The Dissertation Committee for Brian David Vanden Heuvel
Certifies that this is the approved version of the following dissertation:

Molecular Systematics of *Cercocarpus* H.B.K. (Rosaceae)

Committee:

C. Randal Linder: Supervisor

Robert K. Jansen

Beryl B. Simpson

Ulrich Mueller

James Henrickson

Molecular Systematics of *Cercocarpus* H.B.K. (Rosaceae)

by

Brian David Vanden Heuvel, B.S.

Dissertation

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

The University of Texas at Austin
August, 2002

Dedication

For Traci, the love of my life, without which there is nothing

For Mom and Dad

Acknowledgements

First and foremost, I would like to thank my advisor Randy Linder. His help and direction during my graduate career has been superb. Beryl Simpson gave me great comments and support through this dissertation. Bob Jansen gave me great technical support and comments, plus was a good companion on trips to Big Bend. I would also like to thank Ulrich Mueller for his critical reviews of my dissertation. I thank James Henrickson for the trip to Mexico collecting, and his help in the herbarium. I also thank Su Dharmawardhane and David Begun for serving on my prelim committee.

I am indebted to those who helped me obtain leaf material for this dissertation, including James Henrickson, Randy Linder, Bob Jansen, Tim Lowery, Traci Freeman, Tim Chumley, Bonnie Crozier, the Nature Conservancy of Santa Catalina Island, and many state and national forests who gave me collecting permits.

I am grateful for the help of Tom Wendt and the staff of the Plant Resource Center of the University of Texas. I would also like to thank the

herbaria at the University of New Mexico and the University of Nevada-Reno for allowing me to view there collections.

I would like to thank the Mueller lab, including Rachelle Adams, for their help with ETS sequencing on the ABI 3100. Thanks to the Core Facility at the University of Texas for sequencing some of the early ETS amplifications. I would also like to thank Mario McDermitt for his help learning how to use the MJ BaseStation.

I am grateful to the Jansen Lab, particularly Bob Jansen, who taught me how to Southern Blot, Doug Goldman who reminded me how to Southern Blot, and Barbara Goodson for re-teaching me how to Southern Blot.

I am thankful for all of the wonderful support and encouragement from my fellow graduate students, including Andrea Weeks, Sarah Simmons, Jennifer Tate, Leah Larkin, Ruth Timme, Josh McDill, Anneke Schulman, Barbara Goodson, Romey Haberle, Tim Chumley, Mike Mayfield, Carolyn Ferguson, and Jan Barber.

I would finally like to thank my fiancé, Traci Freeman for her undying support through this dissertation process and my parents, who helped me get to where I am today.

This work was funded in part through grants from NSF (Doctoral Dissertation Improvement Award), the American Society of Plant Taxonomists, the Botanical Society of America, and Sigma Xi.

Molecular Systematics of *Cercocarpus* H.B.K. (Rosaceae)

Publication No.____________

Brian David Vanden Heuvel, Ph.D.
The University of Texas at Austin, 2002

Supervisor: C. Randal Linder

Cercocarpus H.B.K. (Rosaceae) is a New-World genus comprised of montane xerophytic shrubs and trees found in deserts, chaparral, and montane regions of western North America. *Cercocarpus* has been shown to be infected by *Frankia*, a nitrogen-fixing soil actinomycete. *Frankia* strains provide their hosts with a source of fixed nitrogen, a nutrient that often controls plant growth. This dissertation set out to create molecular phylogenies for *Cercocarpus* for later studies of coevolutionary relationships between *Cercocarpus* and *Frankia*. Only by knowing the evolutionary histories of both lineages can an in-depth study of coevolution proceed. *Cercocarpus* lacks a consensus classification or phylogeny. Past confusion about the number and circumscription of species within *Cercocarpus* arose from a lack of clear morphological discontinuities between taxa. Variations in leaf morphology show a continuum within and between taxa, making delineation of species difficult. Due to the variation in morphology, extensive sampling was conducted of all described taxa of *Cercocarpus*. The external transcribed spacer (ETS) of the nuclear ribosomal DNA was developed in *Cercocarpus* for the purpose of phylogenetic reconstruction. We document unexpected, deep coalescence of paralogous ETS types in *Cercocarpus*. This is the first record of the maintainence of ETS paralogues that were not produced by recent hybridization and lack of concerted evolution or pseudogene formation. We produced three phylogenies for *Cercocarpus* taxa, one based on ETS sequence data, another on AFLP presence/absence data, and a third based on chloroplast intergenic spacer sequences. Areas of congruence between the three trees supports clear taxonomic conclusions. Most notably, *C. montanus* is distinct from *C. betuloides*, the two newly described species of *Cercocarpus* have affinities to *C. fothergilloides*, and the endangered *C. traskiae* is more closely related to *C. betuloides* than *C. fothergilloides,* as previously thought. Taxon sampling had a great impact on tree topology, highlighting the need to do thorough sampling for closely related plant groups. We also discovered all three independent phylogenies were significantly incongruent, suggesting a complex, reticulating evolutionary history.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

GENERAL INTRODUCTION

Cercocarpus H.B.K. (Rosaceae) is a New World genus composed of montane xerophytic shrubs and trees found in deserts, chaparral, and montane regions of western North America. Considerable attention has been paid to the ecology and management of *Cercocarpus* because the evergreen leaves found on most members of the genus have high levels of protein and are important winter wildlife and livestock forage in the western United States and Mexico (Blauer et al. 1975, Davis and Brotherson 1991). The high level of protein found in leaves of *Cercocarpus* is due to an association with *Frankia*, a nitrogen-fixing soil actinomycete. *Frankia* strains provide their hosts with a source of fixed nitrogen, a nutrient that, when absent, limits plant growth. The host plant, in turn, provides fixed carbon to the *Frankia* strain (Baker and Schwintzer 1990). Much like legumes associated with *Rhizobia*, plants that associate with *Frankia* have a distinct advantage over other plants because their growth is not limited by the absence of nitrogen and can inhabit nitrogen poor soils. Therefore, *Cercocarpus* can play an important role in ecosystem development by securing unstable surfaces and promoting growth of other plants by adding nitrogen into the soil (Klemmedson 1979).

My current and future research interests are to study the coevolutionary history of the mutualistic association between *Cercocarpus* and *Frankia*. *Cercocarpus*, the largest genus of actinorhizal plants in the Rosaceae, is a group of reasonable size (approx. 10 species in 4 species complexes) and geographic distribution (western North America), making it ideally suited for the study of an actinorhizal symbiosis. One of the requirements of a coevolutionary study is to create phylogenetic hypotheses for both members in the mutualistic association. This dissertation addresses the construction of phylogenetic hypotheses for *Cercocarpus*, and will be crucial for a study of the coevolutionary relationships between *Cercocarpus* and *Frankia*.

The systematics of *Cercocarpus* are largely unknown; the genus lacks a consensus classification or phylogeny. Since it was first described, *Cercocarpus* has undergone three revisions (Schneider 1905b, Rydberg 1913, Martin 1950), each varying widely in the ranks assigned to taxa and the number of species. Confusion about the number and circumscription of species within *Cercocarpus* arises primarily from the lack of clear morphological delimitations between taxa (Schneider 1905b, Rydberg 1913, Martin 1950). Variations in leaf morphology show a continuum within and between taxa, making it difficult to define boundaries among species (Brayton and Mooney 1966, Mortenson 1973). In recent years, molecular data have proved useful for resolving relationships among morphologically similar species and for understanding the diversity within species (Soltis and Soltis 1998 and references therein).

This dissertation is composed of three independent chapters, each addressing a different aspect of the evolution and systematics of the genus *Cercocarpus,* using molecular characters. To facilitate discussion of evolutionary relationships in *Cercocarpus*, the second part of this introductory chapter will give a background overview of the taxonomic and systematic work on *Cercocarpus* that has been completed to date.

The remainder of the dissertation (Chapters 2-4) is organized as independent, interrelated papers. Each chapter will be submitted for publication as an independent paper, complete with introduction, materials and methods, results, and discussion.

The second chapter investigates the molecular evolution of the external transcribed spacer (ETS) of the nuclear ribosomal DNA (nrDNA) within *Cercocarpus*. The ETS region was initially developed for use as a phylogenetic marker in *Cercocarpus*. In order to create a phylogenetic tree for a recently diverged lineage, a DNA region must be identified that contains sufficient phylogenetically informative variation (Soltis and Soltis 1998). Preliminary studies found little informative variation in the internal transcribed spacer (ITS) region of the nuclear ribosomal DNA (nrDNA) repeat, a common nuclear marker used for phylogenetic reconstruction at low taxonomic levels in plants (Baldwin et al. 1995). We, therefore, developed the external transcribed spacer (ETS) of the nrDNA repeat for phylogenetic reconstruction in *Cercocarpus* to search for the variation we needed, as well as take advantage of the benefits (concerted evolution and easy amplification) that the ribosomal repeat displays (Nickrent and Soltis 1998).

Chapter two documents four main ETS paralogues that exhibit a deep coalescence in the genus. This was an unexpected result because concerted evolution should be homogenizing the rDNA copies. The observation of multiple rDNA copies is not unprecedented (Dvorak 1990, Suh et al. 1993, Buckler and Holtsford 1996, O'Kane et al. 1996, Waters and Schaal 1996, Wendel et al. 1995, Sang et al. 1995, and Campbell et al. 1997). Instances of multiple rDNA copies have been observed in many plant groups, and explained using hypotheses of hybridization and pseudogene formation (Buckler et al. 1997). Multiple ETS types have also been found in previous studies, yet the multiple ETS types coalesced within species, having little impact on the phylogenetic relationships reconstructed between species (Linder et al. 2000). This was not the case in *Cercocarpus*. We were unable to explain convincingly the multiple types found in *Cercocarpus* using recent hybridization or pseudogene formation. We discuss other possible reasons for the multiple paralogous ETS types with deep coalescence in *Cercocarpus*, the implications that multiple ETS types have for concerted evolution in the rDNA repeat, and hypothesize about the sequence of events that led to the current ETS diversity.

In the third chapter, we report two nuclear phylogenetic hypotheses for *Cercocarpus*. To address correctly previous hypotheses about species and varietal designations, we conducted dense taxon sampling across all recognized taxa (see second half of this introductory chapter). The first phylogenetic hypothesis used amplified fragment length polymorphisms (AFLP) as characters for reconstruction. The second hypothesis was obtained using one of the four ETS types identified in Chapter 2. We found the two phylogenies were significantly incongruent, even though we sampled the same individuals. We address the implications and possible reasons for the incongruency. We also test taxonomic and evolutionary hypotheses made by previous authors for *Cercocarpus* using both the ETS and AFLP trees. Finally, we found that the taxonomic sampling we had done was critical to a correct understanding of the evolution of *Cercocarpus*. Reduced datasets for both AFLP and ETS phylogenies had well supported incongruent tree topologies. This result is important beyond *Cercocarpus*, highlighting the need for dense taxon sampling in other phylogenetic studies in recently evolved plant lineages.

The fourth chapter presents a phylogeny based on chloroplast DNA sequence data. We did not find enough phylogenetically informative variation for the complete resolution of all *Cercocarpus* taxa. The chloroplast phylogeny was significantly incongruent to both the nuclear ETS and amplified fragment length polymorphisms (AFLP) phylogenies, and areas of congruence and incongruence are discussed.

Overall, this dissertation is a model for current limits and difficulties when attempting to reconstruct phylogenies for closely related groups of plants. Figures and Tables will be included in there respective chapters; however, references have been compiled into a single bibliography at the end of the dissertation, along with relevant appendices.

SYSTEMATIC BACKGROUND OF *CERCOCARPUS H.B.K.*

Cercocarpus was first collected near Mexico City by Humboldt and Bonpland and was described and illustrated by Kunth (1823). It was identified as a part of the Rosaceae based on the presence of a tubular calyx and a single one-celled ovary that produced an achene. Because of the presence of one, solitary ovule in the ovary and a lack of petals, it was also recognized as an undescribed genus. The first collection was named *Cercocapus fothergilloides* H.B.K.

During the next 70 years, many more collections were made and new species and varieties described. These included *Cercocarpus montanus* described by Rafinesque in 1832 and individuals currently named *C. betuloides*, collected in California by Nuttall in the 1830's.

There have been three formal revisions of *Cercocarpus*. A comparison of all three taxonomic treatments is illustrated in Figure 1.1. C. K. Schneider

(1905b) conducted the first formal revision of *Cercocarpus* using borrowed material from the Missouri Botanical Garden. Schneider recognized 10 species and 12 varieties (Figure 1.1). P. A. Rydberg re-evaluated *Cercocarpus* in 1913 in his treatment for the *Flora of North America*. Rydberg recognized 21 species and no varieties. Due to the criticism of Rydberg's revision of *Cercocarpus* by later authors, who considered many of the species described by Rydberg not worthy recognition, F. L. Martin, a Masters Student at Pomona College, followed with the third and last revision in 1950. Martin was able to view much more material that had been collected in the 37 years following Rydberg's revision. Martin recognized only six species and 11 varieties (Figure 1.1). Such a drastic reduction in recognized species left many botanists perplexed, preferring instead an intermediate number of species between the extremes of Rydberg's (1913) and Martin's (1950) revisions (Little 1953). Although Martin viewed more material in his revision, his classification has not been widely accepted as evidenced by its absence from current western states floras (Kearny and Peebles 1951, Munz 1959, Hitchcock and Cronquist 1961, Lis 1993).

Much of the disagreement between taxonomic treatments arises from a lack of characters and intergrading morphological variation between taxa. Martin (1950) addressed this difficulty when he noted on page 91:

> All the members of this genus are related very closely; intergradations of all distinguishing characteristics is the rule rather than the exception, and hybridization apparently occurs even between those entities sufficiently distinct to merit specific rank. A feature which seems characteristic of one entity often appears sporadically in other entities (91).

Martin's admission (1950) of a lack of characters and intergradations of forms was then used to justify his drastic reduction in the number of species in *Cercocarpus*. The key, then, for further studies in this group is to find new characters to use to describe taxa.

In order to help clarify the species delineations, Richard Lis (1992) approached *Cercocarpus* using leaf architecture characters. Based on this work, he proposed a re-organized classification that included four species complexes; a *ledifolius* species complex, a *betuloides* species complex, a *montanus* species complex, and a *fothergilloides* (Mexican) species complex. Lis (1993) followed this study by resurrecting the *C. betuloides* name and re-classifying the California taxa for the Jepson manual to reflect his view of the diversity

A list of the recognized *Cercocarpus* taxa used in this study can be found in Table 1.1 and Figure 1.1. The recognized taxa are organized following the species complexes hypothesized by Lis (1992). The following is a description of the diversity, past and current taxonomic entities, and geographic distribution of the four species complexes.

Table 1.1: A list of the recognized *Cercocarpus* taxa for this study. Included is the publication citation.

taxon	Place of publication
montanus complex	
Cercocarpus montanus Raf.	*Atl. J.* pg. 146. 1832.
C. montanus Raf. var. *montanus* F. L. Martin	*Brittonia* 7(2). pg. 101. 1950.
C. montanus Raf. var. *glaber* (S. Wats.)F.L. Martin	*Brittonia* 7(2). pg. 101. 1950.
C. montanus Raf. var. *paucidentatus* (S. Wats.)F.L. Martin	*Brittonia* 7(2). pg. 104. 1950.
C. montanus Raf. var. *argenteus* (Ryd.)F.L. Martin	*Brittonia* 7(2). pg. 104. 1950.
betuloides complex	
Cercocarpus betuloides Nutt.	Torr. and Gray, *Flora of North America.* pg. 427. 1840.
C. betuloides Nutt. var. *betuloides*	*Jepson Manual*, pg 950. 1993.
C. betuloides Nutt. var. *blancheae* (C.K.Schneid.)R.F. Thorne	*Aliso* 9(2). pg. 193. 1978.
Cercocarpus minutiflorus Abrams	*Bull. Torr. Bot. Club* 37. pg. 149. 1910.
fothergilloides complex	
Cercocarpus fothergilloides H.B.K	*Nov. Gen. And Sp*. 6. pg 232. 1832.
C. fothergilloides H.B.K. var. fothergilloides Henrickson	ined.
C. fothergilloides H.B.K. var. *medranoanus* Henrickson	ined.
C. fothergilloides H.B.K. var. *mojadensis* (C.K. Schneid.)Henrickson	ined.
Cercocarpus pringlei (C.K. Schneid.)Ryd.	Rydberg. *North American Flora* 22. pg 420. 1913
Cercocarpus macrophyllus C. K. Schneid.	*Ill. Handb. Laubholzk.* 1. pg 530. 1905a
Cercocarpus traskiae Eastw.	*Proc. Cal. Acad.* 3. pg. 136. 1898.
ledifolius complex	
Cercocarpus ledifolius Nutt.	Torr. and Gray, *Flora of North America.* pg. 427. 1840.
C. ledifolius Nutt. *var. intermontanus* N. Holmgren	*Brittonia* 39(4). pg 424. 1987.

C. ledifolius Nutt. var. *ledifolius*	*Brittonia* 39(4). pg 426. 1987.
Cercocarpus intricatus S. Wats.	*Proc. Amer. Acad. Arts* 10. pg 346. 1875
New taxa	
Cercocarpus mexicanus Henrickson	*Sys. Bot.* 12. pg. 293. 1987.
Cercocarpus rzedowskii Henrickson	*Sys. Bot.* 12. pg. 296. 1987.

Taxonomic background of the *Cercocarpus ledifolius* complex

Members of the *Cercocarpus ledifolius* complex are found distributed in the Great Basin (Figure 1.2), and can be easily identified by their leaves with entire margins (Figure 1.3). The type specimen for *Cercocarpus ledifolius* was collected in Caribou Co., Idaho, near the Bear River of the Timpanagos near the celebrated "Beer Springs' by Nuttall in July, 1834 (see Table 1.1). In 1875, Sereno Watson described *Cercocarpus intricatus* from a collection he made near the mouth of the American Fork Canyon, near Saint George, Utah. In his description he mentioned that *Cercocarpus intricatus* was probably closely allied to *Cercocarpus ledifolius*, but differed by its low, intricately branched growth and by it's smaller, tightly revolute leaves. Sixteen years later, Marcus E. Jones (1891b) subsumed *Cercocarpus intricatus* to a variety of *C. ledifolius* based on his observation that at the type locality, he found a complete transition in morphology between *C. ledifolius* and *C. intricatus* as he decreased in altitude. Also in 1891a, Jones described *C. arizonicus* from a collection in Coconino Co., Arizona, near Willow Springs. Jones recognized that *C. arizonicus* was very similar in morphology to *C. ledifolius var. intricatus*, but differed in flower size, length of persistent style on the achene, and the vestiture of the leaves.

In the first formal revision of the genus in 1905b, Schneider recognized the *Cercocarpus ledifolius* group, which included *Cercocarpus ledifolius* Nutt. and *C. intricatus* S.Wats.. Schneider (1905b) further described *Cercocarpus ledifolius* Nutt. var. *intercedens* C.K. Schneid. as a taxon with smaller, more revolute leaves that were intermediate between *C. ledifolius* and *C. intricatus*, and *Cercocarpus intricatus* S.Wats. var. *villosus* C.K.Schneid. that was similar to *C. intricatus*, yet had a much denser vestiture. *Cercocarpus arizonicus* was not recognized, but was used as a synonym of *C. intricatus* var. *villosus*.

Rydberg (1913) recognized *Cercocarpus ledifolius*, *C. intricatus*, and *C. arizonicus*. He further described *Cercocarpus hypoleucus* Ryd. from a collection near Melrose Montana, to replace *C. ledifolius* var. *intercedens*.

Martin (1950), in the last formal revision, recognized the *Cercocarpus ledifolius* complex as containing three varieties based mostly on leaf size. Individuals with the largest leaves were given the name *Cercocarpus ledifolius* Nutt. var. *typicus* C.K. Schneid. Individuals with intermediate sized leaves were described as *C. ledifolius* var. *intercedens* C.K. Schneid., and *C. ledifolius* var. *intricatus* referred to the small leafed individuals. Holmgren (1987), while preparing a treatment for the *Intermountain Flora*, recognized that the actual type for *Cercocarpus ledifolius* Nutt. was actually *C. ledifolius* var. *intercedens* from Schneider (1905b). This meant that the large-leafed taxon was not named at any rank. Holmgren therefore re-classified the *Cercocarpus ledifolius* complex to include two varieties: *C. ledifolius* var. *intermontanus* and *C. ledifolius* var. *ledifolius*, with variety *intermontanus* being the more common, large-leafed taxon. He did not include *Cercocarpus intricatus* as a variety of *ledifolius*, recognizing it as a distinct species. For this study, we will use the taxonomy from Holmgren's treatment, but have tested if there is indeed molecular support for his classification (Chapter 3 and 4). Holmgren's re-classification is illustrated in Figure 1.1.

Much of the leaf diversity (Figure 1.3) and intergradations in leaf morphology between the two *Cercocarpus ledifolius* varieties (variety *intermontanus* and variety *ledifolius*) and *C. intricatus* (Figure 1.4) has been attributed to hybridization. In the original description of the *Cercocarpus ledifolius* complex, Schneider (1905b) considered his variety *intercedens* to be a hybrid taxon between variety *intricatus* and variety *typicus*. Holmgren (1987) agreed with this view, recognizing that in zones of contact, intermediate forms could be observed. Brayton and Mooney (1966) conducted a detailed study of fifteen populations of *Cercocarpus ledifolus* that displayed leaf variation. They found that individuals of variety *intermontanus* were at higher altitudes, and individuals classified as *C. intricatus* were found at much lower, drier sites.

Although individuals with intermediate leaf morphologies (*C. ledifolius* var. *ledifolius*) occupied the intermediate elevations between the two "parental populations," leaf variation indicative of all three taxa was also found in the intermediate zone. Brayton and Mooney (1966) attributed the elevated degree of leaf variation to hybridization coupled with different selective conditions year-to-year, resulting in the establishment of different leaf types. We have also tested this assumption using molecular data (Chapter 3 and 4).

