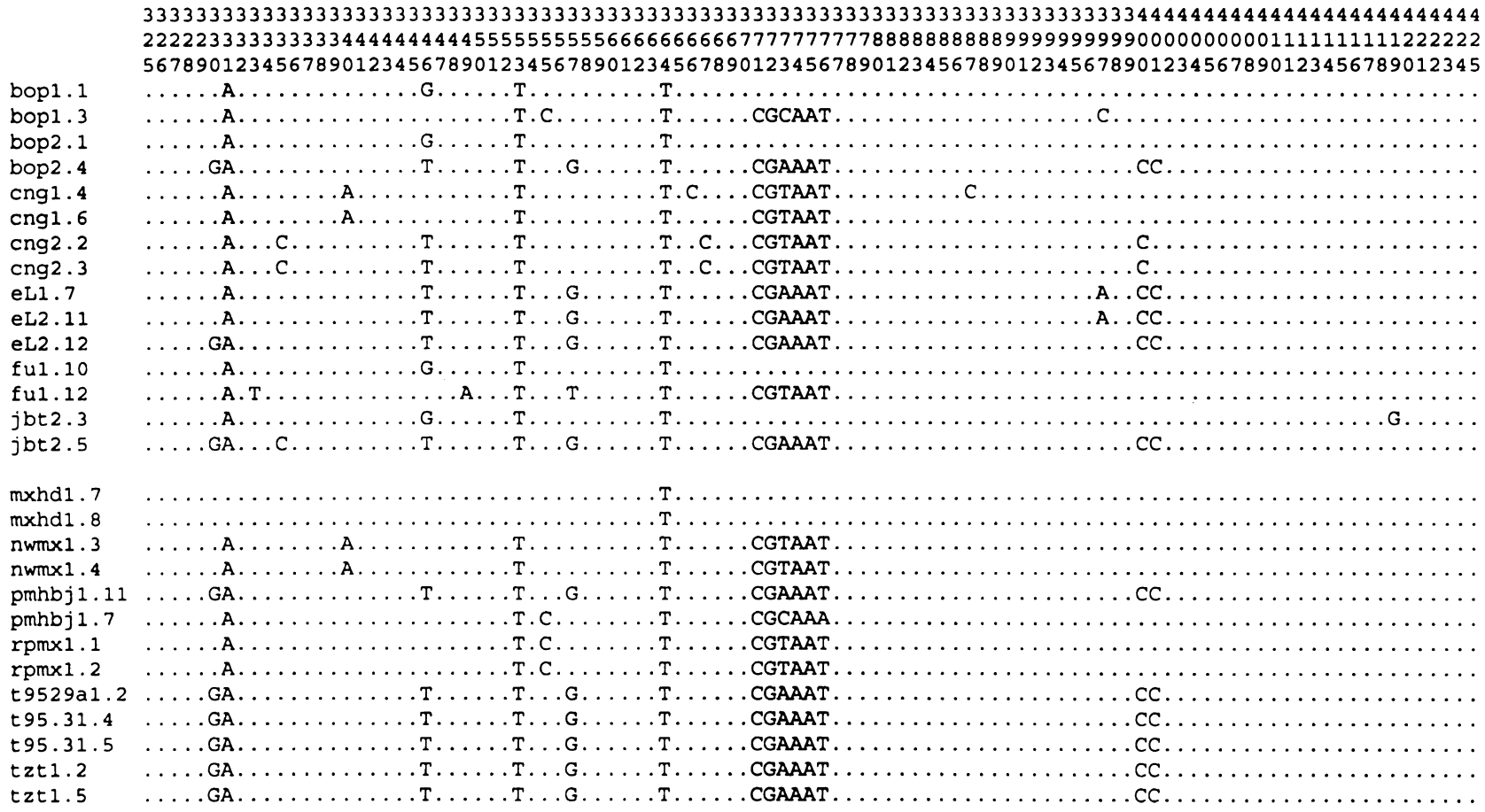

zny4.3
zsb3.2
zsb3.4
zwa2.1
zwa2.8
pon3.2
pny1.1
pny1.3
pny2.1
pny2.2
put1.2
put1.4
put2.4
pwa3.4
c10.3
c10.4
c9.5
f1.6
f6.5
m1.1
m1.4
NC32ma1
NC32ma2
NC44da2
NC44da3
p1.2
p3.1
p3.4
t3.1
t3.3
t4.1
zor5.3
zor5.6
zor6.7
zor6.1
aap1.2
aap1.3
aap2.1
aap2.3


222222222222222222222222222222222222222222222222222222222222222222222222222222233333333333333333333333
111222222222233333333334444444444555555555566666666667777777777888888888999999999900000000001111111111222222
789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234
bop1.1
bop1.3
bop2.1
bop2.4
cng1.4
cng1.6
cng2.2
cng2.3
eL1.7
eL2.11
eL2.12
fu1.10
fu1.12
jbt2.3
jbt2.5
mxhd1.7
mxhd1.8
nwmx1.3
nwmx1.4
pmhbj1.11
pmhbj1.7
rpmx1.1
rpmx1.2
t9529a1.2
t95.31.4
t95.31.5
tzt1.2
tzt1.5


zel2.1 ATTTCTGCCGTCTTGCTATTACTTTCGCGGTCATTTTCTCCTTCAT------AATTTTCTTC-AACTTCATCTTGATCATACTCTTCATCTTCCACTTCAG
zel2.3
zmn1.3
zon2.2
zon2.3
zel3.1
zel3.2
pmi5.1
pmi5.2
pmn2.3
pmn4.1
pmn4.3
pia3.1
pia3.3
pia4.1
pil1.1
pil1.2
pil3.1
pil3.3
zsd1.4
zsd2.5
znd1.3
znd1.2
znd2.2
zpa1.2
zpa1.1
zpa2.4
zmio1.1
zne1.1
zne1.2
ppa2.1
ppa2.3
pma1.2
pma1.3
pon3.1
pne2.3
pel1.1
pel1.2
zca1.6

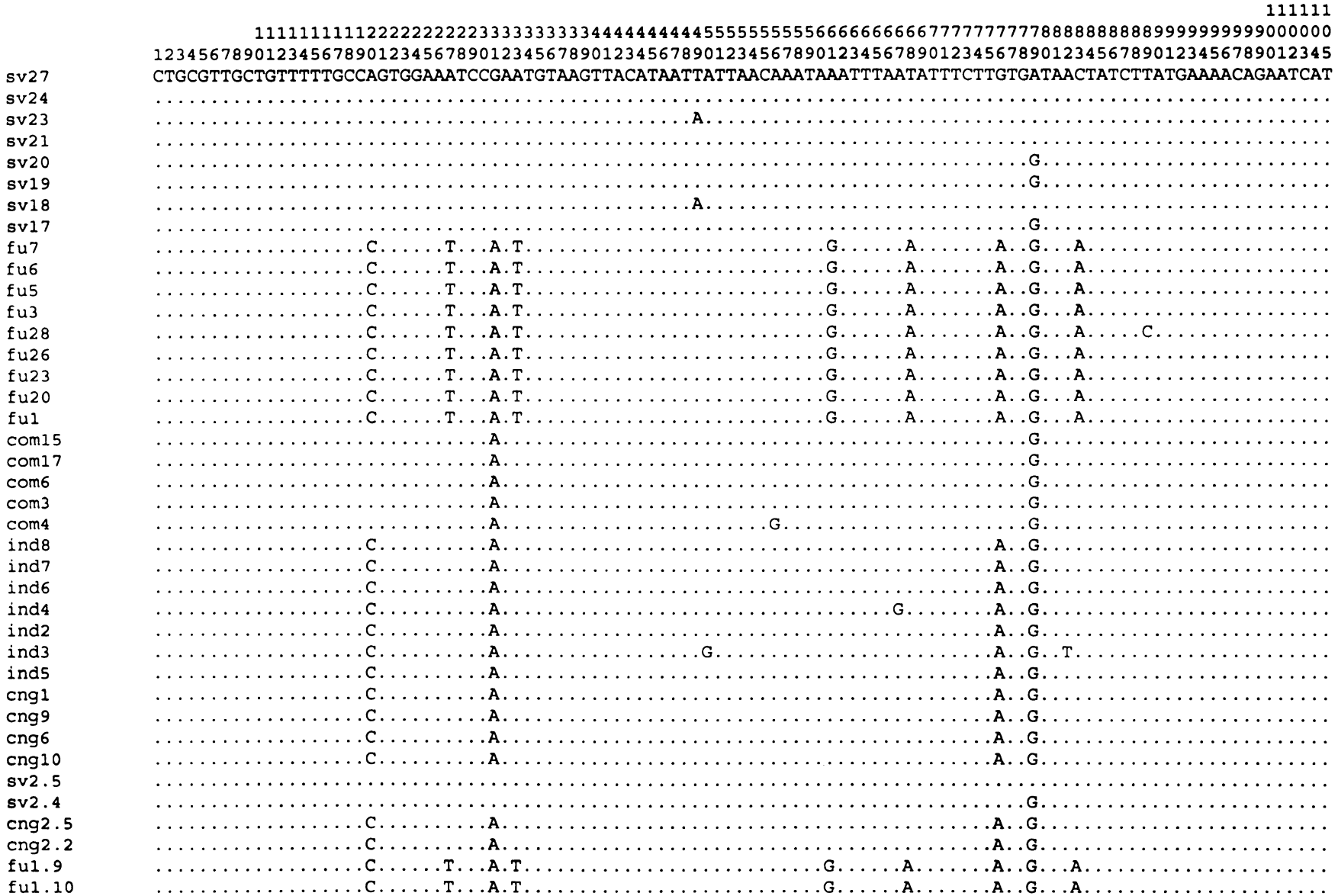

zca1.7 ......A.....A.....T.....T.....CGTAAT.....
zco1.1 .....T.....C.....
zco1.4 .....T.....
zel1.3 ......A.....A.....T.....T.....CGTAAT.....
zel1.5 .....T.....
zid1.3 .....G.....T.....T.....
zid1.7 .....G.....T.....T.....
zmn2.3 .....G.....T.....T.....
zmn2.4 .....G.....T.....
zmt1.12 .....T.....
zmt1.3 .....G.....T.....T.....
zne2.8 .....T.....
zne2.1 ......A.....A.....T.....T.....CGTAAT.....
zny3.12 ......A.....A.....T.....A....T.....CGTAAT.....
zny3.9 ......A.....G.....T.....T.....
zon1.11 .....T.....T.....
zon1.9 .....C.....T.....T.....
zsb4.2 .....G.....T.....T.....
zsb4.3 ......A.....A.....T.....T.....CGTAAT.....
zwa1.1 .....T.....
zwa1.2 .....T.....
zwi1.1 .....GA.....T.....T...G.....T.....CGAAAT.....C.....
zwi1.4 .....T.....A.....
zwy1.4 .....T.....
zwy1.5 .....T.....T.....
zmtB2.4 .....T.....
zca2.2 .....T.....
zca2.5 .....G.....T.....T.....
zmtB2.3 .....G.....T.....T.....
zid2.1 .....G.....T.....T.....
zid2.2 .....T.....T.....
zma1.1 .....T.....
zma1.2 .....G.....T.....T.....
zma2.2 .....T.....
zma2.3 .....T.....
zmt2.3 .....G.....TT.....T.....
zmt2.6 .....G.....TT.....T.....
zmtB1.3 .....T.....
zmtB1.6 .....T.....


zny4.3
zsb3.2
zsb3.4
zwa2.1
zwa2.8
pon3.2
pny1.1
pny1.3
pny2.1
pny2.2
put1.2
put1.4
put2.4
pwa3.4
c10.3
c10.4
c9.5
f1.6
f6.5
m1.1
m1.4
NC32ma1
NC32ma2
NC44da2
NC44da3
p1.2
p3.1
p3.4
t3.1
t3.3
t4.1
zor5.3
zor5.6
zor6.7
zor6.1
aap1.2
aap1.3
aap2.1
aap2.3

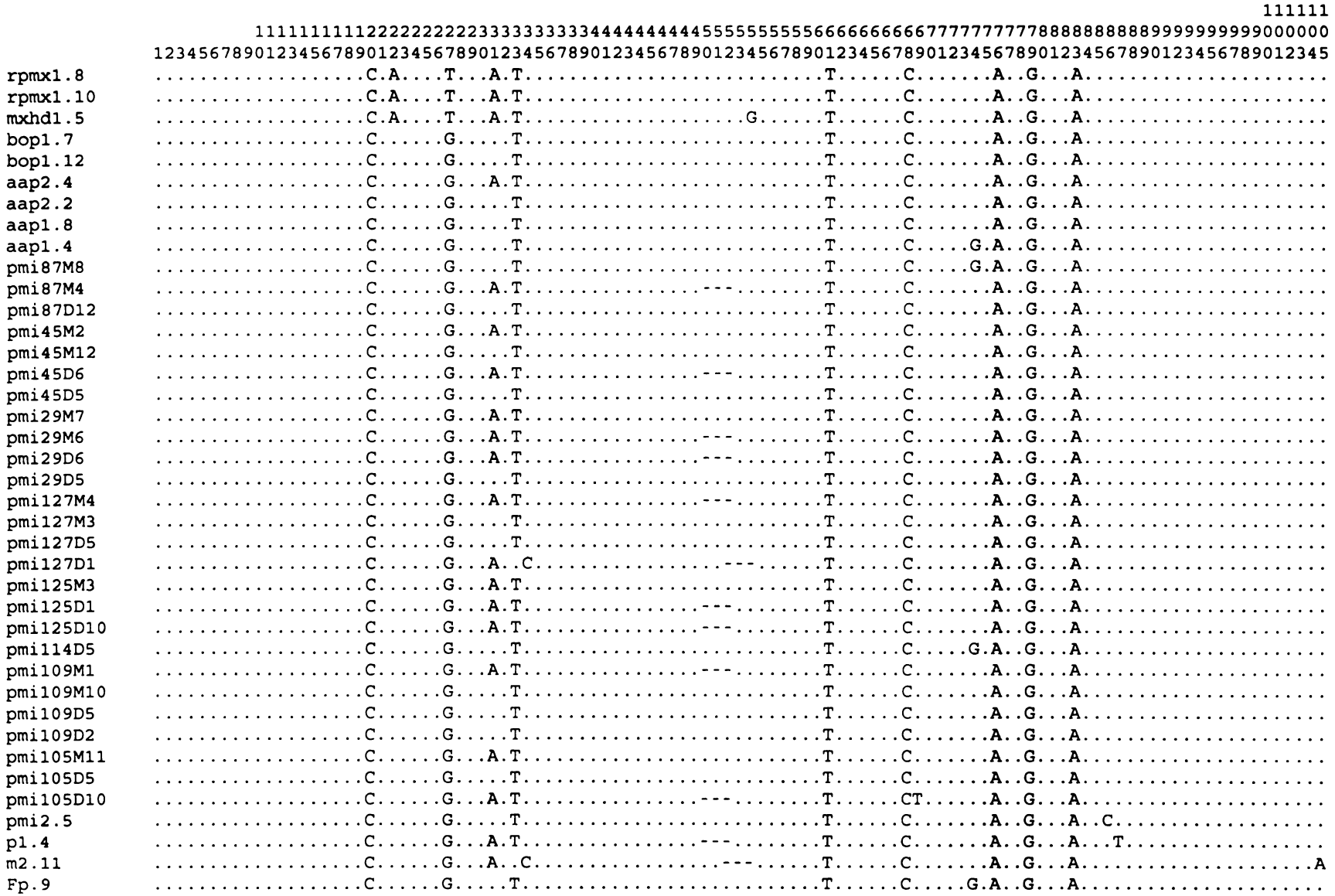

bop1.1 ......A.....G.....T.....T.....
bop1.3 ......A.....T.C.....T.....CGCAAT.....C.....
bop2.1 ......A.....G.....T.....T.....
bop2.4 .....GA.....T.....T...G.....T.....CGAAAT.....CC.....
cng1.4 ......A.....A.....T.....T.C....CGTAAT.....C.....
cng1.6 ......A.....A.....T.....T.....CGTAAT.....
cng2.2 ......A...C.....T.....T.....T..C...CGTAAT.....C.....
cng2.3 ......A...C.....T.....T.....T..C...CGTAAT.....C.....
eL1.7 ......A.....T.....T...G.....T.....CGAAAT.....A..CC.....
eL2.11 ......A.....T.....T...G.....T.....CGAAAT.....A..CC.....
eL2.12 .....GA.....T.....T...G.....T.....CGAAAT.....CC.....
fu1.10 ......A.....G.....T.....T.....
fu1.12 ......A.T.....A...T...T.....T.....CGTAAT.....
jbt2.3 ......A.....G.....T.....T.....G.....
jbt2.5 .....GA...C.....T.....T...G.....T.....CGAAAT.....CC.....
mxhd1.7 .....T.....
mxhd1.8 .....T.....
nwmx1.3 ......A.....A.....T.....T.....CGTAAT.....
nwmx1.4 ......A.....A.....T.....T.....CGTAAT.....
pmhbj1.11 .....GA.....T.....T...G.....T.....CGAAAT.....CC.....
pmhbj1.7 ......A.....T.C.....T.....CGCAAA.....
rpmx1.1 ......A.....T.C.....T.....CGTAAT.....
rpmx1.2 ......A.....T.C.....T.....CGTAAT.....
t9529a1.2 .....GA.....T.....T...G.....T.....CGAAAT.....CC.....
t95.31.4 .....GA.....T.....T...G.....T.....CGAAAT.....CC.....
t95.31.5 .....GA.....T.....T...G.....T.....CGAAAT.....CC.....
tzt1.2 .....GA.....T.....T...G.....T.....CGAAAT.....CC.....
tzt1.5 .....GA.....T.....T...G.....T.....CGAAAT.....CC.....

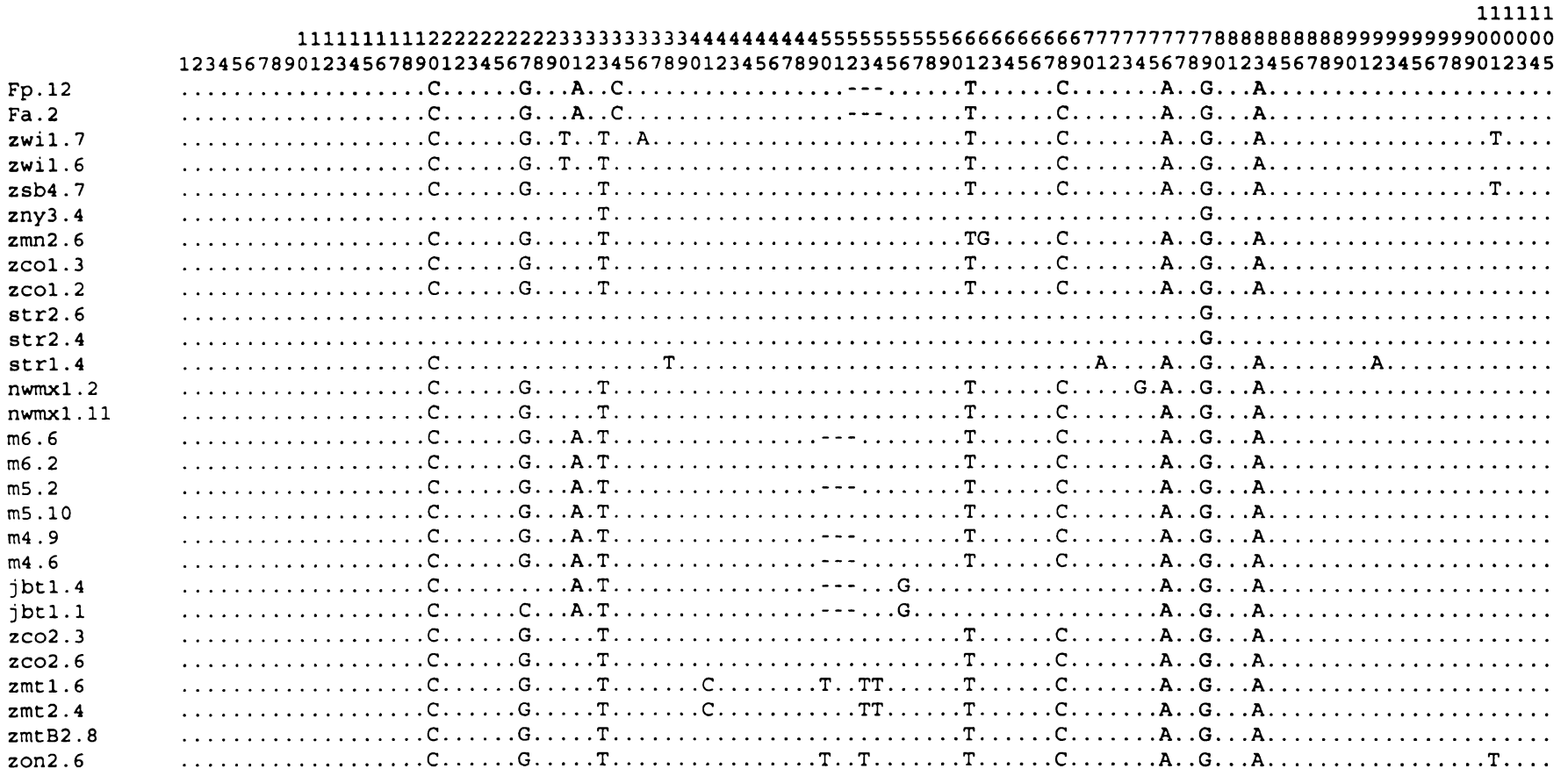
APPENDIX B

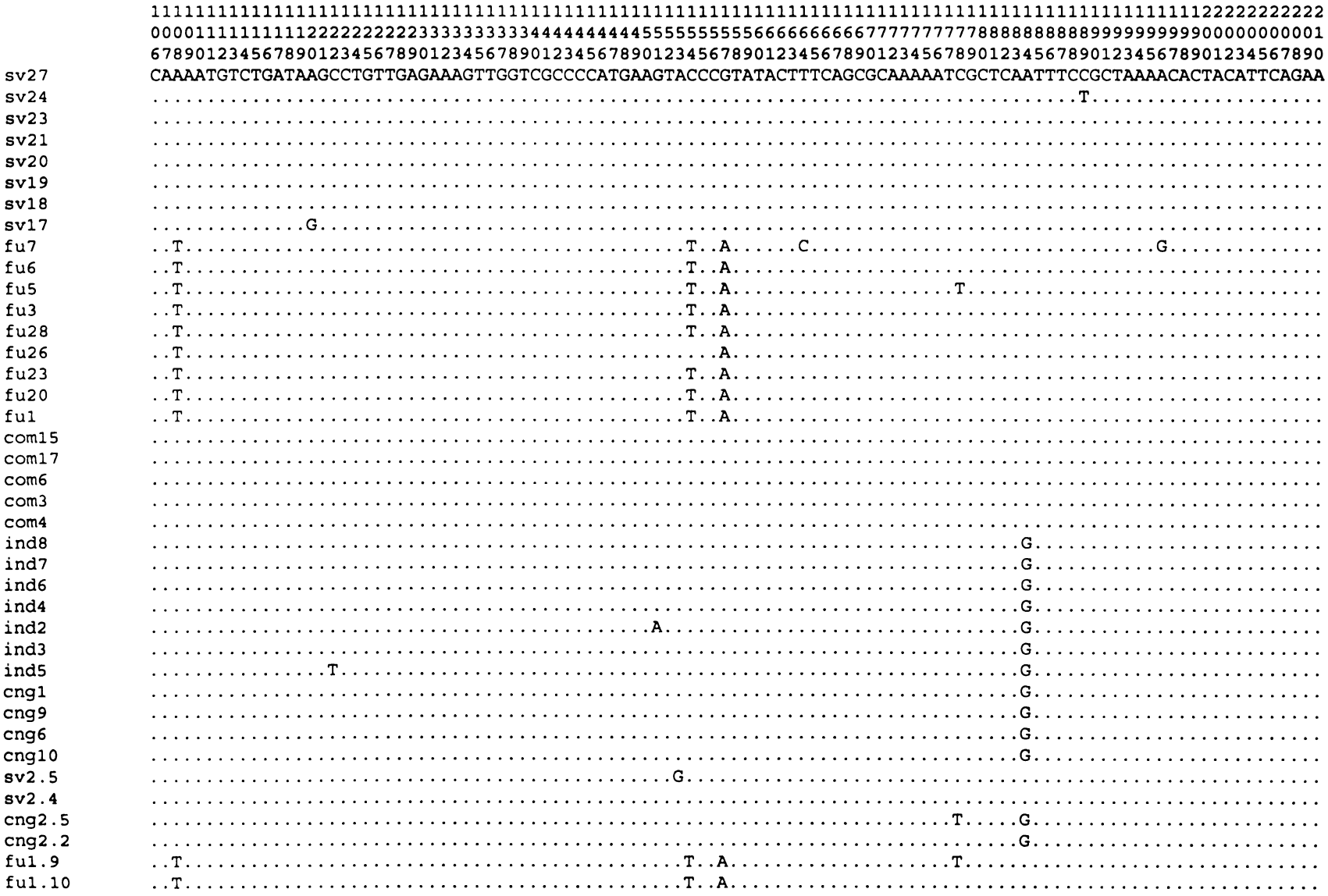
P2956 sequence alignment

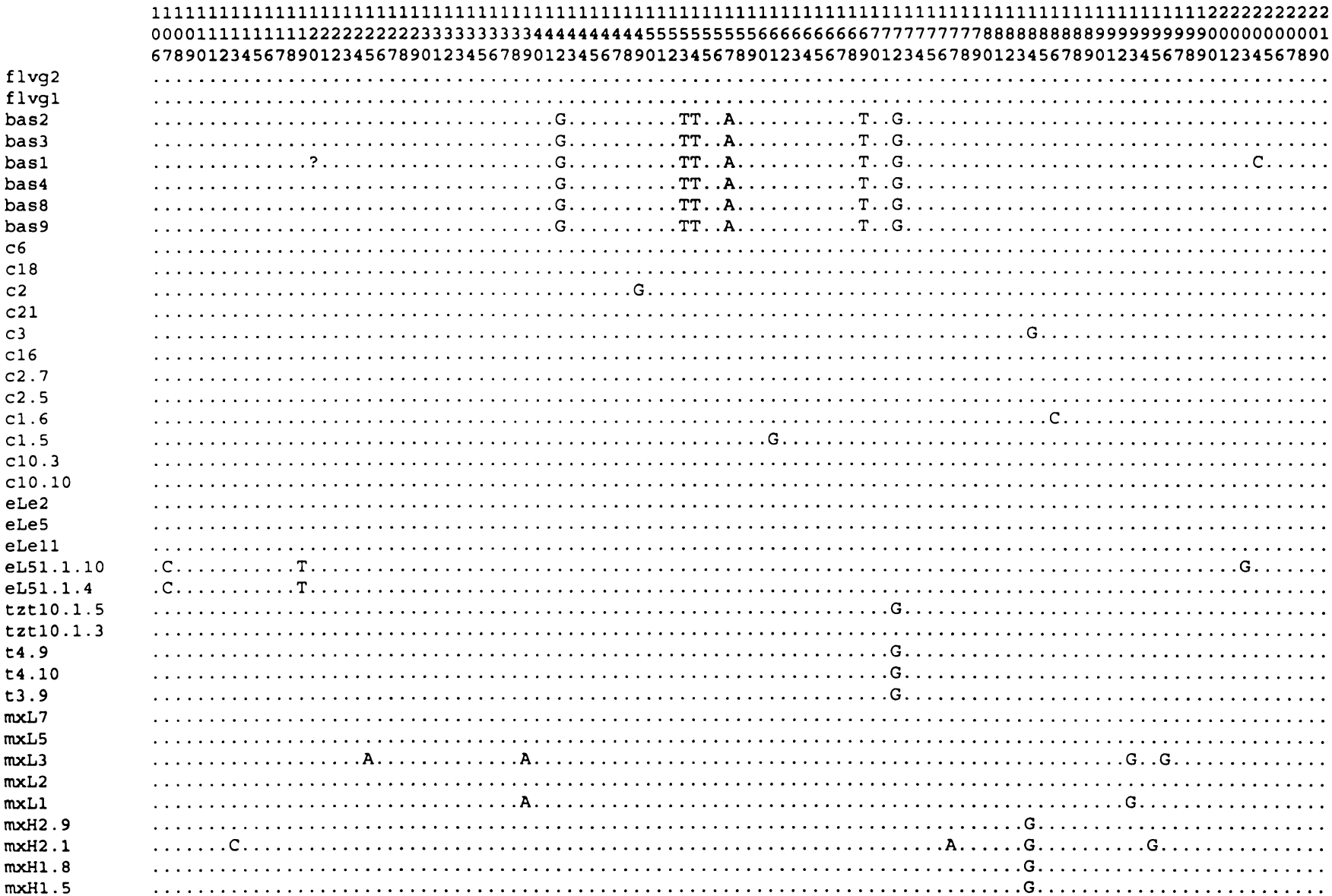

111111
1111111111222222222233333333334444444444555555555566666666667777777777888888888899999999990000000
12345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345
sv27 CTGCGTTGCTGTTTTTGCCAGTGGAAATCCGAATGTAAGTTACATAATTATTAACAAATAAATTTAATATTTCTTGTGATAACTATCTTATGAAAACAGAATCAT
sv24
sv23
sv21
sv20
sv19
sv18
sv17
fu7
fu6
fu5
fu3
fu28
fu26
fu23
fu20
fu1
com15
com17
com6
com3
com4
ind8
ind7
ind6
ind4
ind2
ind3
ind5
cng1
cng9
cng6
cng10
sv2.5
sv2.4
cng2.5
cng2.2
fu1.9
fu1.10

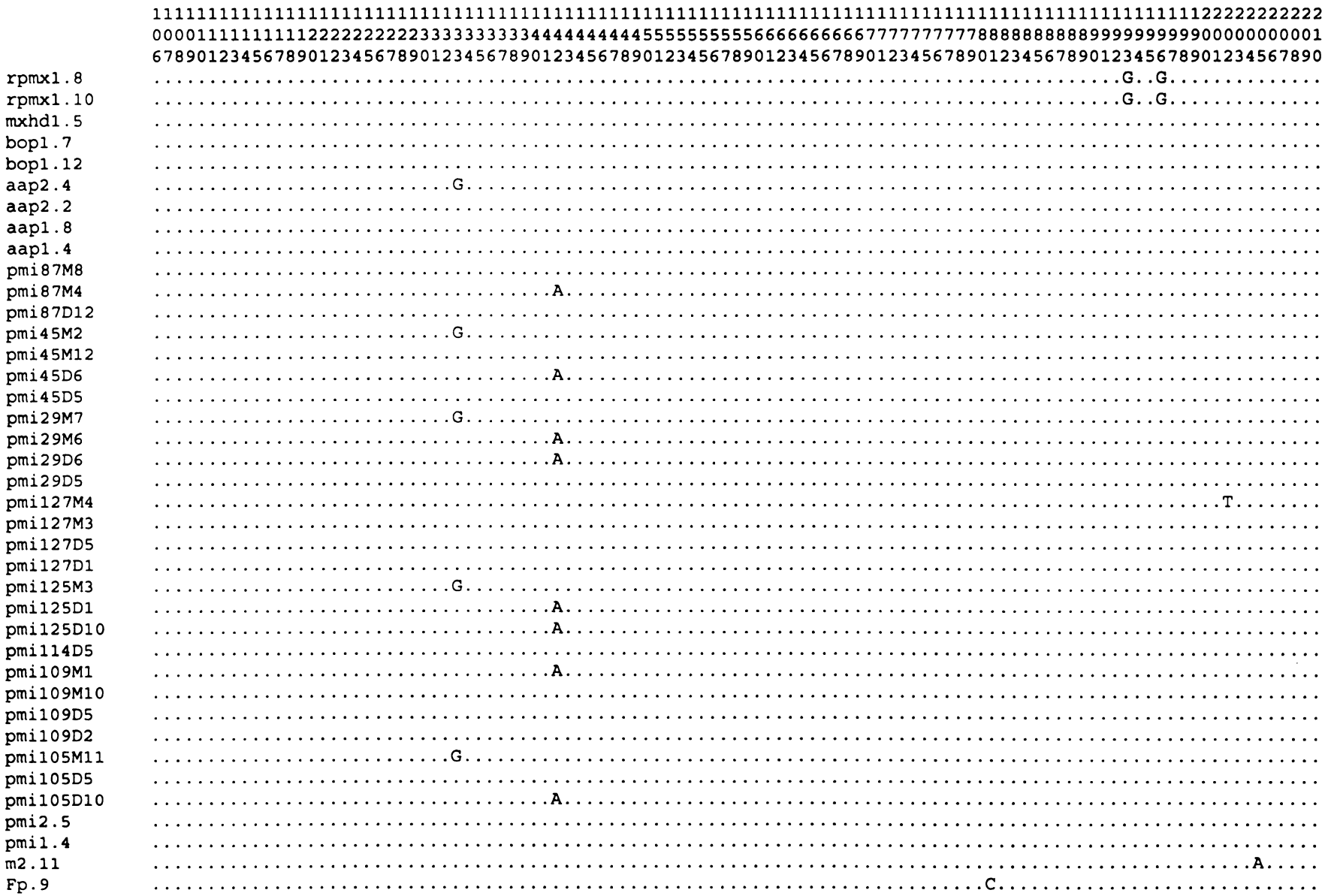
```
                                                                                                             111111
                     1111111111222222222233333333334444444444555555555566666666667777777777888888888899999999990000000
            123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345
flvg2       ..............................................T..........................................................
flvg1       ...........................C..................T..........................................................
bas2        ...................T.........TA.T.........................A.G.A............A.AG.TCA.G.T..................
bas3        ...................T.........TA.T.........................A.G.A............A.AG.TCA.G.T..................
bas1        ...................T.........TA.T.........................A.G.A............A.AG.TCA.G.T..................
bas4        ...................T.........TA.T.........................A.G.A............A.AG.TCA.G.T..................
bas8        ...................T.........TA.T.......................T.A.G.A............A.AG.TCA.G.T..................
bas9        ...................T.........TA.T.........................A.G.A............A.AG.TCA.G.T..........T.......
c6          ...................C..........A.T...........................T....T.........A..G...A.............T........
c18         ...................C..........A.T...........................T....T.........A..G...A.............T........
c2          ...................C..........A.T...........................T....T.........A..G...A.............T........
c21         ...................C..........A.T...........................T.C..T.........A..G...A.............T........
c3          ...................C..........A.T................................T.........A..G
c16         ...................C..........A.T...........................T....T.........A..G...A.............T........
c2.7        ...................C..........A.T...........................T....T.........A..G...A.............T........
c2.5        ...................C..........A.T...........................T....T.........A..G...A.............T........
c1.6        ...................C..........A.T...........................T....T.........A..G...A.............T........
c1.5        ...................C......G...A.T................---........T......C.......A..G...A......................
c10.3       ...................C..........A.T...........................T....T.........A..G...A.......C.....T........
c10.10      ...................C..........A.T...........................T....T.........A..G...A.............T........
eLe2        ...................C............T....T......................T.........A....A..G...A........A.............
eLe5        ...................C..........A.T....T......................T.........A....A..G...A........A.............
eLe11       ...................C..........A.T....T......................T.........A....A..G...A........A.............
eL51.1.10   ...................C..........A.TGTG.T..............T..C....T.........A..C.A..G...A........T.......G.....
eL51.1.4    ...................C..........A.TGTG.T..............T..C....T.........A..C.A..G...A........T.......G.....
tzt10.1.5   ...................C..........A.T....T......................T.........A....A..G...A........A.............
tzt10.1.3   ...................C......G.....T...........................T......C.......A..G...A...........G..........
t4.9        ...................C..........A.T....T.......G..............T.........A....A..G...A........A.............
t4.10       ...................C..........A.T....T......................T.........A....A..G...A........A.............
t3.9        ...................C..........A.T....T......................T.........A....A..G...A........A.............
mxL7        ...................C......G.....T...........................T......C.......A..G...A
mxL5        ...................C......G.....T...........................T......C.......A..G...A
mxL3        ...................C.A....T...A.T.......C............G......T......C.......A..G...A......................
mxL2        ...................C......G.....T...........................T......C.......A..G...A
mxL1        ...................C.A....T...A.T....................G......T......C...C...A..G...A......................
mxH2.9      ...................C..........A............................................A..G..........................
mxH2.1      ...................C..........A............................................A..G..........................
mxH1.8      ...................C..........A............................................A..G..........................
mxH1.5      ...................C..........A..A.........................................A..G..........................
```

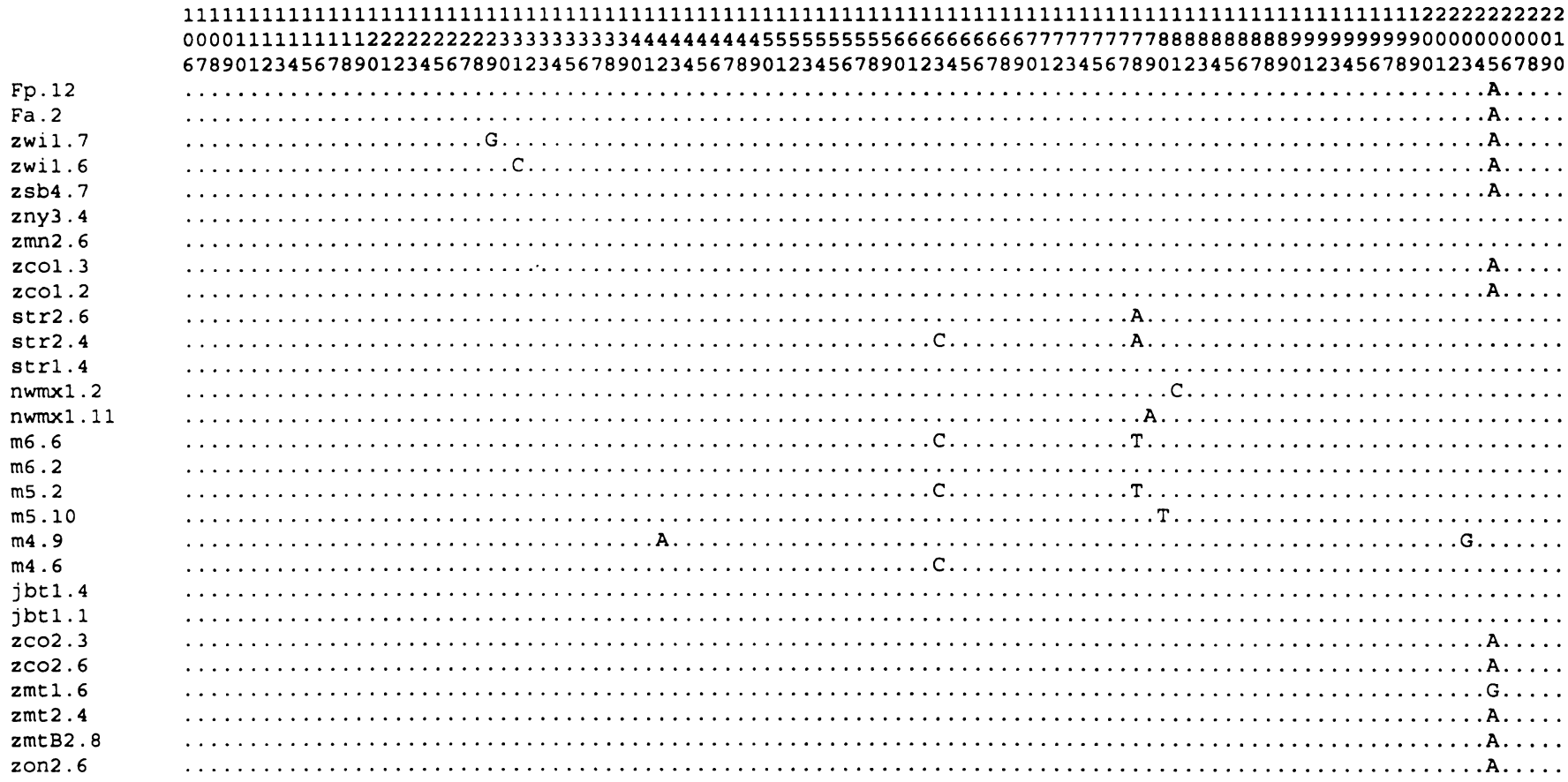

rpmx1.8
rpmx1.10
mxhd1.5
bop1.7
bop1.12
aap2.4
aap2.2
aap1.8
aap1.4
pmi87M8
pmi87M4
pmi87D12
pmi45M2
pmi45M12
pmi45D6
pmi45D5
pmi29M7
pmi29M6
pmi29D6
pmi29D5
pmi127M4
pmi127M3
pmi127D5
pmi127D1
pmi125M3
pmi125D1
pmi125D10
pmi114D5
pmi109M1
pmi109M10
pmi109D5
pmi109D2
pmi105M11
pmi105D5
pmi105D10
pmi2.5
p1.4
m2.11
Fp.9

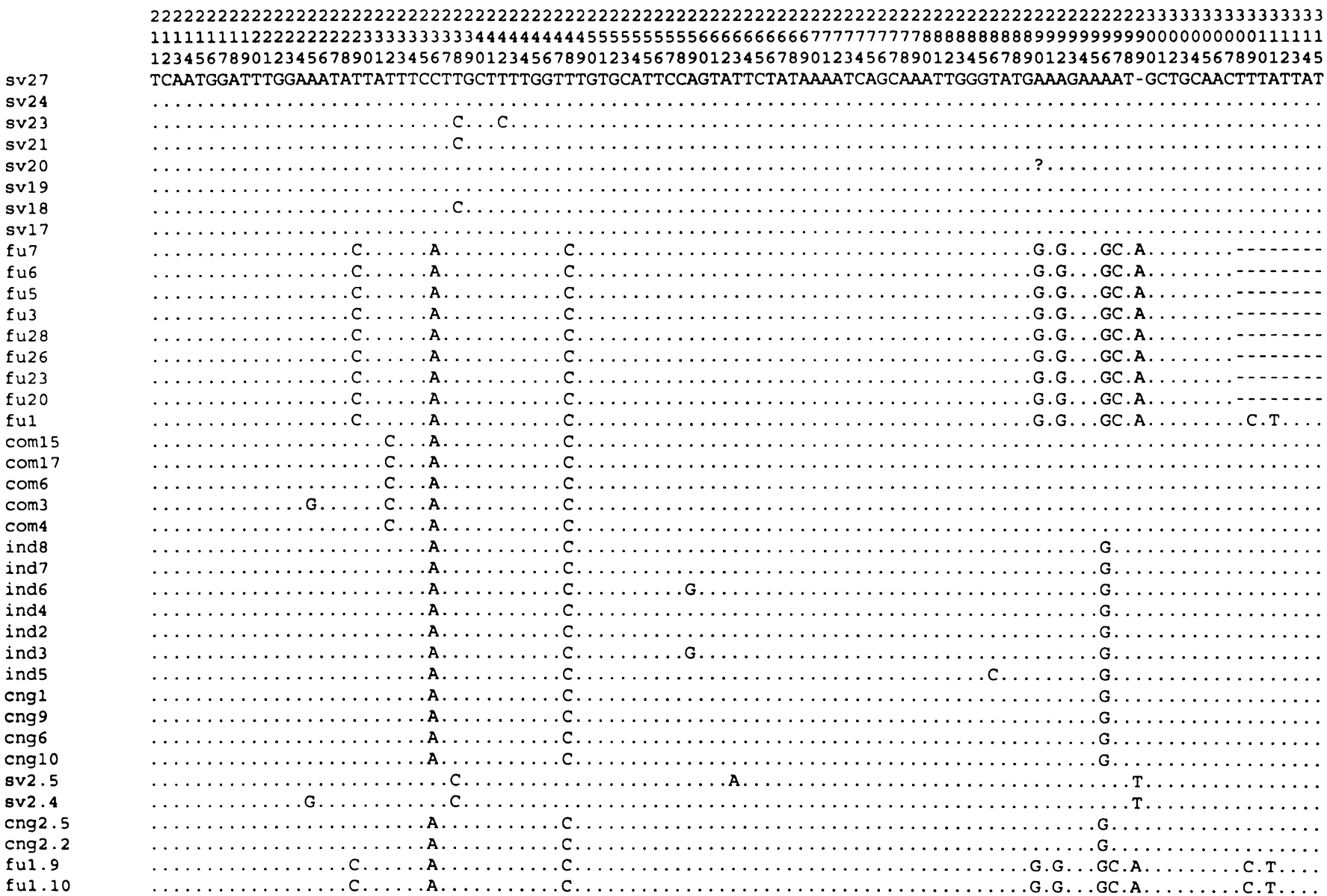

111111
1111111111222222222233333333334444444444555555555566666666667777777777888888888899999999990000000
12345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345
Fp.12 .....C.....G...A..C.....---.....T.....C.....A..G...A.....
Fa.2 .....C.....G...A..C.....---.....T.....C.....A..G...A.....
zwi1.7 .....C.....G..T..T..A.....T.....C.....A..G...A.....T....
zwi1.6 .....C.....G..T..T.....T.....C.....A..G...A.....
zsb4.7 .....C.....G.....T.....T.....C.....A..G...A.....T....
zny3.4 .....T.....G.....
zmn2.6 .....C.....G.....T.....TG.....C.....A..G...A.....
zco1.3 .....C.....G.....T.....T.....C.....A..G...A.....
zco1.2 .....C.....G.....T.....T.....C.....A..G...A.....
str2.6 .....G.....
str2.4 .....G.....
str1.4 .....C.....T.....A....A..G...A.....A.....
nwmx1.2 .....C.....G.....T.....T.....C.....G.A..G...A.....
nwmx1.11 .....C.....G.....T.....T.....C.....A..G...A.....
m6.6 .....C.....G...A.T.....---.....T.....C.....A..G...A.....
m6.2 .....C.....G...A.T.....T.....C.....A..G...A.....
m5.2 .....C.....G...A.T.....---.....T.....C.....A..G...A.....
m5.10 .....C.....G...A.T.....T.....C.....A..G...A.....
m4.9 .....C.....G...A.T.....---.....T.....C.....A..G...A.....
m4.6 .....C.....G...A.T.....---.....T.....C.....A..G...A.....
jbt1.4 .....C.....A.T.....---...G.....A..G...A.....
jbt1.1 .....C.....C...A.T.....---...G.....A..G...A.....
zco2.3 .....C.....G.....T.....T.....C.....A..G...A.....
zco2.6 .....C.....G.....T.....T.....C.....A..G...A.....
zmt1.6 .....C.....G.....T.....C.....T..TT.....T.....C.....A..G...A.....
zmt2.4 .....C.....G.....T.....C.....TT.....T.....C.....A..G...A.....
zmtB2.8 .....C.....G.....T.....T.....C.....A..G...A.....
zon2.6 .....C.....G.....T.....T..T.....T.....C.....A..G...A.....T....

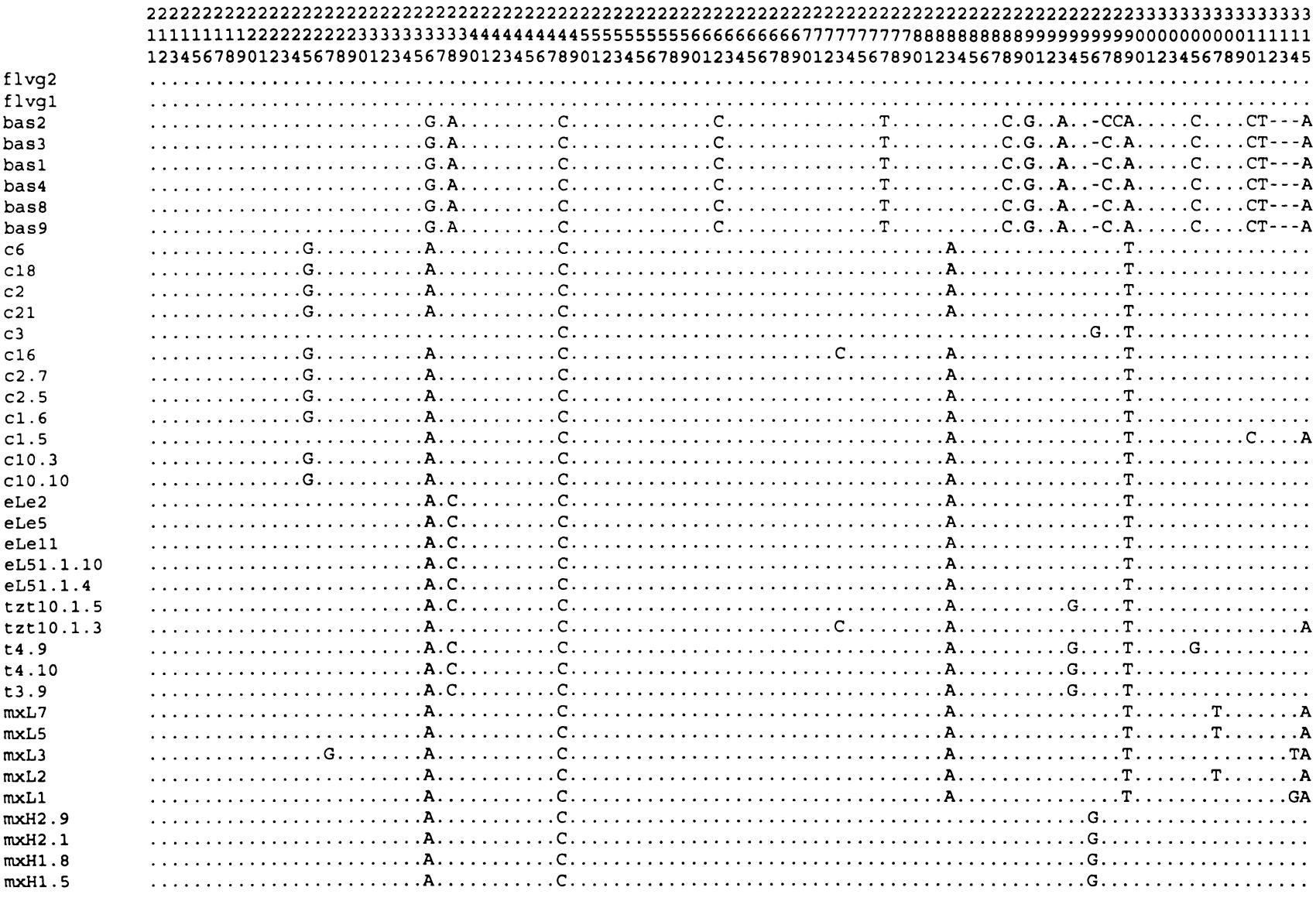

1111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111122222222222
0000111111111112222222222333333333344444444445555555555666666666677777777778888888888999999999900000000001
6789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
sv27 CAAAATGTCTGATAAGCCTGTTGAGAAAGTTGGTCGCCCCATGAAGTACCCGTATACTTTCAGCGCAAAAATCGCTCAATTTCCGCTAAAACACTACATTCAGAA
sv24 .....T.....
sv23 .....
sv21 .....
sv20 .....
sv19 .....
sv18 .....
sv17 .....G.....
fu7 ..T.....T..A.....C.....G.....
fu6 ..T.....T..A.....
fu5 ..T.....T..A.....T.....
fu3 ..T.....T..A.....
fu28 ..T.....T..A.....
fu26 ..T.....A.....
fu23 ..T.....T..A.....
fu20 ..T.....T..A.....
fu1 ..T.....T..A.....
com15 .....
com17 .....
com6 .....
com3 .....
com4 .....
ind8 .....G.....
ind7 .....G.....
ind6 .....G.....
ind4 .....G.....
ind2 .....A.....G.....
ind3 .....G.....
ind5 .....T.....G.....
cng1 .....G.....
cng9 .....G.....
cng6 .....G.....
cng10 .....G.....
sv2.5 .....G.....
sv2.4 .....
cng2.5 .....T.....G.....
cng2.2 .....G.....
fu1.9 ..T.....T..A.....T.....
fu1.10 ..T.....T..A.....

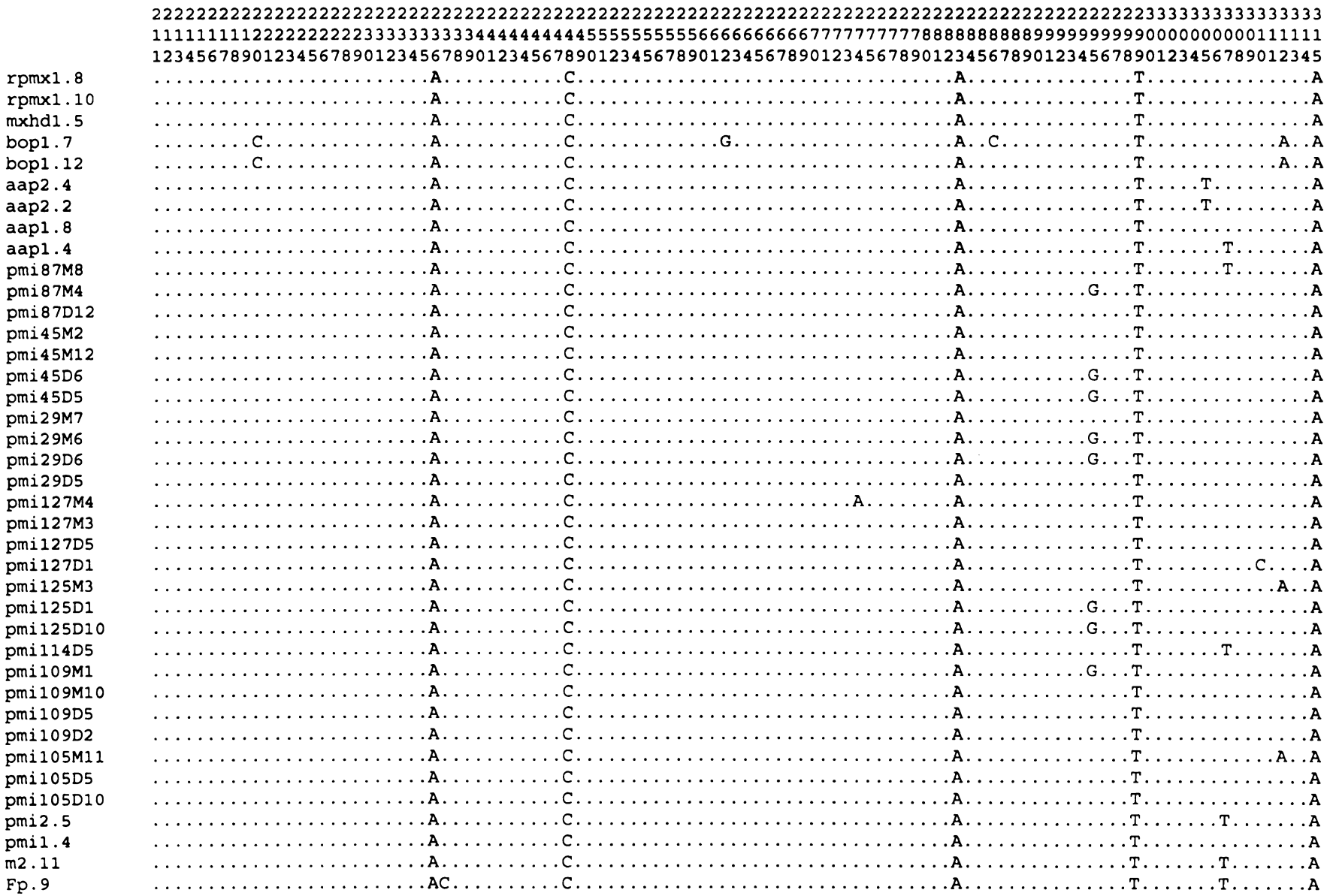

1111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111122222222222
0000111111111122222222223333333333444444444455555555556666666666777777777788888888899999999990000000001
6789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
flvg2
flvg1
bas2 .....G.....TT..A.....T..G.....
bas3 .....G.....TT..A.....T..G.....
bas1 .....?.....G.....TT..A.....T..G.....C.....
bas4 .....G.....TT..A.....T..G.....
bas8 .....G.....TT..A.....T..G.....
bas9 .....G.....TT..A.....T..G.....
c6
c18
c2 .....G.....
c21
c3 .....G.....
c16
c2.7
c2.5
c1.6 .....C.....
c1.5 .....G.....
c10.3
c10.10
eLe2
eLe5
eLe11
eL51.1.10 .C.....T.....G.....
eL51.1.4 .C.....T.....
tzt10.1.5 .....G.....
tzt10.1.3
t4.9 .....G.....
t4.10 .....G.....
t3.9 .....G.....
mxL7
mxL5
mxL3 .....A.....A.....G..G.....
mxL2
mxL1 .....A.....G.....
mxH2.9 .....G.....
mxH2.1 .....C.....A.....G.....G.....
mxH1.8 .....G.....
mxH1.5 .....G.....

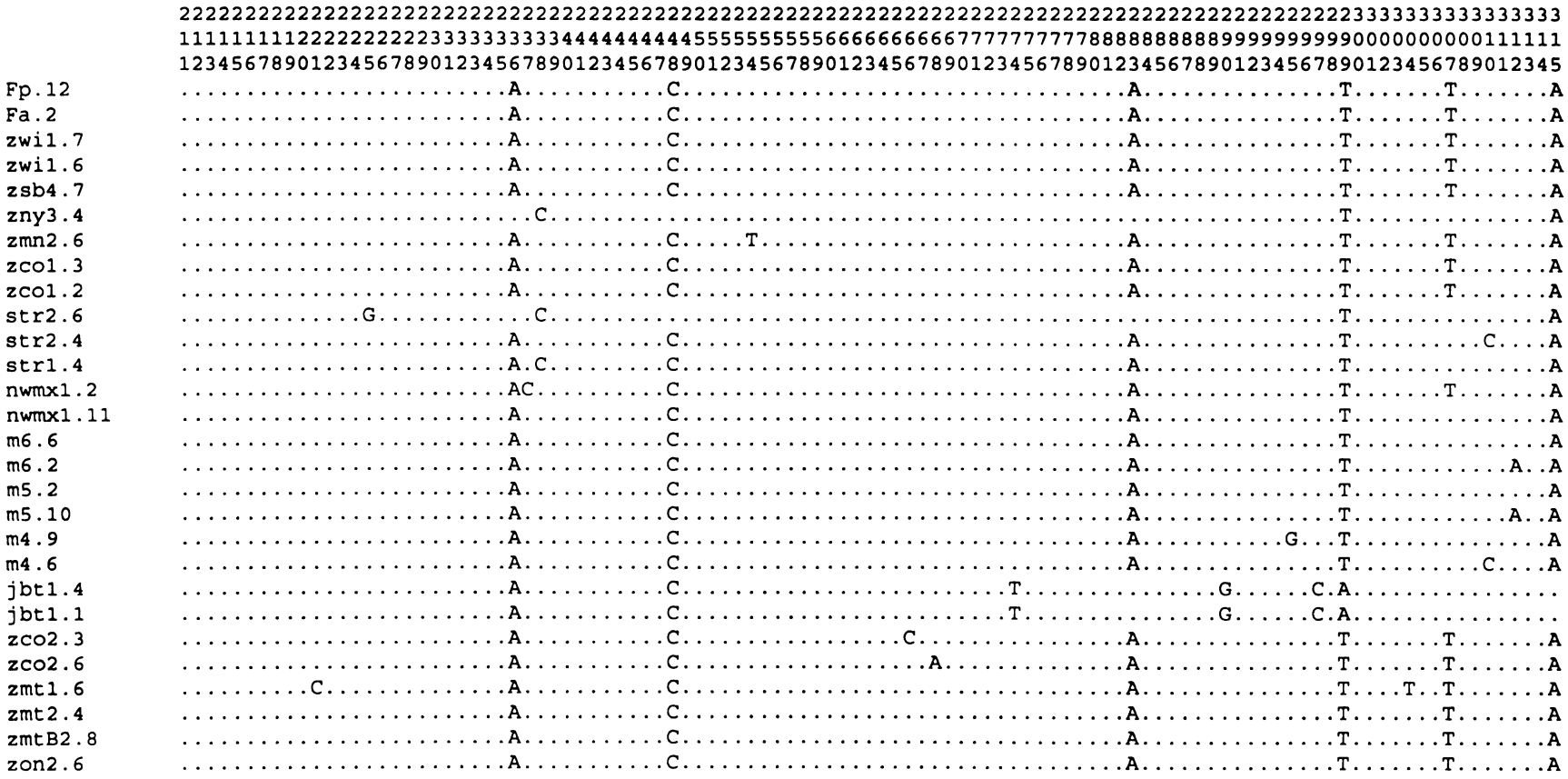

1111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111122222222222
0000111111111122222222223333333333444444444455555555555666666666677777777778888888888999999999900000000001
6789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
rpmx1.8 ..........G..G.....
rpmx1.10 ..........G..G.....
mxhd1.5 ....
bop1.7 ....
bop1.12 ....
aap2.4 ....G.....
aap2.2 ....
aap1.8 ....
aap1.4 ....
pmi87M8 ....
pmi87M4 ....A.....
pmi87D12 ....
pmi45M2 ....G.....
pmi45M12 ....
pmi45D6 ....A.....
pmi45D5 ....
pmi29M7 ....G.....
pmi29M6 ....A.....
pmi29D6 ....A.....
pmi29D5 ....
pmi127M4 ....T.....
pmi127M3 ....
pmi127D5 ....
pmi127D1 ....
pmi125M3 ....G.....
pmi125D1 ....A.....
pmi125D10 ....A.....
pmi114D5 ....
pmi109M1 ....A.....
pmi109M10 ....
pmi109D5 ....
pmi109D2 ....
pmi105M11 ....G.....
pmi105D5 ....
pmi105D10 ....A.....
pmi2.5 ....
pmi1.4 ....
m2.11 ....A.....
Fp.9 ....C.....

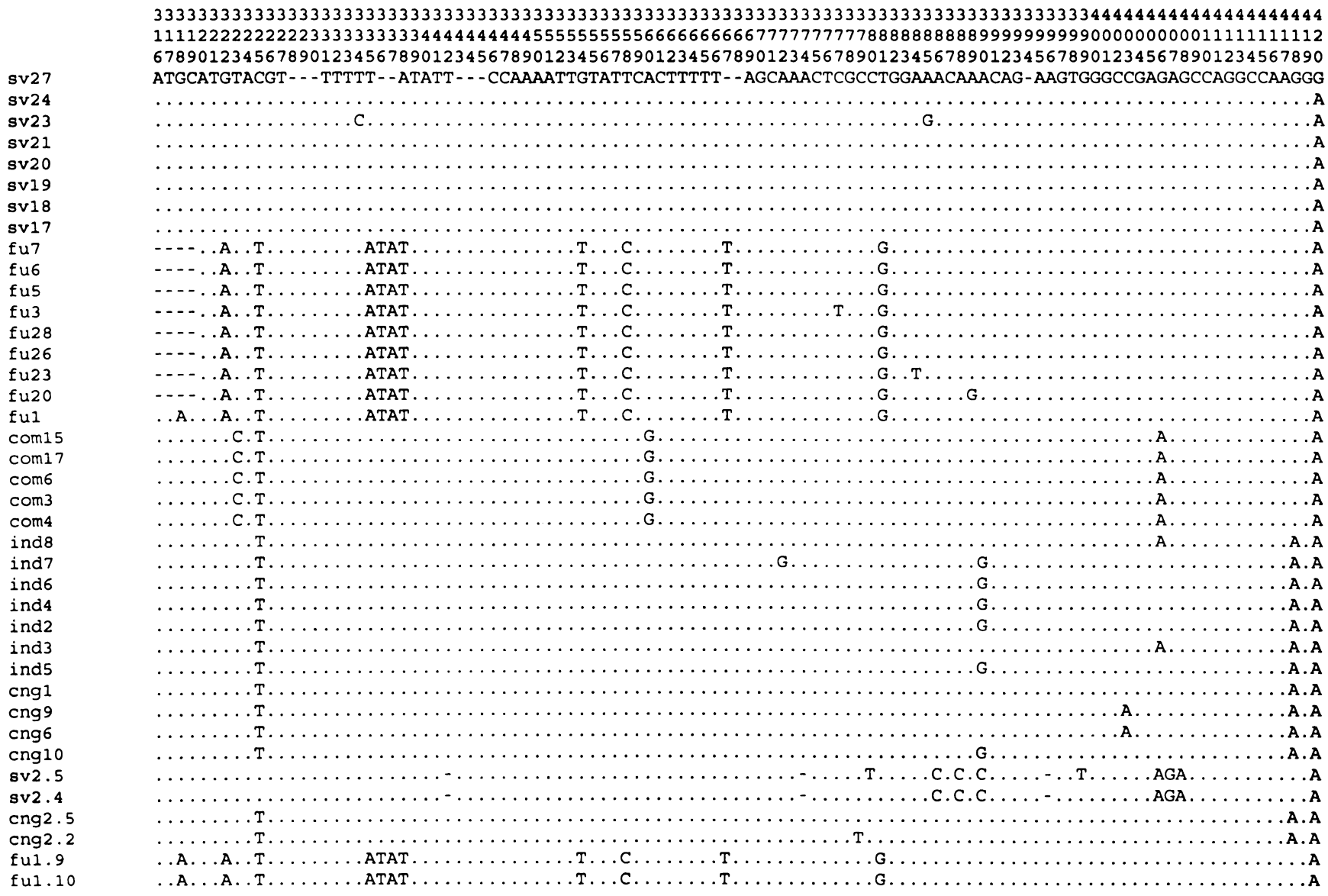

1111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111122222222222
0000111111111122222222223333333333344444444445555555555566666666667777777777888888888899999999990000000001
6789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
Fp.12 ....A.....
Fa.2 ....A.....
zwi1.7 .....G.....A.....
zwi1.6 .....C.....A.....
zsb4.7 ....A.....
zny3.4 ....
zmn2.6 ....
zco1.3 ....A.....
zco1.2 ....A.....
str2.6 .....A..........
str2.4 .....C.....A..........
str1.4 ....
nwmx1.2 .....C..........
nwmx1.11 .....A..........
m6.6 .....C.....T..........
m6.2 ....
m5.2 .....C.....T..........
m5.10 .....T..........
m4.9 .....A.....G.....
m4.6 .....C..........
jbt1.4 ....
jbt1.1 ....
zco2.3 ....A.....
zco2.6 ....A.....
zmt1.6 ....G.....
zmt2.4 ....A.....
zmtB2.8 ....A.....
zon2.6 ....A.....