Taxonomic background of the *Cercocarpus montanus* complex

Rafinesque described *Cercocarpus montanus* in the 1832 from a collection by E. P. James who accompanied Long on his first expedition to Colorado during the summer of 1820. Nuttall later collected a member of the same taxon on the Platte River in Colorado and named it *Cercocarpus parvifolius* in 1839. Other early collections and species descriptions came from Asa Gray (1853), who named *Cercocarpus breviflorus* from a collection by Wright near Frontera, New Mexico, which may be Fronteras, Sonora, Mexico as commented by Rydberg (1913). *Cercocarpus breviflorus* was recognized as distinct from *C. montanus* and *C. parvifolius* by it's spiny, short habit and smaller leaves. Watson (1882) described a similar taxon to *Cercocarpus breviflorus*, who named it *Cercocarpus parvifolius* Nutt. var. *paucidentatus* from a collection in San Luis Potosi in Mexico. These collections had much smaller leaves then the original *C. parvifolius* collections. Jones (1891b) recognized the commonalities between

Cercocarpus parvifolius and *C. brevifolius*, and therefore subsumed *C. breviflorus* as a variety of *parvifolius* under the name *C. parvifolius* Nutt. var. *breviflorus* Colville ex. M.E. Jones.

Schneider (1905b), in the first formal revision, recognized both *C. parvifolius* Nutt. and *C. breviflorus* Gray. He furthur described two new varieties of *C. parviflorus* Nutt.: variety *hirsutissimus* C.K.Schneid. from a collection by Hayden in the Laramie Mountains, Wyoming, and variety *minimus* C.K.Schneid. from an Englemann collection from Utah. Rydberg (1913) later found that *C. parvifolius* var. *minimus* was actually a specimen of *Holodiscus microphyllus*. Schneider also described *C. breviflorus* Gray var. *eximus* C.K. Schneid. from a collection by Earle from the El Captain mountains, Lincoln Co., New Mexico. Schneider furthur described *Cercocarpus treleasei* from a collection in San Luis Potosi, Mexico. He did not recognize *C. parvifolius* var. *paucidentatus* S.Wats.

Rydberg (1913) recognized that *C. montanus* Raf. and *C. parvifolius* Nutt. were the same taxon, and used *C. montanus* in his Flora of North America. He also recognized that *C. treleasei* and *C. parvifolius* var. *paucidentatus* S.Wats. were the same taxon, elevating Watson's variety *paucidentatus* to specific rank. *C. breviflorus* Gray was recognized as being distinct, based on glabrate adaxial leaf surfaces, less tomentum on the abaxial leaf surface, and a more northerly distribution. Rydberg elevated *C. breviflorus* var. *eximus* to specific rank, and furthur described two new species: *Cercocarpus flabellifolius* Ryd. from a collection in southern Utah, and *Cercocarpus argenteus* from Randall Co., Texas.

Martin (1950) subsumed all of the afore-mentioned species and varieties under three varieties of *C. montanus*: *C. montanus* var. *montanus*, *C. montanus* var. *paucidentatus*, and *C. montanus* var. *argenteus* (Figure 1.1). Lis's (1992) view of the *Cercocarpus montanus* complex also included the three varieties as described by Martin (1950), yet he maintained that *Cercocarpus montanus* var. *paucidentatus* displayed sufficient differences in leaf morphology to be recognized at the species level. For this dissertation, we will refer to this taxon at the varietal rank. It differs from variety *montanus* and variety *argenteus* by having much smaller leaves with fewer marginal teeth. It is broadly distributed from Arizona across New Mexico into Texas and south into Mexico. Within this complex, Lis (1992) found *Cercocarpus montanus* var. *montanus* and *C. montanus* var. *argenteus* were very closely related. Both are winter deciduous shrubs, while all other members of *Cercocarpus* are evergreen. *C. montanus* var. *montanus* is broadly distributed from western South Dakota to northern New Mexico and from western Oklahoma to western Utah (Figure 1.6). *C. montanus* var. *argenteus* is found primarily in Palo Duro Canyon in the Texas panhandle, eastern New Mexico, and in the Guadalupe Mountains of west Texas (Figure 1.6).

A fourth taxon recognized as part of the *Cercocarpus montanus* complex for this study is *C. montanus* var. *glaber*, described from a collection from California. Martin (1950) recognized this taxon as a widespread Californian taxon. McVaugh (1952) first misused *Cercocarpus montanus* var. *glaber* (S. Wats)F.L. Martin and later by Correll and Johnston (1970) in the *Manual of the Vascular Plants of Texas*, to describe broad-leafed individuals with approximately the same geographic distribution as variety *paucidentatus*. Since Lis (1992) re-classified the Californian diversity under the name *C. betuloides,* there is no opportunity for confusion. For the sake of convenience, we will continue to use *C. montanus* var. *glaber* to refer to the taxon as McVaugh (1952) and Correll and Johnston (1970) intended and not use it to refer to a Californian taxon.

Because of the lack of discontinuities in leaf morphology between varieties in the *Cercocarpus montanus* complex, which is the primary source of characters for classification, the naturalness of the varieties are in question (Figure 1.7) (Chapter 2 and 3).

Taxonomic background of the *Cercocarpus betuloides* complex

Nuttall collected specimens of *Cercocarpus* from the mountains of Santa Barbara, California during the 1830's. Nuttall sent specimens both to Torrey and Gray, who named it *C. betuloides* Nutt. (*Flora of North America* 1840), and to Hooker, who named it *C. betulaefolius* Nutt. ex Hook., (*Icones plantarum* pl. 322 1840). Watson (1876) did not recognize *Cercocarpus betuloides* Nutt. or *C.*

betulaefolius Nutt. as displaying sufficient morphological differences to warrant specific status, and described *C. parvifolius* Nutt. var. *glaber* S.Wats.

Schneider (1905b) subsumed both these California taxa under one name, *C. betulaefolius*, with three varieties: variety *typicus* C.K. Schneid., variety *blancheae* C.K. Schneid., and variety *minor* C.K. Schneid. (Figure 1.1). Rydberg (1913) noted that *C. betuloides* had been published before *C. betulaefolius*, and recognized only *C. betuloides*. Rydberg did not recognize the varieties of Schneider, but instead described four new species: *C. alnifolius* Ryd. from a collection by Trask near Avalon, Santa Catalina Island; *C. douglasii* Ryd. from a collection by Douglas in California; *C. rotundifolius* Ryd. from a collection by Grant near Los Angelas; and *C. macrourus* Ryd. from a collection by Baker in Modoc Co., California. Additionally, Rydberg recognized *C. minutiflorus*, a species described by Abrams (1910) from San Diego (Figure 1.1).

In the last formal revision of *Cercocarpus*, Martin (1950) grouped all the described, broad-leaved Californian diversity into three varieties of *C. montanus* (Figure 1.1). This taxonomic move was rejected by many of the floras of the Pacific states, including the *Arizona Flora* (Kearny and Peebles 1951) and the *Flora of Southern California* (Munz 1979), all of which retained the name *C. betuloides*. After his study of leaf architecture characters, Lis (1992) recognized a close relationship between Martin's varieties *glaber*, *minutiflorus*, and *marcourus.* Because these three varieties are all located in California (Figure 1.8), Lis (1993) used the name *C. betuloides* and continued the tradition in Pacific states of recognizing *C. betuloides* as a separate entity than *C. montanus*.

For this dissertation, we will recognize the taxonomy of Lis (1993) for the California diversity of *Cercocarpus*. Lis delimited two species, *C. betuloides* with three varieties and *C. minutiflorus* (Figure 1.1, Figure 1.9).

Taxonomic background of the *Cercocarpus fothergilloides* complex

Schneider (1905a,b) described three new taxa similar to *Cercocarpus fothergilloides*: *C. mojadensis* C.K. Schneid., *C. mojadensis* var. *pringlei* C.K. Schneid., and *C. macrophyllus* C.K. Schneid. Rydberg (1913) raised the variety *pringlei* to specific rank, but made no more changes to the taxonomy of this group. Because of a lack of material, Martin (1950) did not reclassify this species complex.

After his study of leaf architectural characters, Lis (1992) reorganized the *fothergilloides* complex to contain *C. fothergilloides*, *C. pringlei, C. mojadensis*, *C. macrophyllus* and *C. montanus* var. *traskiae*. Martin (1950) had placed *C. traskiae* as a variety of *C. montanus* based on overlapping traits between *C. traskiae* and *C. montanus* var. *glaber* and variety *blancheae*. The close association of *C. montanus* var. *traskiae* with the other Mexican taxa based on leaf architecture led Lis (1993) to raise *C. traskiae* to the specific level in his treatment in the Jepson Manual.

James Henrickson (in prep.) is currently revising the *Cercocarpus fothergilloides* complex and his treatment can be seen in Figure 1.1. Henrickson recognizes four species, *C. fothergilloides, C. pringlei*, *C. macrophyllus*, and *C. traskiae*. *Cercocarpus fothergilloides* is described containing three varieties, including the variety *mojadensis*, a taxon that Schneider (1905b) described at the rank of species. Henrickson also recognizes *C. pringlei* with two varieties, but notes that there are strong overlaps in many characteristics with *C. fothergilloides*. *Cercocarpus macrophyllus* is recognized as a distinct species by a large number of characteristics, but particularly by distinctively large leaf blades. Henrickson continues to recognize *Cercocarpus traskiae* as a distinct species, separate from *C. montanus*. The geographic ranges for these taxa are found in Figures 1.10.

Cercocarpus rzedowskii **and** ***C. mexicanus***

Recently, two new species of *Cercocarpus*, *C. rzedowskii* and *C. mexicanus*, were recently described from northern Mexico (Figure 1.11) because of their very distinct leaf morphology (Henrickson 1987). Although Lis (1992) was unable to include these species into his analysis of *Cercocarpus*, Henrickson (1987) recognized similarity between *C. mexicanus* and the *montanus* complex, and noted *C. rzedowskii* had similarities to the *ledifolius* complex.

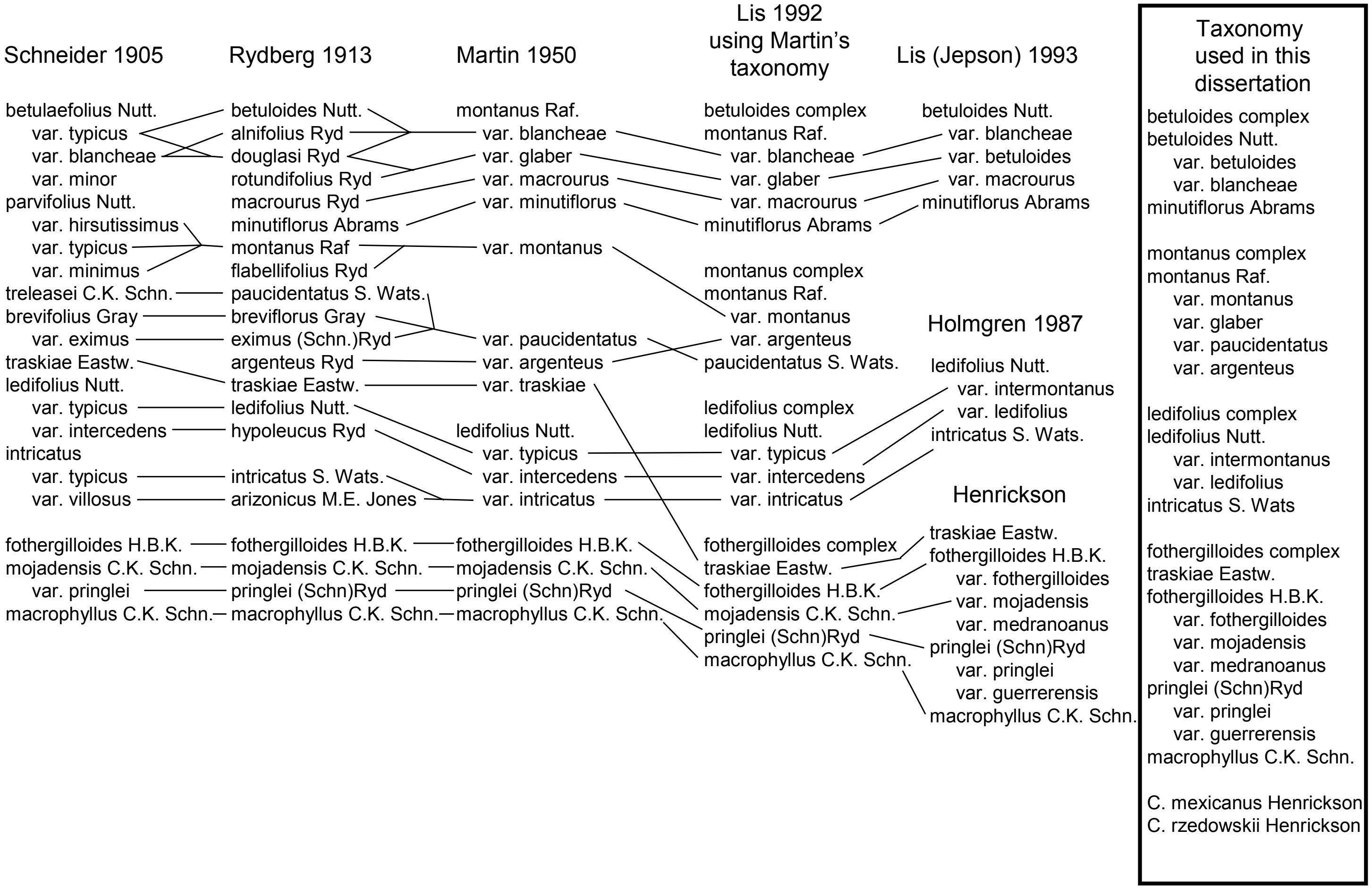
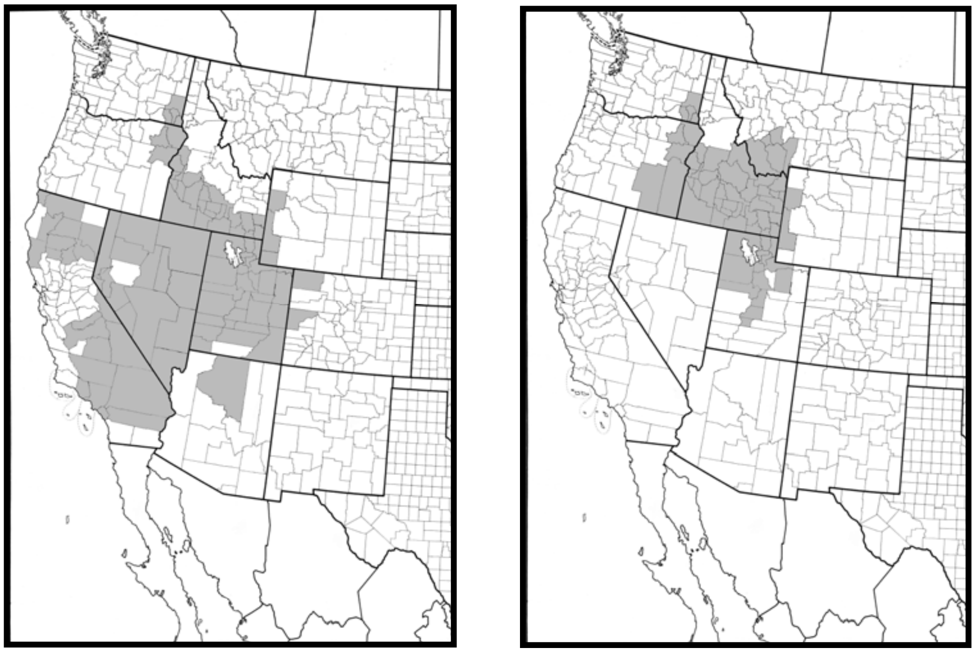
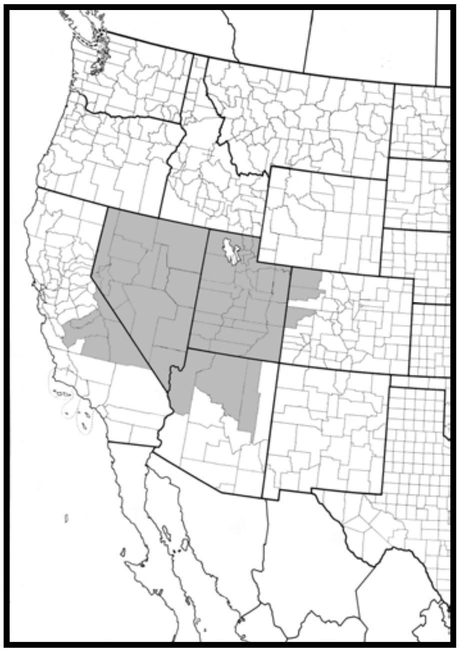


Figure 1.1. Representation of the different treatments of *Cercocarpus* H.B.K. (Rosaceae) and the taxonomy used in this dissertation, including the three complete revisions (Schneider 1905b, Rydberg 1913, and Martin 1950) and infrageneric revisons (Holmgren 1987, Lis 1993, and Henrickson in manuscript).



C. ledifolius var. *intermontanus* *C. ledifolius* var. *ledifolius*



C. intricatus

Figure 1.2. Geographic distribution of the *Cercocarpus ledifolius* complex

The *ledifolius* compex: Varieties of *C. ledifolius* and *C. intricatus*

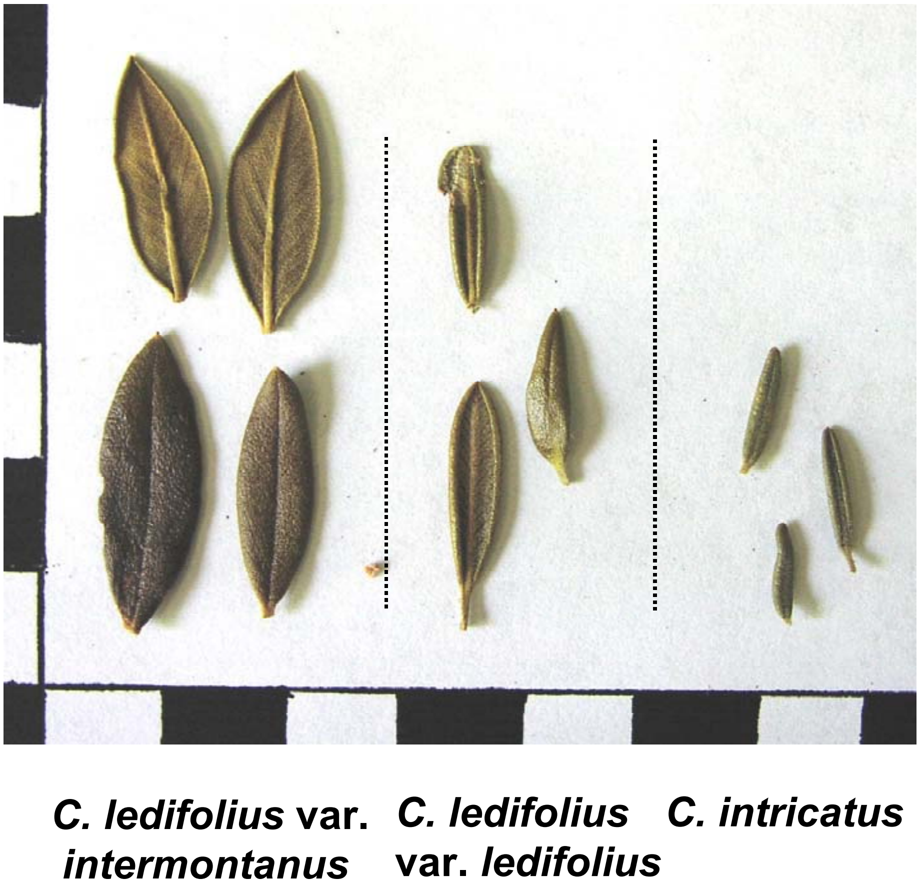


Figure 1.3. A picture of representative leaves for the taxa in the *Cercocarpus ledifolius* complex as delimitated by Holmgren (1987). *C. ledifolius* var. *intermontanus* is found on the left, *C. ledifolius* var. *ledifolius* is found in the middle, and *C. intricatus* is found on the right. The black and white squares found on the bottom and to the left of the picture are square centimeters for scale.

Leaf diversity from sympatric populations

Figure 1.4. A picture showing mature leaf variation found in a single population of *Cercocarpus ledifolius* from White Pine County, Nevada. The individuals displaying leaves on the right would be identified as *C. ledifolius* var. *intermontanus*. Individuals displaying leaves in the middle would be identified as *C. ledifolius* var. *ledifolius*, and individuals displaying leaves on the left would be identified as *C. intricatus*. The black and white squares found on the bottom and to the left of the picture are square centimeters for scale.

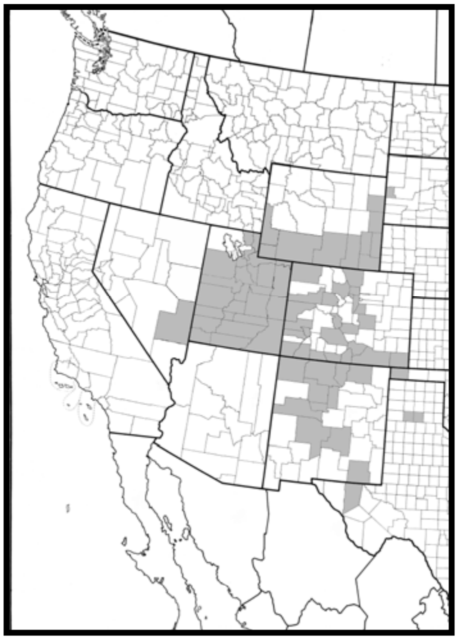
The *montanus* complex: Varieties of *C. montanus*

C. montanus var. *montanus*

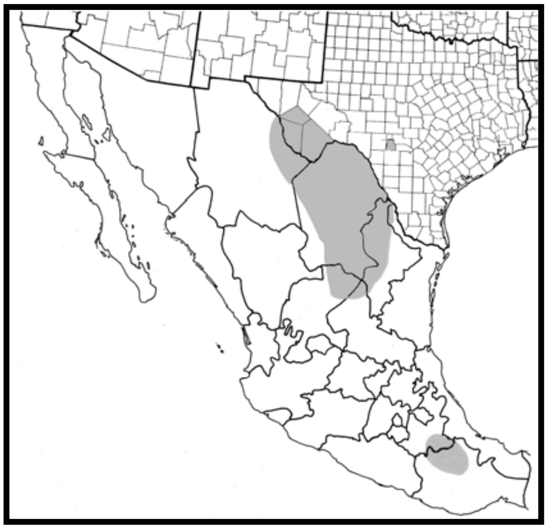
C. montanus var. *glaber*

C. montanus var. *paucidentatus*

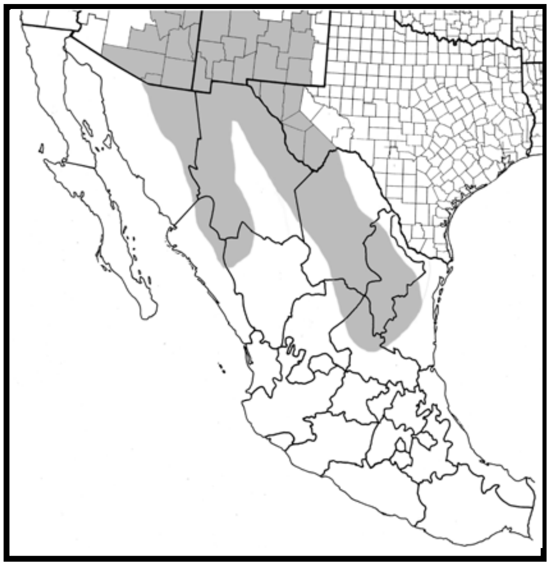
Figure 1.5. A picture of representative leaves for three of the four varieties in the *Cercocarpus montanus* complex as delimitated by Martin (1950). *C. montanus* var. *montanus* is found on the left, *C. montanus* var. *glaber* is found in the middle, and *C. montanus* var. *paucidentatus* is found on the right. The black and white squares found on the bottom and to the left of the picture are square centimeters for scale.