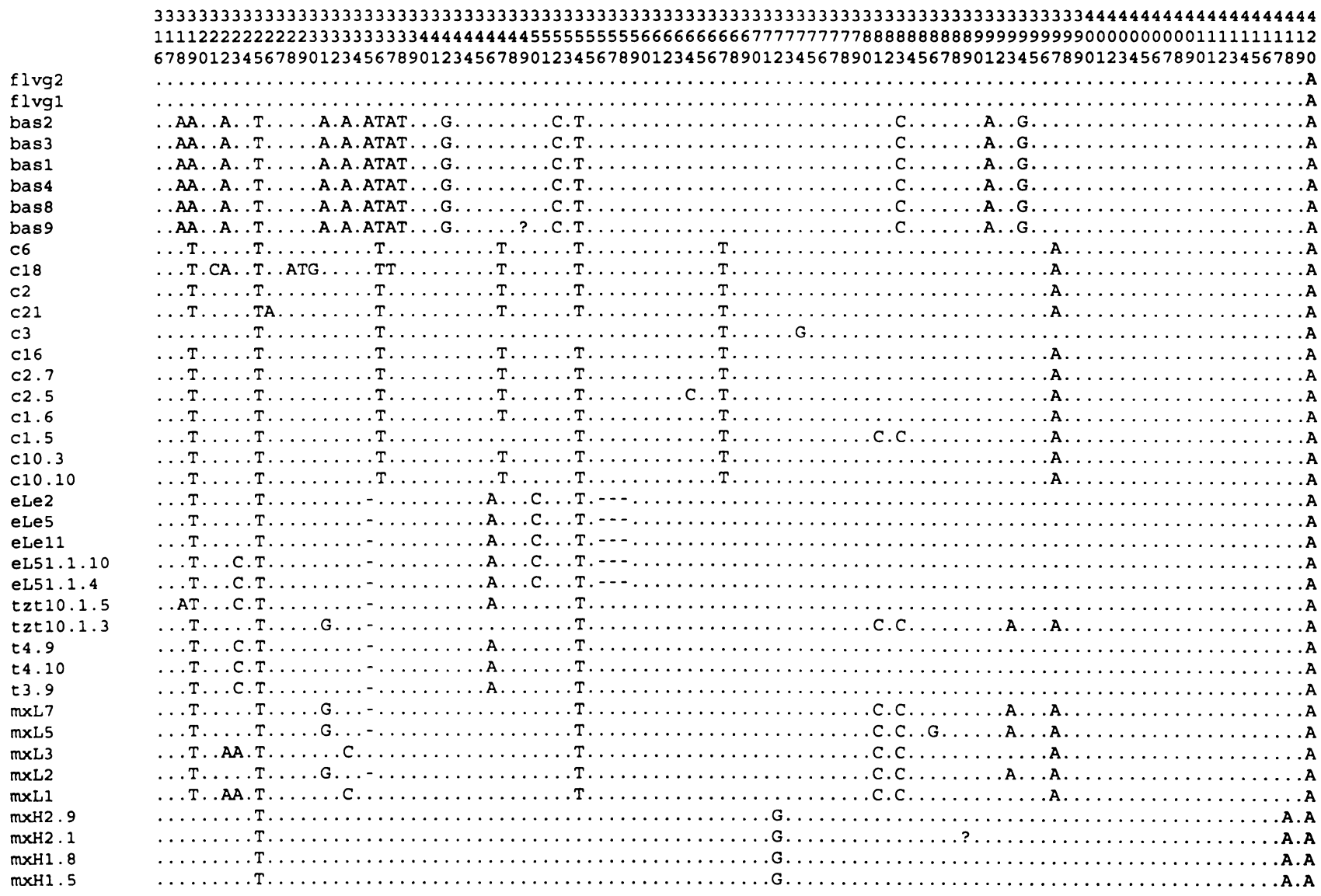

sv27 TCAATGGATTTGGAAATATTATTTCCTTGCTTTTGGTTTGTGCATTCCAGTATTCTATAAAATCAGCAAATTGGGTATGAAAGAAAAT-GCTGCAACTTTATTAT
sv24
sv23
sv21
sv20
sv19
sv18
sv17
fu7
fu6
fu5
fu3
fu28
fu26
fu23
fu20
fu1
com15
com17
com6
com3
com4
ind8
ind7
ind6
ind4
ind2
ind3
ind5
cng1
cng9
cng6
cng10
sv2.5
sv2.4
cng2.5
cng2.2
fu1.9
fu1.10

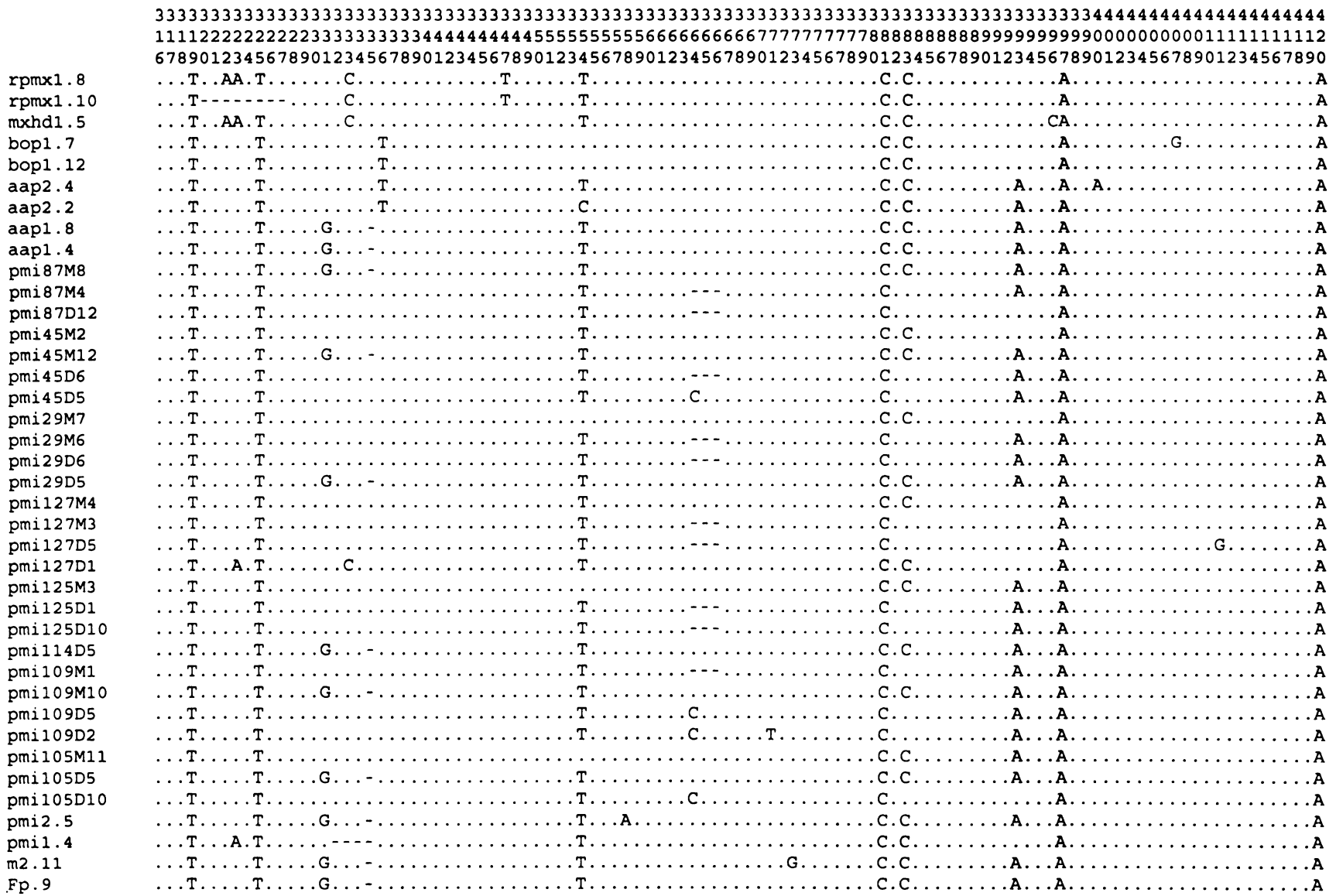

2222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222333333333333333333
1111111112222222222333333333344444444445555555555666666666677777777778888888888999999999900000000001111111
1234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345
flvg2
flvg1
bas2 ..........G.A.....C.....C.....T.....C.G..A..-CCA.....C....CT---A
bas3 ..........G.A.....C.....C.....T.....C.G..A..-C.A.....C....CT---A
bas1 ..........G.A.....C.....C.....T.....C.G..A..-C.A.....C....CT---A
bas4 ..........G.A.....C.....C.....T.....C.G..A..-C.A.....C....CT---A
bas8 ..........G.A.....C.....C.....T.....C.G..A..-C.A.....C....CT---A
bas9 ..........G.A.....C.....C.....T.....C.G..A..-C.A.....C....CT---A
c6 ..........G.....A.....C.....A.....T.....
c18 ..........G.....A.....C.....A.....T.....
c2 ..........G.....A.....C.....A.....T.....
c21 ..........G.....A.....C.....A.....T.....
c3 .....C.....G..T.....
c16 ..........G.....A.....C.....C.....A.....T.....
c2.7 ..........G.....A.....C.....A.....T.....
c2.5 ..........G.....A.....C.....A.....T.....
c1.6 ..........G.....A.....C.....A.....T.....
c1.5 ..........A.....C.....A.....T.....C....A
c10.3 ..........G.....A.....C.....A.....T.....
c10.10 ..........G.....A.....C.....A.....T.....
eLe2 ..........A.C.....C.....A.....T.....
eLe5 ..........A.C.....C.....A.....T.....
eLe11 ..........A.C.....C.....A.....T.....
eL51.1.10 ..........A.C.....C.....A.....T.....
eL51.1.4 ..........A.C.....C.....A.....T.....
tzt10.1.5 ..........A.C.....C.....A.....G....T.....
tzt10.1.3 ..........A.....C.....C.....A.....T.....A
t4.9 ..........A.C.....C.....A.....G....T....G.....
t4.10 ..........A.C.....C.....A.....G....T.....
t3.9 ..........A.C.....C.....A.....G....T.....
mxL7 ..........A.....C.....A.....T.....T.....A
mxL5 ..........A.....C.....A.....T.....T.....A
mxL3 ..........G.....A.....C.....A.....T.....TA
mxL2 ..........A.....C.....A.....T.....T.....A
mxL1 ..........A.....C.....A.....T.....GA
mxH2.9 ..........A.....C.....G.....
mxH2.1 ..........A.....C.....G.....
mxH1.8 ..........A.....C.....G.....
mxH1.5 ..........A.....C.....G.....

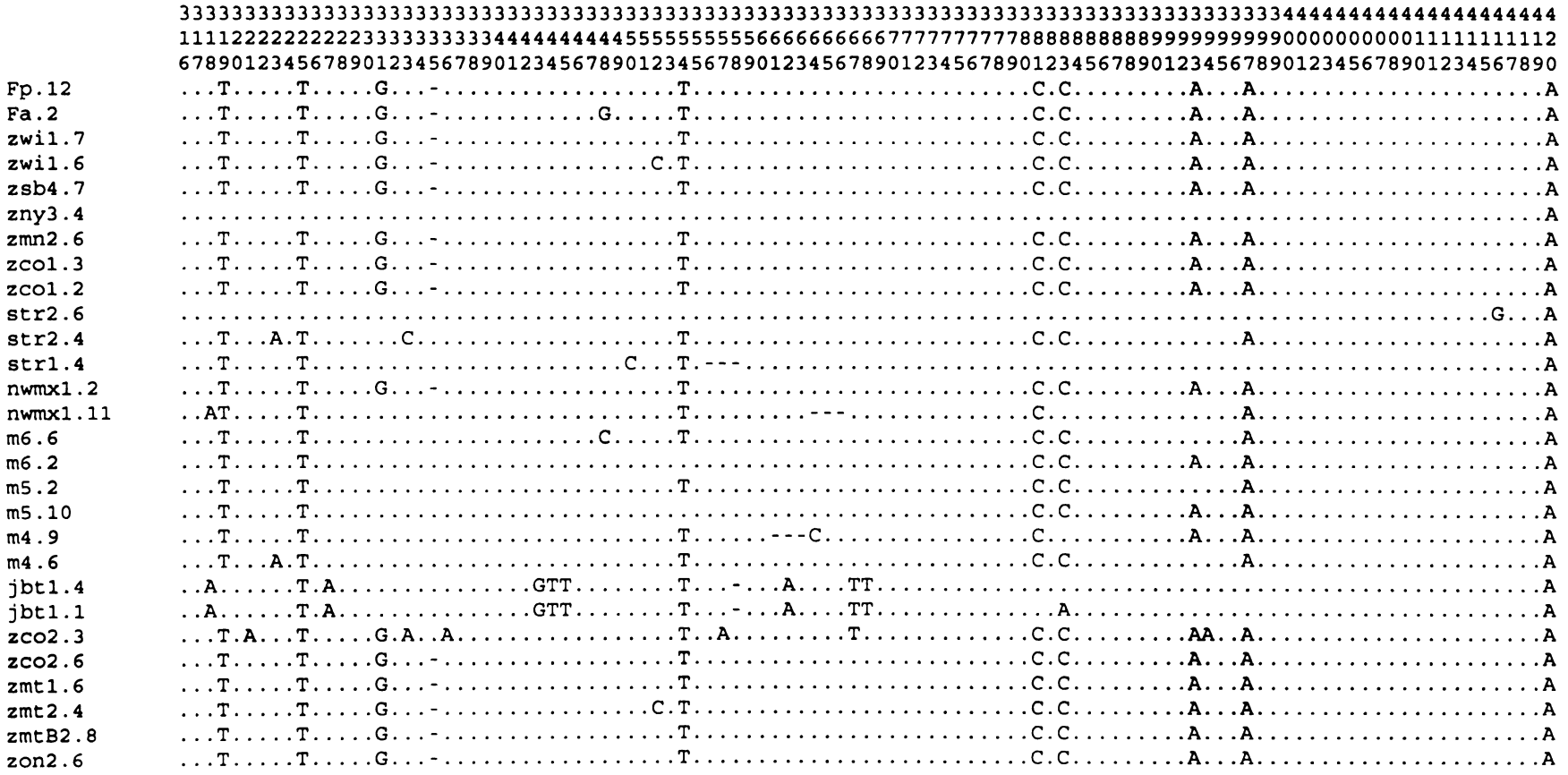

2222222222222222222222222222222222222222222222222222222222222222222222222222222222222233333333333333333
1111111112222222222333333333334444444444555555555566666666667777777777888888888899999999990000000000111111
1234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345
rpmx1.8 .....A.....C.....A.....T.....A
rpmx1.10 .....A.....C.....A.....T.....A
mxhd1.5 .....A.....C.....A.....T.....A
bop1.7 .....C.....A.....C.....G.....A..C.....T.....A..A
bop1.12 .....C.....A.....C.....A.....T.....A..A
aap2.4 .....A.....C.....A.....T.....T.....A
aap2.2 .....A.....C.....A.....T.....T.....A
aap1.8 .....A.....C.....A.....T.....A
aap1.4 .....A.....C.....A.....T.....T.....A
pmi87M8 .....A.....C.....A.....T.....T.....A
pmi87M4 .....A.....C.....A.....G...T.....A
pmi87D12 .....A.....C.....A.....T.....A
pmi45M2 .....A.....C.....A.....T.....A
pmi45M12 .....A.....C.....A.....T.....A
pmi45D6 .....A.....C.....A.....G...T.....A
pmi45D5 .....A.....C.....A.....G...T.....A
pmi29M7 .....A.....C.....A.....T.....A
pmi29M6 .....A.....C.....A.....G...T.....A
pmi29D6 .....A.....C.....A.....G...T.....A
pmi29D5 .....A.....C.....A.....T.....A
pmi127M4 .....A.....C.....A.....A.....T.....A
pmi127M3 .....A.....C.....A.....T.....A
pmi127D5 .....A.....C.....A.....T.....A
pmi127D1 .....A.....C.....A.....T.....C....A
pmi125M3 .....A.....C.....A.....T.....A..A
pmi125D1 .....A.....C.....A.....G...T.....A
pmi125D10 .....A.....C.....A.....G...T.....A
pmi114D5 .....A.....C.....A.....T.....T.....A
pmi109M1 .....A.....C.....A.....G...T.....A
pmi109M10 .....A.....C.....A.....T.....A
pmi109D5 .....A.....C.....A.....T.....A
pmi109D2 .....A.....C.....A.....T.....A
pmi105M11 .....A.....C.....A.....T.....A..A
pmi105D5 .....A.....C.....A.....T.....A
pmi105D10 .....A.....C.....A.....T.....A
pmi2.5 .....A.....C.....A.....T.....T.....A
pmi1.4 .....A.....C.....A.....T.....A
m2.11 .....A.....C.....A.....T.....T.....A
Fp.9 .....AC.....C.....A.....T.....T.....A

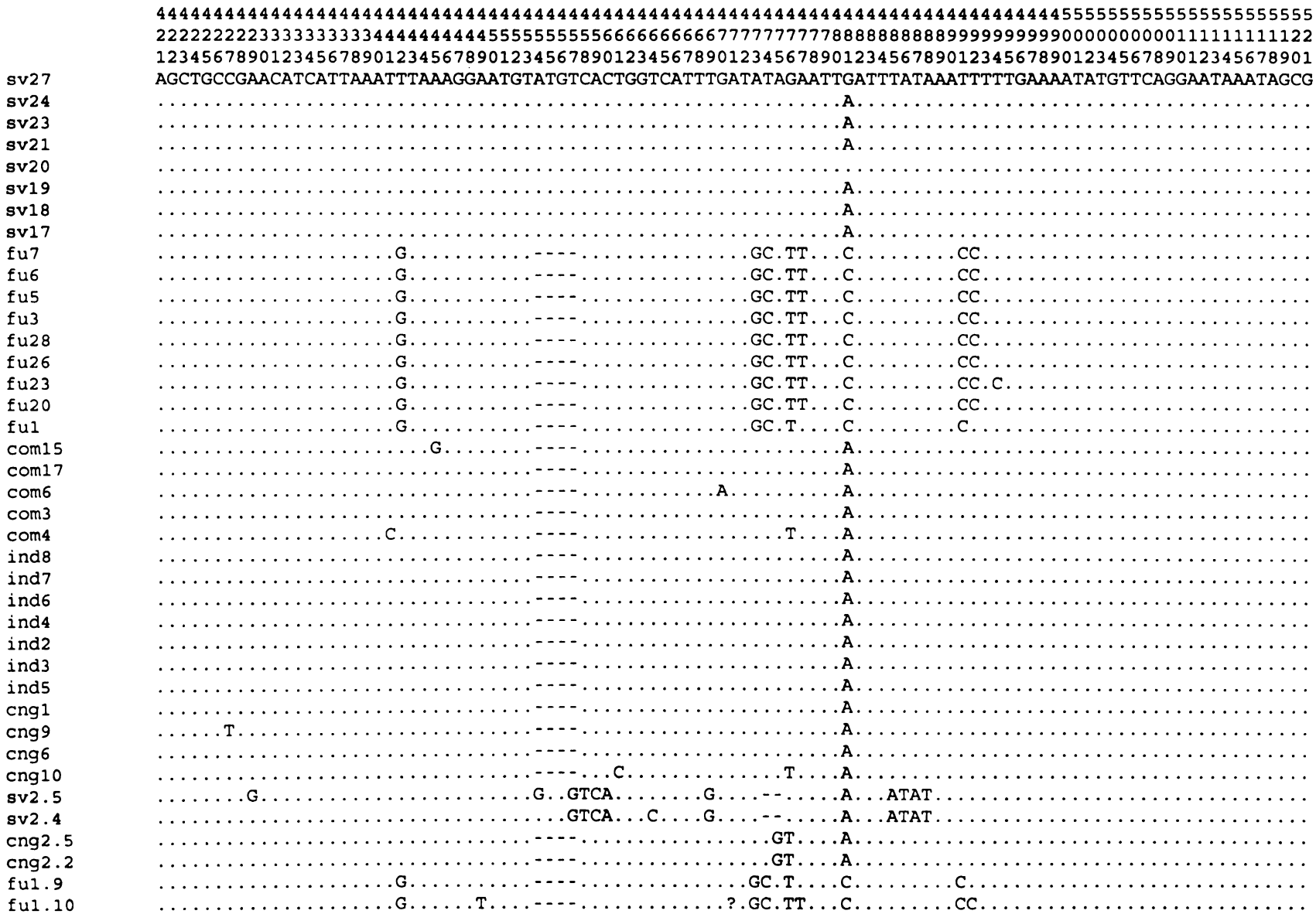

2222222222222222222222222222222222222222222222222222222222222222222222222222222222222233333333333333333
1111111112222222222333333333334444444444555555555566666666667777777777888888888899999999990000000000111111
1234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345
Fp.12 .....A.....C.....A.....T.....T.....A
Fa.2 .....A.....C.....A.....T.....T.....A
zwi1.7 .....A.....C.....A.....T.....T.....A
zwi1.6 .....A.....C.....A.....T.....T.....A
zsb4.7 .....A.....C.....A.....T.....T.....A
zny3.4 .....C.....T.....A
zmn2.6 .....A.....C.....T.....A.....T.....T.....A
zco1.3 .....A.....C.....A.....T.....T.....A
zco1.2 .....A.....C.....A.....T.....T.....A
str2.6 .....G.....C.....T.....A
str2.4 .....A.....C.....A.....T.....C....A
str1.4 .....A.C.....C.....A.....T.....A
nwmx1.2 .....AC.....C.....A.....T.....T.....A
nwmx1.11 .....A.....C.....A.....T.....A
m6.6 .....A.....C.....A.....T.....A
m6.2 .....A.....C.....A.....T.....A..A
m5.2 .....A.....C.....A.....T.....A
m5.10 .....A.....C.....A.....T.....A..A
m4.9 .....A.....C.....A.....G...T.....A
m4.6 .....A.....C.....A.....T.....C....A
jbt1.4 .....A.....C.....T.....G.....C.A.....
jbt1.1 .....A.....C.....T.....G.....C.A.....
zco2.3 .....A.....C.....C.....A.....T.....T.....A
zco2.6 .....A.....C.....A.....A.....T.....T.....A
zmt1.6 .....C.....A.....C.....A.....T...T..T.....A
zmt2.4 .....A.....C.....A.....T.....T.....A
zmtB2.8 .....A.....C.....A.....T.....T.....A
zon2.6 .....A.....C.....A.....T.....T.....A

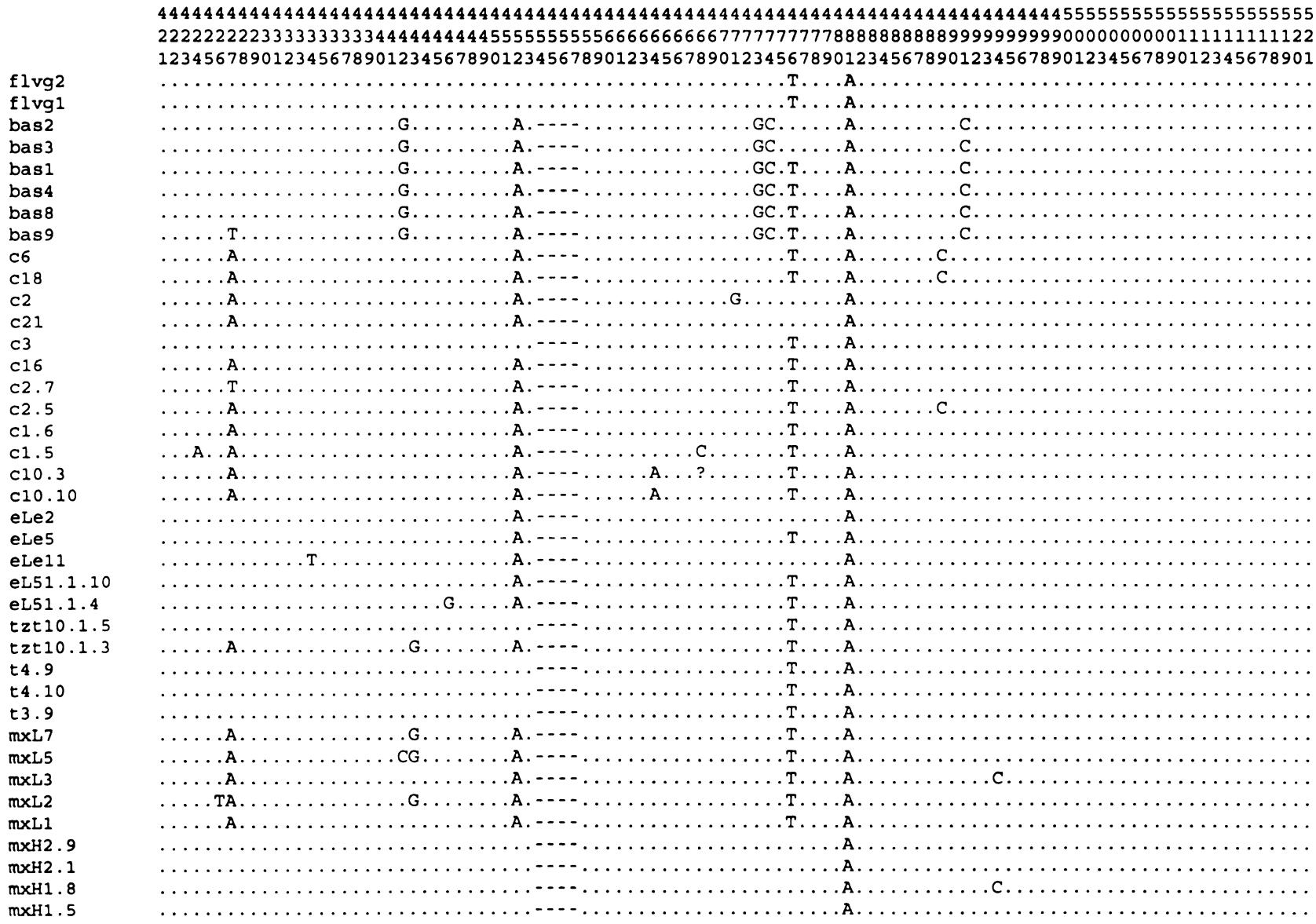

sv27 ATGCATGTACGT---TTTTT--ATATT---CCAAAATTGTATTCACTTTTT--AGCAAACTCGCCTGGAAACAAACAG-AAGTGGGCCGAGAGCCAGGCCAAGGG
sv24
sv23
sv21
sv20
sv19
sv18
sv17
fu7
fu6
fu5
fu3
fu28
fu26
fu23
fu20
fu1
com15
com17
com6
com3
com4
ind8
ind7
ind6
ind4
ind2
ind3
ind5
cng1
cng9
cng6
cng10
sv2.5
sv2.4
cng2.5
cng2.2
fu1.9
fu1.10

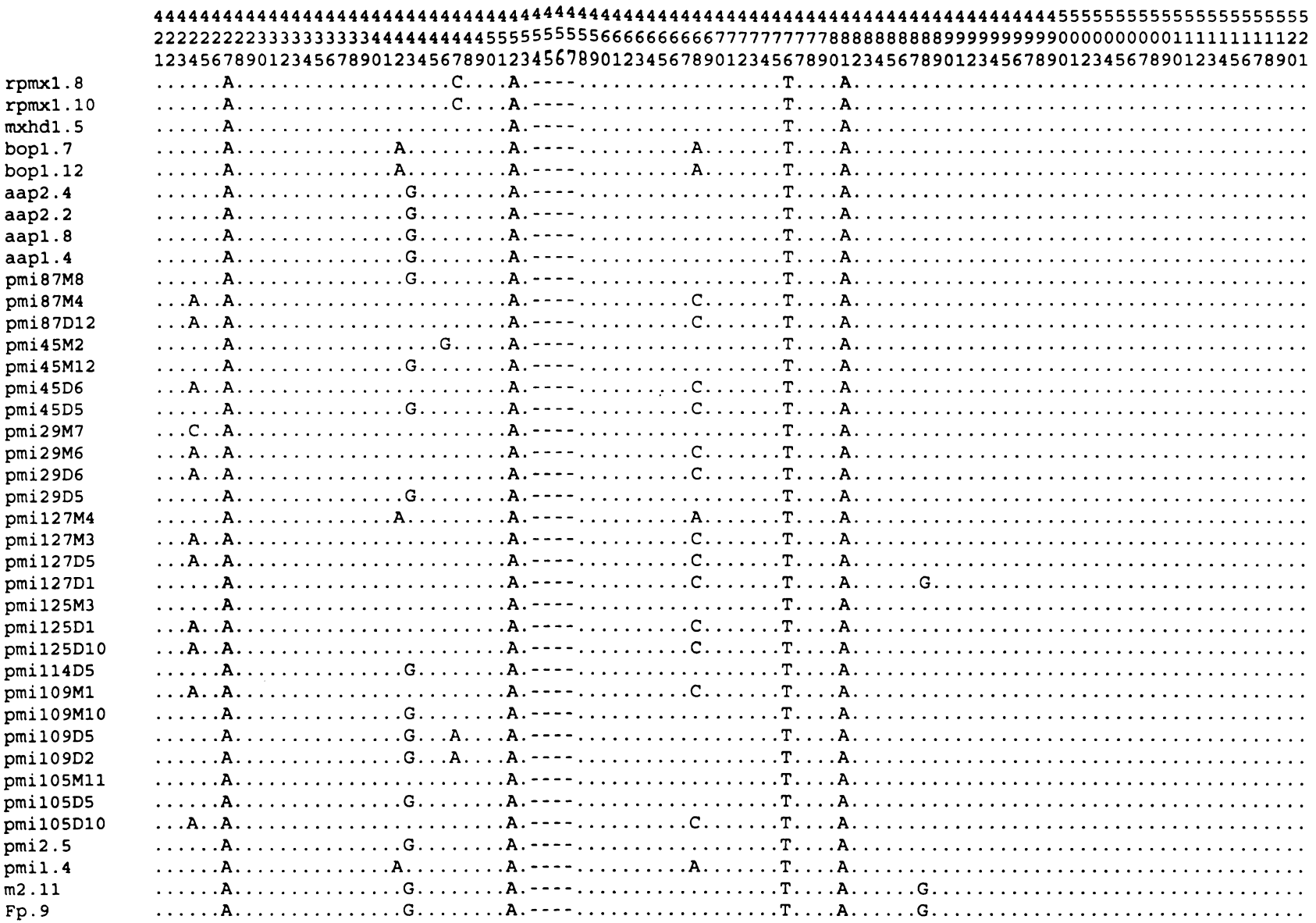

flvg2
flvg1
bas2
bas3
bas1
bas4
bas8
bas9
c6
c18
c2
c21
c3
c16
c2.7
c2.5
c1.6
c1.5
c10.3
c10.10
eLe2
eLe5
eLe11
eL51.1.10
eL51.1.4
tzt10.1.5
tzt10.1.3
t4.9
t4.10
t3.9
mxL7
mxL5
mxL3
mxL2
mxL1
mxH2.9
mxH2.1
mxH1.8
mxH1.5

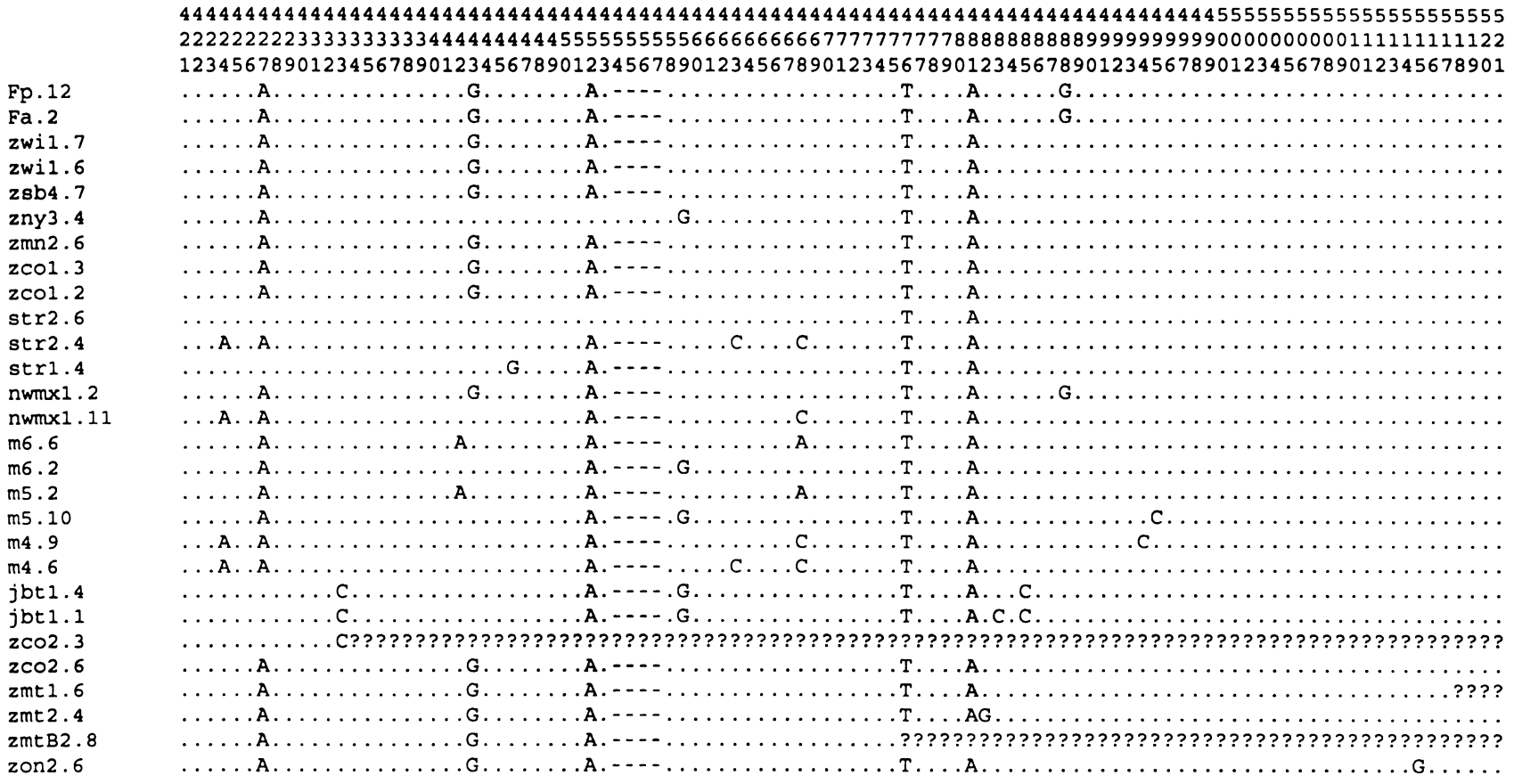

rpmx1.8
rpmx1.10
mxhd1.5
bop1.7
bop1.12
aap2.4
aap2.2
aap1.8
aap1.4
pmi87M8
pmi87M4
pmi87D12
pmi45M2
pmi45M12
pmi45D6
pmi45D5
pmi29M7
pmi29M6
pmi29D6
pmi29D5
pmi127M4
pmi127M3
pmi127D5
pmi127D1
pmi125M3
pmi125D1
pmi125D10
pmi114D5
pmi109M1
pmi109M10
pmi109D5
pmi109D2
pmi105M11
pmi105D5
pmi105D10
pmi2.5
pmi1.4
m2.11
Fp.9


Fp.12
Fa.2
zwi1.7
zwi1.6
zsb4.7
zny3.4
zmn2.6
zco1.3
zco1.2
str2.6
str2.4
str1.4
nwmx1.2
nwmx1.11
m6.6
m6.2
m5.2
m5.10
m4.9
m4.6
jbt1.4
jbt1.1
zco2.3
zco2.6
zmt1.6
zmt2.4
zmtB2.8
zon2.6


sv27 AGCTGCCGAACATCATTAAATTTAAAGGAATGTATGTCACTGGTCATTTGATATAGAATTGATTTATAAATTTTTGAAAATATGTTCAGGAATAAATAGCG
sv24
sv23
sv21
sv20
sv19
sv18
sv17
fu7
fu6
fu5
fu3
fu28
fu26
fu23
fu20
ful
com15
com17
com6
com3
com4
ind8
ind7
ind6
ind4
ind2
ind3
ind5
cng1
cng9
cng6
cng10
sv2.5
sv2.4
cng2.5
cng2.2
ful.9
ful.10

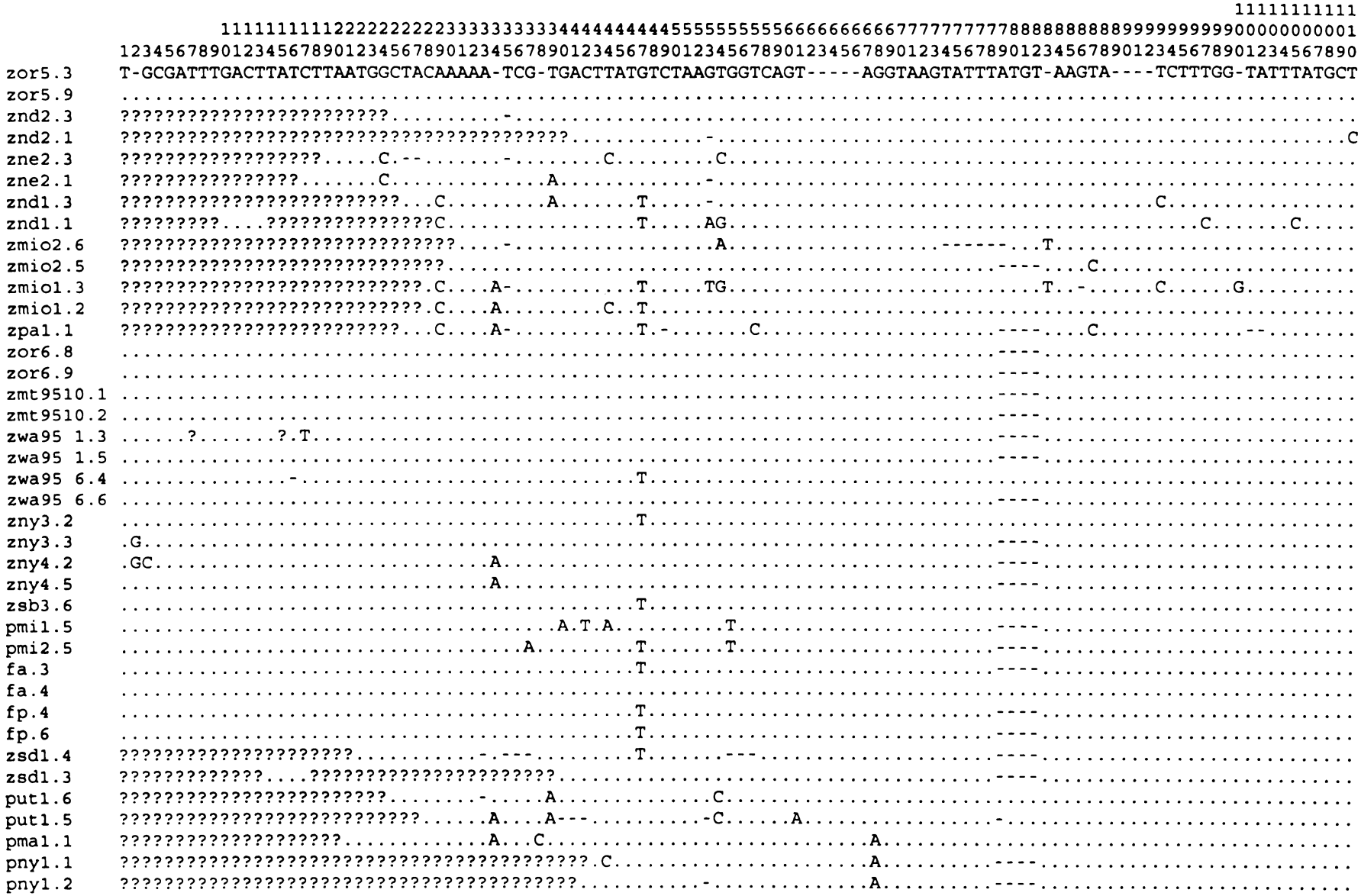

flvg2
flvg1
bas2
bas3
bas1
bas4
bas8
bas9
c6
c18
c2
c21
c3
c16
c2.7
c2.5
c1.6
c1.5
c10.3
c10.10
eLe2
eLe5
eLe11
eL51.1.10
eL51.1.4
tzt10.1.5
tzt10.1.3
t4.9
t4.10
t3.9
mxL7
mxL5
mxL3
mxL2
mxL1
mxH2.9
mxH2.1
mxH1.8
mxH1.5

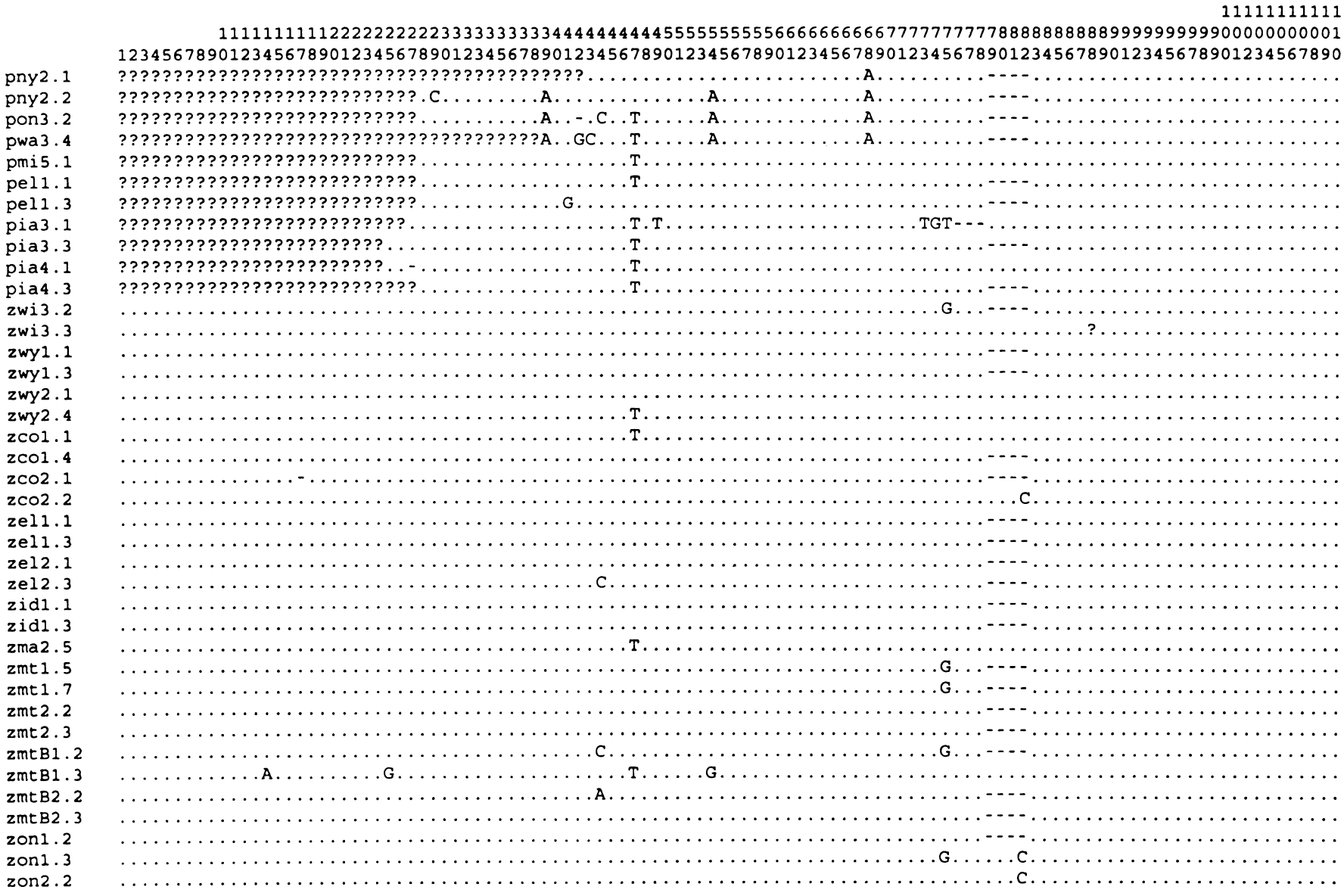

rpmx1.8
rpmx1.10
mxhd1.5
bop1.7
bop1.12
aap2.4
aap2.2
aap1.8
aap1.4
pmi87M8
pmi87M4
pmi87D12
pmi45M2
pmi45M12
pmi45D6
pmi45D5
pmi29M7
pmi29M6
pmi29D6
pmi29D5
pmi127M4
pmi127M3
pmi127D5
pmi127D1
pmi125M3
pmi125D1
pmi125D10
pmi114D5
pmi109M1
pmi109M10
pmi109D5
pmi109D2
pmi105M11
pmi105D5
pmi105D10
pmi2.5
pmi1.4
m2.11
Fp.9

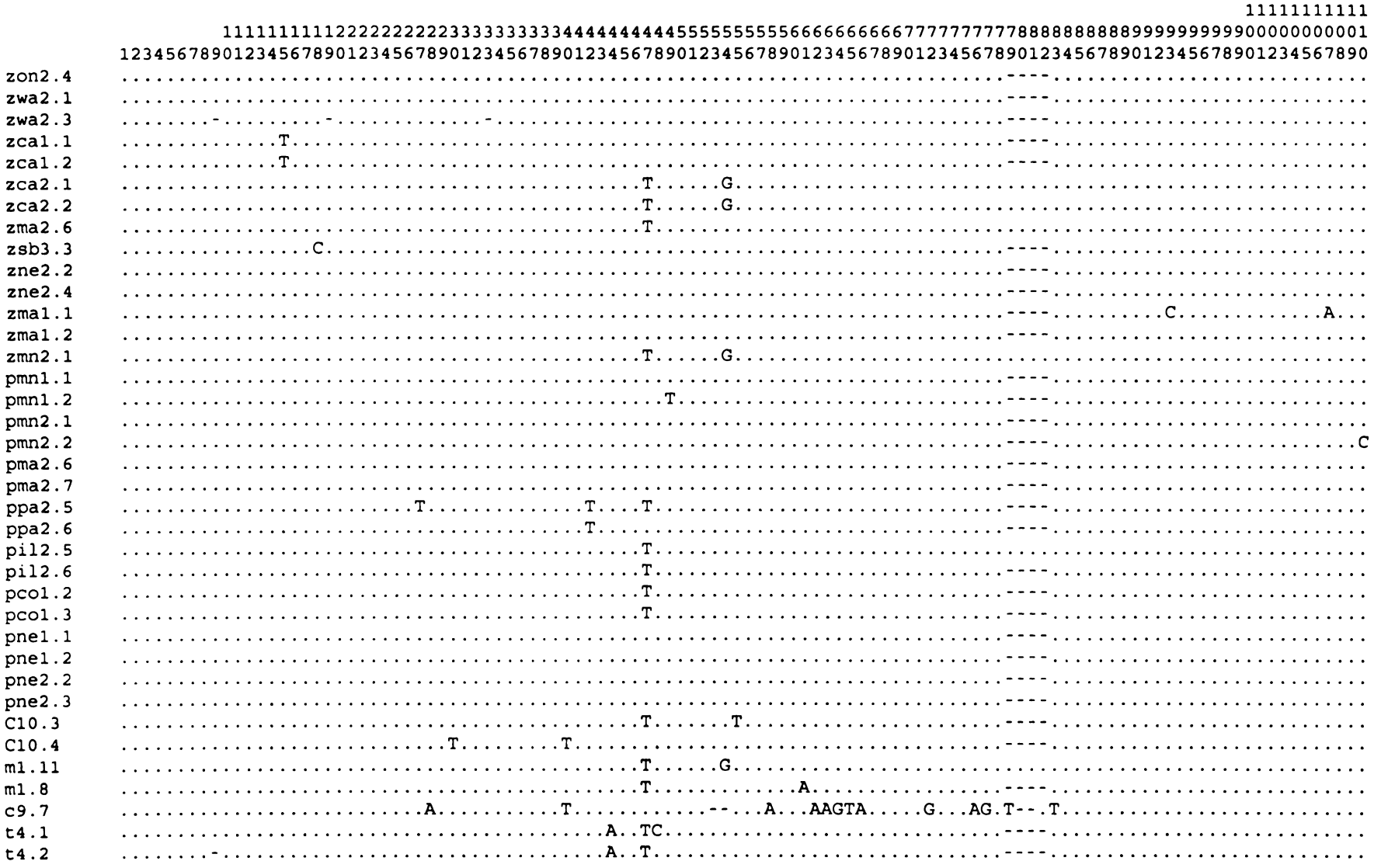

444444444444444444444444444444444444444444444444444444444444444444444444444455555555555555555555
222222222333333333344444444445555555555666666666677777777778888888888999999999900000000001111111112
12345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901
Fp.12 ......A.............G.....A.----.................T....A.....G.....
Fa.2 ......A.............G.....A.----.................T....A.....G.....
zwi1.7 ......A.............G.....A.----.................T....A.....
zwi1.6 ......A.............G.....A.----.................T....A.....
zsb4.7 ......A.............G.....A.----.................T....A.....
zny3.4 ......A.........................G.................T....A.....
zmn2.6 ......A.............G.....A.----.................T....A.....
zco1.3 ......A.............G.....A.----.................T....A.....
zco1.2 ......A.............G.....A.----.................T....A.....
str2.6 ....T....A.....
str2.4 ...A..A.....................A.----.....C....C.....T....A.....
str1.4 ......................G.....A.----.................T....A.....
nwmx1.2 ......A.............G.....A.----.................T....A.....G.....
nwmx1.11 ...A..A.....................A.----.........C.....T....A.....
m6.6 ......A...........A.....A.----.........A.....T....A.....
m6.2 ......A.....................A.-----G.............T....A.....
m5.2 ......A...........A.....A.----.........A.....T....A.....
m5.10 ......A.....................A.-----G.............T....A.....C.....
m4.9 ...A..A.....................A.----.........C.....T....A.....C.....
m4.6 ...A..A.....................A.----.....C....C.....T....A.....
jbt1.4 ...........C.............A.-----G.............T....A...C.....
jbt1.1 ...........C.............A.-----G.............T....A.C.C.....
zco2.3 ....C??
zco2.6 ......A.............G.....A.----.................T....A.....
zmt1.6 ......A.............G.....A.----.................T....A.....????
zmt2.4 ......A.............G.....A.----.................T....AG.....
zmtB2.8 ......A.............G.....A.----.....??
zon2.6 ......A.............G.....A.----.................T....A.....G.....

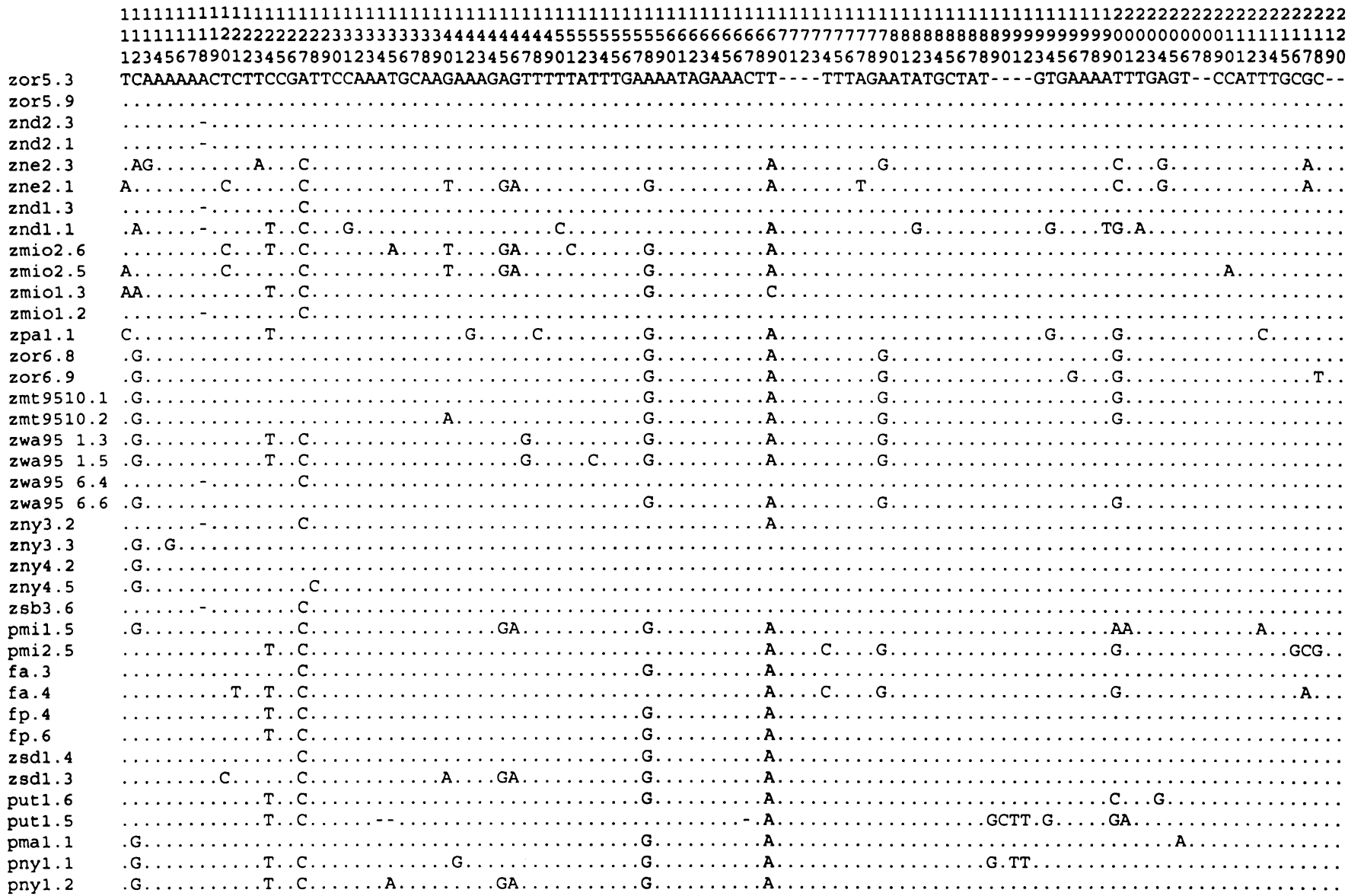
APPENDIX C

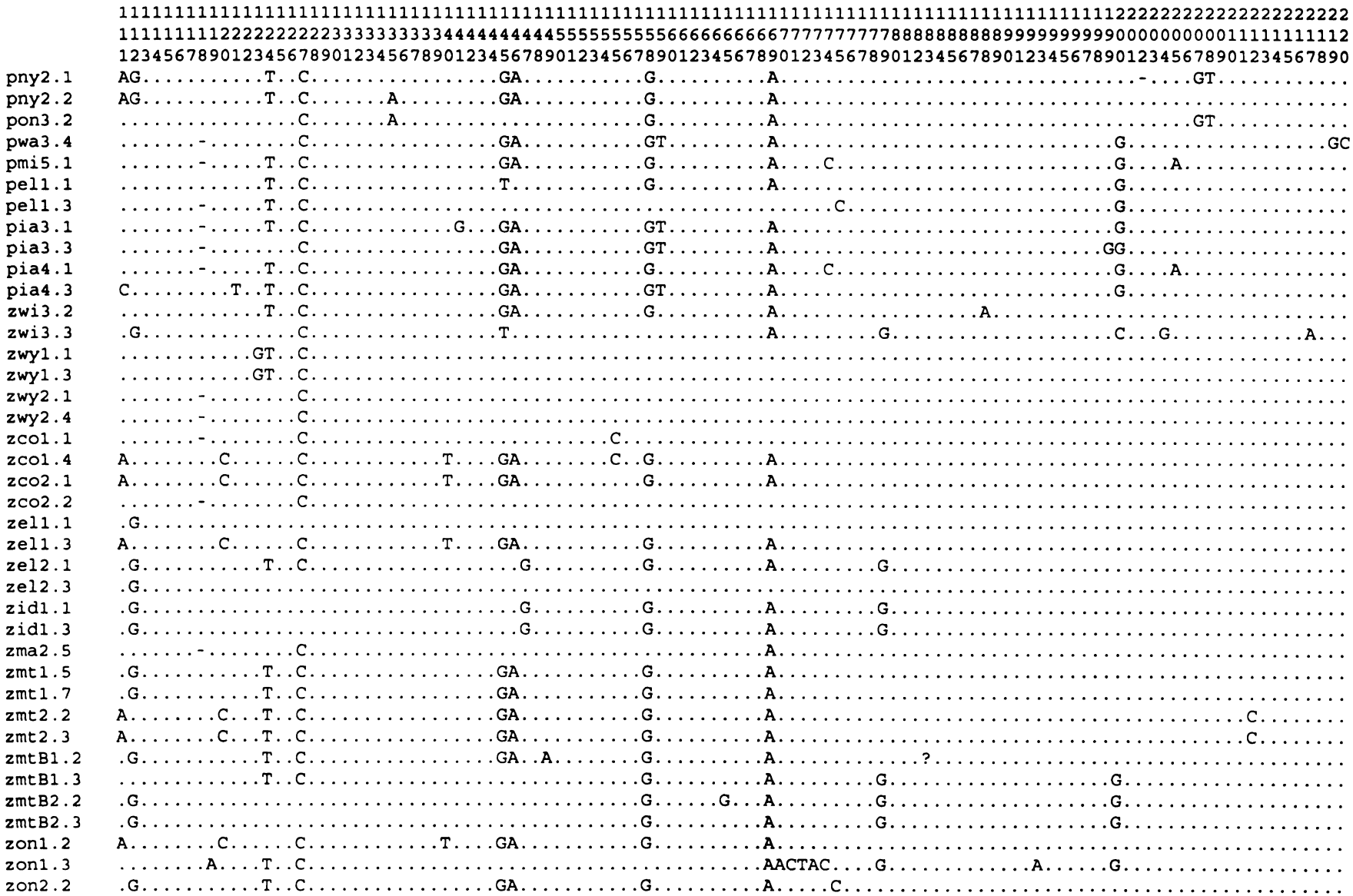
P2480 sequence alignment


zor5.3 T-GCGATTTGACTTATCTTAATGGCTACAAAAA-TCG-TGACTTATGTCTAAGTGGTCAGT-----AGGTAAGTATTTATGT-AAGTA----TCTTTGG-TATTTATGCT
zor5.9
znd2.3
znd2.1
zne2.3
zne2.1
znd1.3
znd1.1
zmio2.6
zmio2.5
zmio1.3
zmio1.2
zpa1.1
zor6.8
zor6.9
zmt9510.1
zmt9510.2
zwa95 1.3
zwa95 1.5
zwa95 6.4
zwa95 6.6
zny3.2
zny3.3
zny4.2
zny4.5
zsb3.6
pmi1.5
pmi2.5
fa.3
fa.4
fp.4
fp.6
zsd1.4
zsd1.3
put1.6
put1.5
pma1.1
pny1.1
pny1.2

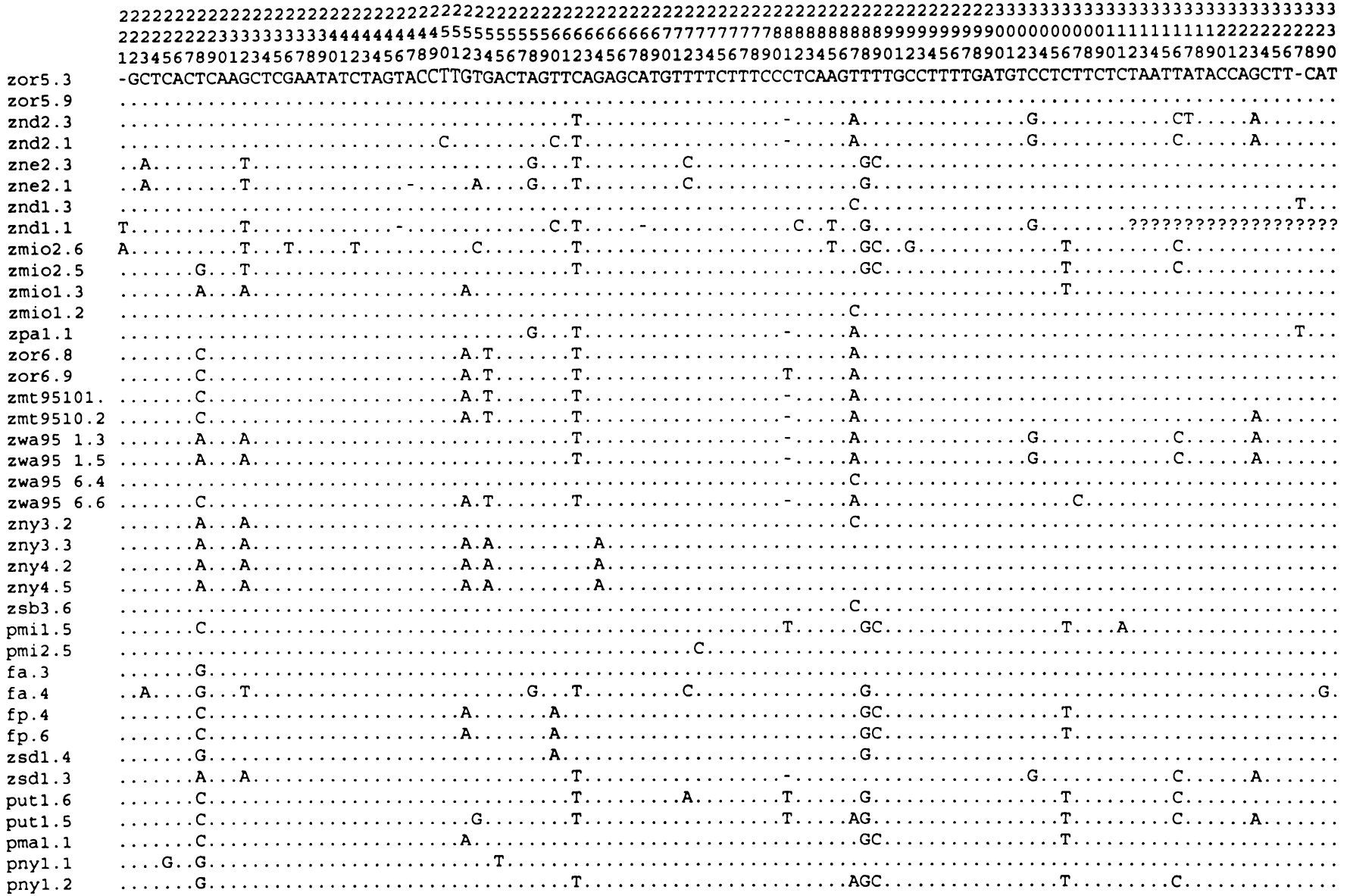

11111111111
1111111111222222222233333333334444444444555555555566666666667777777777888888888899999999990000000001
12345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
pny2.1 ??.....A.....----.....
pny2.2 ??????????????????????????.C.....A.....A.....A.....----.....
pon3.2 ??????????????????????????.....A..-.C..T.....A.....A.....----.....
pwa3.4 ??????????????????????????????????????A..GC...T.....A.....A.....----.....
pmi5.1 ??????????????????????????.....T.....
pel1.1 ??????????????????????????.....T.....----.....
pel1.3 ??????????????????????????.....G.....----.....
pia3.1 ??????????????????????????.....T.T.....TGT---.....
pia3.3 ??????????????????????????.....T.....----.....
pia4.1 ??????????????????????????..-.....T.....
pia4.3 ??????????????????????????.....T.....----.....
zwi3.2 .....G.....----.....
zwi3.3 .....?.....
zwy1.1 .....----.....
zwy1.3 .....----.....
zwy2.1 .....
zwy2.4 .....T.....
zco1.1 .....T.....
zco1.4 .....----.....
zco2.1 ..............-.....----.....
zco2.2 .....C.....
zel1.1 .....----.....
zel1.3 .....----.....
zel2.1 .....----.....
zel2.3 .....C.....----.....
zid1.1 .....----.....
zid1.3 .....----.....
zma2.5 .....T.....
zmt1.5 .....G.....----.....
zmt1.7 .....G.....----.....
zmt2.2 .....----.....
zmt2.3 .....----.....
zmtB1.2 .....C.....G.....----.....
zmtB1.3 .....A.....G.....T.....G.....
zmtB2.2 .....A.....----.....
zmtB2.3 .....----.....
zon1.2 .....----.....
zon1.3 .....G.....C.....
zon2.2 .....C.....