C. montanus var. *montanus*



C. montanus var. *glaber*



C. montanus var. *paucidentatus*

C. montanus var. *argenteus*

Figure 1.6. Geographic distribution of the *Cercocarpus montanus* complex.

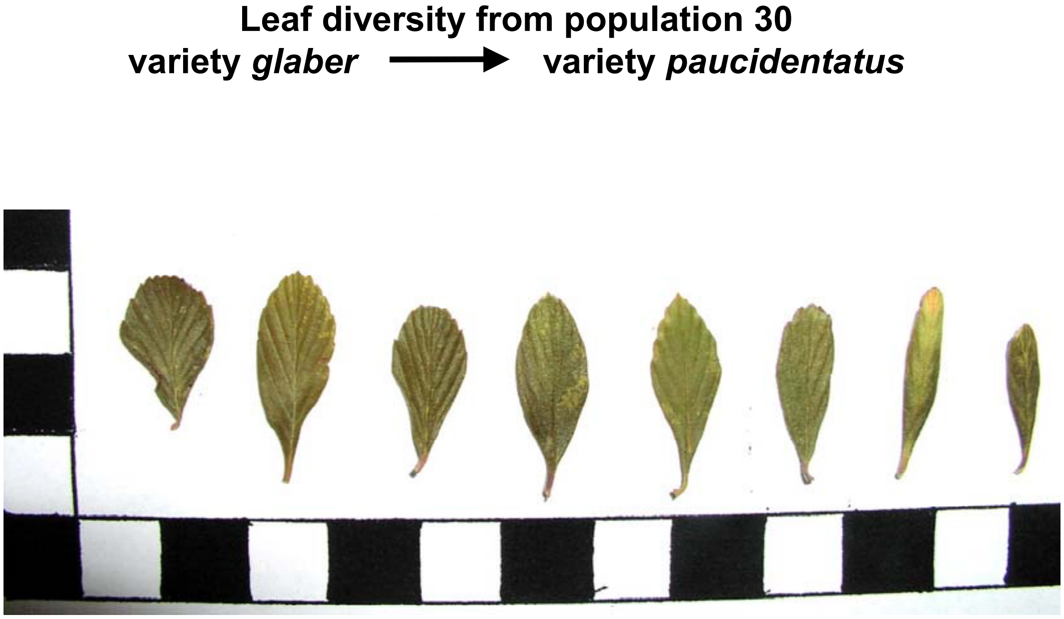
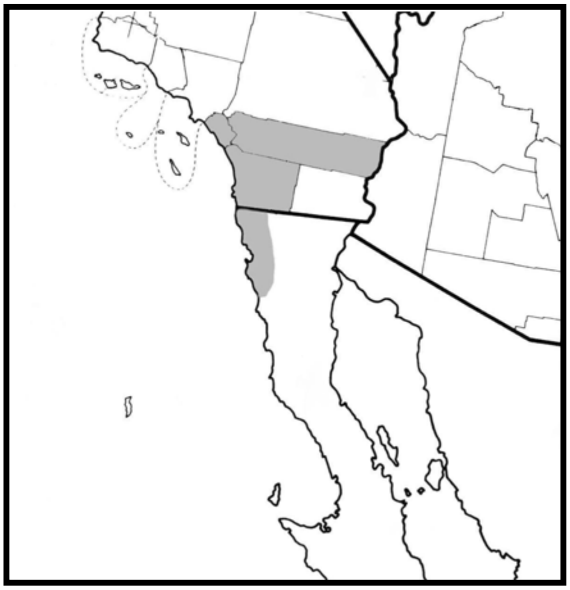


Figure 1.7. A picture showing mature leaf variation found in a single population of *Cercocarpus montanus* var. *paucidentatus* from Catron County, New Mexico. The black and white squares found on the bottom and to the right of the picture are square centimeters for scale.

C. betuloides var. *betuloides*

C. betuloides var. *blancheae*



C. minutiflorus

Figure 1.8. Geographic distribution of the *Cercocarpus betuloides* complex.

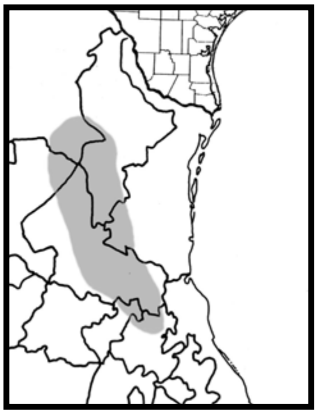
The *betuloides* complex: Varieties of *C. betuloides* and *C. minutiflorus*

C. betuloides var. *betuloides*

C. minutiflorus

C. betuloides var. *blancheae*

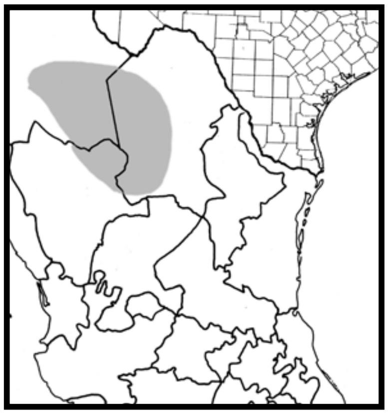
Figure 1.9. A picture of representative leaves for the taxa in the *Cercocarpus betuloides* complex as delimitated by Lis (1993). *C. betuloides* var. *betuloides* is found on the left, *C. betuloides* var. *blancheae* is found in the middle, and *C. minutiflorus* is found on the right. The black and white squares found on the bottom and to the right of the picture are square centimeters for scale.



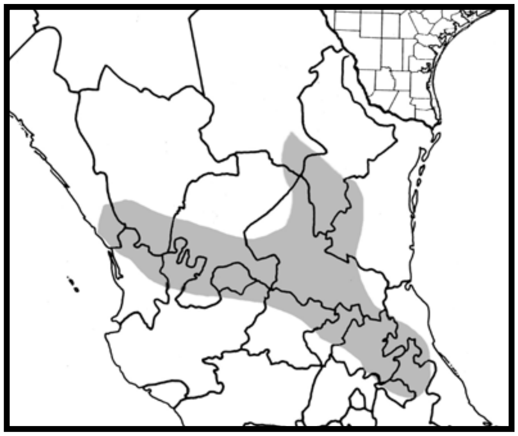
C. fothergilloides
var. *fothergilloides*

C. pringlei

C. fothergilloides
var. *medranoanus*



C. fothergilloides
var. *mojadensis*



C. macrophyllus

C. traskiae
present only on
Santa Catalina Island

Figure 1.10. Geographic ranges for *Cercocarpus fothergilloides* complex.

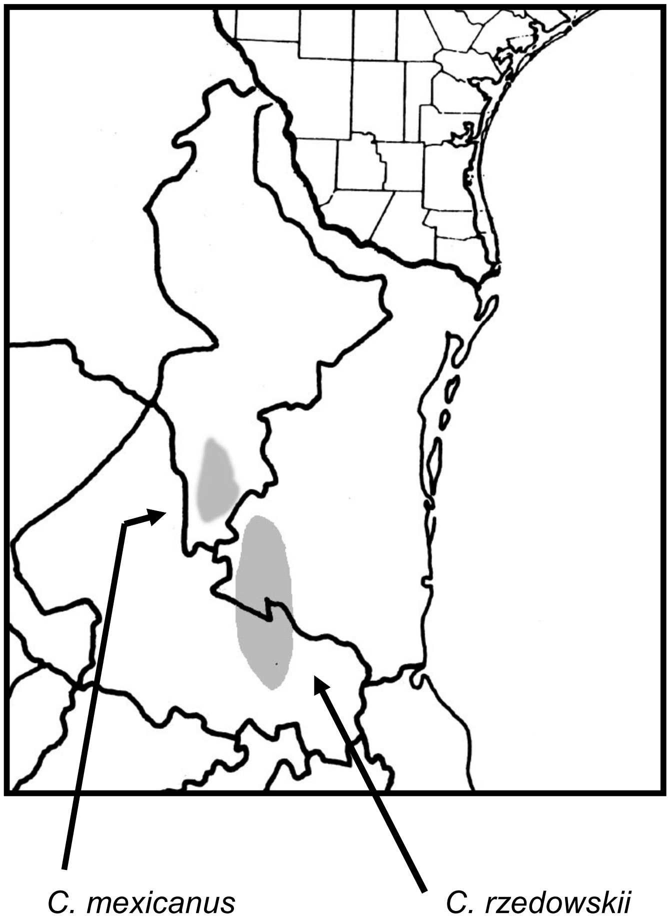


Figure 1.11. Geographic ranges for *Cercocarpus mexicanus* and *C. rzedowskii* (Henrickson 1987).

Chapter 2: Deep Coalescence of Multiple Paralogous Nuclear Ribosomal DNA Copies within Individuals of the Genus *Cercocarpus*

INTRODUCTION

Nuclear ribosomal DNA (nrDNA) is arranged in tandem arrays consisting of 100's of copies, that are generally located in the nucleolar organizing region (NOR), and can be found from one to several regions within the genome (Appels and Honeycutt 1986, Thompson and Flavell 1988). Within a single genome, the sequence similarity of each rDNA repeat is usually extremely high. This production of homogeneity between rDNA repeats is referred to as concerted evolution (Arnheim et al. 1980, Zimmer et al. 1980). Mechanisms which might drive concerted evolution include: gene conversion, slippage, transposition, or unequal crossing over (Arnheim et al. 1980, Dover 1982, Arnheim 1983).

Most attempts at phylogenetic reconstruction in plants using nuclear gene sequences have turned to nuclear rDNA. The reasons for using rDNA are well reviewed (Appels and Honeycutt 1986, Mindell and Honeycutt 1990, Hillis and Dixon 1992, Hamby and Zimmer 1992, Sanderson and Doyle 1992, Nickrent and Soltis 1995). These reasons include: 1) the ease of amplification due to the extremely high copy number and the tandem repeat structure coupled with the availability of universal primers; 2) the heterogeneity of rate of sequence evolution between different regions, which allows study at almost any taxonomic level; and 3) the observation that the rDNA copies undergo rapid concerted evolution, which allow direct sequencing of PCR products, despite the extremely high copy number.

However, if concerted evolution does not eliminate all rDNA variation between rDNA repeats within or between arrays before a taxon speciates (i.e., if the rate of concerted evolution is slower than the rate of speciation) or, if two divergent species come together and form a hybrid, then descendent species will possess multiple rDNA repeats of divergent paralogues until concerted evolution homogenizes the repeat. In recent years the number of reports of paralogous rDNA repeats in certain plant species has increased (Dvorak 1990, Suh et al. 1993, Buckler and Holtsford 1996, O'Kane et al. 1996, Waters and Schaal 1996, Wendel et al. 1995, Sang et al. 1995, and Campbell et al. 1997). Paralogous rDNA repeats in an individual plant have usually been attributed to one of three causes; hybridization, allopolyploidization, and pseudogene formation. With hybridization and allopolyploidization, the paralogous rDNA repeats' origins (i.e., the parental species' rDNA types) can be identified and often lend insight into the diversity and evolutionary history of the taxon being studied. In the case of pseudogene formation, the functional rDNA repeats are usually identified by secondary structure stability analysis and base pair composition (Buckler et al. 1997). In this chapter, we present the first evidence for multiple, paralogous rDNA repeats with a deep coalescence that do not appear to be the product of hybridization, alloploidization, or pseudogene formation. Using the external transcribed spacer sequences (ETS) from the nuclear rDNA present in the taxa of the genus *Cercocarpus* (Rosaceae), we document at least three deep-coalescing copies of the ETS. We also document a complex evolutionary history for the ETS region in *Cercocarpus*.

MATERIALS AND METHODS

Collection of Plant Material and DNA Extraction

The systematics of *Cercocarpus* are largely unknown; the genus lacks a consensus classification or phylogeny. For this report, we followed the suggested taxonomy of Lis (1992) and Holmgren (1987) for the *Cercocarpus* taxa found in the United States. For the Californian taxa of *Cercocarpus*, we used the taxonomy found in the Jepson manual (Lis 1993). For the Mexican taxa of *Cercocarpus*, we used an unpublished treatment by J. Henrickson (Madroño ined.). We collected leaf material for 289 individuals, from 96 populations representing 16 taxa of *Cercocarpus* (Table 2.1). Leaves from five herbarium sheets are included in this total. The remainder of the samples consisted of field-collected leaf material, dried using silica gel for transport back to the lab. For each leaf sample, DNA was extracted from the collections using a modified CTAB procedure (Doyle and Doyle 1987). Genomic DNA extractions were cleaned using the Elu-quik DNA purification kit (Schleicher and Schuell) prior to PCR amplification.

Developing the ETS in *Cercocarpus*

To develop the ETS region in *Cercocarpus* for routine PCR amplification, we amplified the entire intergenic spacer (IGS) of *C. montanus* var. *montanus* (population #5) and *C. ledifolius* var. *intricatus* (population #25) (Table 2.2) using the primers *1M* and *18S-2L* (Linder et al. 2000) (Table 2.3) located in conserved regions of the 26S and 18S genes, respectively (Figure 2.1). A negative control was also used that lacked template DNA. Fifty microliter PCRs were performed using 30mM Tricine, pH 8.4, 2mM $MgCl_2$, 50 mM KCl, 5% acetamide, 100 uM of each dNTP, 1.0 unit of Taq polymerase, and 30 nM of each amplification primer. Twenty nanograms of template DNA were added, and the reactions were overlaid with mineral oil and run on a MJ Research Thermal Cycler programmed for a hot start (95^o C, 5 min; 74^o C, 7 min) and 30 cycles of 94^o C for 30 sec, 50^o C for 45 sec, and 72^o C for 2 min. A final extension of 72^o C for 7 min was also included. Three microliters of the reaction products were run on a 1.5 % agarose gel in 0.5X TBE buffer. PCR products were visualized with ethidium bromide under UV light. The IGS PCR products from *C. montanus* var. *montanus* (population #5) and *C. ledifolius* var. *intricatus* (population #25) were purified using the QIAquick PCR purification kit (QIAGEN) and were sequenced from their 3' end on an ABI Prism 377 automated sequencer using standard reagents. Because the ETS region was too long to sequence with a single primer, we developed an internal primer *Cerc-ETS1r* (Table 2.3) (Figure 2.1) to sequence through the transcription initiation site (TIS). We identified the TIS based on similarity with the TISs in other plant groups (Linder et al. 2000). As a result of its role in transcription of the rDNA repeat, the TIS is under strong selection to stay conserved, which makes it a good choice for priming throughout *Cercocarpus*. This approach has been successful in other studies (Linder et al. 2000). Both sequences were proofread and then aligned using Clustal W in MegAlign (DNAstar, 1998), followed by the usual adjustment. We designed *Cerc-ETS1f* (Table 2.3) to include the TIS and to have high sequence similarity to both taxa (Figure 2.2).

Table 2.1: Taxa and Populations used to study ETS type variation in *Cercocarpus.*

Taxon	Populations sampled	Individuals sampled	Individuals w/ successful amplifications	ETS types and number of individuals
C. montanus var. *montanus*	9	26	26	1(26)
C. montanus var *paucidentatus*	20	57	56	1(50) 3(3) 1 and 3(3)
C. montanus var. *glaber*	4	17	16	1(14) 1 and 2(1) 1 and 3(1)
C. montanus var. *argenteus*	1	1	1	1(1)
C. betuloides var. *betuloides*	14	44	42	1(28) 2(9) 4(2) 1 and 2(3)
C. betuloides var. *blancheae*	5	15	12	1(2) 2(7) 3 and 4 (1) 4(2)
C. minutiflorus	3	7	6	1 and 2 (5) 2(1)
C. traskiae	2	19	19	1(1) 1 and 2(3) 2(12) 3(1) 2 and 4(1) 4(1)
C. fothergilloides var. *fothergilloides*	8	27	23	1(3) 3(18) 3 and 4(2)
C. fothergilloides var. *mojadensis*	1	1	1	2(1)
C. fothergilloides var. *medranoanus*	2	6	6	1 and 3(2) 3(4)
C. macrophyllus	3	4	4	1(1) 1 and 2(1)

				1 and 3(2)
C. rzedowskii	3	12	11	1 and 4(5) 4(6)
C. ledifolius var. *ledifolius*	3	5	4	1(4)
C. ledifolius var. *intermontanus*	13	35	34	1(34)
C. intricatus	5	15	15	1(15)
Total	96	289	275	1(179) 2(29) 3(26) 4(11) 1 and 2(13) 1 and 3(8) 1 and 4(5) 2 and 4(1) 3 and 4(3)

We successfully amplified 9 samples (populations 8, 12, 14, 15, 17, 18, 19, 20, and 21) (Table 2.2) using the *Cerc-ETS1f* and the *18S-2L* primers. After sequencing the 9 amplifications using the procedures outlined above, we aligned all 9 sequences plus *C. montanus* var. *montanus* (population #5) and *C. intricatus* (population #25) and noticed polymorphisms and small indels in the sequences near the *Cerc-ETS1f* priming site (Figure 2.2). We, therefore, designed another forward primer, *Cerc-ETS2f* (Table 2.3) (Figure 2.2), 25 bp downstream of *Cerc-ETS1f* in a region that was completely conserved in all sequences. We used *Cerc-ETS2f* for all further ETS amplifications. We attempted amplification using *Cerc-ETS2f* on all 289 individuals (Table 2.1). Amplification was as described above for the IGS except for a 1 min. 30 sec. extension at 72^{o} C during the 30 cycles.

Three microliters of the reaction mixture were run on a gel and visualized as described above. Of the 289 samples, 57 failed to amplify, so we attempted the PCR again using the same primers with the FailSafe PCR system (Epicentre Technologies). Twenty five microliter reactions were performed using 12.5 uL of premixs A through L, 30 nM of each primer, 1 unit of Taq, and 20 ng of template DNA. Each reaction was overlaid with 30 uL of mineral oil. Amplification was the same as detailed above.

Table 2.2: *Cercocarpus* individuals whose ETS amplification was sequenced. Individuals with * denote the seven individuals used in the probing experiment. † denotes the 14 individuals who were included in the reduced dataset. BVH = Brian Vanden Heuvel, JH = James Henrickson, BC = Bonnie Crozier, Sand = Sanderson

Taxon	Population # and individual	Location and collection number	ETS types
C. montanus var. *montanus*	5	Douglas Co., CO BVH 5	1
	8	Platte Co., WY BVH 8	1
	10	Custer Co., SD BVH 10	1
	28A	Otero Co., NM BVH 28	1
C. montanus var *paucidentatus*	†1	Tamaulipas, MX JH 19218	1
	27C	Eddy Co., NM BVH 27	1
	30A	Grant Co., NM BVH 30	1
	*86B	Bernalillo Co., NM BVH 86	1,3
	87A	Sierra Co., NM BVH 87	4
	87B	Sierra Co., NM BVH 87	3
	90D	Catron Co., NM BVH 90	1,3
C. montanus var. *glaber*	3	Coahuila, MX JH 19218	1
	12	Brewster Co., TX JH 20663	1
	14	Brewster Co., TX JH 20665	1
	15	Brewster Co., TX JH 20666	1
	17	Coahuila, MX JH 20637	1
	58B	Real Co., TX JH 22275	1
	68A	Coahuila, MX BVH and JH 22414	1
	75A	Tamaulipas, MX BVH and JH 22429	1
	80A	Tamaulipas, MX BVH and JH	1

		22457	
	85A	Coahuila, MX JH 22587	1
C. montanus var. *argenteus*	131	Culberson Co., TX BVH 131	1
C. betuloides var. *betuloides*	19	Los Angeles Co., CA BVH 123a	2
	20	Los Angeles Co., CA BVH 123b	2
	†21	Los Angeles Co., CA BVH 123c	2
	62D	Tulare Co., CA BVH 127a	2
	†63B	Tulare Co., CA BVH 127a	4
	†65C	Tulare Co., CA BVH 127b	2
	65D	Tulare Co., CA BVH 127c	2
	*93C	Gila Co., AZ BVH 93	1,2
	118A	Monterey Co., CA BVH 118	2,4
	*120C	San Luis Obisbo Co. CA BVH 120	1
C. betuloides var. *blancheae*	121B	Santa Barbara Co., CA BVH 121	3,4
	121C	Santa Barbara Co., CA BVH 121	2
	122A	Santa Barbara Co., CA BVH 122	2
	123F	Los Angelas Co. CA, BVH 123	2
	123G	Los Angelas Co. CA, BVH 123	1
	126.BC1	Los Angelas Co. CA	2
C. minutiflorus	57B	San Diego Co., CA JH 22311	1,2
	61	San Diego Co., CA JH 22313	1
C. traskiae	52B	Los Angelas Co., CA JH 22312	2
	126A	Los Angelas Co.	2
	126E	Los Angelas Co.	1,2
	†126.29	Los Angelas Co.	2,4
	126.WB3	Los Angelas Co.	4
C. fothergilloides var. *fothergilloides*	66C	Coahuila, MX BVH and JH 22411	3
	66D	Coahuila, MX BVH and JH 22412	3
	67A	Coahuila, MX BVH and JH 22413	3
	*67B	Coahuila, MX BVH and JH 22413	1
	†71A	Coahuila, MX JH and BVH 22423	3
	†83B	Tamaulipas, MX JH and BVH 22464	3
	*†84A	Nuevo Leon, MX JH and BVH 22482	4
C. fothergilloides var. *mojadensis*	†18	Coahuila, MX JH 20639	2
C. fothergilloides var. *medranoanus*	†77B	Tamaulipas, MX BVH and JH 22444	3
	82A	Tamaulipas, MX BVH and JH 22456	1,3
C. macrophyllus	26	Durango, MX BC n/a	1,2
	69A	Coahuila, MX BVH and JH 22415	1
	69B	Coahuila, MX BVH and JH 22415	1,3
C. rzedowskii	76A	Tamaulipas, MX BVH and JH 22434	1,4
	†76E	Tamaulipas, MX BVH and JH	1,4

		22434	
	78A	Tamaulipas, MX BVH and JH 22446	1
	78B	Tamaulipas, MX BVH and JH 22446	4
	*†81A	Tamaulipas, MX BVH and JH 22460	4
C. ledifolius var. *ledifolius*	41B	Emery Co., UT BVH 41	1
	104	White Pine Co., NV BVH 104	1
C. ledifolius var. *intermontanus*	45C	Box Elder Co., UT BVH 45	1
	98A	Sierra Co., CA BVH 98	1
	101A	Lander Co., NV BVH 101	1
	102B	Elko Co., NV BVH 102	1
	108C	Clark Co., NV BVH 108	1
	111A	Mono Co., CA BVH 111	1
C. intricatus	*†25	Garfield Co., UT BVH 25	1
	50D	Uintah Co., UT BVH 50	1
	103A	White Pine Co., NV BVH 103	1
Purshia tridentata	†34	La Plata Co., CO BVH 34	Purshia
Purshia plicata	74	Coahuila, MX BVH and JH	Purshia

Sequencing the ETS

We sequenced 102 ETS amplifications representing 76 individuals (Table 2.2), using a sampling strategy to maximize the geographic distribution and ETS length variants for each *Cercocarpus* taxon (Figure 2.3). One of two strategies was employed for sequencing the ETS for each sample. If the ETS amplification produced a single PCR product, the product was sequenced directly as described above. If the ETS amplification produced multiple-length PCR products, we cloned the PCR products using the TOPO-TA cloning kit (Invitrogen). At least 10 clones per amplification were screened for the ETS insert by direct PCR amplification using the *Cerc-ETS2f* and *18S-2L* primers and the same PCR parameters listed above, except for a 10 min. hot start at 95^{o} C to lyse the cells from the clones. The amplified inserts were visualized on an agarose gel as above. Two to six of these PCR amplifications of cloned material for each ETS length variant were purified (QIAquick PCR purification kit, QIAGEN) and sequenced per sample. Sequencing was performed on an ABI Prism 377 or an ABI Prism 3100 automated sequencer using standard reagents. Sequences were proofread and then aligned using Clustal W in MegAlign (DNAstar, 1998), followed by adjustment by eye.