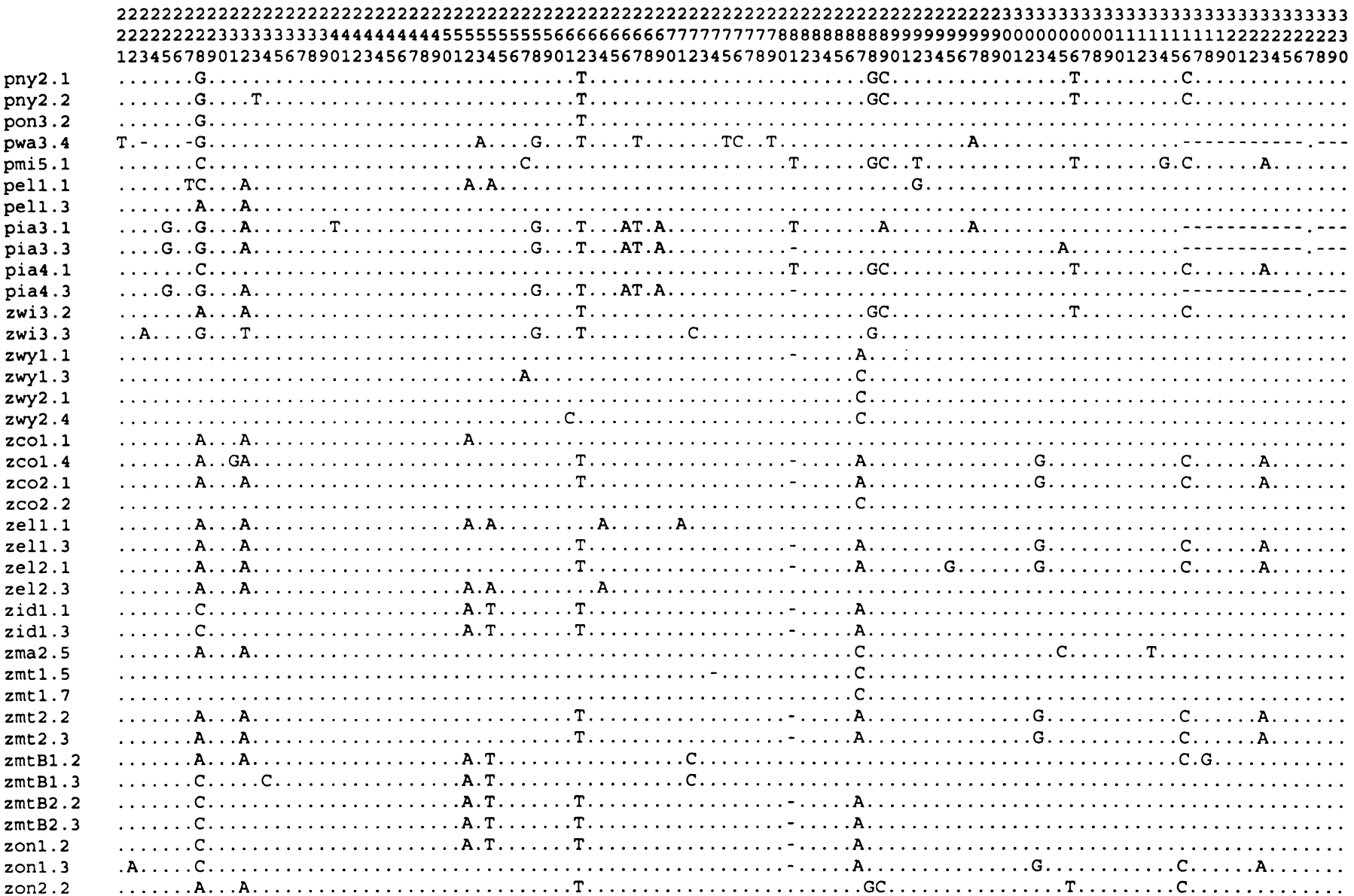

zon2.4
zwa2.1
zwa2.3
zca1.1
zca1.2
zca2.1
zca2.2
zma2.6
zsb3.3
zne2.2
zne2.4
zma1.1
zma1.2
zmn2.1
pmn1.1
pmn1.2
pmn2.1
pmn2.2
pma2.6
pma2.7
ppa2.5
ppa2.6
pil2.5
pil2.6
pco1.2
pco1.3
pne1.1
pne1.2
pne2.2
pne2.3
C10.3
C10.4
m1.11
m1.8
c9.7
t4.1
t4.2

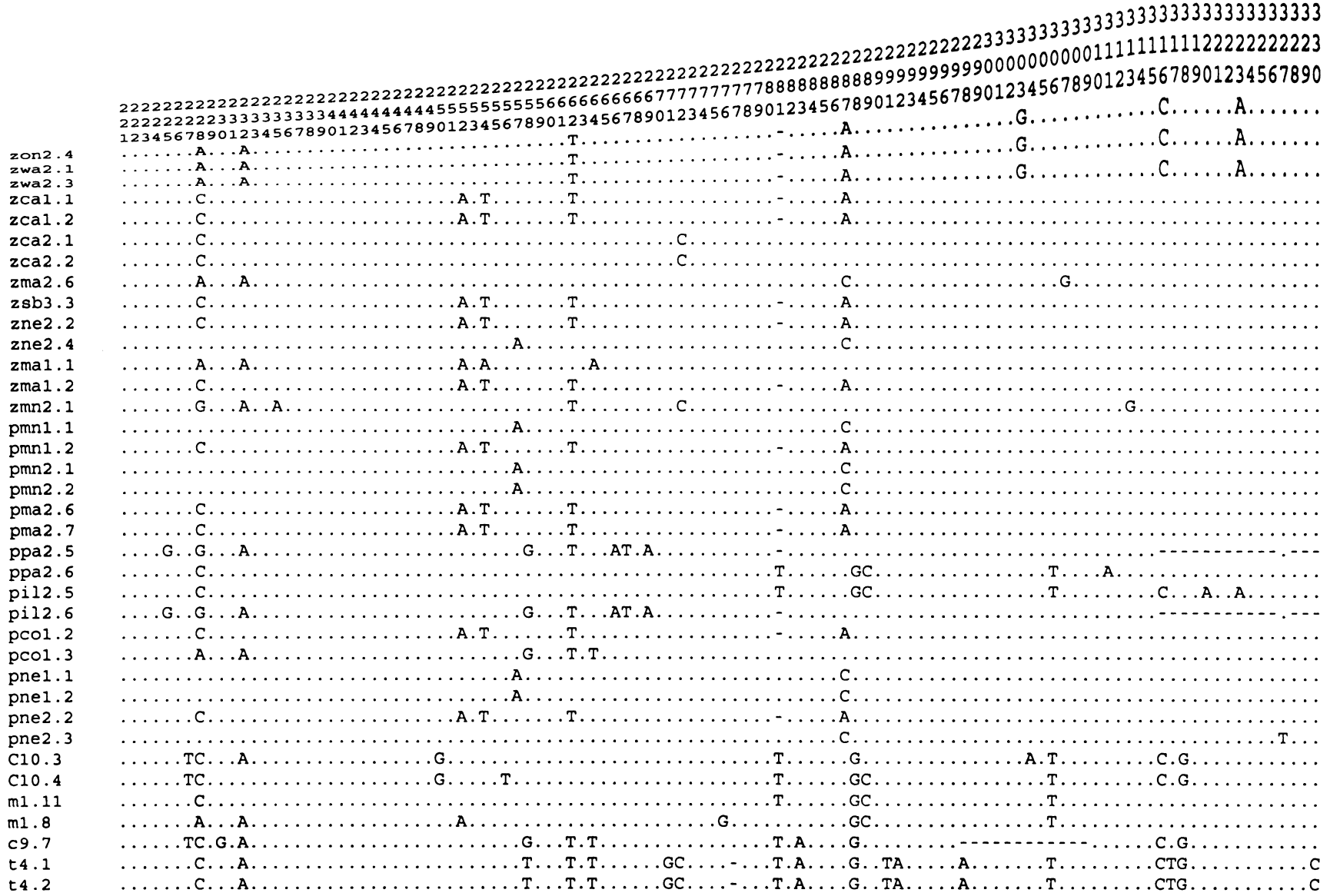

zor5.3 TCAAAAAACTCTTCCGATTCCAAATGCAAGAAAGAGTTTTTATTTGAAAATAGAAACTT----TTTAGAATATGCTAT----GTGAAAATTTGAGT--CCATTTGCGC--
zor5.9
znd2.3
znd2.1
zne2.3
zne2.1
znd1.3
znd1.1
zmio2.6
zmio2.5
zmio1.3
zmio1.2
zpa1.1
zor6.8
zor6.9
zmt9510.1
zmt9510.2
zwa95 1.3
zwa95 1.5
zwa95 6.4
zwa95 6.6
zny3.2
zny3.3
zny4.2
zny4.5
zsb3.6
pmi1.5
pmi2.5
fa.3
fa.4
fp.4
fp.6
zsd1.4
zsd1.3
put1.6
put1.5
pma1.1
pny1.1
pny1.2

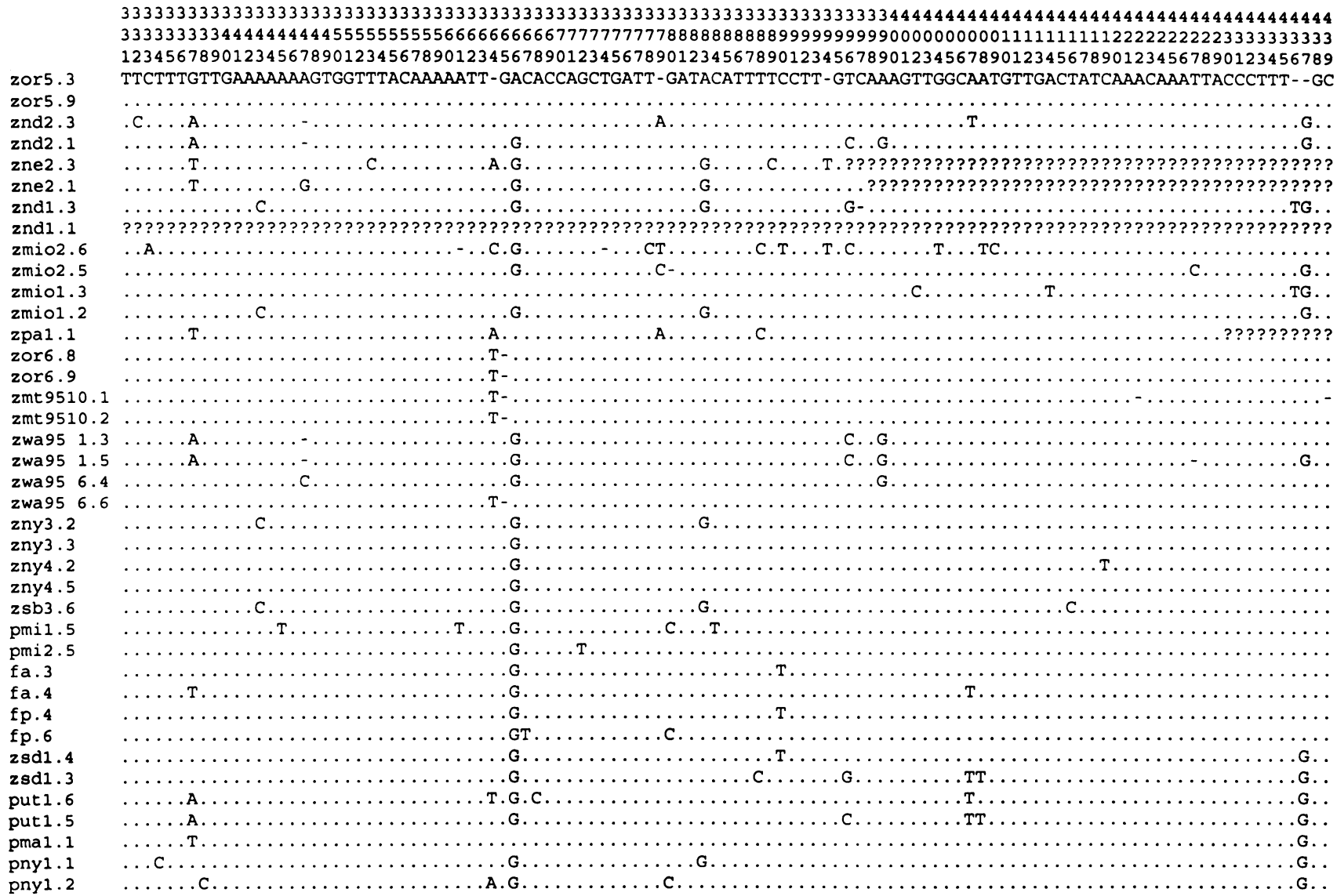

pny2.1 AG.....T..C.....GA.....G.....A.....-....GT.....
pny2.2 AG.....T..C.....A.....GA.....G.....A.....
pon3.2 .....C.....A.....G.....A.....GT.....
pwa3.4 .....-.....C.....GA.....GT.....A.....G.....GC
pmi5.1 .....-.....T..C.....GA.....G.....A....C.....G....A.....
pel1.1 .....T..C.....T.....G.....A.....G.....
pel1.3 .....-.....T..C.....C.....G.....
pia3.1 .....-.....T..C.....G...GA.....GT.....A.....G.....
pia3.3 .....-.....C.....GA.....GT.....A.....GG.....
pia4.1 .....-.....T..C.....GA.....G.....A....C.....G....A.....
pia4.3 C.....T..T..C.....GA.....GT.....A.....G.....
zwi3.2 .....T..C.....GA.....G.....A.....A.....
zwi3.3 .G.....C.....T.....A.....G.....C...G.....A...
zwy1.1 .....GT..C.....
zwy1.3 .....GT..C.....
zwy2.1 .....-.....C.....
zwy2.4 .....-.....C.....
zco1.1 .....-.....C.....C.....
zco1.4 A.....C.....C.....T....GA.....C..G.....A.....
zco2.1 A.....C.....C.....T....GA.....G.....A.....
zco2.2 .....-.....C.....
zel1.1 .G.....
zel1.3 A.....C.....C.....T....GA.....G.....A.....
zel2.1 .G.....T..C.....G.....G.....A.....G.....
zel2.3 .G.....
zid1.1 .G.....G.....G.....A.....G.....
zid1.3 .G.....G.....G.....A.....G.....
zma2.5 .....-.....C.....A.....
zmt1.5 .G.....T..C.....GA.....G.....A.....
zmt1.7 .G.....T..C.....GA.....G.....A.....
zmt2.2 A.....C...T..C.....GA.....G.....A.....C.....
zmt2.3 A.....C...T..C.....GA.....G.....A.....C.....
zmtB1.2 .G.....T..C.....GA..A.....G.....A.....?.....
zmtB1.3 .....T..C.....G.....A.....G.....G.....
zmtB2.2 .G.....G.....G...A.....G.....G.....
zmtB2.3 .G.....G.....A.....G.....G.....
zon1.2 A.....C.....C.....T....GA.....G.....A.....
zon1.3 .....A....T..C.....AACTAC....G.....A.....G.....
zon2.2 .G.....T..C.....GA.....G.....A.....C.....

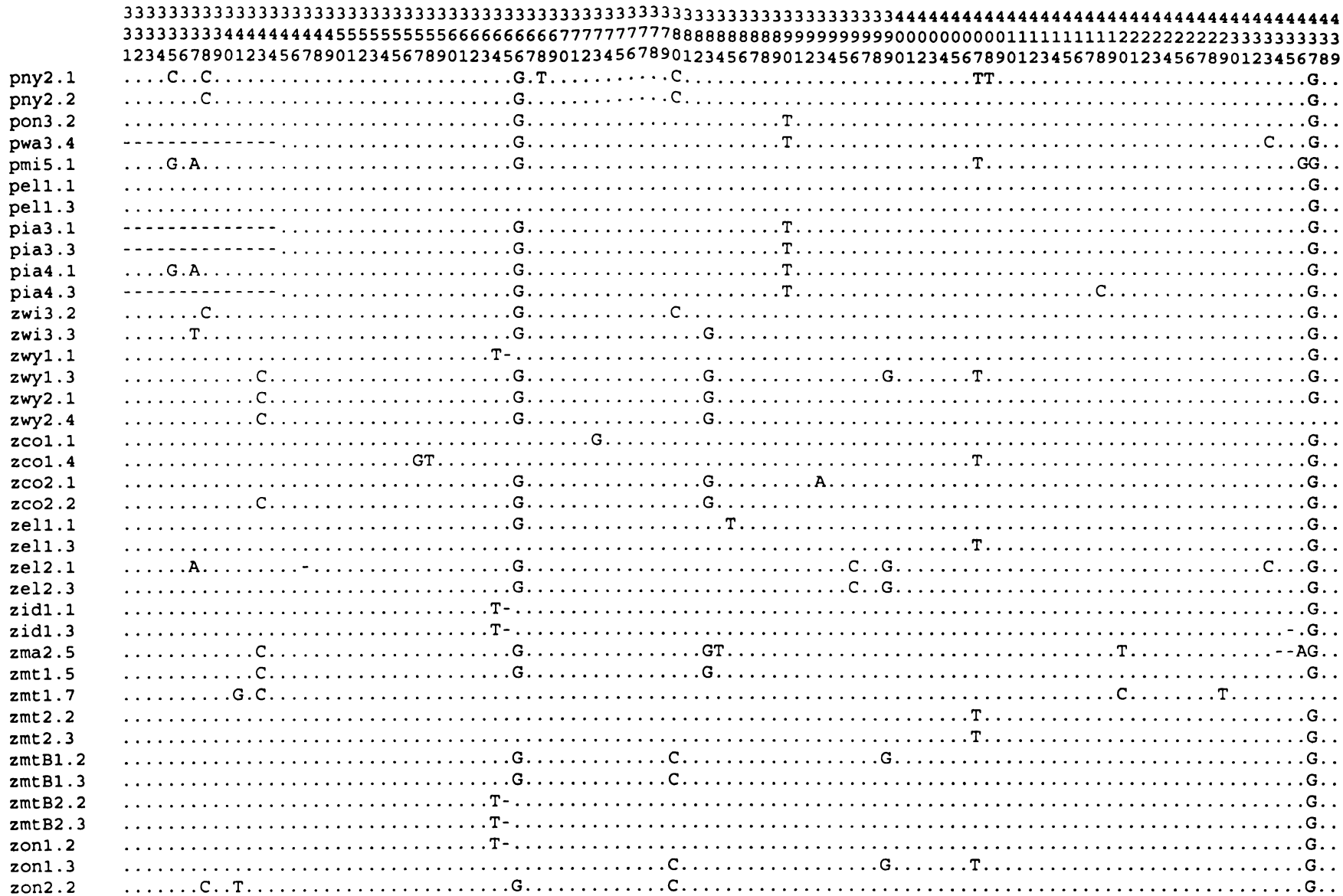
```
        11111111111111111111111111111111111111111111111111111111111111111111111111111111111111111222222222222222222222
        11111111122222222223333333333444444444455555555556666666666777777777788888888889999999999000000000011111111112
        12345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890
zon2.4  A........C......C............T....GA...........G..........A...................................................
zwa2.1  A........C......C............T....GA...........G..........A...................................................
zwa2.3  A........C......C............T....GA...........G..........A...................................................
zca1.1  .G.............................................G..........A.........G....................G....................
zca1.2  .G.............................................G..........A.........G....................G....................
zca2.1  .............T..C..............................G..........A.........G....................G....................
zca2.2  .............T..C..............................G..........A.........G....................G....................
zma2.6  .......-........C.........................................A...................................................
zsb3.3  .G..................................G..........G..........A.........G.........................................
zne2.2  .G..................................G..........G..........A.........G.........................................
zne2.4  ............GT..C.............................................................................................
zma1.1  .G..........A............................................C....................................................
zma1.2  .G.............................................G..........A.........G....................G....................
zmn2.1  .............T..C..............................G..........A.........G....................G........T...........
pmn1.1  ............GT..C.............................................................................................
pmn1.2  .G..................................G..........G..........A.........G.........................................
pmn2.1  ............GT..C.............................................................................................
pmn2.2  ............GT..C.............................................................................................
pma2.6  .G..................................G..........G..........A.........G.........................................
pma2.7  .G..................................G..........G..........A.........G.........................................
ppa2.5  .......-......T.C.................GA...........GT.........A..............................G....................
ppa2.6  .G..............C.................GA...........G..........A..............................AA.T........A........
pil2.5  .......-.....T..C.................GA...........G..........A....C.........................G....A...............
pil2.6  .......-.....T..C.................GA...........G..........A....C.........................G....A...............
pco1.2  .G..................................G..........G..........A.........G.........................................
pco1.3  .......-.....T..C........................................................................G...........G........
pne1.1  ............GT..C.............................................................................................
pne1.2  ............GT..C............................................................A................................
pne2.2  .G..................................G..........G..........A.........G.........................................
pne2.3  ............GT..C.............................................................................................
C10.3   ................C.................T.......................A..............T...............G....................
C10.4   ................C.................G.................A.....A.......G......T...............G...............T....
m1.11   .............T..C............................------.......A.........G....................G....................
m1.8    .............T..C....................................................G........................................
c9.7    ................C.................T...............C.......A..............................G...G...........T....
t4.1    .......-.....T..C............T.............C............A.A...................TAAT.......G..T...........T.....
t4.2    .......-.....T..C............T..G......................TA.A....C..............TAAT.......GC.T...........T.....
```

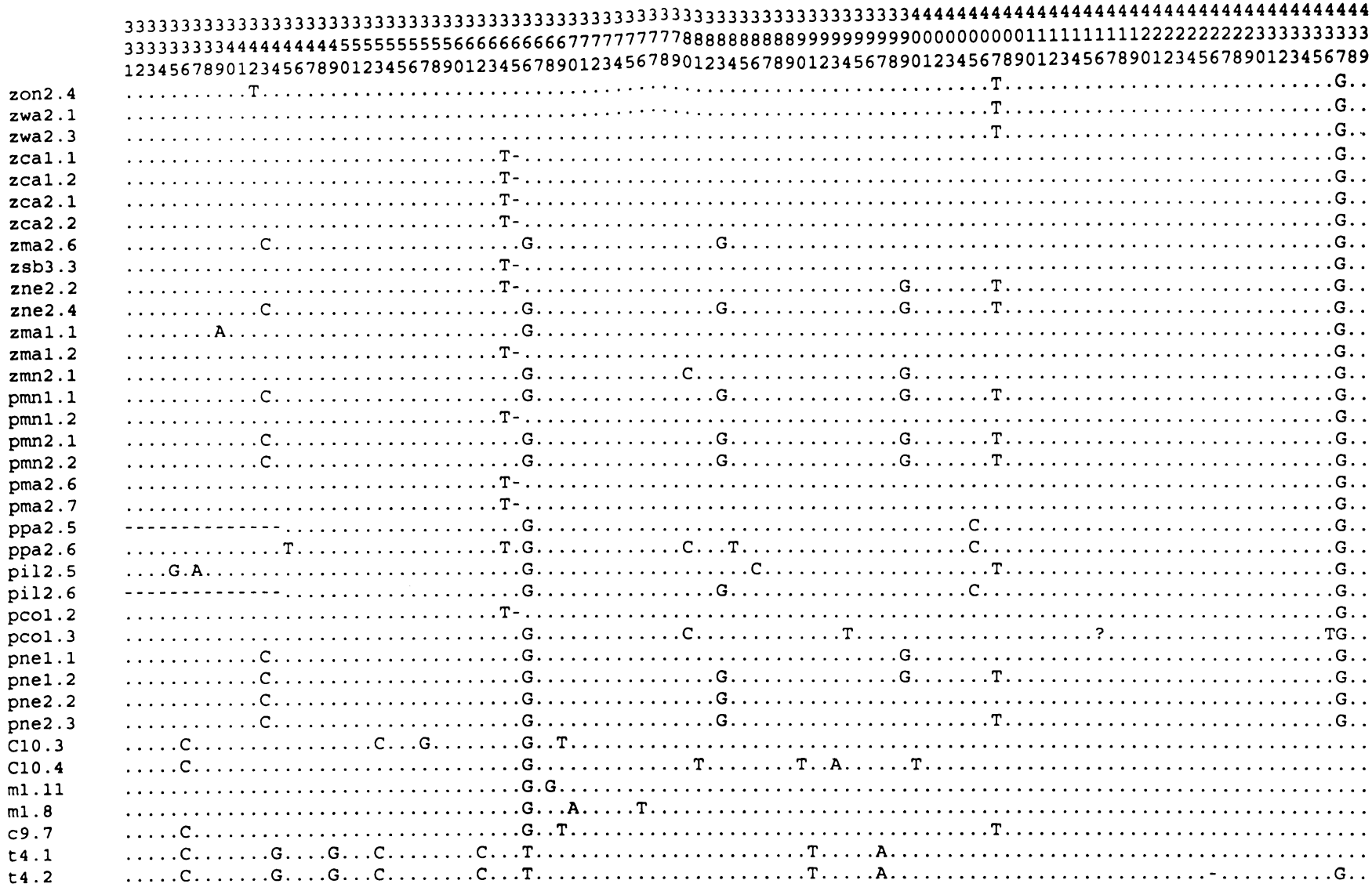

zor5.3 -GCTCACTCAAGCTCGAATATCTAGTACCTTGTGACTAGTTCAGAGCATGTTTTCTTTCCCTCAAGTTTTGCCTTTTGATGTCCTCTTCTCTAATTATACCAGCTT-CAT
zor5.9
znd2.3
znd2.1
zne2.3
zne2.1
znd1.3
znd1.1
zmio2.6
zmio2.5
zmio1.3
zmio1.2
zpa1.1
zor6.8
zor6.9
zmt95101.
zmt9510.2
zwa95 1.3
zwa95 1.5
zwa95 6.4
zwa95 6.6
zny3.2
zny3.3
zny4.2
zny4.5
zsb3.6
pmi1.5
pmi2.5
fa.3
fa.4
fp.4
fp.6
zsd1.4
zsd1.3
put1.6
put1.5
pma1.1
pny1.1
pny1.2


pny2.1
pny2.2
pon3.2
pwa3.4
pmi5.1
pel1.1
pel1.3
pia3.1
pia3.3
pia4.1
pia4.3
zwi3.2
zwi3.3
zwy1.1
zwy1.3
zwy2.1
zwy2.4
zco1.1
zco1.4
zco2.1
zco2.2
zel1.1
zel1.3
zel2.1
zel2.3
zid1.1
zid1.3
zma2.5
zmt1.5
zmt1.7
zmt2.2
zmt2.3
zmtB1.2
zmtB1.3
zmtB2.2
zmtB2.3
zon1.2
zon1.3
zon2.2


zon2.4
zwa2.1
zwa2.3
zca1.1
zca1.2
zca2.1
zca2.2
zma2.6
zsb3.3
zne2.2
zne2.4
zma1.1
zma1.2
zmn2.1
pmn1.1
pmn1.2
pmn2.1
pmn2.2
pma2.6
pma2.7
ppa2.5
ppa2.6
pil2.5
pil2.6
pco1.2
pco1.3
pne1.1
pne1.2
pne2.2
pne2.3
C10.3
C10.4
m1.11
m1.8
c9.7
t4.1
t4.2


zor5.3 TTCTTTGTTGAAAAAAAGTGGTTTACAAAAATT-GACACCAGCTGATT-GATACATTTTCCTT-GTCAAAGTTGGCAATGTTGACTATCAAACAAATTACCCTTT--GC
zor5.9
znd2.3
znd2.1
zne2.3
zne2.1
znd1.3
znd1.1
zmio2.6
zmio2.5
zmio1.3
zmio1.2
zpa1.1
zor6.8
zor6.9
zmt9510.1
zmt9510.2
zwa95 1.3
zwa95 1.5
zwa95 6.4
zwa95 6.6
zny3.2
zny3.3
zny4.2
zny4.5
zsb3.6
pmi1.5
pmi2.5
fa.3
fa.4
fp.4
fp.6
zsd1.4
zsd1.3
put1.6
put1.5
pma1.1
pny1.1
pny1.2

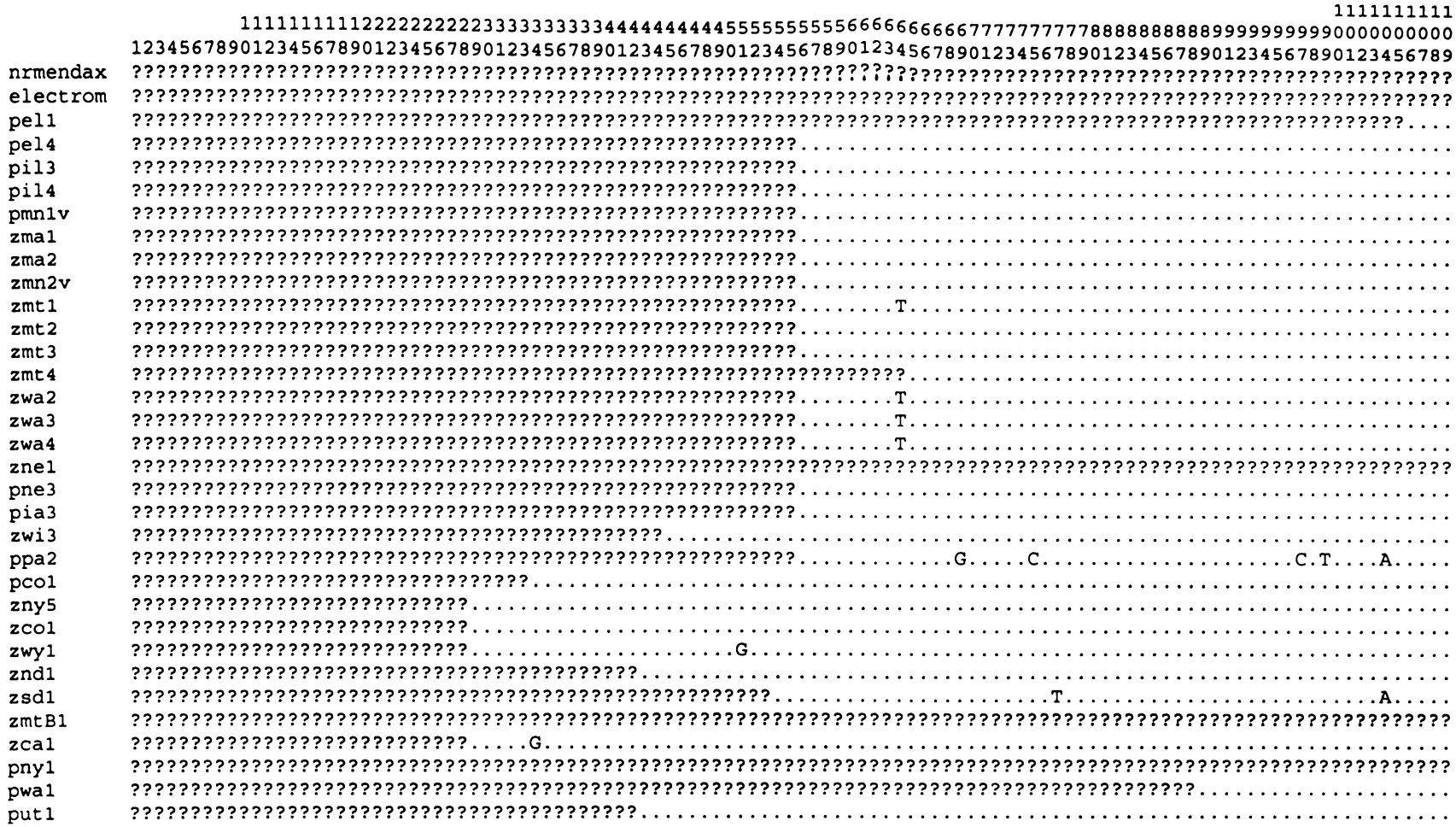

pny2.1
pny2.2
pon3.2
pwa3.4
pmi5.1
pel1.1
pel1.3
pia3.1
pia3.3
pia4.1
pia4.3
zwi3.2
zwi3.3
zwy1.1
zwy1.3
zwy2.1
zwy2.4
zco1.1
zco1.4
zco2.1
zco2.2
zel1.1
zel1.3
zel2.1
zel2.3
zid1.1
zid1.3
zma2.5
zmt1.5
zmt1.7
zmt2.2
zmt2.3
zmtB1.2
zmtB1.3
zmtB2.2
zmtB2.3
zon1.2
zon1.3
zon2.2


zon2.4
zwa2.1
zwa2.3
zca1.1
zca1.2
zca2.1
zca2.2
zma2.6
zsb3.3
zne2.2
zne2.4
zma1.1
zma1.2
zmn2.1
pmn1.1
pmn1.2
pmn2.1
pmn2.2
pma2.6
pma2.7
ppa2.5
ppa2.6
pil2.5
pil2.6
pco1.2
pco1.3
pne1.1
pne1.2
pne2.2
pne2.3
C10.3
C10.4
m1.11
m1.8
c9.7
t4.1
t4.2