Table 2.3: Sequences of the primers used in PCR for this study

Primer	Sequence 5'-3'	Direction
1M	TTGCTGCCACCATCCACTGAGAT	Forward
18S-2L	TGACTACTGGCAGGATCAACCAG	Reverse
Cerc-ETS1f	TATAAAGGGGGAGGCCTCATT	Forward
Cerc-ETS2f	TACAAACATATTCTCCCATATCTC	Forward
Cerc-ETS1r	GGCCCAATCTCCTTTCATGTCACT	Reverse
Pur-ETS1r	CGCACGCACAAGGCAGAAGG	Reverse
Pur-ETS2r	CACCGGTGGGCTGACGAAAGG	Reverse
ETS Type 1f	GTTGGTATCAGAGCAGGATCGATATGTC	Forward
ETS Type 1r	GAATCATCGCAGCTCCTCGGGGTGC	Reverse
ETS type 2f	GCCCACCCATGCAATGCACAAAGC	Forward
ETS type 2r	GATCGAGACACTGGCACCACTATCGC	Reverse
ETS type 3f	GATTGGATCATATTGATGAATACCC	Forward
ETS type 3r	GCGGATGGAGATGGTGGTAGGCTGGC	Reverse

After alignment with Clustal W, we observed four length variants and designated them ETS type 1 (1000 bp), ETS type 2 (800 bp), ETS type 3 (750 bp), and ETS type 4 (500 bp). We found large regions (80-600 bp) that were unalignable between ETS length variant types (Figure 2.3). We searched these putative unalignable regions for conserved alignable regions using pair-wise dot-plot comparisons in MegAlign (minimum window = 20 bp; >70% similarity), and employing representatives for each ETS length variant type in the analyses. We used the ETS sequence from *C. montanus*. var. *paucidentatus* (pop#1) to represent the ETS type 1 sequences, the ETS sequence *C. montanus* var. *blancheae* (pop#21) to represent ETS type 2 sequences, the ETS sequence *C. fothergilloides* var. *mojadensis* (pop#66) to represent ETS type 3 sequences, and the ETS sequence *C. rzedowskii* (pop#81) to represent type 4 sequences (Table 2.2).

The ETS sequences for *Purshia tridentata,* and *Purshia plicata* were obtained to serve as outgroups for this study. *Purshia* has consistently been identified as the sister group to *Cercocarpus* in recent molecular phylogenies (Morgan et al. 1994, Evans et al. 2000). The complete IGS in these species was amplified as described above for *Cercocarpus*. Then we sequenced from the 3' end of the amplified product using *18S-2L* and the reverse internal primers *Pur-ETS1r* (5'-CGCACGCACAAGGCAGAAGG-3') and *Pur-ETS2r* (5'-CACCGGTGGGCTGACGAAAGG-3') until we identified the TIS.

Testing Coalescence of the ETS Types found in *Cercocarpus*

After identifying four ETS types in *Cercocarpus*, we created a gene tree to examine the origin and historical relationships between the ETS types. All alignable regions from the ETS sequences, regardless of type, were aligned using Clustal W, followed by usual adjustment. This included the 5' alignable, 3' alignable, and the **α** and **β** regions (Figure 2.3). For ETS type 3 sequences, only the 5' common region could be included in the aligned sequence matrix. Only the 3' common region of the ETS sequences of *P. tridentata* and *P. plicata* could be included to serve as the outgroups (Figure 2.3). Therefore, ETS type 3 sequences' placement in the phylogeny were based only on their relationship to the other ETS types and not on a direct comparison to the outgroup. All phylogenetic analyses were carried out in PAUP 4.0.0d64 (Swofford 1999). Maximum parsimony (MP) heuristic searches were conducted using the following options: ACCTRAN, gaps treated as missing data, sequences added by stepwise addition, TBR, unordered, and equal weight characters. Support for branches were found using 100 bootstrap replicates. A maximum likelihood (ML) heuristic search was allowed to TBR branch swap for 240 hours, using the HKY85 model, adding sequences by stepwise addition as is and treating gaps as missing data.

Because of the computational time that a ML analysis on a large dataset requires, we created a reduced dataset containing only 14 sequences to examine ETS type evolution fully using the ML optimality criterion. The identity of the 14 sequences can be found in Table 2.2. We conducted two main sets of analyses on this dataset, which differed in the character sets included in the analyses. The first set of analyses included the 5' alignable, 3' alignable, and the α and β regions (511 aligned base pairs). We conducted ML searches using the Jukes-Cantor, Kimura-2, and HKY85 models, adding sequences as-is and using TBR branch swapping. Then we conducted the likelihood ratio test to find the parameter set that gave us the greatest fit. A bootstrap analysis was also conducted using 100 replicates. Maximum parsimony was also conducted on this reduced dataset as described above. The second set of analyses included just the 5' alignable and 3' alignable regions (382 aligned base pairs). We executed the same ML and MP searches as described above.

Probing for the different ETS types

We tested whether the different ETS length types recovered after amplifications were all part of the rDNA repeat rather than an artifact of non-specific priming outside of the rDNA repeat. Seven individual leaf collections were chosen based on the ETS type they contained (Table 2.2). Total DNA was extracted from these seven samples, using 3-5 grams of silica dried leaf material in a modified CTAB extraction procedure (Doyle and Doyle 1987). The samples were further purified using Cesium chloride/ethidium bromide gradient ultracentrifugation (Jansen and Palmer 1987). Total DNA was digested with EcoRI, which is known to cut infrequently in the rDNA repeat (Jourgensen et al. 1987). Restriction enzyme digestion, electrophoresis, bi-directional blotting to Zetabind nylon (AM CUNO), labeling of recombinant plasmids by nick translation, filter hybridizations, and autoradiography were performed following Palmer (1986) and Jansen and Palmer (1987).

To probe for the 5' end of the 18S gene, we used the 1.3 kb BamHI/HindIII rDNA fragment (Jourgensen et al. 1987) from *Pisum* sp. (kindly provided by Bob Jansen). The ETS type 1 probe was constructed by PCR amplifying 600 bp of the unalignable region of *C. betuloides* var. *betuloides* pop # 93C clone 2. Amplification was accomplished using *ETS Type 1f* and *ETS Type 1r* primers (Table 2.3). This PCR product was then placed into a plasmid using the TOPO TA kit as described above. The ETS type 2 probe was constructed in the same manner using a 350 bp fragment from the type 2 unalignable region. It was amplified using *ETS Type 2f* and *ETS Type2r* primers (Table 2.3) from *C. betuloides* var. *betuloides* pop # 93C clone 1. The ETS type 3 probe was constructed using a 250 bp fragment from the ETS type 3 unalignable region. It was amplified using *ETS Type3f* and *ETS Type3r* primers (Table 2.3) from *C. fothergilloides* var. *fothergilloides* pop # 67A clone 3. We did not construct a probe specifically for ETS type 4 because of its sequence-similarity to ETS type 2.

RESULTS

Amplification and Identity of the ETS types in *Cercocarpus*

We successfully amplified the ETS in 275 of the 289 individuals using the *Cerc-ETS2f* primer (Table 2.1) (Table 2.3). We amplified 43 of these samples using premix K from the FailSafe system. Of the 14 samples that failed, 3 were from herbarium sheets and had very low template DNA concentrations, and the remaining 11 samples that failed contained high levels of polysaccharides even after Elu-quik DNA purification. The 14 samples that failed to amplify using both the Tricine/acetimide and FailSafe systems were not taxon specific, but were distributed fairly evenly among all *Cercocarpus* taxa (Table 2.1). They included one individual of *C. montanus* var. *paucidentatus*, one individual of *C. montanus* var. *glaber*, two individuals of *C. betuloides* var. *betuloides*, three individuals of *C. betuloides* var. *blancheae*, one individual of *C. betuloides* var. *minutiflorus*, four individuals of *C. fothergilloides* var. *fothergilloides*, one individual of *C. rzedowskii*, and one individual of *C. ledifolius* var. *ledifolius*.

The distribution of the four different ETS length types (ETS type 1, type 2, type 3, and type 4) amplified can be seen in Table 2.4. ETS type 1 was the most frequent, followed sequentially by type 2, type 3, and type 4. Twenty nine amplifications contained multiple ETS types. The most frequent ETS type combination was ETS type 1 and 2, followed by ETS type 1 and type 3 and ETS type 1 and type 4. The least frequent combination of ETS types was ETS type 2 and type 4.

Table 2.4: Overall distribution of ETS types in individuals sampled in *Cercocarpus*

ETS types	Number of individuals
Only having ETS type 1	180
Only having ETS type 2	29
Only having ETS type 3	26

Only having ETS type 4	11
ETS type 1 and 2	12
ETS type 1 and 3	8
ETS type 1 and 4	5
ETS type 2 and 4	1
ETS type 3 and 4	3

The taxonomic distribution of the ETS types is illustrated in Table 2.5. Five taxa were found to only contain ETS type 1. These five taxa are all found in the Rocky Mountains and the Great Basin. The Californian taxa of *C. betuloides* var. *blancheae* and *C. traskiae* contained all four ETS types. The taxa that contain ETS type 3 all have Mexico as part of there distribution.

Table 2.5: Taxonomic distribution of ETS types in individuals sampled in *Cercocarpus*

ETS types	Cercocarpus taxa
Only having ETS type 1	*C. montanus* var. *montanus, C. montanus* var. *argenteus, C. ledifolius* var. *intermontanus, C. ledifolius* var. *ledifolius, C. intricatus*
Only having ETS type 2	*C. fothergilloides* var. *mojadensis*
ETS type 1 and 2	*C. minutiflorus*
ETS type 1 and 3	*C. montanus* var. *paucidentatus, C. fothergilloides* var. *medranoanus*
ETS type 1 and 4	*C. rzedowskii*
ETS type 1, 2, and 3	*C. macrophyllus, C. montanus* var. *glaber*
ETS type 1, 2, and 4	*C. betuloides* var. *betuloides*
ETS type 1, 3, and 4	*C. fothergilloides* var. *fothergilloides*
ETS type 1, 2, 3, and 4	*C. betuloides* var. *blancheae, C. traskiae*

Structure and Alignment of the ETS types

We found that distinct portions of each ETS type were alignable between ETS types (Figure 2.3). The 5’ alignable region was 174 bp long, and the 3’ alignable region was 207 bp long. ETS type 1, type 2, and type 4 contained both the 5’ and 3’ alignable regions. ETS type 3 only contained the 5’ alignable region. The *Purshia* ETS type contained only the 3’ alignable region.

Dot plot analyses revealed two small alignable regions embedded in the unalignable regions of ETS type 1, ETS type 2, and the *Purshia* ETS type. We designated them as α and β. The α region was 49 bp long in all ETS types which contained it. The β region was 80 bp in ETS type 1 and the *Purshia* ETS type. ETS type 2 had 34 bp of the 3' end of the α region. ETS type 1 contained two copies of the α region while ETS type 2 and the *Purshia* ETS type contained only 1 (Figure 2.3).

The five ETS types also contained different unalignable regions. The unalignable region in ETS type 1 was 499 bp and was upstream of the α and β region (Figure 2.3). The unalignable region in ETS type 2 totaled 314 bp and was found downstream of the α and β region. ETS type 3 contained an unalignable region that totaled 510 bp and stretched from the 5' alignable region all the way to the 18S gene. ETS type 4 contained only 100 bp of the extreme 3' end of the unalignable region found in ETS type 2. The *Purshia* ETS type contained the full ETS type 2 unalignable region and an unalignable region specific to *Purshia* that totaled 729 bp and stretched upstream from the α region to the TIS.

Relationships of the alignable regions of the ETS types

The sequence matrix of 5', 3', and α and β alignable regions totaled 511 aligned base pairs for 102 ETS sequences (Appendix A). The MP analysis resulted in over 10,000 most parsimonious trees of length 530. Figure 2.4 is a strict consensus based upon those trees. The *Cercocarpus* ETS types form a clade with high bootstrap support when compared to the outgroup sequences' indicating that the different *Cercocarpus* ETS types likely had their origin within the genus. ETS type 1 and ETS type 3 each form weakly supported clades, indicating a single coalescence or origin for both types (arrows in Figure 2.4). ETS types 2 and 4 form a "phylogenetic grade," with ETS type 3 basally nested within, suggesting that ETS type 3 may have had an early origin from one of those two ETS types. The ML tree (Figure 2.5.) is a strict consensus tree of five most-likely trees. The ML tree agrees with the MP tree in that the ETS type 1 and ETS type 3 clades are monophyletic, and the ETS type 2 and type 4 sequences form a grade. The ML tree differs from the MP consensus tree in the placement of some of the type 2 and type 4 sequences. The differences between the two reconstructions occur on branches with very little bootstrap support in the MP tree. Bootstrap support for particular reconstructions in the ML trees based on the full dataset were not calculated due to computational limitations.

The results of the ML and MP searches on the reduced dataset of 14 sequences can be seen in Figure 2.6. For the first analysis, where the α and β regions are included, we found that the MP and ML analysis gave the same tree topology. Both analyses supported the basal position of ETS type 1 sequences with moderate to high support. The analyses do not support the monophyly of ETS type 2 or 4 (Figure 2.6). When the α and β regions were excluded from the analyses, the MP and ML analyses resulted in more than 1 tree. The strict consensus trees from the two analyses were congruent. Once again, ETS type 2 and 4 are not monophyletic. In this case, the ETS type 1 clades and the clade including *C. fothergilloides* var. *mojadensis* (18) and *C. fothergilloides* var. *mojadensis* (83B) are unresolved at the base of the tree, sister to the rest of the taxa. In both analyses with the reduced dataset, the members of the ETS type 4 clade found in the basal position in the MP and ML consensus trees from the full dataset are found nested within the type 2 and 3 clade with moderate bootstrap support.

Probing

The results from the probing can be seen in Table 2.6. Only one band was produced on our filters after sequentially hybridizing with three probes. The ETS type 3 probe failed during the labeling reaction and produced negative results. The size of the band was identical for each individual between probes. The size of the band varied between individuals from 8.0 kb to 8.7 kb (Table 2.6).

Table 2.6: Results of the ETS type probing experiment. Listed are four ETS probes hybridized to seven individuals. Numbers in columns refer to the size of the band in kilobases.

Individual	18S probe	ETS type 1 probe	ETS type 2 probe	ETS type 3 probe
C. intricatus (25)	8.3	8.3	8.3	n/a
C. bet. var. *betuloides* (120C)	8.7	8.7	8.7	n/a
C. bet. var. *betuloides* (93C)	8.3	8.3	8.3	n/a
C. foth. var. *fothergilloides* (84A)	8.0	8.0	8.0	n/a
C .mot. var. *paucidentatus* (86B)	8.7	8.7	8.7	n/a
C. foth. var. *fothergilloides* (67B)	8.3(weak)	8.3(weak)	8.3(weak)	n/a
C. rzedowskii (81A)	8.7	8.7	8.7	n/a

DISCUSSION

ETS variation in *Cercocarpus*

We have demonstrated the occurrence of four ETS types in the taxa of *Cercocarpus* that differ from one another both in length and nucleotide substitution. The occurrence of variation, both nucleotide substitutions and length variation, in the ETS region in vascular plants is nothing new (Bena et al. 1998, Markos and Baldwin 1999, Linder et al. 2000). Many previous studies have documented both types of variation occurring both inter-specifically and intra-specifically. In fact, this variation has been harnessed successfully for phylogenetic reconstruction (Bena et al. 1998, Markos and Baldwin 1999, Linder et al. 2000. What is unique about the ETS variation reported here in *Cercocarpus* is the nature of the variation.

In sharp contrast to the lack of variation in length of the coding regions and the ITS in the rDNA within an individual or a species, the length of the ETS has been shown to have inter-specific or even intra-specific variation (Linder et al 2000, Markos and Baldwin 1999). To date, both inter-specific and intra-specific variation in ETS length has been attributed to the presence of varying numbers of tandem sub-repeats. In *Cercocarpus*, we have demonstrated length variation in the ETS that is not due to varying numbers of sub-repeats. Instead, the length variation we have observed is caused by the occurrence of unalignable regions of different length present in the center of the ETS. What is perhaps the most interesting about the placement of the unalignable regions is the conservation of the length of the 5' and 3' alignable regions surrounding the unalignable regions in all four ETS types. ETS type 3 only contains the 5' alignable region, yet the length is still conserved to the same base pair as in the other ETS types.

Why the length of the 5' and 3' alignable regions is so conserved between ETS types is still a mystery. One possibility we explored for the conservation of the 5' and 3' alignable regions is that they are a product of a transposable element, short interspersed elements (SINE), or long interspersed elements (LINE). We were unable to find any nucleotide sequence evidence for footprints of known transposons in the 5' or 3' alignable region or in the unalignable regions. We also conducted BLAST searches in GenBank for all of the different regions in all the ETS types and did not uncover any probable matches to known transposons. Another possability for the conserved nature of the alignable regions may be that the inherent secondary structure of the ETS is maintaining the length of the alignable regions in a stem, and the unalignable regions are found in a loop and therefore less constrained. To explore this possibility, we submitted ETS sequences of each type to folding programs that use minimum free energy. We were unable to find any evidence for stems or loops that were correlated with the base pair positions of the alignable and unalignable regions.

Are some of the ETS types pseudogenes?

Plant ribosomal pseudogenes have been identified in the Winteraceae (Suh 1993), *Gossypium* (Wendel et al. 1995), *Tripsacum* (Buckler and Holtsford 1996), *Zea* (Buckler and Holtsford 1996), and *Nicotiana* (Buckler et al. 1997),. Pseudogenes are identified in rDNA most often using ITS sequences that have low GC content, high substitution rate in conserved motifs, and low thermodynamic stability where certain conserved secondary structures are predicted (Buckler et al. 1997). Paralogues due to pseudogene formation are also diagnosed by large indel events that radically change the secondary structure (Mayol and Rossello 2001). As described above, we were unable to find any obvious evidence for secondary structure features in the ETS sequences in

Cercocarpus. There is no known secondary structure for the ETS to test against or a predicted base composition for the ETS. The only sequence motif that is known to be present in a functional ETS sequence is the TIS, which all four ETS types we recovered do possess.

Based on the probing data and the priming location of the *18S-2L* primer within the 5' end of the 18S gene, we have good evidence that at least three of the four ETS types recovered are located in the same fragment as the 5'end of the 18S gene and are not the product of non-specific priming outside the rDNA. Also, we identified individuals who possess only one of each of the four ETS types. If any one of the ETS types were a pseudogene, we would not expect to find individuals possessing only that ETS type. We would expect those individuals not to survive due to non-functional ribosomes. Therefore, it seems that each ETS type is part of a functional rDNA copy.

Are the multiple ETS types a product of hybridization or allopolyploidization?