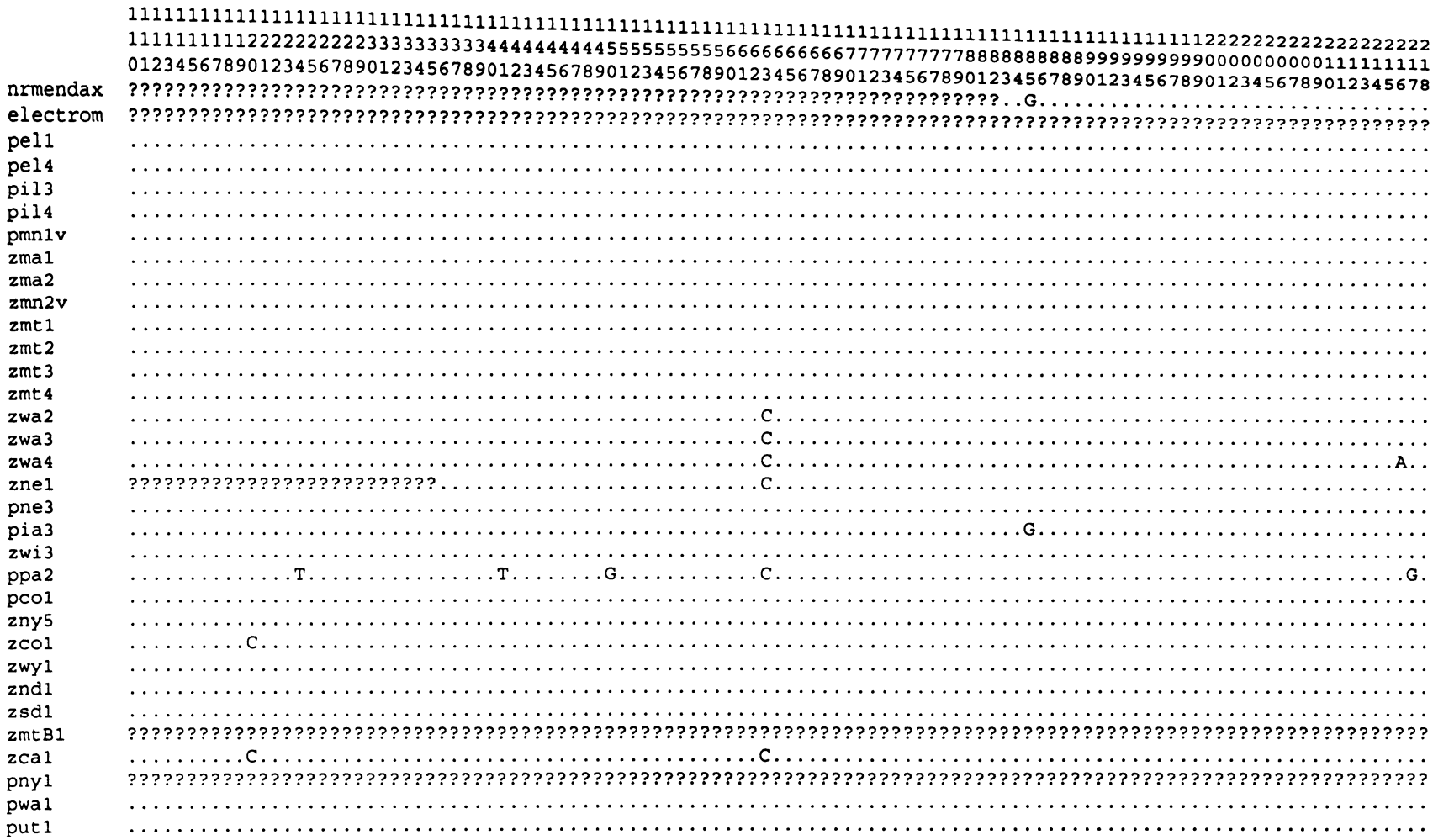
APPENDIX D

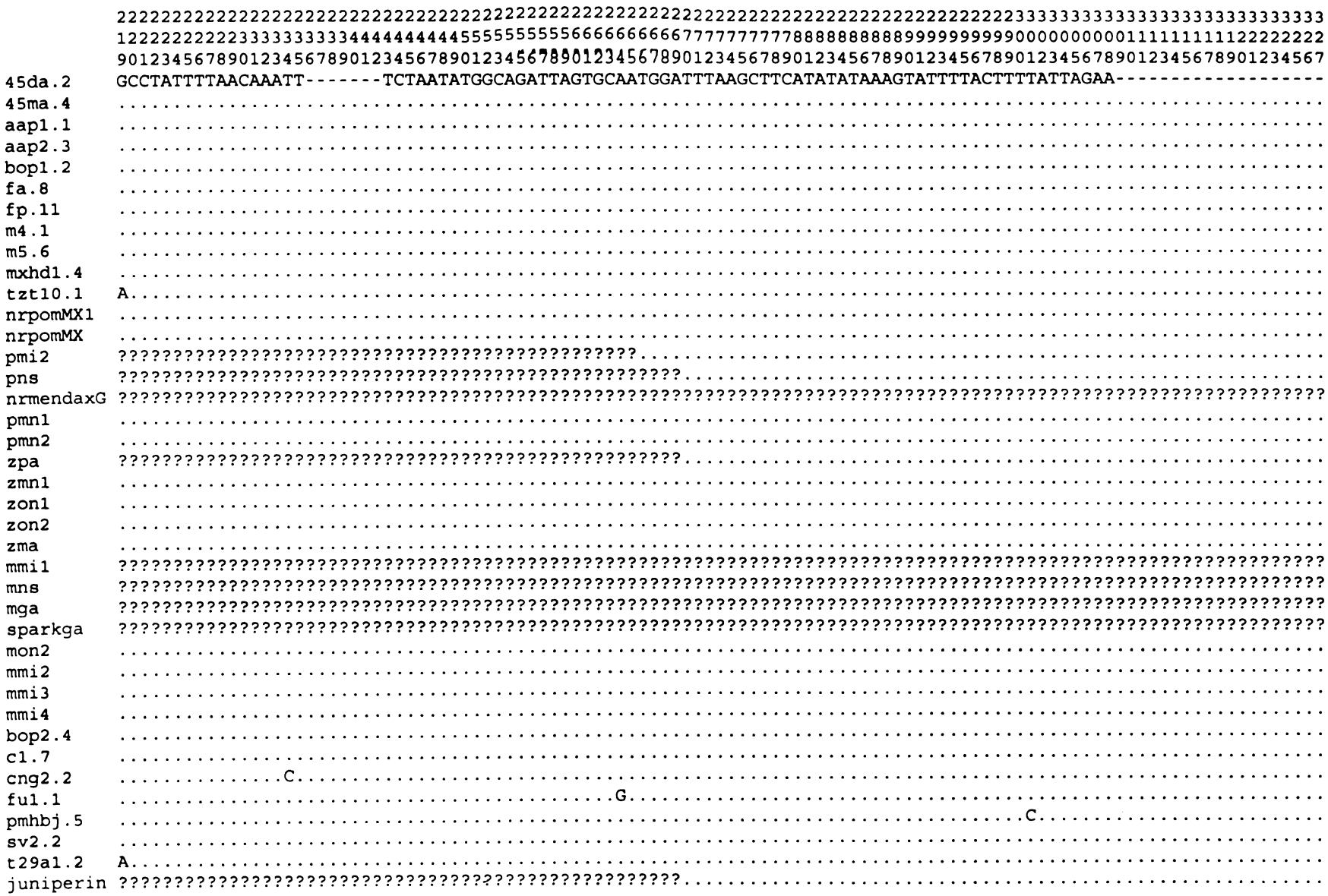
COI/COII sequence alignment

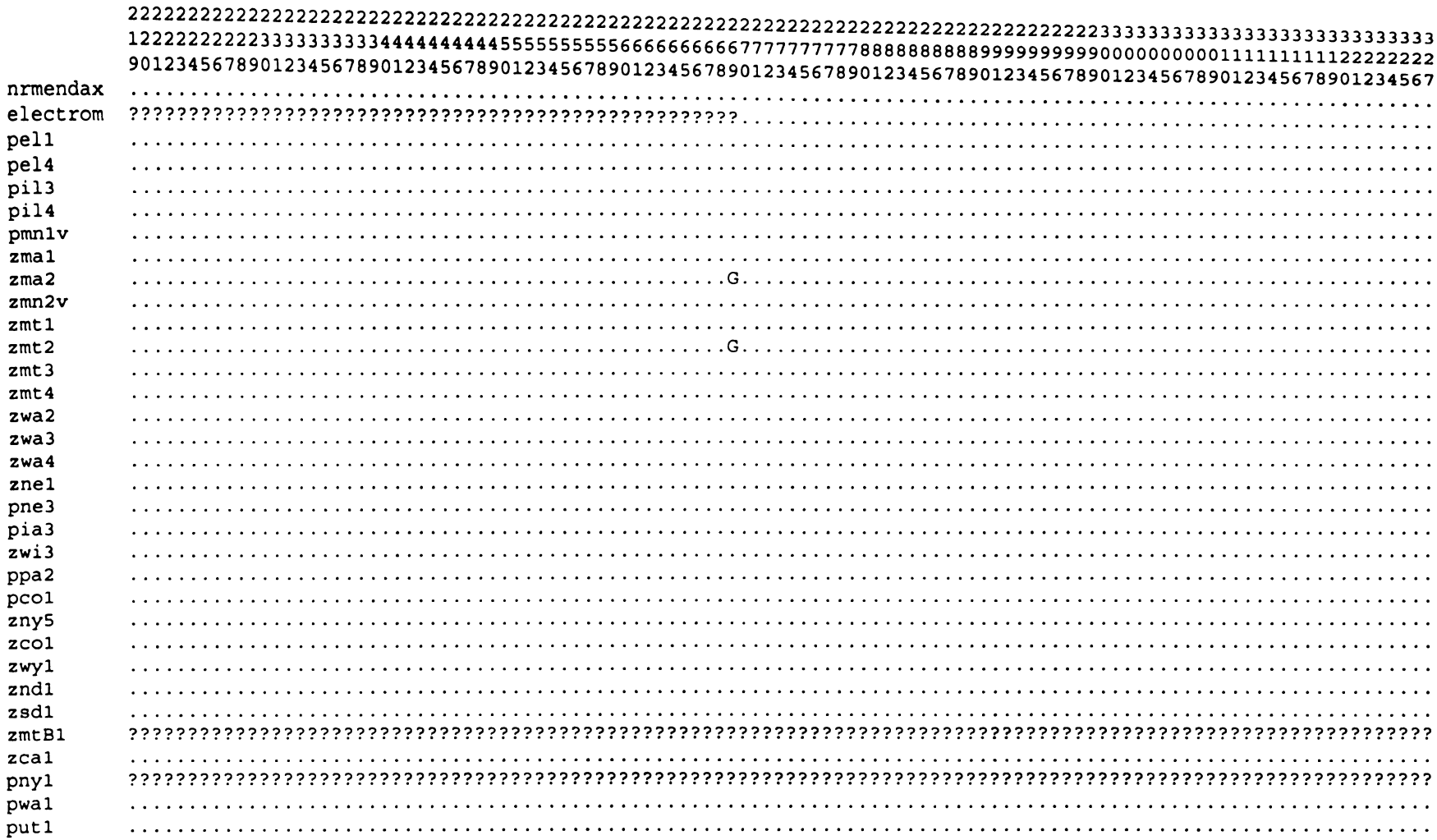
```
                                                                                                                      1111111111
                    1111111111222222222233333333334444444444555555555566666666667777777777888888888899999999990000000000
           1234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789
45da.2     ATACCTCGACGTTATTCAGACTATCCAGATGCTTACACTACATGAAATGTAGTTTCAACTATTGGATCTTCTATTTCACTACTAGGAATTTTATTTTTCTTATTTATTA
45ma.4     .............................................................................................................
aap1.1     .............................................................................................................
aap2.3     .............................................................................................................
bop1.2     .............................................................................................................
fa.8       .............................................................................................................
fp.11      .............................................................................................................
m4.1       .............................................................................................................
m5.6       .............................................................................................................
mxhd1.4    .............................................................................................................
tzt10.1    ....................T...........................A..A.C........................T..T.......C........T..........
nrpomMX1   ???????????????????????????????????????????????????????......................................................
nrpomMX    ???????????????????????????????????????????????????????..........G...............T....G......................
pmi2       ?????????????????????????????????????????????????????????????????????????????????????????????????????????????
pns        ?????????????????????????????????????????????????????????????????????????????????????????????????????????????
nrmendaxG  ?????????????????????????????????????????????????????????????????????????????????????????????????????????????
pmn1       ???????????????????????????????????????????????????????......................................................
pmn2       ???????????????????????????????????????????????????????......................................................
zpa        ?????????????????????????????????????????????????????????????????????????????????????????????????????????????
zmn1       ???????????????????????????????????????????????????????......................................................
zon1       ???????????????????????????????????????????????????????......................................................
zon2       ???????????????????????????????????????????????????????......................................................
zma        ???????????????????????????????????????????????????????......................................................
mmi1       ?????????????????????????????????????????????????????????????????????????????????????????????????????????????
mns        ?????????????????????????????????????????????????????????????????????????????????????????????????????????????
mga        ?????????????????????????????????????????????????????????????????????????????????????????????????????????????
sparkga    ?????????????????????????????????????????????????????????????????????????????????????????????????????????????
mon2       ???????????????????????????????????????????????????????......................................................
mmi2       ???????????????????????????????????????????????????????......................................................
mmi3       ???????????????????????????????????????????????????????......................................................
mmi4       ???????????????????????????????????????????????????????????????..............................................
bop2.4     ???????????????????????????????????????????????????????......................................................
c1.7       ???????????????????????????????????????????????????????.......................T..T...........................
cng2.2     ???????????????????????????????????????????????????????.............A.........T..T................T..........
fu1.1      ???????????????????????????????????????????????????????.......................T..T................A..........
pmhbj.5    ???????????????????????????????????????????????????????......................................................
sv2.2      ???????????????????????????????????????????????????????.............A.........T...................T..........
t29a1.2    ???????????????????????????????????????????????????????.......................T..T.......C........T..........
juniperin  ?????????????????????????????????????????????????????????????????????????????????????????????????????????????
```

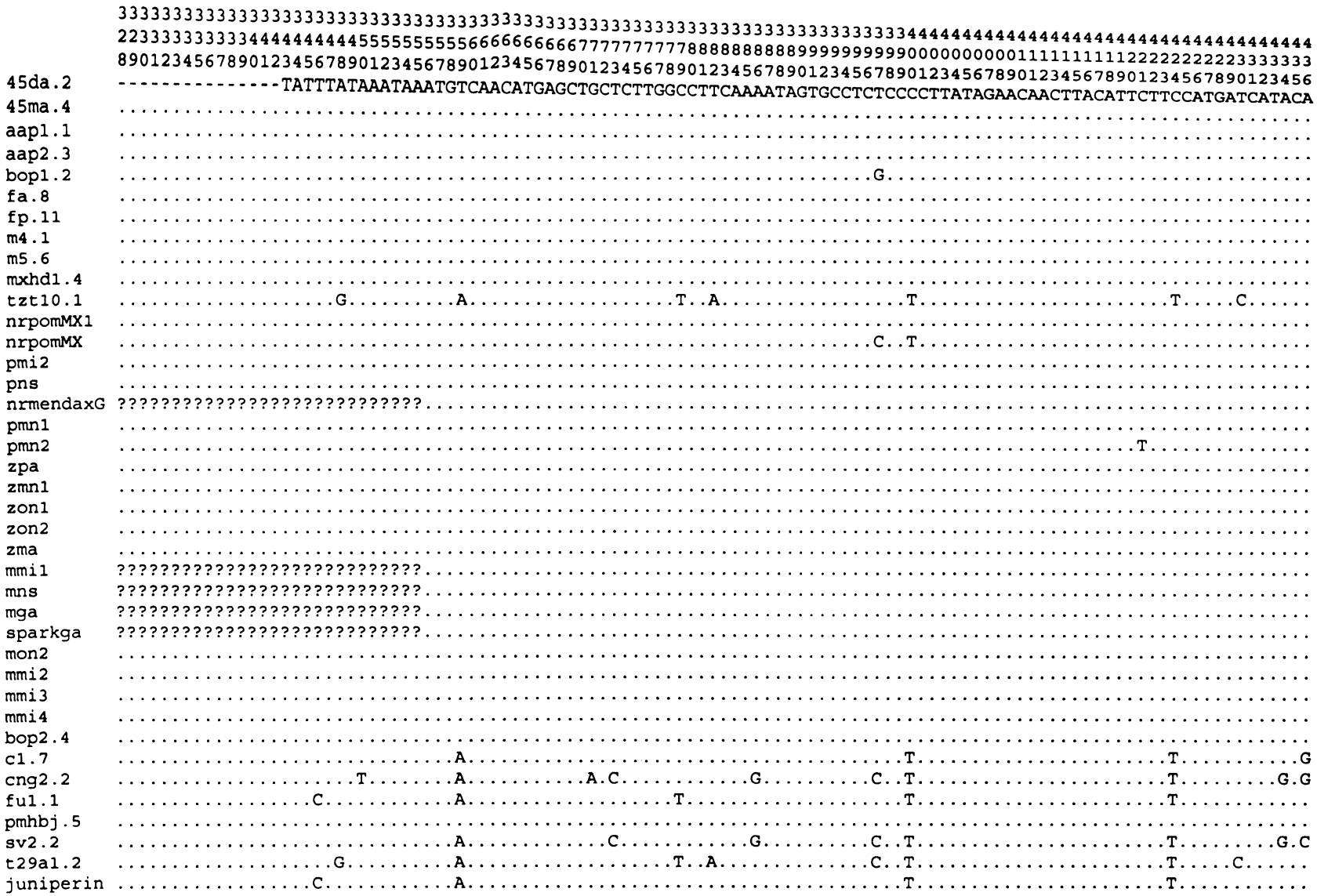

1111111111222222222233333333334444444444555555555566666666677777777788888888899999999990000000000
1234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789
nrmendax
electrom
pel1
pel4
pil3
pil4
pmn1v
zma1
zma2
zmn2v
zmt1
zmt2
zmt3
zmt4
zwa2
zwa3
zwa4
zne1
pne3
pia3
zwi3
ppa2
pco1
zny5
zco1
zwy1
znd1
zsd1
zmtB1
zca1
pny1
pwa1
put1

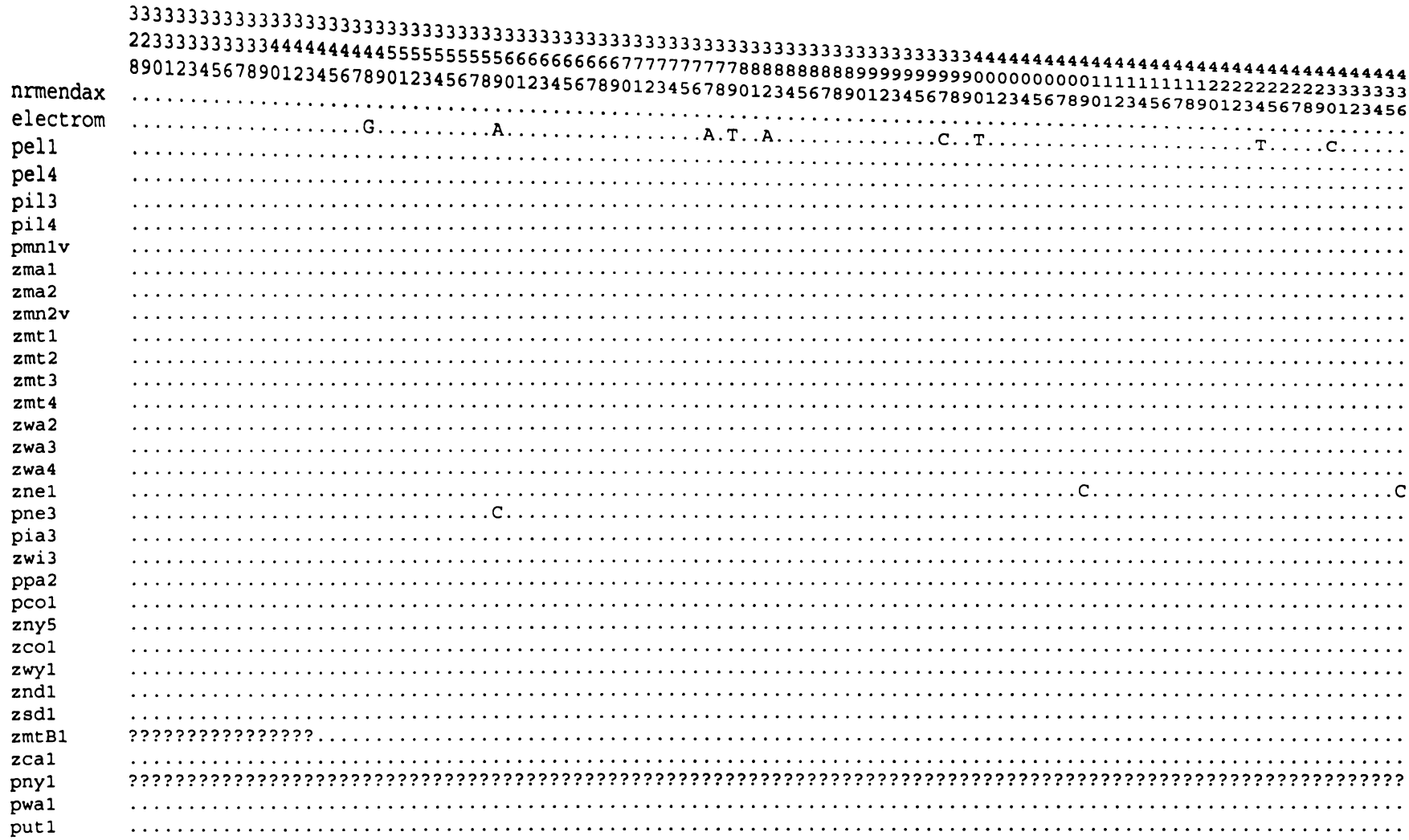
```
            1111111111111111111111111111111111111111111111111111111111111111111111111111111111111111112222222222222222222
            1111111111222222222233333333334444444444555555555566666666667777777777888888888899999999990000000000111111111
            0123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678
45da.2      TTTGAGAAAGTTTAATTTCACAACGACAAGTAATTTATCCAATACAATTAAATTCTTCAATTGAATGATTACAAAATACACCTCCAACAGAACATAGATACTCTGAGTT
45ma.4      .............................................................................................................
aap1.1      .............................................................................................................
aap2.3      .............................................................................................................
bop1.2      .............................................................................................................
fa.8        .............................................................................................................
fp.11       .............................................................................................................
m4.1        .............................................................................................................
m5.6        .............................................................................................................
mxhd1.4     .............................................................................................................
tzt10.1     ..........C.......T......T......................................G.....................G.............T.....A..
nrpomMX1    .............................................................................................................
nrpomMX     ..........C..........................C.....................................G..........G..........C..T.....A..
pmi2        ?????????????????????????????????????????????????????????????????????????????????????????????????????????????
pns         ?????????????????????????????????????????????????????????????????????????????????????????????????????????????
nrmendaxG   ?????????????????????????????????????????????????????????????????????????????????????????????????????????????
pmn1        .............................................................................................................
pmn2        .............................................................................................................
zpa         ?????????????????????????????????????????????????????????????????????????????????????????????????????????????
zmn1        .............................................................................................................
zon1        .............................................................................................................
zon2        .............................................................................................................
zma         .............................................................................................................
mmi1        ?????????????????????????????????????????????????????????????????????????????????????????????????????????????
mns         ?????????????????????????????????????????????????????????????????????????????????????????????????????????????
mga         ?????????????????????????????????????????????????????????????????????????????????????????????????????????????
sparkga     ?????????????????????????????????????????????????????????????????????????????????????????????????????????????
mon2        .............................................................................................................
mmi2        ........................................T....................................................................
mmi3        .............................................................................................................
mmi4        .............................................................................................................
bop2.4      .............................................................................................................
c1.7        ......................................................................................G..........C........A..
cng2.2      ....................A.................................................................G.............T.....A..
ful.1       ..........C...........................................................................G.............T.....A..
pmhbj.5     ................................G............................................................................
sv2.2       .....................G..............................................C.................G.............T.....A..
t29a1.2     ..........C.......T......T......................................G.....................G.............T.....A..
juniperin   ?????????????????????????????????????????????????????????????????????????????????????????????????????????????
```

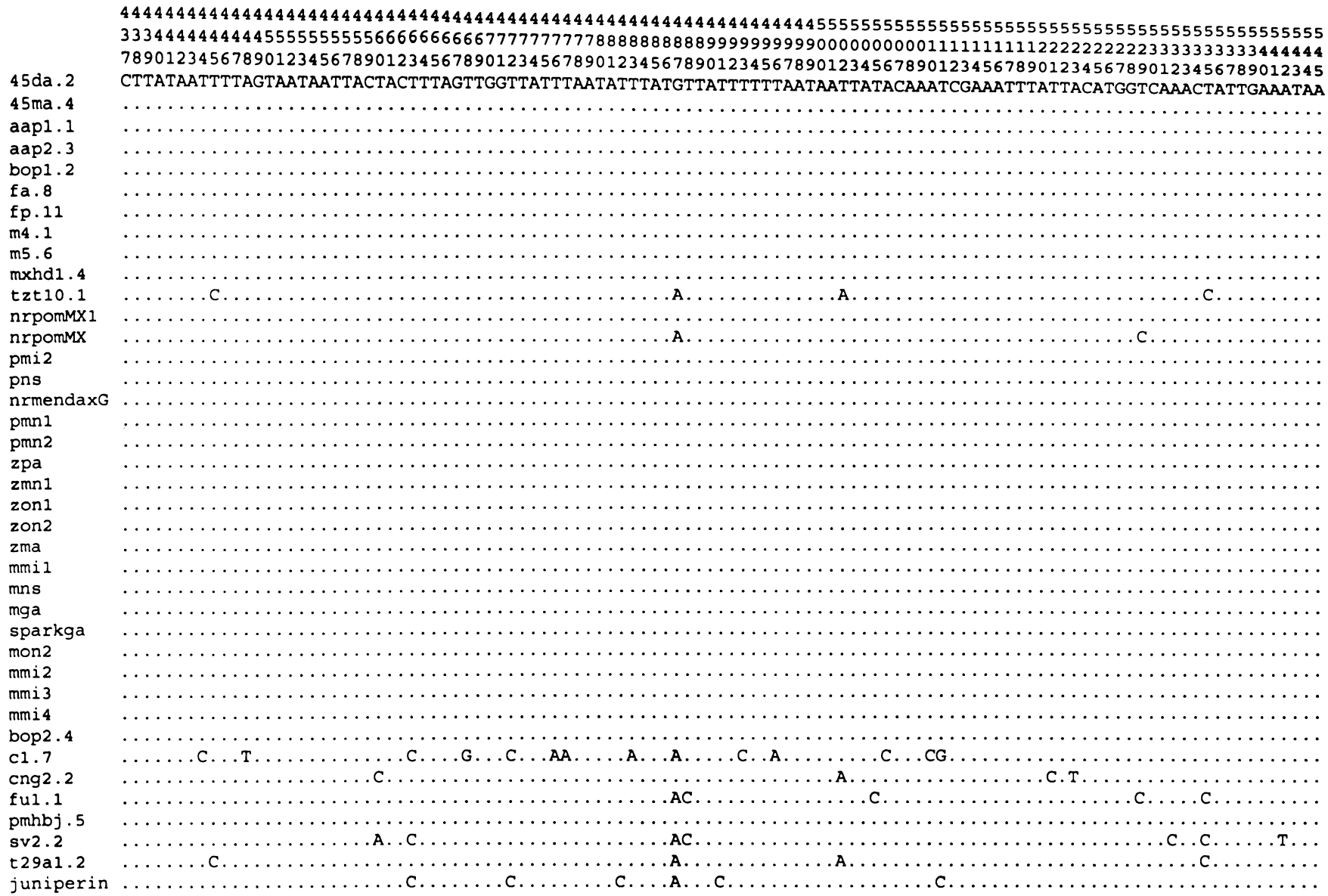

nrmendax
electrom
pel1
pel4
pil3
pil4
pmn1v
zma1
zma2
zmn2v
zmt1
zmt2
zmt3
zmt4
zwa2
zwa3
zwa4
zne1
pne3
pia3
zwi3
ppa2
pco1
zny5
zco1
zwy1
znd1
zsd1
zmtB1
zca1
pny1
pwa1
put1

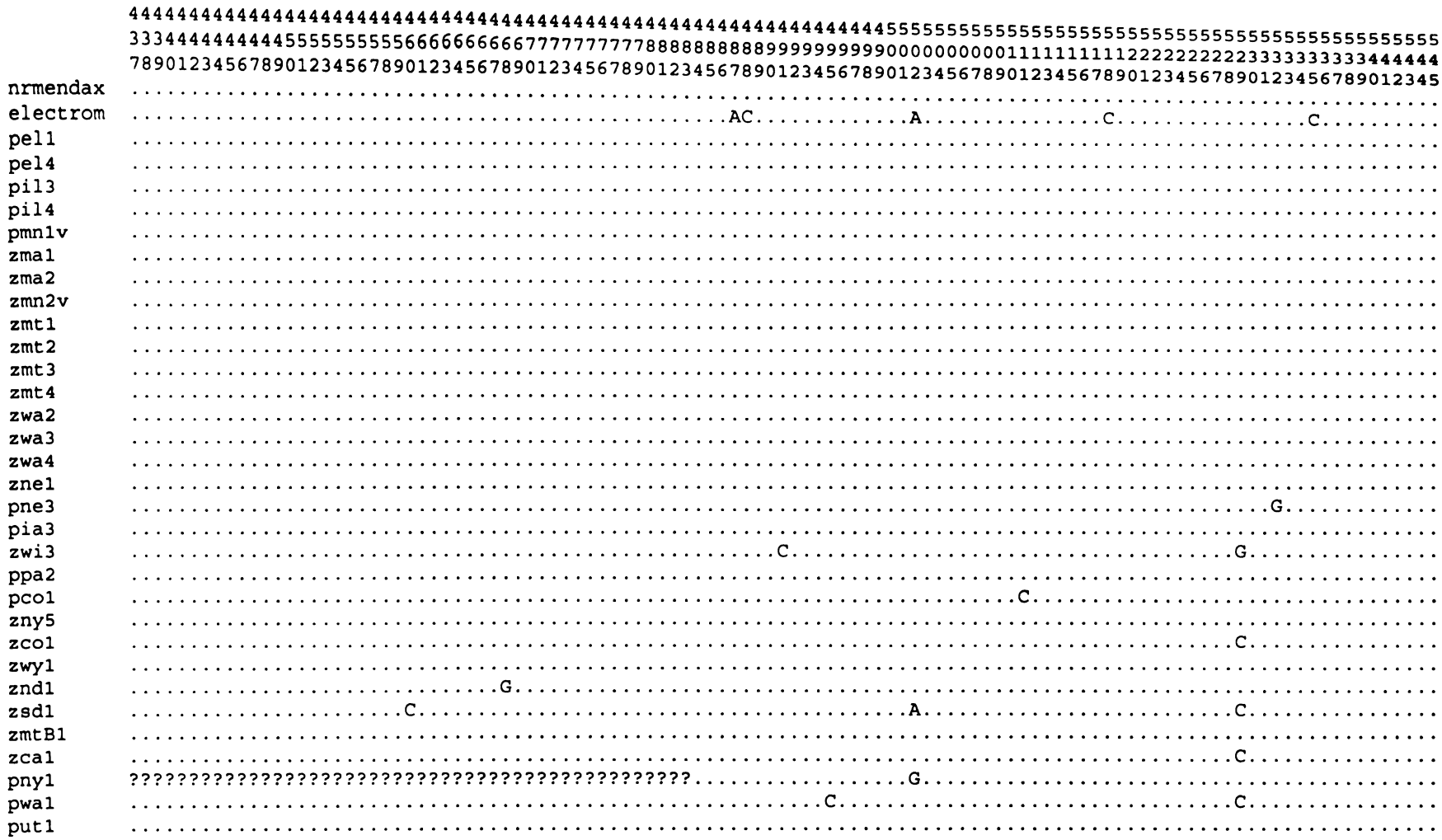

45da.2 GCCTATTTTAACAAATT-------TCTAATATGGCAGATTAGTGCAATGGATTTAAGCTTCATATATAAAGTATTTTACTTTTATTAGAA----------
45ma.4
aap1.1
aap2.3
bop1.2
fa.8
fp.11
m4.1
m5.6
mxhd1.4
tzt10.1 A
nrpomMX1
nrpomMX
pmi2
pns
nrmendaxG
pmn1
pmn2
zpa
zmn1
zon1
zon2
zma
mmi1
mns
mga
sparkga
mon2
mmi2
mmi3
mmi4
bop2.4
c1.7
cng2.2 C
fu1.1 G
pmhbj.5 C
sv2.2
t29a1.2 A
juniperin

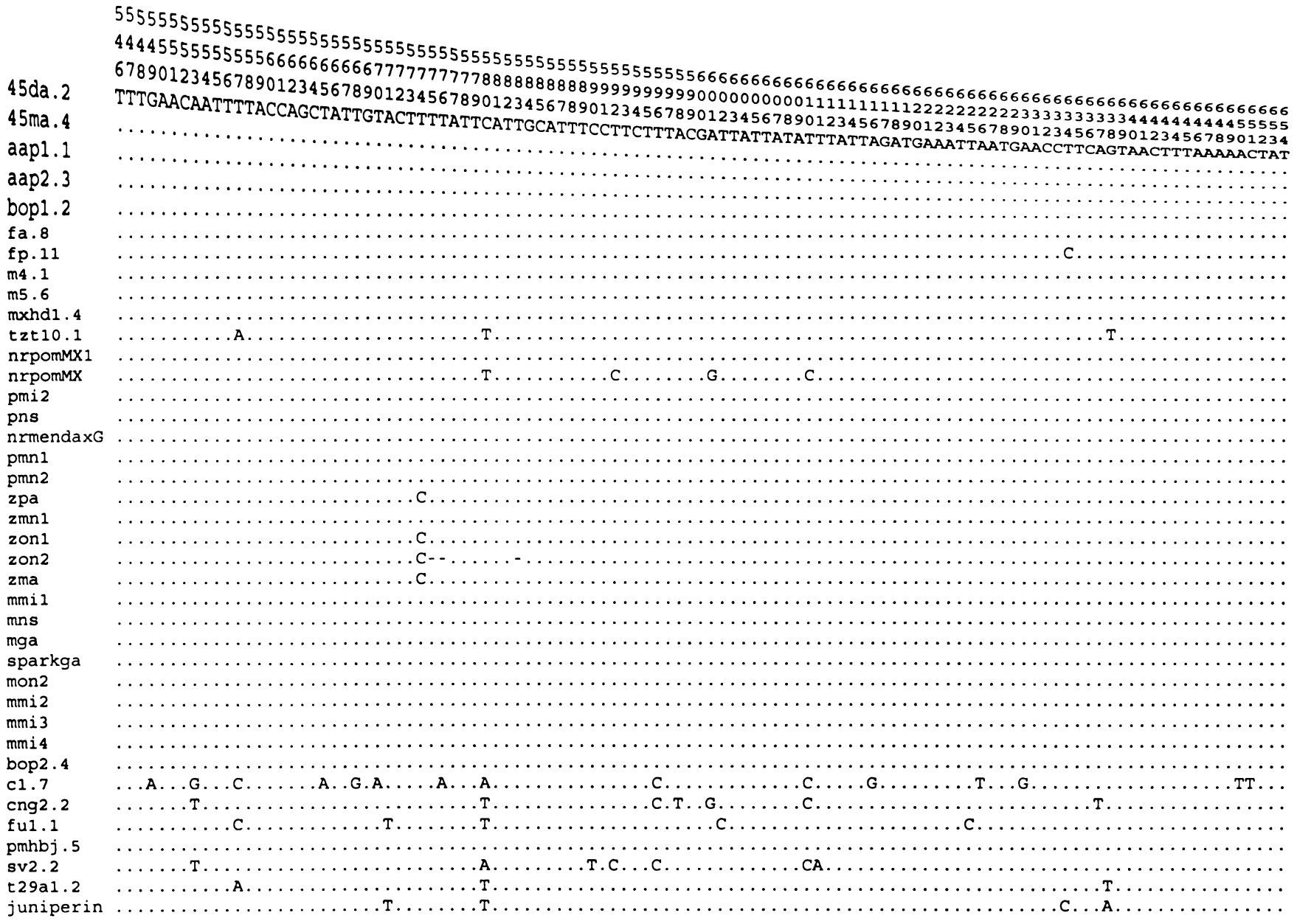

2222222222222222222222222222222222222222222222222222222222222222222222222222233333333333333333333333333
1222222222333333333344444444445555555555666666666677777777778888888889999999990000000000111111111122222222
9012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567
nrmendax
electrom
pel1
pel4
pil3
pil4
pmn1v
zma1
zma2
zmn2v
zmt1
zmt2
zmt3
zmt4
zwa2
zwa3
zwa4
zne1
pne3
pia3
zwi3
ppa2
pco1
zny5
zco1
zwy1
znd1
zsd1
zmtB1
zca1
pny1
pwa1
put1