In past studies, when multiple paralogous rDNA repeats in individuals were attributed to hybridization or allopolyploidization, each parental rDNA type was identified by comparison to "pure" parental individuals (Rieseberg et al. 1990, Soltis et al. 1991, Sang et al. 1995). These "pure" parental individuals can be identified because concerted evolution has homogenized the rDNA repeats within the individuals of the parental taxon. The occurrence of multiple rDNA types in the hybrid individual is due to the relatively recent creation of the hybrid, where concerted evolution has not had enough time to fully homogenize the paralogous rDNA repeats. In the data presented here, the different ETS types were not taxon specific, as one would expect in a “pure” taxon, where concerted evolution has homogenized all the rDNA repeats. The ETS types were distributed throughout the majority of the *Cercocarpus* taxa sampled. Although hybridization is known to occur within *Cercocarpus* (Rieseberg 1989, 1994), to explain the distribution of ETS types throughout the *Cercocarpus* taxa using recent hybridization events, we would have to postulate a high level of long-distance hybridization. For example, *C. macrophyllus* (collection 26 from Durango, Mexico) is sister to *C. ledifolius* var. *intermontanus* (collection 45C from Utah) within the ETS type 1 clade, yet *C. macrophyllus* (collection 26) is sister to *C. betuloides* var. *betuloides* (collection 20 from California). The distances in this example exceed 1600 km. Gene flow in plants can occur through the movement of seeds or pollen. *Cercocarpus* displays both seed and pollen characteristics that favor dispersal. *Cercocarpus* achenes have a persistent, plume-like style that is thought to promote dispersal (Russel 1997). In addition, *Cercocarpus* has a floral morphology (apetalous and excerted stamens) and pollen morphology (small with smooth exine) that suggests wind pollination, although no formal studies of the pollination biology of *Cercocarpus* have been completed to date. Yet, when studying the patterns of seed dispersal in *C. ledifolius,* Russel and Schupp (1998) found that both seed density and seed dispersal were highly influenced by microhabitat conditions, with most if not all the seeds being trapped in the leaf litter near the location of dispersal. Also, although the small, light pollen grains in *Cercocarpus* do disperse readily, the probability of a pollen grain landing on a receptive stigmatic surface decreases sharply as both the density of individuals goes down and distance increases, both of which occur between *Cercocarpus* populations. Therefore, we do not believe that long distance gene flow explains the distribution of ETS types in *Cercocarpus*.

The results we have obtained are also inconsistent with allopolyploidization as the process by which we have multiple, paralogous rDNA repeats. There have been eight chromosome counts spanning the four *Cercocarpus* taxa *C. betuloides* var. *betuloides*, *C. ledifolius* var. *ledifolius*, *C. ledifolius* var. *intermontanus*, and *C. montanus* var. *montanus*, and all have been constant at X=9 (Morley 1949, McArthur and Sanderson 1985). The chromosome counts for species in other genera closely related to *Cercocarpus* are also constant at X=9. There has been no evidence for recent polyploidization events within *Cercocarpus* or the genera allied with *Cercocarpus*.

Deep Coalescence of multiple ETS types in Cercocarpus

Using the alignable regions present in each ETS type, we showed that all of the ETS types found in *Cercocarpus* form a clade with high bootstrap support when compared to the outgroup *Purshia*. The monophyletic nature of the ETS types in *Cercocarpus* could change with broader sampling in the genus *Purshia* and other genera closely allied with *Cercocarpus* in the Rosaceae. The ETS type 1 sequences are monophyletic, rooted at (Figure 2.5, 2.6A, 2.6B) or near the base of the gene trees (Figure 2.4), suggesting an early origin within the genus. ETS types 2, 3 and 4 also show a deep coalescence and are found sister to ETS type 1 clade (Figure 2.5, 2.6A). ETS type 2 and ETS type 4 sequences occur polyphyletic in both the MP and ML reconstructions in both the complete and reduced dataset, with ETS type 3 being monophyletic and nested within the phylogenetic grade. We believe that these results cannot be explained by simply inter-specific variation. In other studies of rDNA diversity within a species using internal transcribed spacer (ITS) sequences, the divergence between two different types did not exceeded 5% (Baldwin et al. 1995, Vargas et al. 1998, Steane et al. 1999). Yet we found the different ETS types within taxa and from single individuals differed by both length and sequence content due to the presence of long, unalignable regions. Multiple inter-specific ETS sequences from *Argyranthemum* and *Helianthus* coalesced within a species, leading to the conclusion that concerted evolution was working, but not as rapid as in the ITS or coding regions of the rDNA repeat (Linder et al. 2000). On occasions when multiple rDNA types have been discovered, which do not coalesce within a species and are not associated with hybridization or allopolyploidization, one of the rDNA types has been shown to feature pseudogene qualities (Buckler et al.

1997, Mayol and Rossello 2001). Although we cannot rule out that some of the different ETS types are pseudogenes, we have presented evidence above for why we believe the ETS types found in *Cercocarpus* are functional. If the different ETS types are functional, which is likely, it is clear that the different ETS repeats are not being homogenized due to concerted evolution, even when they are present in the same individual.

The unique nature of a deep coalescence of ETS types implies there is a lack of concerted evolution in the ETS region in *Cercocarpus*. The phenomenon of concerted evolution has been documented, yet the mechanism by which it proceeds is still not truly known. If concerted evolution indeed proceeds by unequal crossing over, then large mutations that interfere with successful pairing between strands may break down concerted evolution. It may be that the unalignable regions we found in the ETS types of *Cercocarpus* are examples of this.

Origin of multiple ETS types in *Cercocarpus*

Using the phylogenies we created for the ETS types and the structure and location of different alignable and unalingable regions found in each ETS type, we developed a hypothesis for the steps that might have occurred to create the current ETS diversity (Figure 2.7). To create an ancestral ETS sequence, we must account for the presence of the 3' alignable, the ETS type 2 unalignable, and the α

and β alignable regions in both the extant *Purshia* ETS type and *Cercocarpus* ETS types 1 and 2. Although other evolutionary scenarios with the same number of steps exist, the ancestral ETS sequence must contain these three regions. We propose, as the first step in this scenario, the insertion of the *Purshia* unalignable region into the *Purshia* ETS type and the insertion of the 5' alignable region into an ancestral *Cercocarpus* ETS. Whether or not the type 1 unalignable region was present in the putative ancestral ETS or was inserted along with the 5' alignable region has the same number of steps in this scenario. Based on the gene tree topology and the shared features of ETS type 1 and ETS type 2 (mainly the position of the α and β regions) we propose that ETS type 1 and type 2 were each created by the deletion of different regions from this proposed ancestral *Cercocarpus* ETS type. We do not have any evidence from the gene tree to make inferences about the sequence of the type 1 and type 2 creations. In the MP and ML analyses using the reduced dataset including the α and β regions, ETS type 1 is most basal, yet when we excluded the α and β regions, ETS type 1 and type 2 are unresolved at the base of the tree. The ETS type 3 clade is found nested within the ETS type 2/type 4 grade, suggesting an origin from ETS type 2. Based on the sequence and structural similarity, as well as their placement in the gene tree, we believe ETS type 4 is a derivative of ETS type 2 by way of a deletion event which excises the α and β regions and a portion of the ETS type 2 unalignable region. According to the gene tree, it appears this deletion event has occurred 3 times. The ETS type 4 sequences take a more basal position to all the ETS type 2 sequences in the MP tree (Figure 2.4). Although this is contradictory to the arrangement one would expect based on our phylogenies (Figure 2.4 and 2.5), our other analyses (Figure 2.6) show at least one individual of ETS type 2 (*C. fothergilloides* var. *mojadensis* 18) is found at the base of the tree. This arrangement is most likely an artifact of the dataset in the MP analysis. The ETS type 2 sequences within the aligned dataset contain more information due to the inclusion of the α and β regions, which the ETS type 4 sequences do not contain. The lack of the α and β regions in ETS type 4 sequences forces them to a more basal position due to the absence of informative variation contained in the α and β regions. Also, there would be a low probability of a ETS type 2 arising from a type 4 multiple times to include all of the alignable regions in common with ETS type 1 and the outgroup *Purshia*.

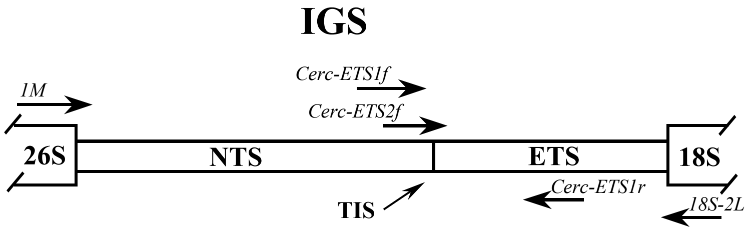


Figure 2.1. A representation of the intergenic spacer (IGS) between the 18S and 26S ribosomal genes in the nuclear genome. The genes 26S and 18S are shown by the large boxes. Transciption of the rDNA repeat begins at the transcription initiation site (TIS). The general location and direction of the main primers used in this study are shown by the arrows and the italized text. The size of the IGS in *Cercocarpus* is approximately 2.5 kb. The size of the ETS in *Cercocarpus* ranges from 1000 bp to 500 bp.

```
          Transcription Initiation site    TATAAAGGGGG
                             Cerc-ETS1f    TATAAAGGGGGAGGCCTCATT
          Consensus               GGAATTGTGTATAAAGGGGGAGGCCTCATTTTGGCAAA-C
                                          10        20        30        40

1. C. mon. var. mon (pop #5)      ······································CC
2. C. led. var. int (pop #25)     ······································-C
3. C. mon. var. gla. (pop #17)                                         ·-·
4. C. foth. var. moj. (pop #18)                                   ·G····-C
5. C. bet. var. bet. (pop #19)                                    ·G····-C
6. C. bet. var. bet. (pop #20)                                    ·G····-C
7. C. bet. var. bet. (pop #21)                                    ·G····-C
8. C. mon. var. gla. (pop #14)                                    ······-·
9. C. mon. var. gla. (pop #15)                                    ·G····-C
10. C. mon. var. mon. (pop #8)                                    ·K····-C
11. C. mon. var. gla. (pop #12)                                   ·C····CC

                                Cerc-ETS2f    TACAAACATATTCTCCCATATCT
          Consensus               ATCT-GAGAAA-TACAAACATATTCTCCCATATCT
                                          50        60        70

1. C. mon. var. mon (pop #5)      G···-······-·······················
2. C. led. var. int (pop #25)     ····-······-·······················
3. C. mon. var. gla. (pop #17)    ····-······-·······················
4. C. foth. var. moj. (pop #18)   ····G······-·······················
5. C. bet. var. bet. (pop #19)    ····G······-·······················
6. C. bet. var. bet. (pop #20)    ····-······-·······················
7. C. bet. var. bet. (pop #21)    ····-······-·······················
8. C. mon. var. gla. (pop #14)    ····-······-·······················
9. C. mon. var. gla. (pop #15)    G···-······-·······················
10. C. mon. var. mon. (pop #8)    G···-······-·······················
11. C. mon. var. gla. (pop #12)   R···-·····SA·······················
```

Figure 2.2. Alignment of ETS sequences in the region of the TIS and the placement of the *Cerc-ETS1f* and *Cerc-ETS2f* primers. The consensus sequence is shown on top. The dots refer to the same base as found in the consensus sequence, the dashes refer to gaps. The identity of each sequence is represented by an abbreviation. The abbreviation begins with C, referring to *Cercocarpus*. The three letters after the first period are the first three letters of the species name. var. refers to variety. The last three letters are an abbreviation of the first three letters of the variety name. The number in parentheses refers to the collection number. C. mon. var. mon (5) and C. int (25) were generated using the *Cerc-ETS1r* primer. Also aligned are the 9 samples successfully amplified and sequenced using the *Cerc-ETS1f* and *18S-2L* primers.

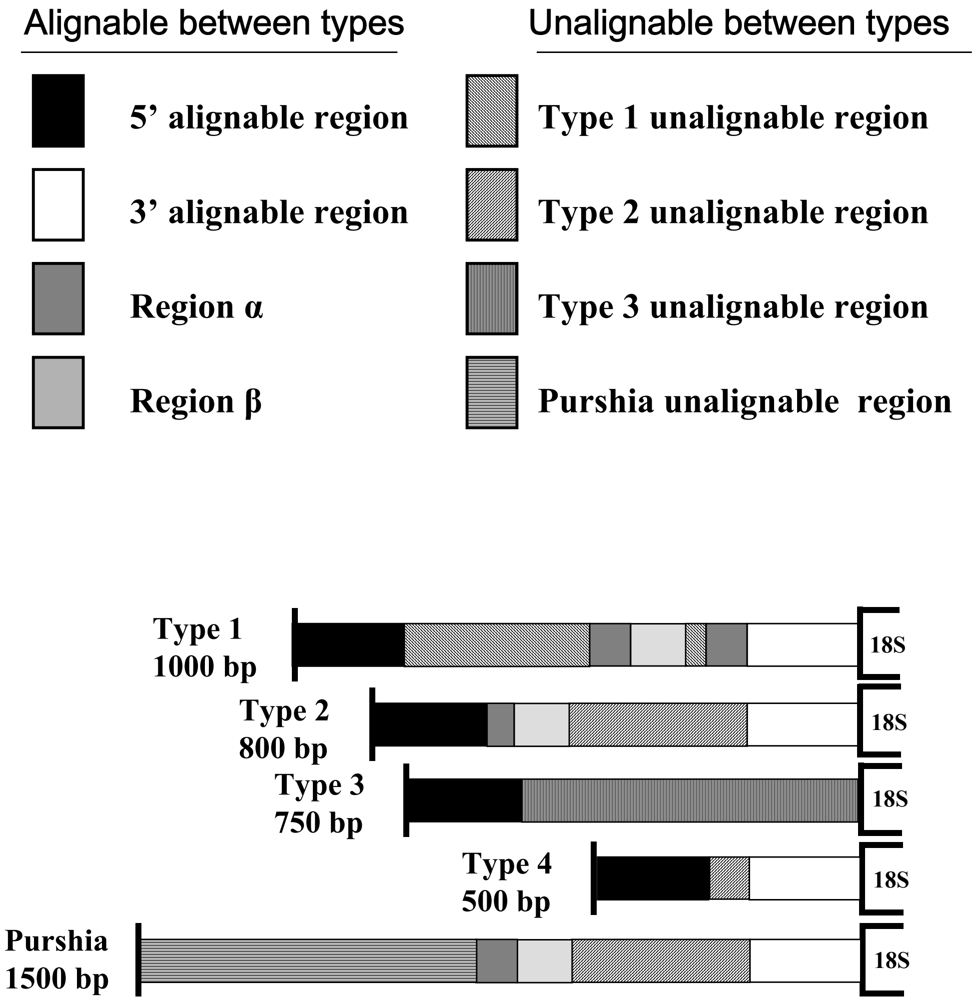


Figure 2.3. Representation of all four ETS types found in *Cercocarpus* and the ETS found in *Purshia*. Figure is not to scale. Representations of the ETS region begin at the transcription initiation site (TIS), shown by the vertical black bar on the left, and continue through to the 18S gene. Alignable and unalignable regions are shown by different hatching patterns, the identities are defined in the key.

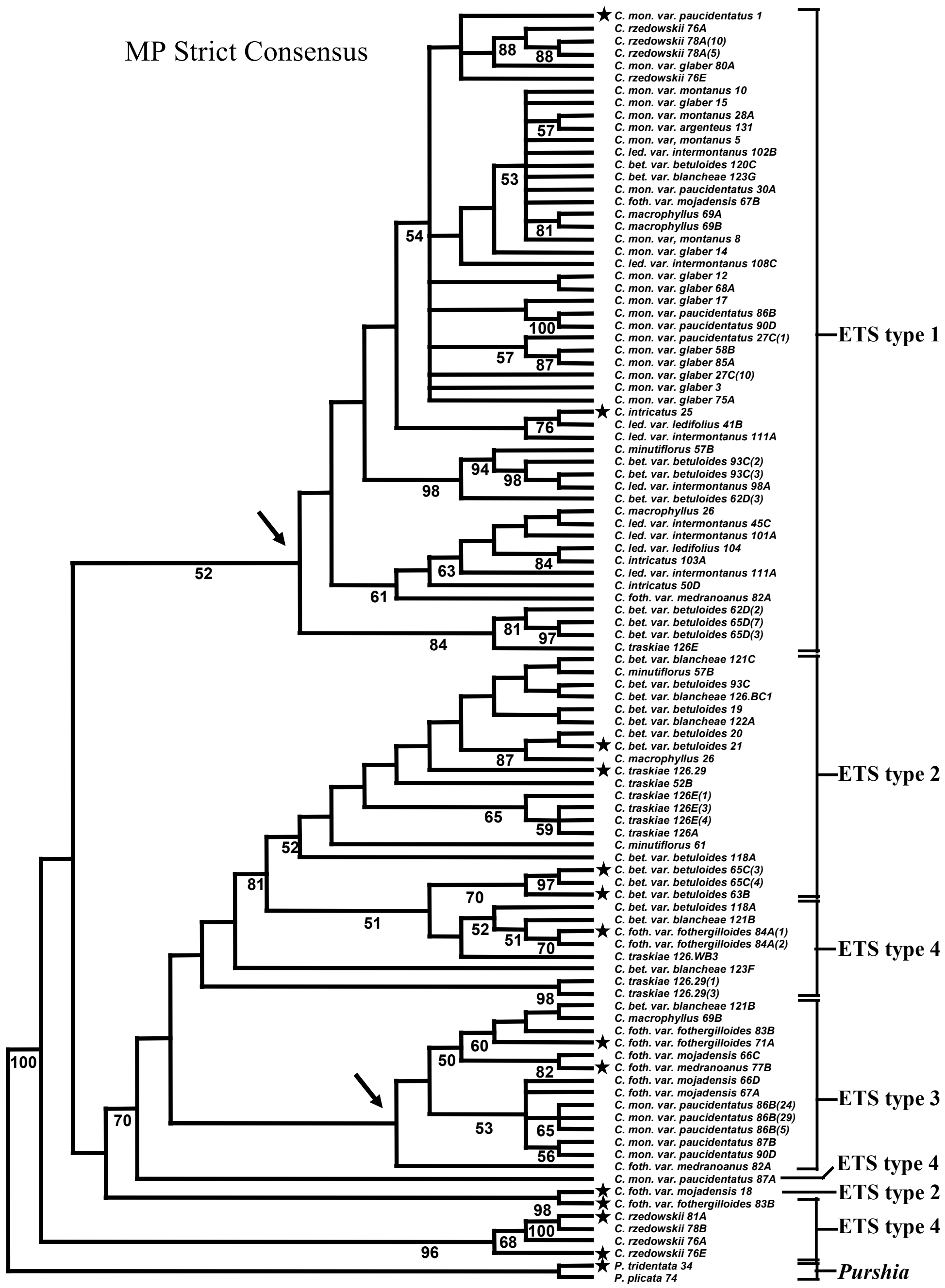


Figure 2.4. A maximum parsimony (MP) strict consensus tree including all ETS types of over 10,000 most parsimonious trees of length 530. Numbers below the branches indicate bootstrap support. The identity of the individual that produced each ETS sequence is abbreviated and shown on the right. The abbreviation begins with either C, referring to *Cercocarpus*, or P referring to *Purshia*. The three letters after the first period are the first three letters of the species name. var. refers to variety. The numbers in parentheses refer to the collection numbers. The ETS type designation of each sequence is shown on the right (Figure 2.3). The stars located to the left of the abbreviated name of the individual show the 14 sequences used in further analysis (Figure 2.6). The large arrows show the monophyletic nature of ETS type 1 and 3.

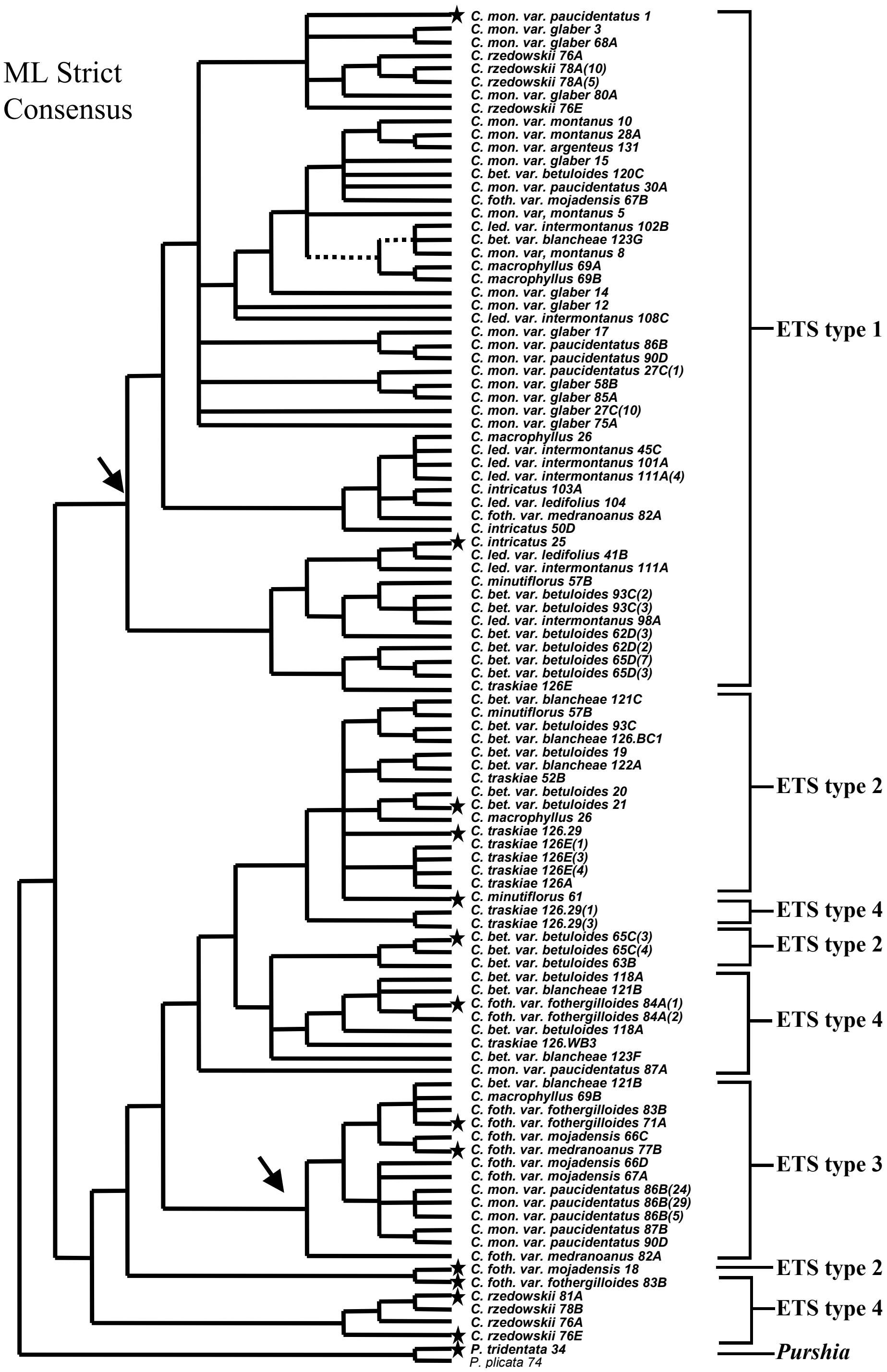
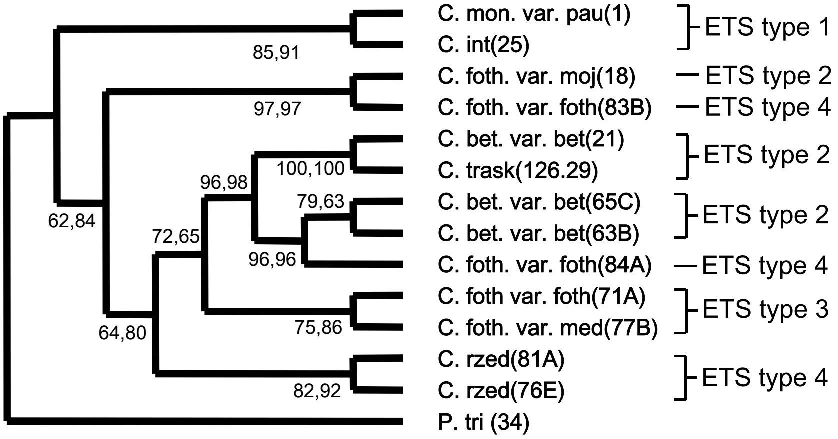


Figure 2.5. A maximum likelihood (ML) cladogram for all ETS types with a score of -Ln 3851.04666 produced from an aligned dataset of 512 bp of ETS sequence. This is one of five most likely trees. The dotted lines represent the branches that collapse in the strict consensus. The identity of the source of the ETS sequence and the ETS type designations are as in Figure 2.4. The small stars refer to the 14 ETS sequences used in further analyses. The large arrows highlight the monophyletic nature of ETS type 1 and 3.