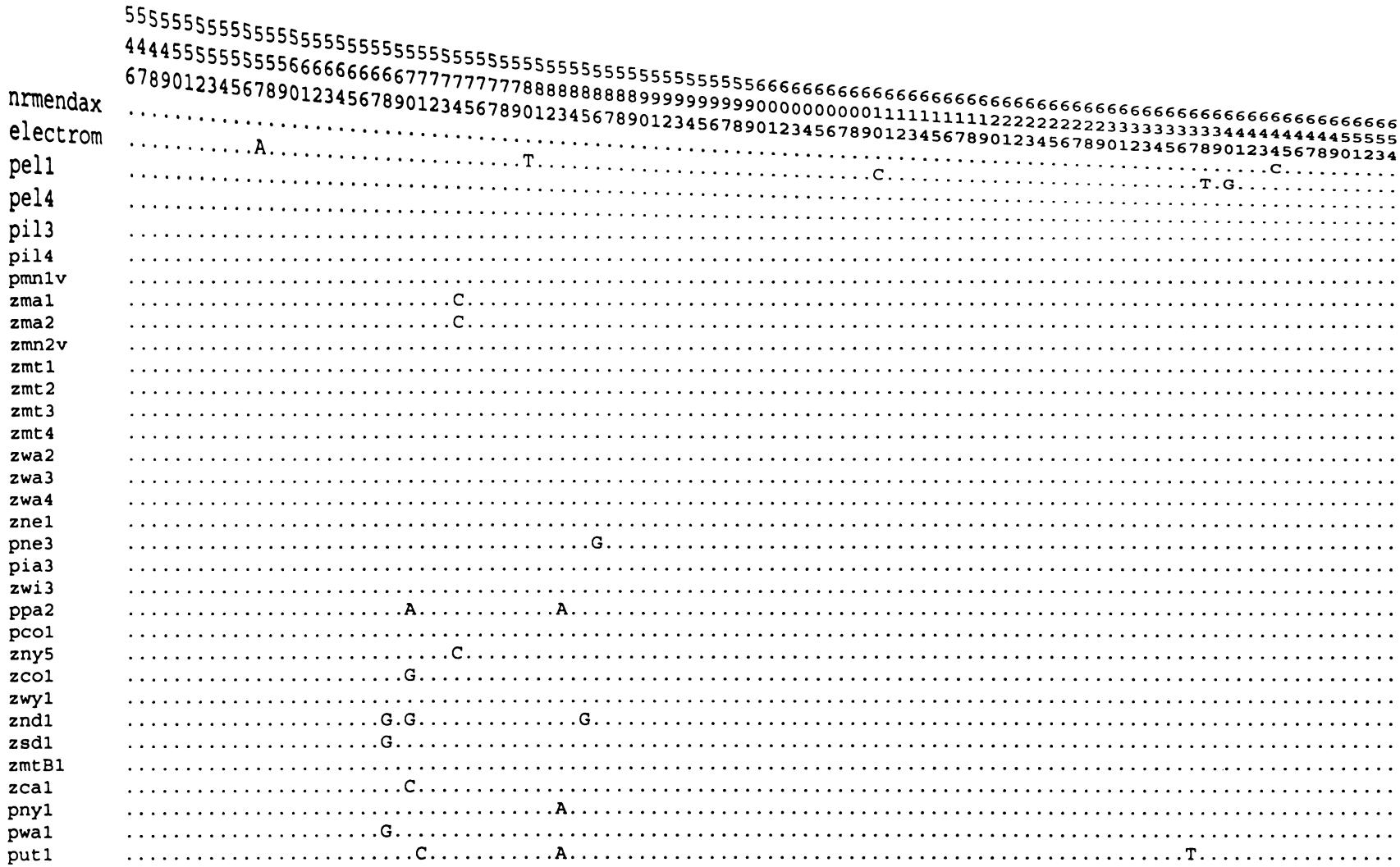

45da.2 ----------TATTTATAAATAAATGTCAACATGAGCTGCTCTTGGCCTTCAAAATAGTGCCTCTCCCCTTATAGAACAACTTACATTCTTCCATGATCATACA
45ma.4
aap1.1
aap2.3
bop1.2
fa.8
fp.11
m4.1
m5.6
mxhd1.4
tzt10.1
nrpomMX1
nrpomMX
pmi2
pns
nrmendaxG
pmn1
pmn2
zpa
zmn1
zon1
zon2
zma
mmi1
mns
mga
sparkga
mon2
mmi2
mmi3
mmi4
bop2.4
c1.7
cng2.2
fu1.1
pmhbj.5
sv2.2
t29a1.2
juniperin


nrmendax
electrom
pel1
pel4
pil3
pil4
pmn1v
zma1
zma2
zmn2v
zmt1
zmt2
zmt3
zmt4
zwa2
zwa3
zwa4
zne1
pne3
pia3
zwi3
ppa2
pco1
zny5
zco1
zwy1
znd1
zsd1
zmtB1
zca1
pny1
pwa1
put1

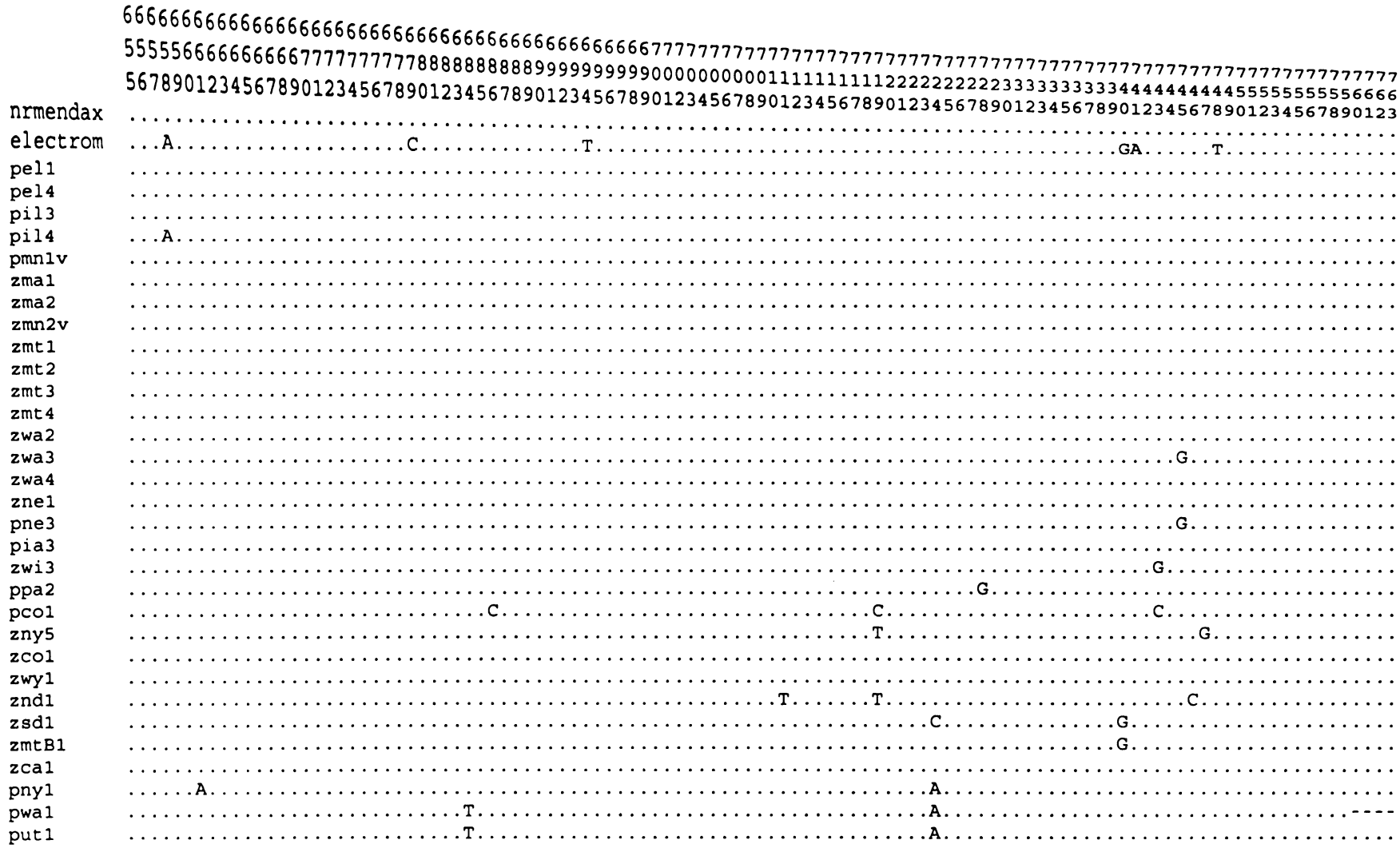

45da.2 CTTATAATTTTAGTAATAATTACTACTTTAGTTGGTTATTTAATATTTATGTTATTTTTTAATAATTATACAAATCGAAATTTATTACATGGTCAAACTATTGAAATAA
45ma.4
aap1.1
aap2.3
bop1.2
fa.8
fp.11
m4.1
m5.6
mxhd1.4
tzt10.1
nrpomMX1
nrpomMX
pmi2
pns
nrmendaxG
pmn1
pmn2
zpa
zmn1
zon1
zon2
zma
mmi1
mns
mga
sparkga
mon2
mmi2
mmi3
mmi4
bop2.4
c1.7
cng2.2
fu1.1
pmhbj.5
sv2.2
t29a1.2
juniperin

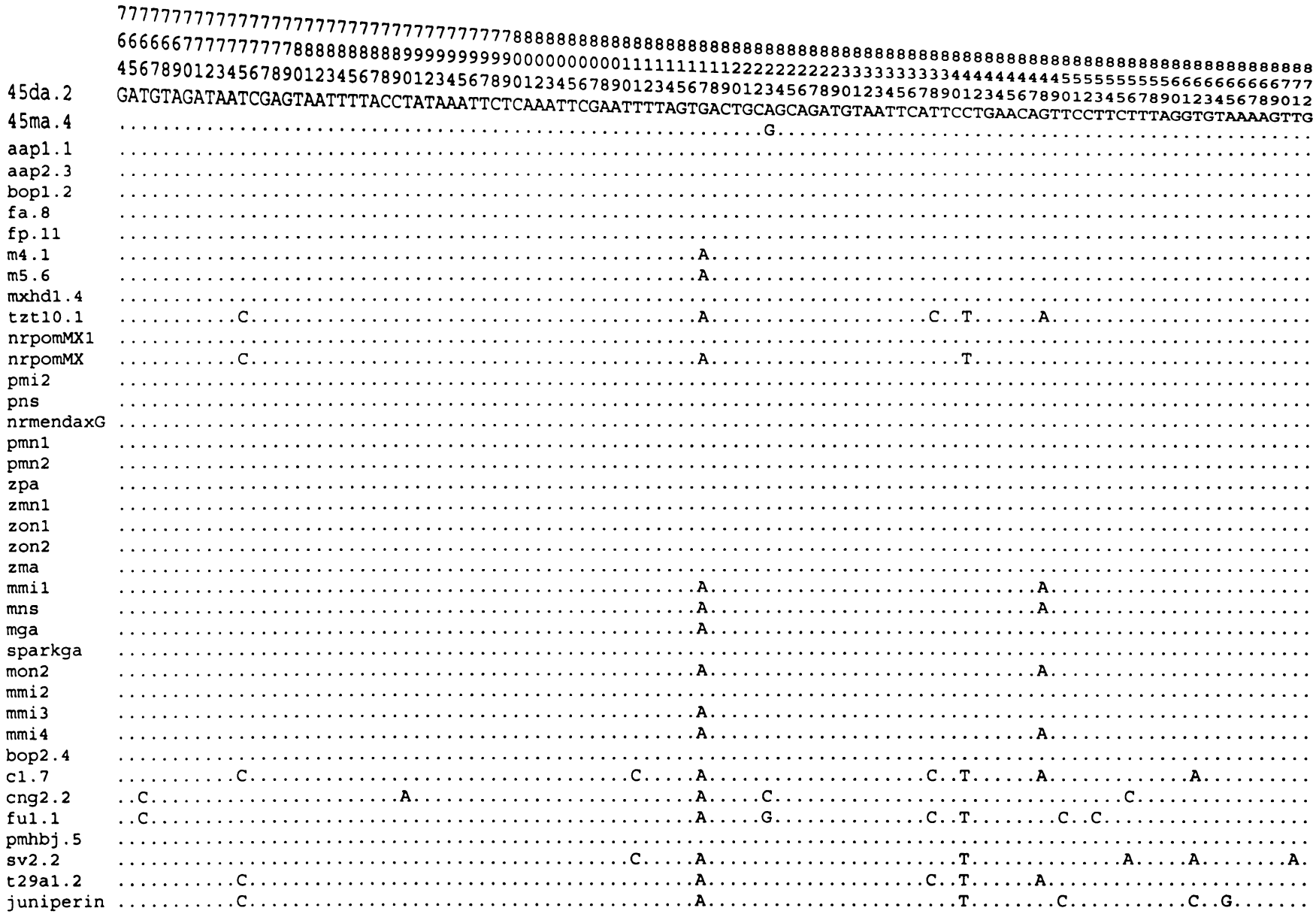

444444444444444444444444444444444444444444444444444444444444455555555555555555555555555555555555555555
333444444444455555555556666666666777777777788888888899999999900000000001111111111222222222233333333334444444
789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345
nrmendax
electrom AC A C C
pel1
pel4
pil3
pil4
pmn1v
zma1
zma2
zmn2v
zmt1
zmt2
zmt3
zmt4
zwa2
zwa3
zwa4
zne1
pne3 G
pia3
zwi3 C G
ppa2
pco1 C
zny5
zco1 C
zwy1
znd1 G
zsd1 C A C
zmtB1
zca1 C
pny1 ??? G
pwa1 C C
put1

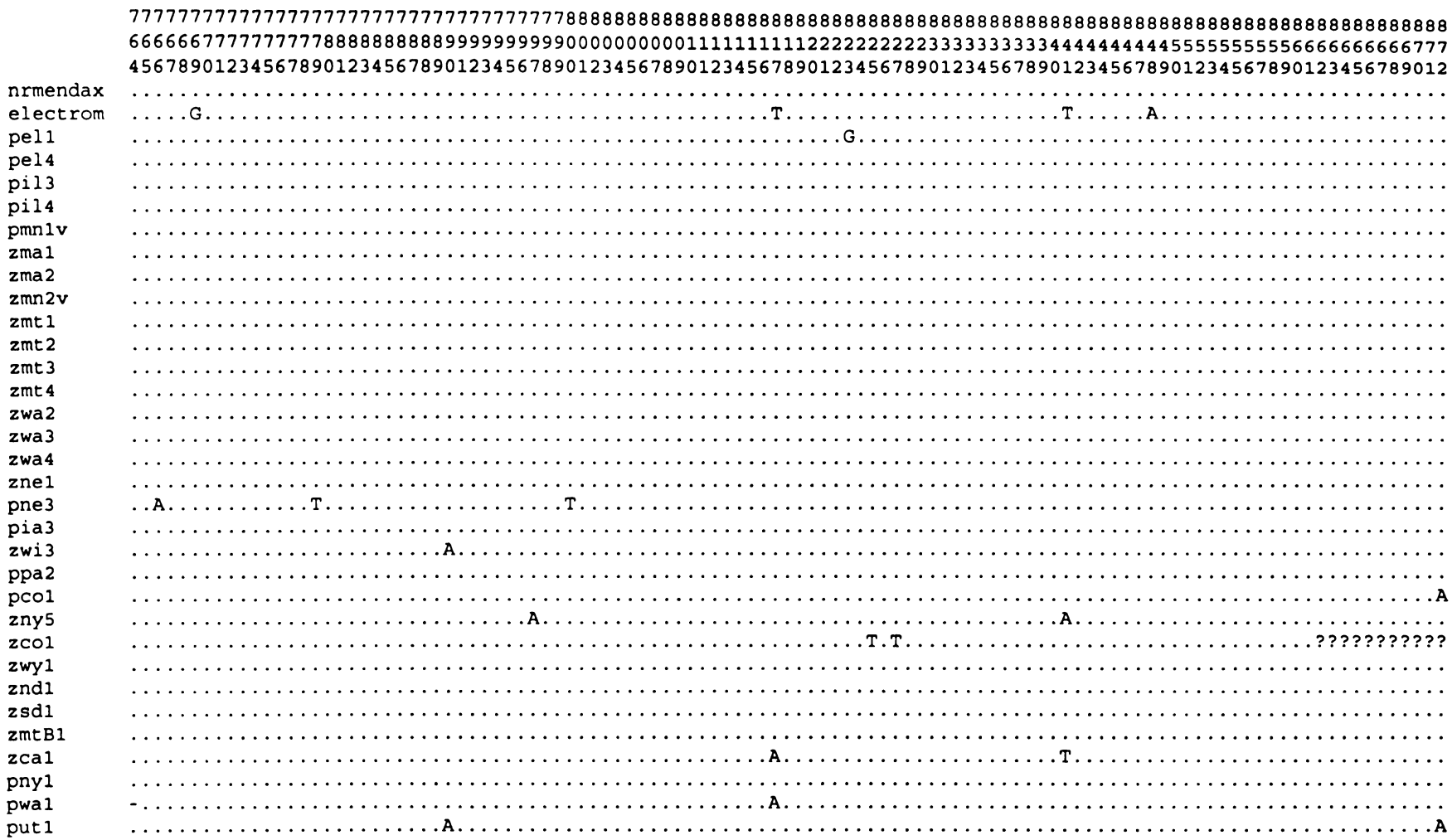

45da.2
45ma.4
aap1.1
aap2.3
bop1.2
fa.8
fp.11
m4.1
m5.6
mxhd1.4
tzt10.1
nrpomMX1
nrpomMX
pmi2
pns
nrmendaxG
pmn1
pmn2
zpa
zmn1
zon1
zon2
zma
mmi1
mns
mga
sparkga
mon2
mmi2
mmi3
mmi4
bop2.4
c1.7
cng2.2
fu1.1
pmhbj.5
sv2.2
t29a1.2
juniperin

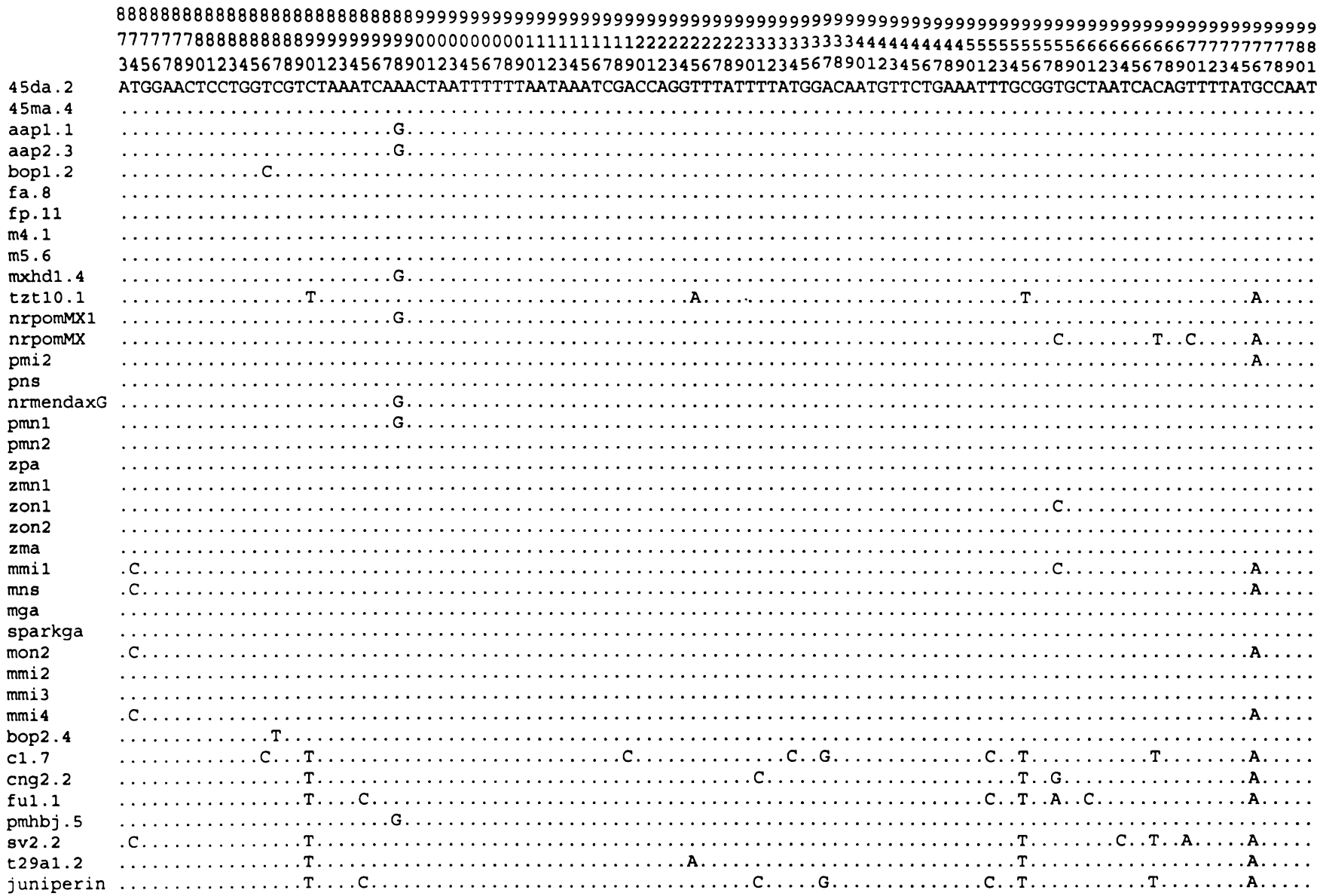

nrmendax
electrom
pel1
pel4
pil3
pil4
pmn1v
zma1
zma2
zmn2v
zmt1
zmt2
zmt3
zmt4
zwa2
zwa3
zwa4
zne1
pne3
pia3
zwi3
ppa2
pco1
zny5
zco1
zwy1
znd1
zsd1
zmtB1
zca1
pny1
pwa1
put1

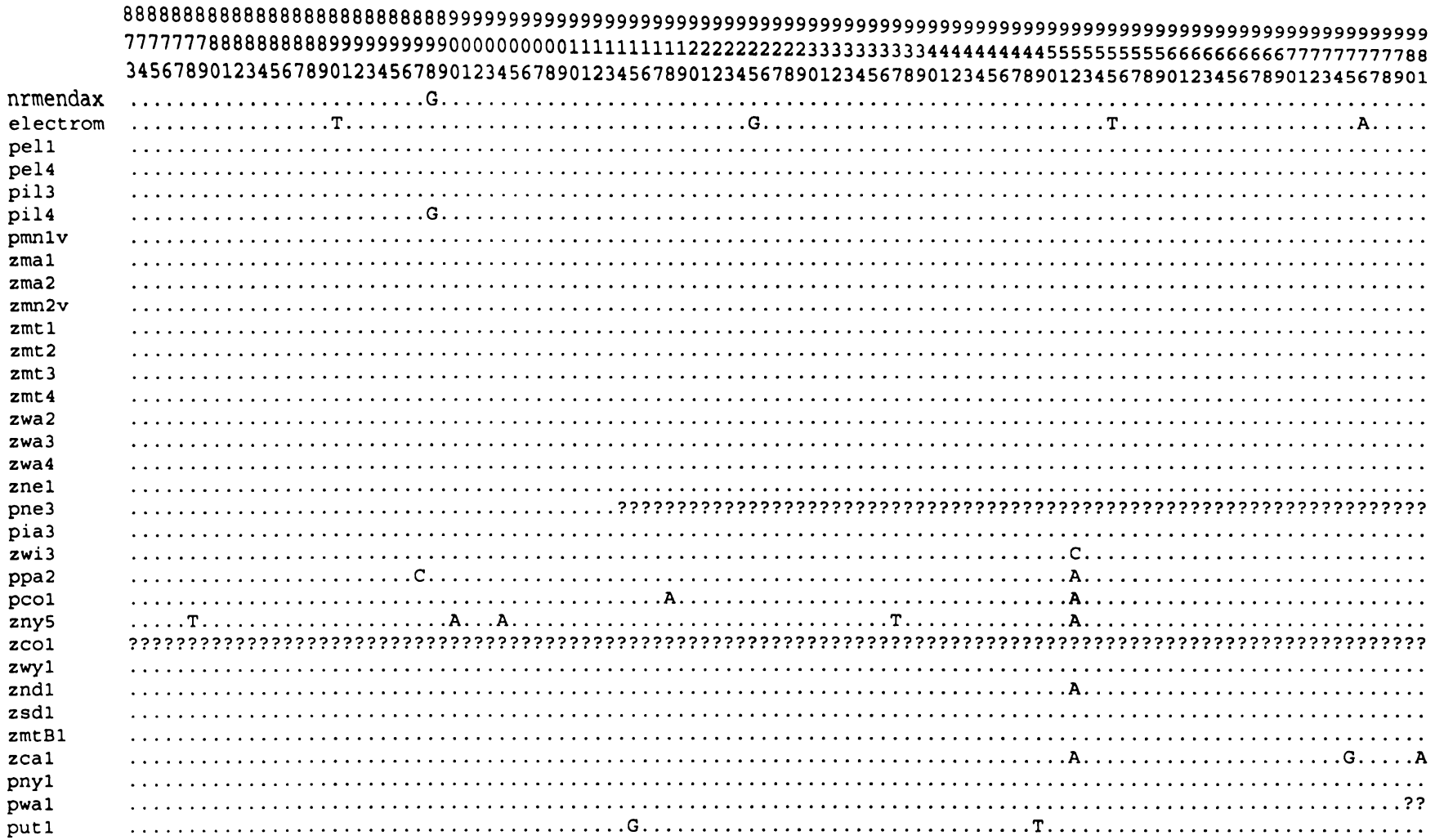
```
           6666666666666666666666666666666666666666666667777777777777777777777777777777777777777777777777777777777777777
           5555566666666667777777777888888888899999999990000000000111111111122222222223333333333444444444455555555556666
           5678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123
45da.2     TGGGCATCAATGATATTGAAGTTATGAATATTCAGATTTCATAAATGTAGAATTTGATTCTTATATAATTCCAACAAATGAATTAACTACAGACGGATTTCGATTATTA
45ma.4     .............................................................................................................
aap1.1     .............................................................................................................
aap2.3     .............................................................................................................
bop1.2     ....................................................................................................G........
fa.8       .............................................................................................................
fp.11      .............................................................................................................
m4.1       .............................................................................................................
m5.6       .............................................................................................................
mxhd1.4    .............................................................................................................
tzt10.1    ...A....................C..............T..............................................G......T........G......
nrpomMX1   .............................................................................................................
nrpomMX    ...A..........................C........T...........G.........................................................
pmi2       .............................................................................................................
pns        .............................................................................................................
nrmendaxG  .............................................................................................................
pmn1       .............................................................................................................
pmn2       .............................................................................................................
zpa        .............................................................................................................
zmn1       .............................................................................................................
zon1       .............................................................................................................
zon2       .............................................................................................................
zma        .............................................................................................................
mmi1       .............................................................................................................
mns        .............................................................................................................
mga        ...A.........................................................................................................
sparkga    ...A.........................................................................................................
mon2       .............................................................................................................
mmi2       .............................................................................................................
mmi3       .............................................................................................................
mmi4       .............................................................................................................
bop2.4     .............................................................................................................
c1.7       ............T..C...............C.......A........................T.......................T....T...............
cng2.2     .......................................T..........................G....................GGT....CA............G
ful.1      ...A....................C..............T.............................C..................G....T...............
pmhbj.5    .............................................................................................................
sv2.2      ......C................................T.C....A........................................AG....TCA.............
t29a1.2    ...A....................C..............T..............................................G......T........G......
juniperin  ...A...........C........C.....C........T.....................................................T............C..
```

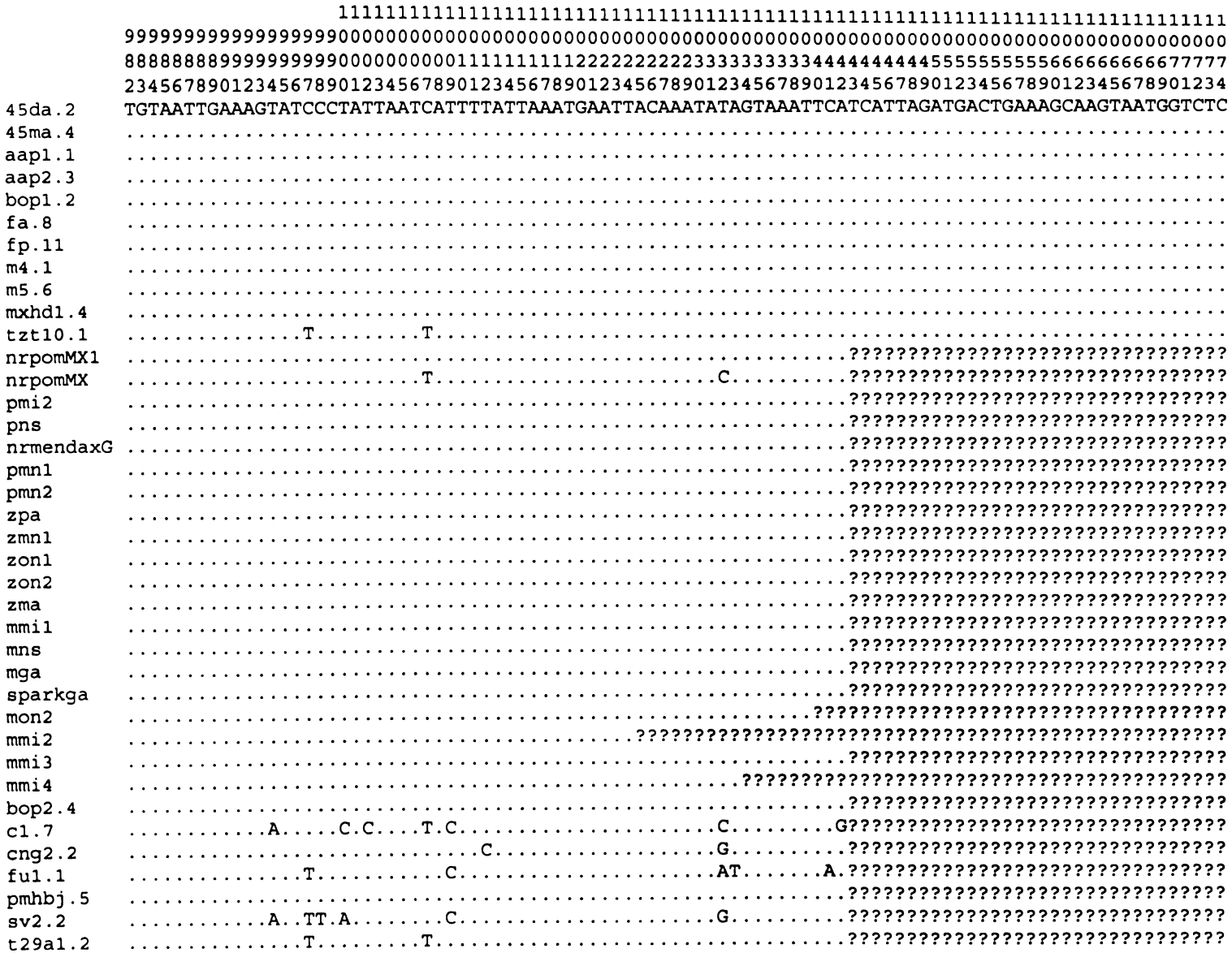

6666666666666666666666666666666666666666666666677777777777777777777777777777777777777777777777777777
5555566666666667777777777888888888899999999900000000001111111111222222222233333333334444444444555555555566666
5678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123
nrmendax
electrom ...A.....C.....T.....GA.....T.....
pel1
pel4
pil3
pil4 ...A.....
pmn1v
zma1
zma2
zmn2v
zmt1
zmt2
zmt3
zmt4
zwa2
zwa3 .....G.....
zwa4
zne1
pne3 .....G.....
pia3
zwi3 .....G.....
ppa2 .....G.....
pco1 .....C.....C.....C.....
zny5 .....T.....G.....
zco1
zwy1
znd1 .....T.....T.....C.....
zsd1 .....C.....G.....
zmtB1 .....G.....
zca1
pny1 ......A.....A.....
pwa1 .....T.....A.....----
put1 .....T.....A.....