A. With α and β



B. Without α and β

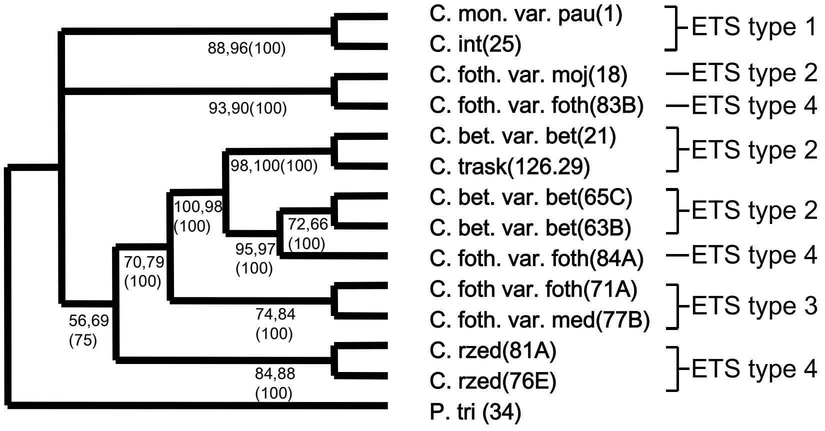


Figure 2.6. Cladograms representing both MP and ML analyses for the two reduced ETS datasets. **A**. A cladogram representing both MP and ML analyses using the 5' alignable, 3' alignable, and α and β regions. The two analyses resulted in the same topology. The MP tree had a length 182. The ML had a likelihood score of -Ln = 1616.56253. The numbers below the branches indicate bootstrap support. The numbers before the comma are bootstrap percentages from the MP analysis, the number after the comma bootstrap percentages from the ML analysis. The identities of the sequences are abbreviated as in Figure 2.2. The ETS type is designated as in Figure 2.4 and 2.5. **B**. A strict consensus cladogram representing both MP and ML analyses using just the 5' alignable and 3' alignable regions. The MP analyses produced 4 most parsimonious trees with a length of 141. The ML analysis resulted in 2 most likely trees with a score of -Ln = 1246.24172. Numbers below the branches are measures of bootstrap support. The number before the comma is the bootstrap values from the MP analysis; the number after the comma is the bootstrap value from the ML analysis. The numbers in the parentheses are the majority rule consensus values from the MP analysis.

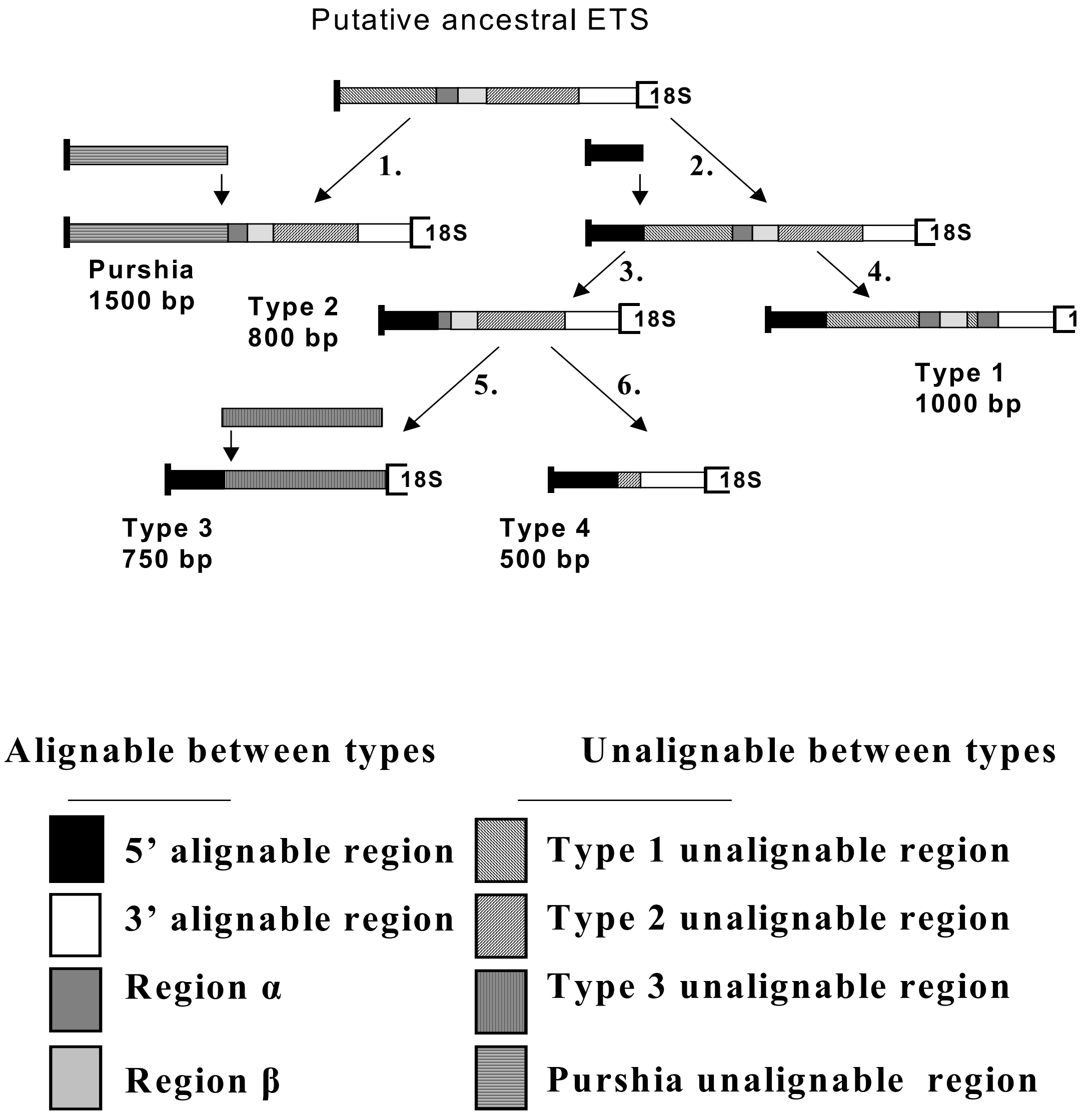


Figure 2.7. A proposed series of steps for the evolution of multiple ETS types in *Cercocarpus* and *Purshia*. Alignable and unalignable regions are illustrated as in Figure 2.3, the identities are defined in the key. 1) From the proposed ancestral sequence, the *Purshia* unalignable region is inserted. 2) From the proposed ancestral sequence, the 5' alignable region is inserted. 3) ETS type 2 is created by the deletion of the type 1 unalignable region. 4) ETS type 1 is created by the deletion of the type 2 unalignable region. 5) ETS type 3 is created by the insertion of the type 3 unalignable region. 6) ETS type 4 is created by the deletion of the central part of the ETS type 2 unalignable region.

Chapter 3: A nuclear phylogeny for *Cercocarpus* H.B.K. (Rosaceae) using AFLP and ETS sequence data

INTRODUCTION

Cercocarpus H.B.K. (Rosaceae) is a small genus made up of shrubs and trees found in chaparral and arid montane regions of western North America. Currently, *Cercocarpus* does not have a consensus classification or phylogeny. Since it was first described, *Cercocarpus* has undergone three revisions (Schneider 1905b, Rydberg 1913, Martin 1950), each varying widely in the ranks assigned to taxa and the number of species. The first revision (Schneider 1905b) described 10 species with 12 varieties. Rydberg (1913), in the second revision, described 21 species. Rydberg never recognized sub-specific taxa. The last revision by Martin (1950) described only six species with 11 varieties. Confusion about the number and circumscription of species within *Cercocarpus* results from the absence of clear morphological delimitations between taxa and, possibly, the lack of molecular phylogenetic treatments. Historically, the taxonomy in *Cercocarpus* has used leaf characteristics as the major source of characters for definition. Variations in leaf morphology show a continuum within and between taxa making it difficult to define boundaries among species (Brayton and Mooney 1966, Mortenson 1973), and making a phylogeny based on these characters misleading due to severe homoplasy (see Figure 1.4). It is clear that new characters, not based on gross morphology, are needed to resolve evolutionary relationships among species and the diversity within species, which have otherwise remained elusive.

The predominant hypothesis (Schneider 1905b, Brayton and Mooney 1966) for the origin of intergradations in leaf characters between putative distinct taxa of *Cercocarpus* has been hybridization and introgression. Individuals displaying intermediate leaf forms are thought to be F1 hybrids, and the intergradations of leaf form between the hybrids and each parental taxon arises from varying amounts of introgression of leaf traits into both parental taxa at the zone of contact. A second hypothesis, mentioned briefly by Brayton and Mooney (1966) describes leaf variation in *Cercocarpus* populations as a plastic response to varying habitats, or a mixture of different ecotypes, where similar genotypes have been selected to have different leaf shapes, based on microhabitat selection.

Molecular phylogenetic treatments may be useful to test these hypotheses for the origin of leaf-based variation. If leaf variation is indeed due to hybridization, we might expect the two “parental taxa“ to unite in two, distinct monophyletic groups. The placement of the individuals with intermediate leaf morphology may differ, uniting with either “parental taxon” depending on whether the marker was inherited uni-parentally (chloroplast markers) or if concerted evolution has converted all copies of the marker in the putative hybrid to one “parental taxon”. Individuals with intermediate morphology may also be found in an unresolved, basal position if there has been recombination between the two parental marker types in the hybrid individuals.

If leaf variation in populations of *Cercocarpus* is a product of plastic characters responding to different conditions or ecotypal variation, we might expect different tree topologies than those presented above based on hybridization. In this case, we would expect a single lineage for all leaf shapes with no clear clades representing the two "parental taxa." We might expect instead to see clades corresponding to geographic/ecological position and not leaf shape.

A phylogeny for *Cercocarpus* can also help test three specific hypotheses about the evolution and taxonomic affinities of particular *Cercocarpus* taxa. These hypotheses are listed below. 1) Stutz (1974) claimed that *Cercocarpus ledifolius* is the ancestral species from which both *C. montanus* and *C. intricatus* originated by peripheral isolation. 2) Lis (1992) claimed that there are similarities between *Cercocarpus traskiae*, found only on Santa Catalina Island off the coast of southern California and *C. fothergilloides* in Mexico based on leaf architectural characters. This is a surprising hypothetical relationship due to the historical view that *Cercocarpus traskiae* is closely related to *C. betuloides* of California (Martin 1950, Rieseberg 1989, 1994). 3) We can also test if *Cercocarpus rzedowskii* and *C. mexicanus*, two recently described species, are related to *C. ledifolius* and *C. montanus*, respectively (Henrickson 1987).

In this study, we developed the first phylogenetic hypotheses for *Cercocarpus* using nuclear molecular characters. We constructed phylogenies for *Cercocarpus* using amplified fragment length polymorphisms (AFLP) markers (Vos et al. 1995) and sequence data from the external transcribed spacer (ETS) of the nuclear ribosomal DNA (nrDNA). AFLPs (Vos et al. 1995) have been used extensively to examine relationships among closely related groups of plants, including cultivars, varieties, and species (Mueller and Wolfenbarger 1999, Zhang et al. 2001, Koopman et al. 2001, Hedren et al. 2001). In this study, we used AFLPs to construct phylogenies for *Cercocarpus*, concentrating on the *C. montanus* and *C. ledifolius* species complexes, to elucidate species and variety delimitations and relationships among *Cercocarpus* taxa. Initially, we explored the internal transcribed spacer (ITS) region of the nrDNA as a potential marker for phylogenetic reconstruction within *Cercocarpus*. The ITS region has been used with success for phylogenetic reconstruction at the generic and specific level in other plant groups (Baldwin et al. 1995). After finding the ITS region lacking in informative variation, we developed the ETS region as a DNA sequence marker for phylogenetic reconstruction. The details of the ETS's development for phylogenetic reconstruction and its molecular evolution are reported in Chapter 2. The ETS has been shown to be phylogenetically informative in very closely related groups, including island taxa (Linder et al. 2000, Baldwin and Markos

1998). Here, we present the results of the phylogenetic reconstruction using the ETS and their implications for the relationships within *Cercocarpus*.

MATERIALS AND METHODS

Collection and classification of Plant Material

We followed the proposed taxonomy of Lis (1992) with the exception of the California taxa - for which we used the taxonomy in the Jepson manual (Lis 1993) - and the Mexican taxa, for which we used an unpublished treatment by J. Henrickson (personal communication). We used 153 accessions representing 16 taxa of *Cercocarpus* (Table 3.1). Leaves from five herbarium sheets were included in this total. The remaining 148 accessions were field-collected leaf material, dried using silica gel for transport to the lab. The morphological characters used to distinguish species and sub-specific categories are listed in Table 3.2. Members of the genus *Purshia* were used as the outgroup for this study. *Purshia* has been shown to be sister to *Cercocarpus* in recent molecular phylogenies based on both chloroplast (Morgan et al. 1994) and nuclear genes (Rodgers et al. 1998).

Table 3.1: *Cercocarpus* taxa included in this study, including codes used in this paper, voucher numbers, collectors, and geographic position. BVH = Brian Vanden Heuvel, JH = James Henrickson. * denote individuals sampled in the ETS dataset. Taxa have been grouped by species complex according to Lis (1992).

taxon	code	location	Collector and Voucher
montanus complex			
C. montanus var. montanus	*5	Castle Rock, Douglas Co., CO USA	BVH 5
	*8	Hartville, Platte Co., WY USA	BVH 8
	*10	Dewey, Custer Co., SD USA	BVH 10
	*28A	Bailey Canyon, Otero Co., NM USA	BVH 28
C. montanus var. glaber	*3	Jame, Coahuila, MX	JH 20439B
	*12	Chisos Mts., Brewster Co., TX USA	BVH and JH 20662
	13	Chisos Mts., Brewster Co., TX USA	BVH and JH 20663
	*14	Chisos Mts., Brewster Co., TX USA	BVH and JH 20664
	*16	Chisos Mts., Brewster Co., TX USA	BVH and JH 20665
	*17	Sierra de la Modera, Coahuila, MX	JH 20637
	*58B	Leaky, Real Co., TX USA	JH 22275
	58D	Leaky, Real Co., TX USA	JH 22275
	S.7	Guadalupe Mts., Culberson Co., TX USA	Correll 18482
C. montanus var. *paucidentatus*	*1	La Perdida, Tamaulipas, MX	JH 19218
	27A	Guadalupe Mts., Eddy Co., NM USA	BVH 27
	27B	Guadalupe Mts., Eddy Co., NM USA	BVH 27
	*27C	Guadalupe Mts., Eddy Co., NM USA	BVH 27
	*30A	Pinos Altos, Grant Co., NM USA	BVH 30A
	30D	Pinos Altos, Grant Co., NM USA	BVH 30D
	30E	Pinos Altos, Grant Co., NM USA	BVH 30E
	30F	Pinos Altos, Grant Co., NM USA	BVH 30F
	31	Gila Mts., Grant Co., NM USA	BVH 31
	32B	Gila Mts., Grant Co., NM USA	BVH 32
	68A	Jame, Coahuila, MX	BVH and JH 22414
	68B	Jame, Coahuila, MX	BVH and JH 22414
	68C	Jame, Coahuila, MX	BVH and JH 22414
	70A	Jame, Coahuila, MX	BVH and JH 22416
	70B	Jame, Coahuila, MX	BVH and JH 22417
	72A	Cerro Potosi, Coahuila, MX	BVH and JH 22424
	72B	Cerro Potosi, Coahuila, MX	BVH and JH 22424
	*75A	Inturbide, Coahuila, MX	BVH and JH 22429
	*80A	Valle Hermosa, Tamaulipas, MX	BVH and JH 22457
	80C	Valle Hermosa, Tamaulipas, MX	BVH and JH 22457
	80D	Valle Hermosa, Tamaulipas, MX	BVH and JH 22457
	*85A	Sierra de la Encantada, Coahuila, MX	BVH and JH 22587
	85B	Sierra de la Encantada, Coahuila, MX	BVH and JH 22587
	85D	Sierra de la Encantada, Coahuila, MX	BVH and JH 22587
	85E	Sierra de la Encantada, Coahuila, MX	BVH and JH 22587
	86A	Tunnel Springs, Bernalillo Co., NM USA	BVH 86
	*86B	Tunnel Springs, Bernalillo Co., NM USA	BVH 86
	87A	Kingston, Sierra Co., NM USA	BVH 87
	89A	Mogollon Mts., Catron Co., NM USA	BVH 89
	89B	Mogollon Mts., Catron Co., NM USA	BVH 89
	90A	San Francisco Mts., Catron Co., NM USA	BVH 90
	90B	San Francisco Mts., Catron Co., NM USA	BVH 90
	*90C	San Francisco Mts., Catron Co., NM USA	BVH 90
	92A	Rose Peak, Greenlee Co., AZ USA	BVH 92

	92B	Rose Peak, Greenlee Co., AZ USA	BVH 92
	95A	Tonto State Park, Gila Co., AZ USA	BVH 95
	*96A	Strawberry, Gila Co., AZ USA	BVH 96
	96B	Strawberry, Gila Co., AZ USA	BVH 96
	HEN1	San Antonio Pena Nevada, Nuevo Leon, MX	JH HEN1
	S.3	Guadalupe Mts., Culberson Co., TX USA	Correll 24804
C. montanus var. *argenteus*	*131	Guadalupe Mts., Culberson Co., TX USA	McVaugh 7424

betuloides complex

C. betuloides var. betuloides	19	La Conada, Los Angelas Co., CA USA	JH 20731
	20	La Conada Los Angelas Co., CA USA	JH 20732
	*62D	Porterville,Tulare Co., CA USA	BVH 127a
	63B	Lake Hume,Tulare Co., CA USA	BVH 127b
	65C	Camp Nelson, Tulare Co., CA USA	BVH 127c
	*65D	Camp Nelson, Tulare Co., CA USA	BVH 127c
	*93C	Miami, Gila Co., AZ USA	BVH 93
	112	Yreka, Siskiyou Co., CA USA	BVH 112
	113A	Callahan, Siskiyou Co., CA USA	BVH 113
	116	Weaverville, Trinity Co., CA USA	BVH 116
	118A	Tularcitos Ridge, Monterey Co., CA USA	BVH 118
	*120C	American Canyon, San Luis Obispo Co., CA USA	BVH 120
	127A	Camp Wishone, Tulare Co., CA USA	BVH 127
	*128A	Oakhurst, Madera Co., CA USA	BVH 128
	130A	Hathaway Pines, Calaveras Co., CA USA	BVH 130
C. betuloides var. *blancheae*	52B	Rancho Santa Ana Botanic Garden, Los Angeles Co., CA USA	JH 22312
	121A	Parisima Hills, Santa Barbara Co., CA USA	BVH 121
	122A	Solvang, Santa Barbara Co., CA USA	BVH 122
	123F	Santa Monica Mts., Los Angeles Co., CA USA	BVH 123
	*123G	Santa Monica Mts., Los Angeles Co., CA USA	BVH 123
	125A	Santa Catalina Island, Los Angeles Co., CA USA	BVH 125
C. minutiflorus	*57B	Del Mar, San Diego Co., CA USA	JH 22311
	61	Fallbrook, San Diego Co., CA USA	BVH 127a

fothergilloides complex

C. traskiae	126A	Santa Catalina Island, Los Angeles Co., CA USA	Thorne 36737
	*126E	Santa Catalina Island, Los Angeles Co., CA USA	Thorne 36737
C. fothergilloides var. *fothergilloides*	66C	Los Lirios, Coahuila, MX	BVH and JH 22411
	66D	Los Lirios, Coahuila, MX	BVH and JH 22412
	67A	Los Lirios, Coahuila, MX	BVH and JH 22413
	*67B	Los Lirios, Coahuila, MX	BVH and JH 22413
	71A	Cerro Potosi, Coahuila, MX	BVH and JH 22423
	83A	Tula, Tamaulipas, MX	BVH and JH 22464
	84G	San Roberto Junction, Nuevo Leon, MX	BVH and JH 22482
C. fothergilloides var. *medranoanus*	77B	Valle Hermosa, Tamaulipas, MX	BVH and JH 22444
	*82A	Valle Hermosa, Tamaulipas, MX	BVH and JH 22456b
C. fothergilloides var. *mojadensis*	18	Sierra Encantada, Coahuila, MX	JH 20639
C. rzedowskii	*76A	Miquihuana, Tamaulipas, MX	BVH and JH 22434
	*76E	Miquihuana, Tamaulipas, MX	BVH and JH 22434
	*78A	Valle Hermosa, Tamaulipas, MX	BVH and JH 22446
	81A	Valle Hermosa, Tamaulipas, MX	BVH and JH 22460

C. mexicanus	134A	Coahuila, MX	JH 19295
C. macrophyllus	*26	Durango, MX	Crozier n/a
	*69A	Sierra de la Viga, Coahuila, MX	BVH and JH 22415
	*69B	Sierra de la Viga, Coahuila, MX	BVH and JH 22415
	79	Valle Hermosa, Tamaulipas, MX	BVH and JH 22458
ledifolius complex			
C. ledifolius var. *intermontanus*	37A	Casto Canyon, Garfield Co., UT USA	BVH 37
	37B	Casto Canyon, Garfield Co., UT USA	BVH 37
	37C	Casto Canyon, Garfield Co., UT USA	BVH 37
	39	Gooseberry, Sevier Co., UT USA	BVH 39
	45A	Mantua, Box Elder Co., UT USA	BVH 45
	45B	Mantua, Box Elder Co., UT USA	BVH 45
	*45C	Mantua, Box Elder Co., UT USA	BVH 45
	46A	Preston, Franklin Co., ID USA	BVH 46
	46B	Preston, Franklin Co., ID USA	BVH 46
	46C	Preston, Franklin Co., ID USA	BVH 46
	*98A	Verni, Sierra Co., CA USA	BVH 98
	98B	Verni, Sierra Co., CA USA	BVH 98
	98C	Verni, Sierra Co., CA USA	BVH 98
	99A	Santa Rose Mts., Humboldt Co., NV USA	BVH 99
	99B	Santa Rose Mts., Humboldt Co., NV USA	BVH 99
	99C	Santa Rose Mts., Humboldt Co., NV USA	BVH 99
	*101A	Toiyabe Range, Lander Co., NV USA	BVH 101
	102A	Ruby Mts., Elko Co., NV USA	BVH 102
	*102B	Ruby Mts., Elko Co., NV USA	BVH 102
	102C	Ruby Mts., Elko Co., NV USA	BVH 102
	105	Ely, White Pine Co., NV USA	BVH 105
	108A	Spring Mts., Clark Co., NV USA	BVH 108
	108B	Spring Mts., Clark Co., NV USA	BVH 108
	*108C	Spring Mts., Clark Co., NV USA	BVH 108
	110	Whits Mts., Inyo Co., CA USA	BVH 110
	*111A	West Walker river, Mono Co., CA USA	BVH 111
	111B	West Walker river, Mono Co., CA USA	BVH 111
	111C	West Walker river, Mono Co., CA USA	BVH 111
C. ledifolius var. *ledifolius*	41A	Emery, Emery Co., UT USA	BVH 41
	41B	Emery, Emery Co., UT USA	BVH 41
	*41C	Emery, Emery Co., UT USA	BVH 41
	*104	Ely, White Pine Co., NV USA	BVH 104
	106A	Cannons pass, White Pine Co., NV USA	BVH 106
	106B	Cannons pass, White Pine Co., NV USA	BVH 106
	106C	Cannons pass, White Pine Co., NV USA	BVH 106
	114	Noyes valley, Siskiyou Co., CA USA	BVH 114
C. intricatus	*25	Escalante, Garfield Co., UT USA	BVH 25
	43	Emery, Emery Co., UT USA	BVH 43
	48A	Flaming Gorge, Dagget Co., UT USA	BVH 48
	48B	Flaming Gorge, Dagget Co., UT USA	BVH 48
	48C	Flaming Gorge, Dagget Co., UT USA	BVH 48
	48D	Flaming Gorge, Dagget Co., UT USA	BVH 48
	50A	Vernal, Uintah Co., UT USA	BVH 50
	50B	Vernal, Uintah Co., UT USA	BVH 50
	50C	Vernal, Uintah Co., UT USA	BVH 50
	*50D	Vernal, Uintah Co., UT USA	BVH 50
	*103A	Ely, White Pine Co., NV USA	BVH 103
	103B	Ely, White Pine Co., NV USA	BVH 103
	103C	Ely, White Pine Co., NV USA	BVH 103

Purshia tridentata	*36B	Casto Canyon, Garfield Co., UT USA	BVH 36
	100B	Santa Rose Mts., Humboldt Co., NV USA	BVH 100
Purshia stansburiana	54A	Coliseum Canyon, San Bernardino Co., CA USA	JH 22307-5
Purshia glandulosa	55A	Cima Rd., San Bernardino Co., CA USA	JH 22303-5
Purshia plicata	*74D	Interbide, Coahuila, MX	Nesom 6198

Table 3.2. Eight morphological characters used to define taxa in *Cercocarpus*. The characters listed for *C. montanus* var. *montanus*, *C. montanus* var. *argenteus*, *C. montanus* var. *paucidentatus*, and *C. montanus* var. *glaber* are from Martin (1950). The characters of *C. betuloides* var. *betuloides*, *C. betuloides* var. *blancheae, C. minutiflorus*, and *C. traskiae* are from Lis (1993). The characters defining *C. rzedowskii* and *C. mexicanus* are from Henrickson (1987). *C. fothergilloides* var. *fothergilloides*, *C. fothergilloides* var. *mojadensis*, *C. fothergilloides* var. *medranoanus*, and *C. macrophyllus* are from an unpublished treatment by Henrickson (personal comunication).

taxon	**leaf size (cm)**	**leaf thickness**	**leaf shape**	**abaxial leaf pubescence**	**# of lateral veins**	**leaf margin**	**length of persistent style (cm)**	**geography**
C. montanus var. *montanus*	1.5 - 3 long 0.4-2 wide	thin	obovate, rotund, flabelliform	short villous, appressed silky	3 - 6	coarse ovate to triangular teeth present on distal half of blade	5 - 6	Great Basin and Rocky Mountains
C. montanus var. *glaber*	1 - 3 long 0.6 - 2 wide	thin	obovate, elliptical, rotund	glabrate	4 - 6	short, broad triangular apiculate tipped teeth present on distal half of blade	5 - 9	Chihuahuan Desert
C. montanus var. *paucidentatus*	0.7 - 2 long 0.3 - 1 wide	thin	lanceolate, oblanceolate, narrow obovate	short villous, appressed silky	3 - 5	entire to 3 - 5 short teeth at apex	1.5 - 5	Chihuahuan Desert
C. montanus var. *argenteus*	2 - 4 long 0.9 - 1.7 wide	thin	oblanceolate, obovate, rotund	appressed silky to wholly	5 - 6	short, broad triangular apiculate tipped teeth present on distal half of blade	5 - 6	TX and NM
C. betuloides var. *betuloides*	1 - 4 long 0.6 - 2 wide	thin	obovate, rotund	sparsely silky	4 - 7	finely toothed to serrate	5 - 9	CA and AZ
C. betuloides var. *blancheae*	2.5 - 6 long 2 - 3.5 wide	coriaceous	widely elliptic, ovate	sparsely silky	6 - 10	finely toothed to serrate, margins revolute	3 - 6	CA
C. minutiflorus	1 - 2.5 long 0.7 - 1.5 wide	thin	ovate, rotund	glabrous	3 - 5	entire to short, broad triangular apiculate tipped teeth present on distal half of blade	3 - 7	CA and Baja California

C. traskiae	3 - 6 long 2 - 4 wide	coriaceous	rotund, broadly obovate	woolly	6 - 8	revolute with short, broad triangular apiculate tipped teeth present on distal half of blade	4.5 - 6	Santa Catalina Island, CA
C. macrophyllus	4 - 11 long 2.5 - 7.8 wide	coriaceous to thin	oblong-ovate, oblong-elliptical	glabrous	12 - 18	teeth ovate to acute	5 - 8.5	Mexico
C. fothergilloides var. *fothergilloides*	2.5 - 4.5 long 1.5 - 3 wide	coriaceous	ovate, broadly ovate	sparsely to moderately villous	5 - 8	serrate to crenate on distal two-thirds of blade	4 - 8.5	Mexico
C. fothergilloides var. *mojadensis*	1.6 - 2.8 long 1.2 - 2.2 wide	coriaceous	broadly ovate to orbicular	dense, white tomentose	4 - 6	crenate on distal two-thirds of blade	4 - 6.5	Mexico
C. fothergilloides var. *medranoanus*	1.5 - 2.5 long 1.1 - 1.6 wide	coriaceous	ovate to orbicular	uniform, tight coiled hairs	4 - 6	crenate on distal two-thirds of blade	3 -4	Mexico
C. rzedowskii	0.4 - 1.5 long 0.4 - 0.6 wide	coriaceous	elliptical	thick, white tomentose	5 - 6	revolute, entire or with 2 - 4 spine tipped teeth on distal one-third	4.5 - 5	Mexico
C. mexicanus	1.7 - 3.1 long 1 - 1.5 wide	coriaceous	narrowly ovate	sparingly silky with large stomatal crypts	4 - 7	serrated on distal two-thirds	4.5. - 5	Mexico
C. ledifolius var. *intermontanus*	2.5 - 6 long 0.5 - 0.8	coriaceous	elliptical-lanceolate	glabrate	n/a	entire, revolute	4 - 10	Great Basin

	wide							
C. ledifolius var. *ledifolius*	1 - 2.5 long 0.1 - 0.4 wide	coriaceous	narrowly elliptical-lanceolate	densely-woolly	n/a	entire, revolute	3 - 6	Great Basin
C. intricatus	0.3 - 1 long 0.1 - 0.2 wide	coriaceous	linear	densely-woolly	n/a	entire, revolute to midrib	1 - 2	Great Basin

DNA Extraction

For all leaf samples, DNA was extracted using a modified CTAB procedure (Doyle and Doyle 1987). Genomic DNA extractions were cleaned using the Elu-quik DNA purification kit (Schleicher and Schuell), prior to further work.

AFLP methods

We followed the general protocol of Vos et al. (1995) to produce AFLP fragments. Briefly, sample DNA was restriction digested using the endonucleases EcoRI and MseI and ligated to the appropriate adapters. Two rounds of PCR amplification followed. The first preselective amplification used primers with a 1 base pair (bp) extension, and the second selective amplification used primers with 3 bp extensions. We used three different selective primer combinations having extensions -CCC/-AGC, -CCC/-ATA, and -CGG/-AGC, where the extensions refer to the EcoRI/MseI adapters, respectively. The -CCC EcoRI selective primers were labeled with the flourochrome FAM and the -CGG EcoRI selective primer was labeled with the flourochrome TAMRA (IDT industries).

The amplified fragments were separated and visualized on a MJ BaseStation (MJ Research), and sized using the size standard GS500 (ABI). Fragment scoring and lane tracking were done automatically within the software package Cartographer (MJ Research), and checked and adjusted by eye.

Fragments between 35 and 500 bp were scored. The three primer combinations were combined into one dataset containing 376 presence/absence characters (Appendix B and C).

The data were analyzed in two steps. First, analyses were performed on the complete dataset, containing all 153 accessions. The complete AFLP dataset was analyzed using both distance and parsimony criteria. The distance analysis consisted of NJ clustering and a heuristic search using minimum evolution (ME) distance estimates as implemented in PAUP∗ 4.09b (Swofford 1999). The ME method searches for tree topologies where the absolute difference between observed distances and path length distances are minimized. The parsimony analysis was performed in PAUP∗ 4.09b (Swofford 1999). A heuristic search strategy was employed with tree bisection/reconnection (TBR) branch swapping and applied the saving multiple parsimonious trees at each step (MULPARS) in effect. In order to explore a large amount of tree space without running out of memory trying to swap all possible trees to completion, 10,000 addition replicates were performed, using equally weighted characters and saving only 10 trees per replicate. A histogram of tree lengths for all 100,000 trees saved was obtained which allows the identification of multiple, equally parsimonious trees. The validity of using maximum parsimony as an optimality criterion in analyses using AFLP data has been challenged recently (see Koopman et al. 2001 for discussion). The dispute centers the assumption that equal-sized AFLP markers are homologous and the homoplasies that may be introduced into parsimony analysis due to this assumption. Therefore, we have compared the MP and distance trees to look for incongruence and assessed the strength of support for particular branches with 100 bootstrap replicates using the same parameters as described above.

The second set of analyses used a presence/absence AFLP data subset that matched the same accessions used for the ETS analysis (see below). This reduced dataset contained 49 individuals (Table 3.1) and was used to compare directly to the ETS dataset. Parsimony analyses were performed on this smaller dataset as described above, except 1000 bootstrap replicates were performed for a more accurate indication of branch support.

ETS amplification and sequencing

The methods and details of the development for amplifying the ETS in *Cercocarpus* are found in the preceding chapter (Chapter 2). Briefly, following the methods outlined in Linder et al. (2000), we developed a forward primer located near the transcription initiation site (TIS) that reliably amplified the ETS in individuals of *Cercocarpus*. We discovered five different ETS paralogues (types) in our survey, four within *Cercocarpus* and one type in the outgroup *Purshia*. The four ETS types in *Cercocarpus* have deep coalescences within the genus and none of the ETS types was taxon specific. Only two of the four ETS types (type 1 and type 3) were monophyletic in analyses of all the ETS types, and

ETS type 1 had a deeper coalescence point than ETS type 3 in our analysis. Also, of the 275 successful ETS amplifications, ETS type 1 was found in 180 of these (Chapter 2). Due to the broad distribution, monophyletic nature, and deep coalescence of ETS type 1, we used ETS type 1 sequences for phylogeny reconstruction in *Cercocarpus*.

The total data set included 55 ETS sequences with an aligned length of 977 bp. We included all 53 ETS type 1 sequences, generated from 47 individuals representing 16 taxa of *Cercocarpus,* and two ETS sequences from *Purshia plicata* and *Purshia tridentata* served as outgroups for this analysis (Table 3.1). Six *Cercocarpus* individuals (populations 27C, 62D, 65D, 78A, 93C, and 111A) had two cloned sequences each in the data matrix. .

ETS type 1 products were amplified using the forward primer *Cerc-ETS1f*, located at the transcription initiation site (TIS), or the *Cerc-ETS2f* forward primer, located 14 bp downstream of *Cerc-ETS1f*, and the reverse primer *18S-2L* located in the 5' end of the 18S gene (Chapter 2). Fifty microliter PCRs were performed, as described in Chapter 2, using twenty nanograms of template DNA. The PCRs were overlaid with mineral oil and run on a MJ Research Thermal Cycler. Three microliters of the reaction products were run on a 1.5 % agarose gel in 0.5X TBE buffer. PCR products were visualized with ethidium bromide under UV light and purified using the QIAquick PCR purification kit (QIAGEN). DNA sequencing was performed on an ABI Prism 377 or an ABI Prism 3100 automated sequencer using standard reagents. Sequences were proofread and then aligned using Clustal W in MegAlign (DNAstar, 1998), followed by adjustment by eye.

Phylogenetic analysis of the ETS

Maximum parsimony and maximum likelihood analyses were performed using the software package PAUP∗ version 4.0b8 (Swofford 1999). For our maximum parsimony analysis, we employed a heuristic search strategy using 10 random replicates, TBR branch-swapping and the MULTREES optimization. We set MAXTREES to 15,000. Support for branches was evaluated using bootstrap analysis (Felsenstein 1985) and random sequence addition for 100 replicates, using the same parameters.

For our maximum likelihood analysis, we used the HKY85 model with two substitution types, a transition-transversion ratio of 2, use of empirical base frequencies derived from the alignment, and no *a priori* site-specific variation. We used a heuristic search strategy using 10 random replicates and TBR branch-swapping.

Tests for incongruence

To assess the incongruence between the ETS and AFLP datasets, we conducted the "compare-2" test (Faith et al. 1991), the Templeton test (Templeton 1983), and the Kishino-Hasegawa test as implemented in PAUP∗ version 4.0b8 (Swofford 1999). To compare the AFLP dataset to the ETS dataset, we had to account for the two cases of paraphyletic ETS type 1 clones from the individuals *C. ledifolius* var. *intermontanus* 111A (clone 4 and 5) and *C. betuloides* var. *betuloides* 62DA (clones 2 and 3) (see Results below). Therefore, we constructed 4 ETS datasets, which differed in the clones that represented the individuals 111A and 62D in all possible combinations (Table 3.3). The compare-2 test examines the number of steps required by a dataset on two different constraint trees. For this study, the first constraint tree was the MP bootstrap consensus tree containing only branches with greater than 50% support from the dataset being tested. The second constraint tree, the test constraint tree, was the rival bootstrap consensus tree using the other dataset.

Table 3.3: A description of the clones included in each of the ETS datasets tested against the cpDNA dataset. 111A refers to the individual *C. ledifolius* var. *intermontanus* collection 111A (Table 3.1). 62D refers to *C. betuloides* var. *betuloides* collection 62D (Table 3.1).

	111A clone 4	111A clone 5	62D clone 2	62D clone 3
ETS dataset 1	included	not included	included	not included
ETS dataset 2	included	not included	not included	included
ETS dataset 3	not included	included	included	not included
ETS dataset 4	not included	included	not included	included

The Templeton test compares the number of steps each individual character needs to fit on two rival tree topologies, and then uses a Wilcoxian sign ranks (WSR) test to ascertain significance. The Kishino-Hasegawa is a parametric test comparing the difference in tree lengths of the dataset on the two rival tree topologies to the variance in the tree length difference over many sites. Significance is determined by comparing the original length difference to a distribution of tree length differences. The distribution is created from examining all the sites where the tree length differs between the two constraint trees, subtracting a mean difference and squaring. If the dataset supports both topologies equally, we would expect tree length differences to be distributed normally with a mean of zero. The trees used in the Templeton and the Kishino-Hasegawa tests are the same as the constraint trees used with the compare-2 test.

Testing the effect of different combinations of individuals representing taxa on tree topology

We constructed MP trees based on reduced AFLP and ETS datasets to determine if reduced sampling within taxa would have significantly altered our conclusions about tree topology. For the AFLP data, we created three reduced datasets (A, B, and C), each including 6 individuals, one for each of the taxa (Table 3.4). For the MP analyses on the three reduced AFLP datasets, we conducted exhaustive searches in PAUP∗ version 4.0b8 based on the 376 presence/absence characters. We also measured branch support with a bootstrap analysis using 1000 replicates.

Table 3.4: The identity of the individuals used in the reduced AFLP datasets A, B, and C. The numbers in parentheses are the collection number. The collection location and voucher number can be found in Table 3.1.

Species	AFLP Dataset A	AFLP Dataset B	AFLP Dataset C
C. montanus	*C. montanus* var. *montanus* (5)	*C. montanus* var. *montanus* (5)	*C. montanus* var. *glaber* (17)
C. betuloides	*C. betuloides* var. *blancheae* (123G)	*C. betuloides* var. *betuloides* (65D)	*C. betuloides* var. *blancheae* (123G)
C. ledifolius	*C. ledifolius* var. *intermontanus* (102B)	*C. ledifolius* var. *intermontanus* (102B)	*C. ledifolius* var. *intermontanus* (102B)
C. macrophyllus	*C. macrophyllus* (26)	*C. macrophyllus* (26)	*C. macrophyllus* (26)
C. fothergilloides	*C. fothergilloides* var. *medranoanus* (82A)	*C. fothergilloides* var. *medranoanus* (82A)	*C. fothergilloides* var. *medranoanus* (82A)
Purshia tridentata	*P. tridentata* (34)	*P. tridentata* (34)	*P. tridentata* (34)

For the ETS data, we created two reduced datasets (D and E), each including 15 individuals, one for each of the taxa (Table 3.5). Five of the 15 individuals (individuals 131, 123G, 57B, 126E, and 34) were the same in each dataset because we had no other ETS sequences to represent that taxon. We conducted a MP heurisitic search analysis on the two ETS reduced datasets based on 977 aligned bases. Bootstrap support was determined using 100 replicates.

Table 3.5: The identity of the individuals used in the reduced ETS datasets D and E. The numbers in the columns are the collection numbers. The collection location and voucher number can be found in Table 3.1.

taxon	ETS reduced dataset D	ETS reduced dataset E
C. montanus var. *montanus*	28A	28A
C. montanus var. *paucidentatus*	1	30A
C. montanus var. *glaber*	3	14
C. montanus var. argenteus	131	131
C. betuloides var. *betuloides*	93C	128C
C. betuloides var. *blancheae*	123G	123G
C. minutiflorus	57B	57B
C. traskiae	126E	126E
C. fothergilloides	var. *fothergilloides* 67B	var. *medranoanus* 82A
C. ledifolius var. *intermontanus*	102B	101A
C. ledifolius var. *ledifolius*	104	41B
C. intricatus	25	25
Purshia tridentata	34	34

RESULTS

Relationships in *Cercocarpus* based on AFLP characters

NJ analysis

The NJ analysis produced a phenogram that contained four clusters, corresponding with some exceptions to the four species complexes proposed by Lis (1992) (Figure 3.1). Cluster 1 consisted of members identified as belonging to the *Cercocarpus montanus* complex, with the exception of the inclusion of five individuals of the *C. ledifolius complex* highlighted by arrows in Figure3.1. There was no support for the unity of any of the *Cercocarpus montanus* varieties described by Martin (1950), although there was support for the clustering of three of the four *C. montanus* var. *montanus* accessions (5, 8, and 10). Cluster 2 contained individuals identified as belonging to the *Cercocarpus betuloides* complex, with the inclusion of one individual of *C. fothergilloides* var. *fothergilloides* (18), one individual of *C. mexicanus* (134A), two individuals of *C. montanus* (variety *glaber* 17 and variety *paucidentatus* 75A), and two individuals of *C. traskiae* (126A and E) all identified by arrows in the Figure 3.1. The two individuals of *Cercocarpus minitiflorus* clustered together with moderate bootstrap support.

Cluster 3 consisted exclusively of individuals identified as belonging to the *Cercocarpus ledifolius* species complex. There was no support for the clustering of *Cercocarpus intricatus* or the varieties of *C. ledifolius*. Cluster 4 consisted exclusively of individuals representing the *Cercocarpus fothergilloides* complex and the newly described species *C. rzedowskii.*

The only cluster in the distance analysis that had bootstrap support above 50 was cluster 4 and the outgroup *Purshia* (Figure 3.1). The *Cercocarpus montanus* cluster 1 was found in the most derived position, sister to the *C. betuloides* cluster 2. The *Cercocarpus ledifolius* cluster 3 was found sister to both cluster 1 and 2. Cluster 4 sits sister to cluster 1, 2, and 3.