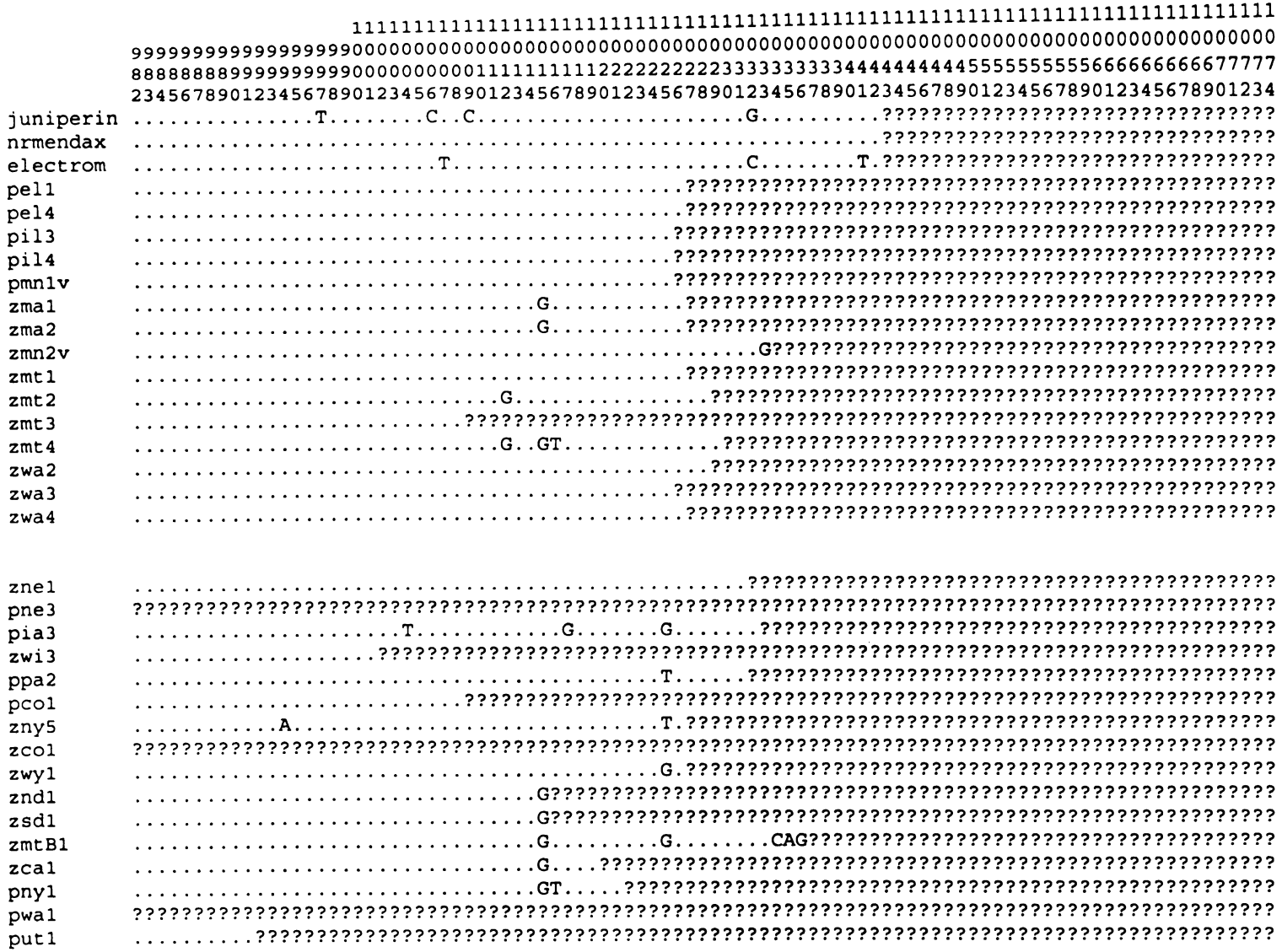

45da.2 GATGTAGATAATCGAGTAATTTTACCTATAAATTCTCAAATTCGAATTTTAGTGACTGCAGCAGATGTAATTCATTCCTGAACAGTTCCTTCTTTAGGTGTAAAAGTTG
45ma.4
aap1.1
aap2.3
bop1.2
fa.8
fp.11
m4.1
m5.6
mxhd1.4
tzt10.1
nrpomMX1
nrpomMX
pmi2
pns
nrmendaxG
pmn1
pmn2
zpa
zmn1
zon1
zon2
zma
mmi1
mns
mga
sparkga
mon2
mmi2
mmi3
mmi4
bop2.4
c1.7
cng2.2
fu1.1
pmhbj.5
sv2.2
t29a1.2
juniperin


nrmendax
electrom
pel1
pel4
pil3
pil4
pmn1v
zma1
zma2
zmn2v
zmt1
zmt2
zmt3
zmt4
zwa2
zwa3
zwa4
zne1
pne3
pia3
zwi3
ppa2
pco1
zny5
zco1
zwy1
znd1
zsd1
zmtB1
zca1
pny1
pwa1
put1


45da.2 ATGGAACTCCTGGTCGTCTAAATCAAACTAATTTTTTAATAAATCGACCAGGTTTATTTTATGGACAATGTTCTGAAATTTGCGGTGCTAATCACAGTTTTATGCCAAT
45ma.4
aap1.1
aap2.3
bop1.2
fa.8
fp.11
m4.1
m5.6
mxhd1.4
tzt10.1
nrpomMX1
nrpomMX
pmi2
pns
nrmendaxG
pmn1
pmn2
zpa
zmn1
zon1
zon2
zma
mmi1
mns
mga
sparkga
mon2
mmi2
mmi3
mmi4
bop2.4
c1.7
cng2.2
ful.1
pmhbj.5
sv2.2
t29a1.2
juniperin


888888888888888888888888889999999999999999999999999999999999999999999999999999999999999999999999999999999999999999
777777788888888889999999999000000000011111111112222222222333333333344444444445555555555666666666677777777788
345678901234567890123456789012345678901234567890123456789012345678901234567890123456789012345678901234567890123
nrmendax .....G.....
electrom .....T.....G.....T.....A.....
pel1 .....
pel4 .....
pil3 .....
pil4 .....G.....
pmn1v .....
zma1 .....
zma2 .....
zmn2v .....
zmt1 .....
zmt2 .....
zmt3 .....
zmt4 .....
zwa2 .....
zwa3 .....
zwa4 .....
zne1 .....
pne3 .....??
pia3 .....
zwi3 .....C.....
ppa2 .....C.....A.....
pco1 .....A.....A.....
zny5 .....T.....A..A.....T.....A.....
zco1 ??
zwy1 .....
znd1 .....A.....
zsd1 .....
zmtB1 .....
zca1 .....A.....G.....A
pny1 .....
pwa1 .....??
put1 .....G.....T.....


45da.2 TGTAATTGAAAGTATCCCTATTAATCATTTTATTAAATGAATTACAAATATAGTAAATTCATCATTAGATGACTGAAAGCAAGTAATGGTCTC
45ma.4
aap1.1
aap2.3
bop1.2
fa.8
fp.11
m4.1
m5.6
mxhd1.4
tzt10.1
nrpomMX1
nrpomMX
pmi2
pns
nrmendaxG
pmn1
pmn2
zpa
zmn1
zon1
zon2
zma
mmi1
mns
mga
sparkga
mon2
mmi2
mmi3
mmi4
bop2.4
c1.7
cng2.2
fu1.1
pmhbj.5
sv2.2
t29a1.2


juniperin
nrmendax
electrom
pel1
pel4
pil3
pil4
pmn1v
zma1
zma2
zmn2v
zmt1
zmt2
zmt3
zmt4
zwa2
zwa3
zwa4
zne1
pne3
pia3
zwi3
ppa2
pco1
zny5
zco1
zwy1
znd1
zsd1
zmtB1
zca1
pny1
pwa1
put1

REFERENCES

Albertson RC, Markert JA, Danley PD, Kocher TD (1999): Phylogeny of a rapidly evolving clade: The cichlid fishes of Lake Malawi, East Africa. Proc Natl Acad Sci USA 96: 5107-5110

AliNiazee MT, Brunner JF (1986): Apple maggot in the western United States: A review of its establishment and current approaches to management. J Entomol Soc Brit Columbia 83: 49-53

Alphen JJM van, Visser ME (1990): Superparasitism as an adaptive strategy for insect parasitoids. Ann Rev Entomol 35: 59-79

Avise JC (1994): Molecular markers, natural history and evolution. Chapman & Hall, NY.

Avise JC (2000): Phylogeography: The history and formation of species. Harvard University Press, Cambridge, MA

Avise JC, Arnold J, Ball RM Jr., Bermingham E, Lamb T, Neigel JE, Reeb A, Saunders NC (1987): Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Ann Rev Ecol Syst 18: 489-522.

Barraclough TG, Vogler AP (2000): Detecting the geographic pattern of speciation from species-level phylogenies. Am Nat 155: 419-434

Berlocher SH (1995): Population structure of *Rhagoletis mendax,* the blueberry maggot. Heredity 74: 542-555.

Berlocher SH (1998): Can sympatric speciation via host or habitat shift be proven from phylogenetic and biogeographic evidence? In: Howard DJ, Berlocher SH (eds): Endless Forms. Oxford University Press, New York. pp 99-113.

Berlocher SH (2000): Radiation and divergence in the *Rhagoletis pomonella* species group: Inferences from allozymes. Evolution 54: 543-557

Berlocher SH, Bush GL (1982): An electrophoretic analysis of *Rhagoletis* (Diptera: Tephritidae) phylogeny. Syst Zool 31: 136-155.

Berlocher SH, McPheron BA, Feder JL & Bush GL (1993): Genetic differentiation at allozyme loci in the *Rhagoletis pomonella* (Diptera: Tephritidae) species complex. Ann Entomol Soc Amer 86: 716-727.

Bernardi G, Sordino P, Powers DA (1993): Concordant mitochondrial and nuclear DNA phylogenies for populations of the teleost fish *Fundulus heteroclitus*. Proc Natl Acad Sci USA 90: 9271-9274.

Bierbaum TJ, Bush GL (1990): Genetic differentiation in the viability of sibling species of *Rhagoletis* fruit flies on host plants, and the influence of reduced hybrid viability on reproductive isolation. Entomol Exp Appl 55: 105-118.

Bogdanowicz SM, Wallner WE, Bell J, Odell TM, Harrison RG (1993): Asian Gipsy Moths (Lepidoptera: Lymantriidae) in North America: Evidence from molecular data. Ann Entomol Soc Am 86: 710-715

Brigneti G, GarciaMas J, Baulcombe DC (1997): Molecular mapping of the potato virus Y resistance gene Ry(sto) in potato. Theor Appl Genet 94: 198-203

Britten RJ (1986): Rates of DNA sequence evolution differ between taxonomic groups. Science 231: 1393-1398

Brower AVZ (1994): Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. Proc Natl Acad Sci USA 91: 6491-6495

Brown JM, Abrahamson WG, Way PA (1996): Mitochondrial DNA phylogeny of the goldenrod ball gallmaker, *Eurosta solidaginis* (Diptera: Tephritidae). Evolution 50: 777-786

Brown JM, McPeek MA, May ML (2000): A phylogenetic perspective on habitat shifts and diversity in the north American *Enallagma* damselflies. Syst Biol 49: 697-712

Brown WL Jr. (1957): Centrifugal speciation. Quart Rev Biol 32: 247-277

Brown WM, George M Jr., Wilson AC (1979): Rapid evolution of animal mitochondrial DNA. Proc Acad Natl Sci USA 76: 1967-1971

Burton RS, Lee B-N (1994): Nuclear and mitochondrial gene genealogies and allozyme polymorphism acros a major phylogeographic break in the copepod *Tigriopus californicus*. Proc Natl Acad Sci USA 91: 5197-5201.

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

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

Bush GL (1975): Modes of animal speciation. Ann Rev Ecol Syst 6: 339-363

Bush GL (1992): Host race formation and sympatric speciation in *Rhagoletis* fruit flies (Diptera: Tephritidae). Psyche 99: 335-357

Bush GL (1994): Sympatric speciation in animals: New wine in old bottles. Trends Ecol Evol 9: 285-288

Bush GL, Smith JJ (1998): The genetics and ecology of sympatric speciation: A case study. Res Pop Ecol 40: 175-187.

Caillaud MC, Via S (2000): Specialized feeding behavior influences both ecological specialization and assortative mating in sympatric host races of pea aphids. Am Nat 156: 606-621

Cardoso SRS, Eloy NB, Provan J, Cardoso MA, Ferreira PCG (2000): Genetic differentiation of *Euterpe edulis* Mart. populations estimated by AFLP analysis. Mol Ecol 9: 1753-1760

Caroll SP, Dingle H, Klassen SP (1997): Genetic differentiation of fitness associated traits among rapidly evolving populations of the soapberry bug. Evolution 51: 1182-1188

Castiglioni P, Ajmone-Marsan P, van Wijk R, Motto M (1999): AFLP markers n a molecular linkage map of maize: Codominant scoring and linkage group distribution. Theor Appl Gen 99: 425-431

Catling PM, Catling VR (1993): Floristic composition, phytogeography and relationships of prairies, savannas and sand barrens along the Trent River, Eastern Ontario. Can Field Nat 107: 24-45

Charlesworth B, Morgan MT, Charlesworth D (1993): The effect of deleterious mutations on neutral molecular variation. Genetics 134: 1289-1303

Cho YG, Blair MW, Panaud O, McCouch SR (1996): Cloning and mapping of variety-specific rice genomic DNA sequences: Amplified fragment length polymorphisms (AFLP) from silver-stained polyacrylamide gels. Genome 39 (2): 373-378

Chown SL, Smith VR (1993): Climate change and the short-term impact of feral house mice at the sub-Antarctic Prince Edward Islands. Oecologia 96: 508-516

Clark AG (1997): Neutral behavior of shared polymorphism. Proc Natl Acad Sci USA 94: 7730-7734

Coyne JA (1992): Genetics and speciation. Nature 355: 511-515

Danley PD, Markert JA, Arnegard ME, Kocher TD (2000): Divergence with gene flow in the rock-dwelling Cichlids of Lake Malawi. Evolution 54: 1725-1737

Doak P (2000): Habitat patchiness and the distribution, abundance, and population dynamics of an insect herbivore. Ecology 81: 1842-1857

Doebeli M, Dieckmann U (2000): Evolutionary branching and sympatric speciation caused by different types of ecological interactions. Am Nat 156: S77-S101

Emelianov I, Mallet J, Baltensweiler W (1995): Genetic differentiation in *Zeiraphera diniana* (Lepidoptera: Tortricidae), the larch budmoth: polymorphism, host races or sibling species. Heredity 75: 416-424.

Estoup A, Solignac M, Cornuet JM, Goudet J, Scholl A (1996): Genetic differentiation of continental and island populations of *Bombus terrestris* (Hymenoptera: Apidae) in Europe. Mol Ecol 5: 19-31

Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS (1998): Investigation of the bottleneck leading to the domestication of maize. Proc Natl Acad Sci USA 95: 4441-4446

Feder JL, Berlocher SH, Opp SB (1997): Sympatric host-race formation and speciation in *Rhagoletis* (Diptera: Tephritidae): a tale of two species for Charles D. In Mopper S, Strauss SY, eds. Genetic structure and local adaptation in natural insect populations: Effects of ecology, life history and behavior. Chapman & Hall, New York. pp. 408-441

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

Feder JL, Chilcote CA & Bush GL (1988): Genetic differentiation between sympatric host races of the apple maggot fly *Rhagoletis pomonella*. Nature 336: 61-64.

Feder JL, Filchak KE (1999): It's about time: The evidence for host plant-mediated selection in the apple maggot fly, *Rhagoletis pomonella*, and its implications for fitness trade-offs in phytophagous insecs. Entomol Exp Appl 91: 211-225

Feder JL, Roethele JB, Wlazlo B, Berlocher SB (1997): Selective maintenenace of allozyme differences between sympatric host races of the apple maggot fly. Proc Natl Acad Sci USA 94: 11417-11424.

Feder JL, Williams SM, Berlocher SH, McPheron BA, Bush GL (1999): The population genetics of the apple maggot fly, *Rhagoletis pomonella* and the snowberry maggot, *R. zephyria*: implications for models of sympatric speciation. Entomol Exp Appl 90: 9-24.

Felsenstein J (1981): Skepticism toward Santa Rosalia, or why are there so few kinds of animals? Evolution 35: 124-138

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

Filchak KE, Feder JL, Roethele JB, Stolz U (1999): A field test for host-plant dependent selection on larvae of the apple maggot fly, *Rhagoletis pomonella*. Evolution 53: 187-200

Foote RH, Blanc FL, Norrbom AL (1993): Handbook of the Fruit Flies (Diptera: Tephritidae) of America North of Mexico. Cornell University Press, Ithaca, NY

Frey JE, Feder JL, Palm J, Bush GL (1998): Differences in the electroantenal responses of apple and hawthorn-infesting races of *Rhagoletis pomonella* to host fruit volatile compounds. Chemoecology 8: 175-186

Fry JD (1996): The evolution of host specialization: Are trade-offs overrated? Am Nat 148: S84-S107

Fu Y-X, Li W-H (1993): Statistical tests of neutrality of mutations. Genetics 133: 693-709

Funk DJ (1999): Molecular systematics of cytochrome oxidase I and 16S from *Neoclamisus* leaf beetles and the importance of sampling. Mol Biol Evol 16: 67-82.

Funk DJ, Futuyma DJ, Ortí G, Meyer A (1995): Mitochondrial DNA sequences and multiple data sets: a phylogenetic study of phytophagous beetles (Chrysomelidae: *Ophraella*). Mol Biol Evol 12: 627-640.

Futuyma DJ (1997): Evolutionary Biology, 3rd ed. Sinauer Associates, Inc, Sunderland, MA.

Futuyma DJ, Moreno G (1998): The evolution of ecological speciation. Annu Rev Ecol Syst 19: 207-234

Gavrilets S, Li H, Vose M (2000): Patterns of parapatric speciation. Evolution 54: 1126-1134

Geritz SAH, Kisdi E, Meszena G, Metz JAJ (1998): Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. Evol Ecol 12: 35-57

Grant PR, Abbott I, Schluter D, Curry RL, Abbott LK (1985): Variation in the size and shape of Darwin's finches. Biol J Linn Soc 25: 1-39

Gilbert OL (1995): *Symphoricarpos albus* (L.) S.F. Blake (*S. rivularis* Suksd., *S. racemosus* Mihaux). J Ecol 83: 159-166

Guldemond JA, Mackenzie A (1994): Sympatric speciation in aphids. I. Host race formation by escape from gene flow. In: Leather SR,Walters KFA, Mills NJ, Watt AD (eds): Individuals, populations and patterns in ecology. Intercept, Andover, Hampshire, pp 367-378.

Han HY, McPheron, BA (1997): Molecular phylogenetic study of Tephritidae (Insecta : Diptera) using partial sequences of the mitochondrial 16S ribosomal DNA. Mol Phylogenet Evol 7: 17-32

Hare MP, Avise JC (1998): Population structure in the American oyster as inferred by nuclear gene genealogies. Mol Biol Evol 15: 119-128.

Harrison S (1991): Local extinction in a metapopulation context: an empirical evaluation. Biol J Linn Soc 42: 73-88

Hatfield T, Schluter D (1999): Ecological speciation in sticklebacks: Environment-dependent hybrid fitness. Evolution 53: 866-873.

Hendry AP, Kinnison MP (1999): Perspective: The pace of modern life: Measuring rates of contemporary microevolution. Evolution 53: 1637-1653

Hendry AP, Wenburg JK, Bentzen P, Volk EC, Quinn TP (2000): Rapid evolution of reproductive isolation in wild: Evidence from introduced salmon. Science 290: 516-519

Hillis DM, Mabel BK, Moritz C (1996): Applications of molecular systematics. In: Hillis D, Moritz C, Mable B (eds): Molecular Sysytematics. Sinauer, Sunderland, MA. pp 515-543

Hudson RR, Kreitman M, Aguade M (1987): A test of neutral molecular evolution based on nucleotide data. Genetics 116: 153-159

Jeffreys AJ, Wilson V, Thein SL (1985): Individual-specific "fingerprints" of human DNA. Nature 316: 76-79

Jenkins J (1996): Systematic studies of *Rhagoletis* and related genera. Ph.D. Thesis, Michigan State University.

Johnson PA, Hoppensteadt FC, Smith JJ, Bush GL (1996): Conditions for sympatric speciation: A diploid model incorporating habitat fidelity and non-habitat assortative mating. Evol Ecol 10: 187-205

Jones GN (1940): A monograph of the genus *Symphoricarpos*. J Arnold Arbor 21: 201-252

Jukes TH, Cantor CR (1969): Evolution of protein molecules. In: Munro NH (ed): Mammalian Protein Metabolism. Academic Press, New York. pp 21-132

Kawecki TJ (1996): Sympatric speciation driven by beneficial mutations. Proc R Soc Lond B, Biol Sci 263: 1515-1520

Keohavong P, Thilly WG (1989): Fidelity of DNA polymerases in DNA amplification. Proc Natl Acad Sci USA 86: 9253-9257

Kiester AR, Lande R, Schemske DW (1984): Models of coevolution and speciation in plants and their pollinators. Am Nat 124: 220-243

Kimura M (1968): Evolutionary rate at the molecular level. Nature 217: 624-626

Kimura M (1983): The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge, MA

Kimura M, Ohta T (1971): Protein polymorphism as a phase of molecular evolution. Nature 229: 467-469

Kitching IJ, Forey PL, Humphries CJ, Williams DM (1998): Cladistics: the Theory and Practice of Parsimony Analysis, 2nd ed. Oxford University Press, New York

Klein J (1986): Natural history of the Major Histocompatibility Complex. John Wiley, New York

Klein J, Takahata N, Ayala FJ (1993): MHC polymorphism and human origins. Sci Am 269: 78-83

Kliman RM, Andolfatto P, Coyne JA, Depaulis F, Kreitman M, Berry AJ, McCarter J, Wakeley J, Hey J (2000): The population genetics of the origin and divergence of the *Drosophila simulans* complex species. Genetics 156: 1913-1931

Kondrashov AS, Kondrashov FA (1999): Interactions among quantitative traits in the course of sympatric speciation. Nature 400: 351-354

Laird CD, McConaughy BL, McCarthy BJ (1969): Rate of fixation of nucleotide substitutions in evolution. Nature 224: 149-154

Li W-H (1997): Molecular evolution. Sinauer Associates, Sunderaland, MA

Liu ZJ, Li P, Kucuktas H, Nichols A, Tan G, Zheng XM, Argue BJ, Dunham RA (1999): Development of amplified fragment length polymorphism (AFLP) markers suitable for genetic linkage mapping of catfish. T Am Fish Soc 128: 317-327

Lenteren JC van (1981): The potential of entomophagous parasites for pest control. Ag Ecos Env 10: 143-158

Losos JB, Jackman TR, Larson A, de Queiroz K, Rodiguez-Schettino L (1998): Contingency and determinism in replicated adaptive radiations of island lizards. Science 279: 2115-2118

Losos JB, Warheit KB, Schoener TW (1997): Adaptive differeniation following experimental island colonization in *Anolis* lizards. Nature 387: 70-73

Lynch M, Crease TJ (1990): The analysis of population survey data on DNA sequence variation. Mol Biol Evol 7: 377-394

Lynch M, Milligan BG (1994): Analysis of population genetic structure with RAPD markers. Mol Ecol 3: 91-99

Maddison DR (1991): The discovery and importance of multiple islands of most-parsimonious trees. Syst Zool 40: 315-328

Maddison WP (1995): Phylogenetic histories within and between species. In: Hoch PC, Stephanson AG (eds): Experimental and molecular approaches to plant biosystematics. Monogr Syst Bot Mo Bot Gard 53, pp 273-287

Maddison WP (1997): Gene trees in species trees. Syst Biol 46: 523-536

Maddison DR, Maddison WP (2000): MacClade 4. Analysis of Phylogeny and Character Evolution. Sinauer Associates, Sunderland, MA

Mantel N (1967): The detection o disease clustering and a generalized regression approach. Cancer Res 27: 209-220

Martin AP, Palumbi SR (1993): Body size, metabolic rate, generation time, and the molecular clock. Proc Natl Acad Sci USA 90: 4087-4091

Masta SE (2000): Phylogeography of the jumping spider *Habronattus pugillis* (Aranea: Salticidae): Recent vicariance of sky island populations? Evolution 54: 1699-1711

Matter SF (1996): Interpatch movement of the red milkweed beetle, *Tetrapoes tetraophthalamus*: individual responses to patch size and isolation. Oecologia 105: 447-453

Mayr E (1963): Animal species and evolution. Harvard University Press, Cambridge, MA.

McDade L (1992): Hybrids and phylogenetic systematics II. The impact of hybrids on cladistic analysis. Evolution 46: 1329-1346

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

McPheron BA, Smith DC, Berlocher SH (1988): Genetic differences between *Rhagoletis pomonella* host races. Nature 336: 64-66.

McPheron BA, Han HY (1997): Phylogenetic analysis of North American Rhagoletis (Diptera : Tephritidae) and related genera using mitochondrial DNA sequence data. Mol Phylogenet Evol 7: 1-16.

Metz JAJ, Nisbet RM, Geritz SAH (1992): How should we define fitness for general ecological scenarios? Trends Ecol Evol 7: 198-202

Meyer A (1993): Phylogenetic relationships and evolutionary processes in East African cichlid fishes. Trends Ecol Evol 8: 279-284

Miller MP (1997): Tools for population genetic analyses (TFPGA) 1.3: A Windows program for the analysis of allozyme and molecular population genetic data. Computer software distributed by author. Available from http://www.public.asu.edu/~mmile8/index.html, accessed March 19, 2001

Mitter C, Farrell B, Futuyma DJ (1991): Phylogenetic studies of insect-plant interactions: Insights into the genesis of diversity. Trends Ecol Evol 6: 290-293.

Moericke V, Prokopy RJ, Berlocher SH, Bush GL (1975): Visual stimuli eliciting attraction of *Rhagoletis pomonella* (Diptera: Tephritidae) flies to trees. Entomol Exp Appl 18: 497-507

Moran P, Kornfield I (1993): Retention of an ancestral polymorphism in the Mbuna species flock (Teleostei, Cichlidae) of lake Malawi. Mol Biol Evol 10: 1015-1029

Mueller UG, Wolfenbarger LL (1999): AFLP genotyping and fingerprinting. Trends Ecol Evol 14: 389-394

Myashita NT, Kawabe A, Innan H (1999): DNA variation in the wild plant *Arabidopsis thaliana* revealed by amplified fragment length polymorphism analysis. Genetics 152: 1723-1731

Nagel L, Schluter D (1998): Body size, natural selection, and speciation in sticklebacks. Evolution 52: 209-218

Nei M (1972): Genetic distance between populations. Am Nat 106: 283-292

Nei M (1978): Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89: 583-590

Neigel JE, Avise JC (1986): Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. In: Karlin S, Nevo E (eds): Evolutionary processes and theory. Academic Press, New York. pp. 515-534

Nijman IJ, Bradley DG, Hanote O, Otsen M, Lenstra AJ (1999): Satellite DNA polymorphisms and AFLP correlate with *Bos indicus-taurus* hybridization. An Genet 30: 265-273

Orr MR, Smith TB (1998): Ecology and speciation. Trends Ecol Evol 13: 502-506

Palumbi SR, Baker CS (1994): Contrasting population structure from nuclear intron sequences and mtDNA of humpback whales. Mol Biol Evol 11: 426-435

Price PW (1977): General concepts on the evolutionary biology of parasites. Evolution 31: 405-420

Prokopy RJ (1972): Evidence for a marking pheromone deterring repeated oviposition in apple maggot flies. Env Entomol 1: 326-332

Prokopy RJ, Bennett EW, Bush GL (1971): Mating behavior in *Rhagoletis pomonella* (Diptera: Tephritidae). I. Site of assembly. Can Entomol 100: 318-329

Prokopy RJ, Bennett EW, Bush GL (1972): Mating behavior in *Rhagoletis pomonella* (Diptera: Tephritidae). II. Temporal organization. Can Entomol 103: 1405-1409

Prokopy RJ, Bush GL (1972): Mating behavior in *Rhagoletis pomonella* (Diptera: Tephritidae). III. Male aggregation in response to an arrestant. Can Entomol 104: 275-283

Prokopy RJ, Bush GL (1973): Mating behavior in *Rhagoletis pomonella* (Diptera: Tephritidae). IV. Courtship. Can Entomol 105: 873-891

Prokopy RJ, Diehl SR, Cooley SS (1988): Behavioral evidence for host races in *Rhagoletis pomonella* flies. Oecologia 76: 138-147

Prokopy RJ, Moericke V, Bush GL (1973): Attraction of apple maggot flies to odor of apples. Env Entomol 2: 743-749

Questiau S, Eybert MC, Taberlet P (1999): Amplified fragment length polymorphism (AFLP) markers reveal extra-pair parentage in a bird species: the bluethroat (*Luscinia svecica*). Mol Ecol 8: 1331-1339

Rademaker JLW, Hoste B, Louws FJ, Kersters K, Swings J, Vauterin L, Vauterin P, de Bruijn FJ (2000): Comparison of AFLP and rep-PCR genomic fingerprinting with DNA-DNA homology studies: *Xanthomonas* as a model system. Int J Syst Evol Micr 50: 665-677

Rambaut A (1996): Sequence Alignment Editor v 1.0. Available at http://evolve.zps.ox.ac.uk/software/Se-Al/main.html, accessed April 18, 2001

Randen AJ van, Roitberg BD (1996): The effect of egg load on superparasitism by the snowberry fly. Entomol Exp Appl 79: 241-245

Rausher MD (1984): The evolution of habitat preference in subdivided populations. Evolution 38: 596-608

Reineke A, Karlovsky P, Zebitz CPW (1998): Amplified fragment length polymorphism analysis of different geographic populations of the gypsy moth, *Lymantria dispar* (Lepidoptera: Lymatriidae). Bull Entomol Res 89: 79-88

Reznick DN, Shaw FH, Rodd FH, Shaw RG (1997): Evaluation of the rate of evolution in natural populations of guppies (*Poecilia reticulata*). Science 275: 1934-1937

Rice WR, Hostert EE (1993): Laboratory experiments on speciation: What have we learned in 40 years? Evolution 47: 1637-1653

Rice WR, Salt GW (1990): The evolution of reproductive isolation as a correlated character under sympatric conditions: experimental evidence. Evolution 44: 1140-1152

Roethele JB, Feder JL, Berlocher SH, Kreitman ME, Lashkari DA (1997): Toward a molecular genetic linkage map for the apple maggot fly (Diptera : Tephritidae): Comparison of alternative strategies. Ann Entomol Soc Amer 90: 470-479.

Rozas J, Rozas R (1999): DnaSP version 3: An integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15: 174-175

Schliewen U, Tautz D, Pääbo S (1994): Sympatric speciation suggested by monophyly of crater lake cichlids. Nature 368: 629-632.

Schluter D (1988): Character displacement and the adaptive divergence of finches on islands and continents. Am Nat 131: 799-824

Schluter D (1994): Experimental evidence that competition promotes divergence in adaptive radiation. Science 266: 798-801

Schluter D (1996): Ecological causes of adaptive radiation. Am Nat 148: S40-S64

Schluter D (1998): Ecological causes of speciation. In: Howard DJ, Berlocher SH (eds): Endless Forms. Oxford University Press, NY. pp. 114-129.

Simon C, Frati F, Bekenbach A, Crespi B, Liu H, Flook P (1994): Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation conserved polymerase chain-reaction primers. Ann Entomol Soc Am 87: 651-701

Smith DC (1986): Genetics and reproductive isolation of *Rhagoletis* flies. Ph. D. Thesis. University of Illinois at Urbana-Champaign.

Smith DC, Lyons SA, Berlocher SH (1993): Production and electrophoretic verifiction of F_1 hybrids between the sibling species *Rhagoletis pomonella* and *R. cornivora.* Entomol Exp App 69: 209-213

Smith JJ, Bush, GL (1997): Phylogeny of the genus *Rhagoletis* (Diptera : Tephritidae) inferred from DNA sequences of mitochondrial cytochrome oxidase II. Mol Phylogenet Evol 7: 33-43

Smith JJ, Bush GL (2000): Phylogeny of the subtribe Carpomyina (Trypetinae), emphasizing relationships of the genus *Rhagoletis*. In: Aluja M, Norrbom AL (eds): Fruit Flies (Tephritidae): Phylogeny and Evolution of Behavior. pp 187-217. CRC Press LLC

Soper JH, Heimburger ML (1994): Shrubs of Ontario. A life sciences miscellaneous publications of the Royal Ontario Museum, Toronto, ON

Sota T, Vogler AP (2001): Incongruence of mitochondrial and nuclear gene trees in the carabid beetles *Ohomopterus*. Syst Biol 50: 39-59

Stephan W, Wiehe THE, Lenz MW (1992): The effect of strongly selected substitutions on neutral polymorphism: Analytical results based on diffusion theory. Theor Pop Biol 41: 237-254

Sturmbauer C (1998): Explosive speciation in cichlid fishes of the African Great Lakes: a dynamic model of adaptive radiation. J Fish Biol 53: 18-36 Suppl. A

Sültmann H, Mayer WE, Figueroa F, Tuchy H, Klein J (1995): Ohylogenetic nalysis of Ccihlid fishes using nuclear DNA markers. Mol Biol Evol 12: 1033-1047

Swofford D L (1999): PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts

Tajima F (1989): Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585-595

Takahata N, Nei M (1985): Gene genealogy and variance of interpopulational nucleotide differences. Genetics 110: 325-344

Thompson DQ, Smith RH (1970): The forest primeval in the Northeast – a great muth? Proc Tall Timbers Fire Ecol Conf 10: 255-265

Travisano M, Rainey PB (2000): Studies of adaptive radiation using model microbial systems. Am Nat 156: S35-S44

Tregenza T, Pritchard VL, Butlin RK (2000): The origins of premating reproductive isolation: Testing hypothess in the grasshopper *Chorthippus parallelus*. Evolution 54: 1687-1698

U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fire Sciences Laboratory (2001, February). Fire Effects Information System, [Online]. Available at http://www.fs.fed.us/database/feis/ , accessed April 18, 2001

Via S, Bouck AC, Skillman S (2000): Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in parental environments. Evolution 54: 1626-1637

Voorrips RE, Jongerius MC, Kanne HJ (1997): Mapping of two genes for resistance to clubroot (*Plasmodiophora brassicae*) in a population of doubled haploid lines of *Brassica oleracea* by means of RFLP and AFLP markers. Theor Appl Genet 94: 75-82

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995): AFLP: A new technique for DNA fingerprinting. Nucl Acid Res 23: 4407-4414

Vuylsteke M, Mank R, Antonise R, Bastiaans E, Senior ML, Stuber CW, Melchinger AE, Lübberstedt T, Xia XC, Stam P, Zabeau M, Kuiper M (1999): Two high-density AFLP-linkage maps of Zea mays L.: Analysis of distribution of AFLP markers. Theor Appl Gen 99: 921-935

Wang RL, Wakeley J, Hey J (1997): Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. Genetics 147: 1091-1106

Wayne RK, Meyer A, Lehman N, Van Valkenburgh B, Kat PW, Fuller TK, Girman D, O'Brien SJ (1990): Large sequence divergence among mitochondrial DNA genotypes within populations of eastern African black-backed jackals. Proc Natl Acad Sci USA 87: 1772-1776

Weber JL, May PE (1989): Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Gen 44: 388-396

Weir B, Cockerham CC (1984): Estimating F-statistics for the analysis of population structure. Evolution 38: 1358-1370

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

White MD (1978): Modes of Speciation. WH Freeman, San Francisco.

Whitlock MC (1996): The Red Queen beats the jack-of-all-trades: The limitations on the evolution of phenotypic plasticity and niche breadth. Am Nat 148: S65-S77

Wilson AB, Noack-Kunnmann, Meyer A (2000): Incipient speciation in sympatric Nicaraguan crater lake cichlid fishes: sexual selection versus ecological diversification. Proc Roy Soc Lond B Biol Sci 268: 2133-2141

Wright S (1968-1978): Evolution and Genetics of Populations (4 volumes). University of Chicago Press, Chicago.

Zuckerkandl E, Pauling L (1965): Evolutionary divergence and convergence in proteins. Pp 97-166 in: V Bryson, HJ Vogel, eds. Evolving genes and proteins. Academic Press, New York