ME analysis

The minimum evolution distance analysis also yielded four clusters that corresponded to the four species complexes, although the relationship between the clusters was different (Figure 3.2). Cluster 1 contained individuals identified exclusively as belonging to the *Cercocarpus montanus* complex, although I did not find any support for the varieties as described by Martin (1950). Cluster 2, found sister to cluster 1, contained individuals of the *Cercocarpus ledifolius* complex and one individual identified as *C. montanus* var. *paucidentatus* (HEN1) identified by an arrow in Figure 3.2. As in the NJ analysis, we found no support for the varieties of *Cercocarpus ledifolius* (Holmgren 1987) or *C. intricatus*. Cluster 3, which was found sister to clusters 1 and 2, contained individuals of the *Cercocarpus betuloides* complex, with the inclusion of one individual of *C. fothergilloides* var. *fothergilloides* (18), one individual of *C. mexicanus* (134A), two individuals of *C. montanus* (variety *glaber* 17 and variety *paucidentatus*

75A), and two individuals of *C. traskiae* (126A and E) all identified by arrows in the Figure 3.2. The two individuals of *Cercocarpus minutiflorus* once again group together. The final cluster, cluster 4, contained individuals of the *Cercocarpus fothergilloides* complex. We found *Cercocarpus rzedowskii* grouped in this cluster with high bootstrap support.

MP analysis

The parsimony analysis resulted in one most-parsimonious tree of length 2643, consistency index of 0.134, and retention index of 0.484 (Figure 3.3). The parsimony tree contained four main groups. Clade 1 contained almost all members of *Cercocarpus montanus*, but did not support the monophyly of any of the varieties. Two individuals of *Cercocarpus montanus* (var. *paucidentatus* 75A and var. *glaber* 17) were found nested within the *C. betuloides* grade (Figure 3.3). Clade 1 also contained two individuals identified as *Cercocarpus ledifolius* var. *intermontanus* (individuals 37C and 98C), highlighted by arrows in Figure 3.3. Members of the *Cercocarpus betuloides* complex were found in a phylogenetic grade basal to clade 1. The two individuals of *Cercocarpus traskiae* were found in this grade, but not in a monophyletic group. The varieties of *Cercocarpus betuloides* were also not supported. Only *Cercocarpus minutiflorus* was supported as a monopyletic group with weak bootstrap support. The *Cercocarpus betuloides* grade also included one individual of *C. fothergilloides* var. *mojadensis* (individual 18), one individual of *C. mexicanus* (individual 134A), and the two individuals of the *C. montanus* complex, all highlighted by arrows in Figure 3.3. Clade 2 was well supported, containing only individuals representing the Mexican taxa *Cercocarpus rzedowskii*, *C. fothergilloides*, and *C. macrophyllus,* but not *C. traskiae*. The species *Cercocarpus rzedowskii* was supported as a monophyletic group. *Cercocarpus macrophyllus* and *C. fothergilloides* were found to be paraphyletic. Clade 2 was the only major clade that had a bootstrap score above 50. Clade 3 contained individuals representing the *Cercocarpus ledifolius* complex. There was no support for the monophyly of the species *Cercocarpus intricatus* or the varieties of *C. ledifolius.*

In the MP analysis, the *Cercocarpus montanus* clade and *C. betuloides* grade form a monophyletic clade, but very weakly supported (bootstrap score less than 50). The *Cercocarpus fothergilloides* clade 2 was sister to the *C. ledifolius* clade 3, which together was sister to the *C. montanus* clade 1 and *C. betuloides* grade.

The three analyses (NJ, ME and MP) of the full AFLP dataset that included all 153 individuals support the four species complexes as described by Lis (1992) (Figure 3.4). All three analyses also supported the *C. fothergilloides* complex clustering at the base of the tree. The three analyses differed in the overall relationships between clusters. The NJ cluster analysis and the MP analysis supported a close relationship between the *Cercocarpus montanus* complex and the *C. betuloides* complex in the most derived position. The ME analysis, in contrast, grouped the *Cercocarpus montanus* and *C. ledifolius* clusters in the most derived position.

In all three analyses, the *Cercocarpus betuloides* cluster consistently contained three members of the *C. fothergilloides* complex as identified by Lis (1992) (*C. fothergilloides* var. *mojadensis* (18), *C. traskiae* 126A and 126E), *C. mexicanus* (134A), and two individuals identified as *C. montanus* (variety *glaber* 17 and variety *paucidentatus* 75A).

Relationships in *Cercocarpus* based on ETS characters

MP analysis

The aligned dataset of 977 characters contained 147 parsimony informative characters for the 53 sequences included (Appendix D). The heuristic search resulted in 15,000 equally most parsimonious trees of 456 steps (Figure 3.5). Clade A contained a diverse set of taxa including representatives of *Cercocarpus montanus* and all the individuals of the Mexican taxa *C. rzedowskii*. Clade B contained multiple members of *Cercocarpus montanus*, *C. betuloides,* and *C. macrophyllus* and one individual of *C. fothergilloides* (67B) and one individual of *C. ledifolius* (individual 108C). Clade C was highly supported, consisting of two individuals of the *Cercocarpus ledifolius* complex (individual 25 and 41B). Clades D and E were also highly supported, containing individuals of the *Cercocarpus betuloides* complex. Clade F was weakly supported, containing individuals of the *Cercocarpus ledifolius* species complex, one individual of *C. macrophyllus* (26) and one individual of *C. fothergilloides* (82A). The ETS sequence from one individual of *Cercocarpus ledifolius* var. *intermontanus* (111A clone 5) was found in an unresolved position with clades C, D, and E.

Clades A and B were sister to each other and are in the most derived position in the MP analysis. Clades C, D, and E were in unresolved positions, sister to clades A and B. Clade F was found in the most basal position. The only deep node that was supported was the branch leading to clades A and B (Figure 3.5).

ML analysis

The maximum likelihood analysis did not swap to completion, but was stopped with two most likely trees with likelihood scores of -Ln 4223.56020 in memory (Figure 3.5). The dashed line shows where the two most likely trees differed. Overall, the clades (clades A -F) recovered in the maximum parsimony analysis were supported by the ML analysis, however, the relationships between the clades were not consistent between the MP and ML reconstructions. In the ML analysis, clades A and B were again found in the most derived position. Clade F was recovered as sister to clades A and B, and included the ETS sequence from *Cercocarpus ledifolius* var. *intermontanus* (111A clone 5). Clade C was found more basal to clade F followed by clade E and D. Although the MP and

ML trees differed in topology, the branches that differ between reconstructions were either unresolved or have bootstrap support below 50 in the MP tree. This suggests there is little reason to prefer one topology over another.

Comparison of ETS and AFLP trees

The maximum parsimony analysis on the AFLP presence/absence data subset, which included just the 49 accessions matching the individuals sampled for the ETS dataset, resulted in 24 equally parsimonious trees of length 802. The strict consensus of these 24 trees is in Figure 3.6 along side the strict consensus tree of over 10,000 most parsimonious trees recovered from the ETS dataset. The AFLP MP strict consensus tree contained 4 major clades. Clade A contained almost all the individuals identified as part of *Cercocarpus montanus*, but did not support the varieties as described by Martin (1950). Clade A also contained two individuals of the *Cercocarpus betuloides* complex (*C. betuloides* var. *betuloides* 93C, and *C. minutiflorus* 57B). Clade B contained all of the members of the *Cercocarpus ledifolius* complex, but did not support the monophyly of the species *C. intricatus* or the varieties *of C. ledifolius* (Holmgren 1992). Clade B also contained two individuals of the *Cercocarpus betuloides* complex (*C. betuloides* var. *betuloides* 65D, and *C. betuloides* var. *blancheae* 123G). Clade C contained all of the strictly Mexican taxa of *Cercocarpus*. Clade D contained 2 individuals of *Cercocarpus montanus* (*C. montanus* var. *glaber* 17 and *C. montanus* var *paucidentatus* 75A), 2 individuals of the *C. betuloides* complex (*C. betuloides* var. *betuloides* 128A and 120C), *C. traskiae* 126E, and *C mexicanus* 134A.

The MP AFLP strict consensus was not congruent with the MP or ML ETS strict consensus phylogenies (Figure 3.6). Members of clade A from the MP AFLP strict consensus tree are found in 3 clades of the MP ETS strict consensus tree. Members of clade B are found in 5 separate ETS clades. Members of clade C were found 2 of the ETS clades and members of clade D were found in 4 clades from the ETS strict consensus tree.

Incongruence results

The results of the compare-2, the Templeton, and the Kishino-Hasegawa tests can be seen in Table 3.6. For the compare-2 test, 100 permutations of each dataset were tested on each constraint tree. We found significant results using both datasets, where the difference in fit for the dataset on its own tree versus the constraint tree was significantly better than the difference with permuted data. The Templeton and Kishino-Hasegawa tests yielded highly significant results for each dataset, signaling significant incongruence between both datasets tested.

Table 3.6: Results of the incongruence tests conducted on the AFLP and ETS datasets. ETS 1, ETS 2, ETS 3, and ETS 4 datasets refer to Table 3.5. The constraint trees used for these tests were bootstrap consensus trees containing only branches that had support over 50%.

COMPARE-2		**KISHINO-HASEGAWA**				**TEMPLETON**			
	p-value	tree length difference	s.d.	t	p-value	N	+ sign ranks	- sign ranks	p-value
AFLP vs ETS 1									
AFLP dataset	p=0.01	133	24.73017	5.3780	p=0.0001	88	3188.5	727.5	p=0.0001
ETS 1 dataset	p=0.01	290	37.78070	7.6759	p=0.0001	98	4833.5	17.5	p=0.0001
AFLP vs ETS 2									
AFLP dataset	p=0.01	126	23.81775	5.2902	p=0.0001	88	3176.5	739.5	p=0.0001
ETS 2 dataset	p=0.01	287	37.60504	7.6320	p=0.0001	96	4639.5	16.5	p=0.0001
AFLP vs ETS 3									
AFLP dataset	p=0.01	117	24.04435	4.8660	p=0.0001	88	3069	847	p=0.0001
ETS 3 dataset	p=0.01	283	37.36917	7.5731	p=0.0001	95	4544	16	p=0.0001
AFLP vs ETS 4									
AFLP dataset	p=0.01	119	24.87939	4.7831	p=0.0001	87	2993.5	834.5	p=0.0001
ETS 4 dataset	p=0.01	283	37.87471	7.4720	p=0.0001	94	4448.5	16.5	p=0.0001

The effect of taxon sampling on tree topology

The three AFLP reduced datasets produced three different tree topologies (Figure 3.7 A, B, C), which differed with respect to the relationships between *Cercocarpus montanus*, *C. betuloides*, and *C. ledifolius,* and for which all possible topological relationships were recovered. The MP tree produced from AFLP reduced dataset A was the only one that did not have bootstrap support above 50 for all branches. The two ETS reduced datasets produced different tree topologies as well. We found no common clades between the trees produced from ETS datasets D and E.

DISCUSSION

The goals of this project were to generate molecular characters to help define species and varietal boundaries in *Cercocarpus* and to test some hypotheses generated by other authors about the evolution and taxonomic affinities of certain taxa. The three analyses of the full AFLP dataset generated support for four species complexes in *Cercocarpus* (Lis 1992). The ETS analyses did not support these complexes. Phylogenetic trees produced from the AFLP and ETS datasets were significantly incongruent. We also found that taxon sampling could have had a large effect on tree topology.

Taxonomic implications

The *betuloides* complex

Martin (1950) considered the taxa *Cercocarpus betuloides*, *C. minutiflorus* to be varieties of *C. montanus*. Lis (1992) reorganized these taxa after a detailed study of the leaf architecture. Lis recognized *Cercocarpus betuloides* as having three varieties, two of which are included in this study (*C. betuloides* var. *betuloides* and *C. betuloides* var. *blancheae*). He also recognized *Cercocarpus minutiflorus* at the specific level. We found that these taxa were distinct from the

C. montanus complex in the three AFLP analyses, however, they were found at the base of a clade including the *C. montanus* complex in the AFLP MP analysis and sister to the *C. montanus* cluster in the NJ and suggesting a close relationship. We only had two samples of *Cercocarpus minutiflorus* in our AFLP data, yet in all three AFLP analyses we found those two samples united with moderate bootstrap support. Lis (1992) found that *Cercocarpus minutiflorus* shared leaf architecture character states with both *C. traskiae* and *C. betuloides*, but was reluctant to assign it as a variety of either species. *Cercocarpus minutiflorus* consistently grouped with *C. betuloides* in our analyses. *Cercocarpus minutiflorus* differs from *C. betuloides* in its distribution (restricted range from San Diego to northern Baja California, Mexico) and glabrous leaf characteristics. The data presented here would suggest *Cercocarpus minutiflorus* may be a distinct group, but appears to be closely related to *C. betuloides*.

We found no support for the varieties of *Cercocarpus betuloides* defined by Lis (1993). This is a surprising result because Searcy (1967) found that these taxa retained their leaf traits in a common garden experiment, and hypothesized distinct genotypes for the two groups.

The *ledifolius* complex

Holmgren (1987) recognized two varieties of *Cercocarpus ledifolius*, *C. ledifolius* var. *intermontanus* and *C. ledifolius* var. *ledifolius*. The characters used to distinguish them consisted of leaf size, geographic location, and leaf vestiture

(Table 3.2). *Cercocarpus intricatus* has been raised to specific status in many of the western floras, a decision Holmgren (1987) agreed with, based on the divergent leaf characteristics and ecological requirements. Based on the AFLP and ETS data presented above, we have no support for recognizing the two varieties of *Cercocarpus ledifolius* or *C. intricatus*.

The continuous variation seen between *Cercocarpus ledifolius* and *C. intricatus* has been postulated as evidence of hybridization and hybrid swarms between the two taxa (Schneider 1905b, Brayton and Mooney 1966, Davis and Brotherson 1991). In his revision of *Cercocarpus*, Schneider (1905b) believed the intermediate forms may have been of hybrid origin, and he designated them as *Cercocarpus ledifolius* var. *intercedens*. This taxon was later synomnized with *Cercocarpus ledifolius* var. *ledifolius*. If we accept this hypothesis, we would expect each parental type (*Cercocarpus ledifolius* var. *intermontanus* and *C. intricatus*) to cluster independently in the NJ and ME phenograms. The intermediate leaf shape (*Cercocarpus ledifolius* var. *ledifolius*) would be expected to cluster in between or at the base of the parental taxa due to the presence of both sets of AFLP markers in the hybrid individual. Further, instead of clustering in between, the individual with intermediate morphology may group with one of the parents if considerable backcrossing had occurred. We do not see this distribution of individuals in any of our AFLP or ETS reconstructions. Instead we see a single cluster (with some exceptions) in the NJ and ME AFLP trees, and a single clade in the MP AFLP tree, which would the suggest leaf diversity seen in *Cercocarpus* might be the result of ecotype variation, where microhabitat factors have selected certain leaf shapes.

The ***montanus*** complex

Three of the four individuals of *Cercocarpus montanus* var. *montanus* (5, 8, and 10) form a moderately supported cluster (bootstrap value 65) in the AFLP NJ and ME trees and a moderately supported clade (bootstrap value 66) in the AFLP MP tree (Figure 3.1, 3.2, and 3.3). The fourth individual of *Cercocarpus montanus* var. *montanus,* individual 28A, was found in the larger *C. montanus* cluster in the NJ and ME trees and the larger clade (*C. montanus* clade 1) in the AFLP MP tree, but is not directly a part of a monophyletic group. In the phylogeneies based on the ETS dataset, all members of *Cercocarpus montanus* var. *montanus* were found in an unresolved clade (part of clade B) in both the MP and ML analyses. These results suggest that *Cercocarpus montanus* var. *montanus* may indeed be a real taxonomic entity, although the support for this claim is weak. *Cercocarpus montanus* var. *montanus* is winter deciduous, a character not listed in Table 3.2, which separates it from the other varieties of *C. montanus*.

Cercocarpus montanus var. *paucidentatus* and *C. montanus var. glaber* do not unite in monophyletic groups in any analysis. As illustrated in Table 3.2, these taxa overlap in many of the characters used to differentiate them. Although not studied in *Cercocarpus montanus*, leaf characters have been shown to vary based on environment in *C. betuloides* (Searcy 1977) and *C. ledifolius* (Davis and Brotherson 1991). We argue, based on the data presented here, that the varieties currently defined by Martin (1950) are not real taxonomic entities.

The *fothergilloides* complex

Richard Lis (1992) described a relationship between the taxa *Cercocarpus fothergilloides*, *C. mojadensis, C. pringlei*, *C. macrophyllus* and *C. traskiae* based on leaf characters. He proposed that these taxa were all part of one species complex he called the *C. fothergilloides* complex. James Henrickson is currently revising this group and has recognized the same overall taxa, making a few taxonomic changes (Madroño ined.), including the recognition of *Cercocarpus mojadensis* as a variety of *C. fothergilloides*. In our study, based on the AFLP data, *Cercocarpus fothergilloides* var. *mojadensis* (18) and *C. traskiae* (126A and E) consistently grouped with the *betuloides* complex. The other members of this complex, *Cercocarpus fothergilloides* var. *medranoanus*, variety *fothergilloides*, and *C. macrophyllus* consistently grouped together with high bootstrap support. This result gives strength to Lis' claim that the *Cercocarpus fothergilloides* complex exists, if one removes *C. traskiae*, and possibly, *C. fothergilloides* var. *mojadensis*, but more expanded sampling is needed before any taxonomic decisions are warranted.

The ETS dataset did not support this complex as currently defined. *Cercocarpus fothergilloides* var. *fothergilloides* (67B) and *C. macrophyllus* (69A and 69B) grouped with members of *C. montanus. Cercocarpus fothergilloides* var. *medranoanus* (82A) and *C. macrophyllus* (26) grouped with *C. ledifolius* and *C. intricatus*.

Cercocarpus rzedowskii and *C. mexicanus*

In 1987, Henrickson described two new species of *Cercocarpus*, *C. rzedowskii* from southern Tamaulipas, and *C. mexicanus* from southern Coahuila (see Figure 1.12). He suggested a close relationship between *Cercocarpus rzedowskii* and *C. ledifolius,* based on overall leaf morphology. He also postulated a possible relationship between *Cercocarpus mexicanus* and *C. montanus*. We did not find any support for these relationships. In the ETS phylogenies, *Cercocarpus rzedowskii* is nested deeply in a clade with members of the *C. montanus* complex, while in the AFLP trees, we find *C. rzedowskii* united with members of the *C. fothergilloides* complex. The one individual of *Cercocarpus mexicanus* (134A) included in this study is found deeply nested within the *C. betuloides* complex in all three analyses using the AFLP data.

ETS considerations

Within the ETS data set, we had six individuals that were represented by two cloned ETS sequences (Figure 3.4). In four of the six, the two clones united in a monophyletic group in both the MP and ML analyses. The other two, *Cercocarpus betuloides* var. *betuloides* (62D clone 2 and 3) and *C. ledifolius* var. *intermontanus* (111A clone 4 and 5) did not unite, although in the ML analysis, the two clones from 111A did fall out in the same clade. Explaining these results depends on how concerted evolution and gene flow (hybridization) are operating in the ETS region of the nrDNA in *Cercocarpus*. If concerted evolution is operating quickly (Wendel 1989), the divergent clones are most likely evidence of recent gene flow, especially if the two clones reside in different, highly supported clades. If concerted evolution is proceeding more slowly, the divergent clones could be evidence of multiple ETS types (orthologous) due to mutation. In this case, we would expect the different clones to inhabit the same clade, but not necessarily as sister to each other.

The two clones of *Cercocarpus betuloides* var. *betuloides* (62D) resided in clades (clades D and E) that were almost entirely composed of members of the *C. betuloides* complex. The relationship between the two clades is not resolved using our ETS dataset, so it is possible that the two clades (clades D and E) are sister clades. However, each clade is highly supported as independent and, therefore, it could be that the position of the two clones is a result of gene flow between different geographic populations of *Cercocarpus betuloides* present in clades D and E. Members of *Cercocarpus betuloides* var. *betuloides* present in clade D are located in the foothills of the Sierra, while members found in clade E are located in Arizona (Table 3.1) (please see Figure 1.10).

The two clones of *Cercocarpus ledifolius* var. *intermontanus* (111A clone 4 and 5) do not unite in the ETS MP tree, yet in the ETS ML analysis, the two clones reside in the same clade. Although unexpected, this result does not suggest hybridization as strongly as the results above. Multiple intra-specific ETS types have been reported in other studies (Linder et al. 2000) from *Argyranthemum* and *Helianthus*. The ETS types coalesced within a species, leading to the conclusion that concerted evolution was working, but not as rapid as in the ITS or coding regions of the rDNA repeat (Linder et al. 2000). The two clones of *Cercocarpus betuloides* var. *betuloides* (62D) appear to coalesce with *C. betuloides* species complex, but the clones from *C. ledifolius* var. *intermontanus* do not. We might explain the two clones recovered from *Cercocarpus ledifolius* var. *intermontanus* as evidence of multiple ETS types (orthologous) due to mutation, and the two clones were not homogenized by concerted evolution.

ETS versus AFLP

Direct comparison of the AFLP and ETS phylogenetic reconstructions for the same individuals showed a high degree of incongruence (Figure 3.5) (Table 3.6). Although the characters used to reconstruct the phylogenies came from different sources, such a high degree of incongruence was unexpected. The AFLP reconstruction recovered, for the most part, each of the species complexes that are currently defined by morphology and geography (Lis 1992), although it did not recover the varieties. The ETS phylogenies did not recover monophyletic species or varieties. We have identified three possible sources for this incongruence between the ETS and AFLP analyses. First, it may just be an artifact of the lack of real support for most of the branches in either analysis. Few branches in either the NJ, ME, or MP reconstructions of the AFLP dataset and some branches in the ETS MP reconstruction have bootstrap support over 50.

Second, the incongruence may be due to an excess of homoplasy in the AFLP dataset. Since homoplasy can be generated in an AFLP analysis if non-identical markers of equal length are assumed to be homologous - this homoplasy should be magnified the greater the evolutionary distance is between samples - incongruence between the AFLP and ETS trees could have been caused by these factors. We found that when we compared the NJ, ME, and MP reconstructions of the AFLP dataset in this study, the topologies were incongruent. Koopman et al. (2001) suggests that in these cases, it is important to examine the strength of support for branches in the MP tree by bootstrap support, assuming the incongruence is due to homoplasy and will be evident by having low bootstrap scores. Branches that are incongruent between trees that have low support should be discarded and not used for conclusions about species relationships. Based on this approach, only the *Cercocarpus fothergilloides* clade could be supported (Figure 3.4). It is clear that the AFLP dataset contains homoplasy and also contains only a few characters that are supporting the deeper branches in the MP reconstruction. The combination of these two traits leads to low bootstrap scores, and may have led to the incongruence we have seen.

Third, the incongruence between the ETS and AFLP trees could be a result of hybridization and lack of concerted evolution, in which ETS sequences we assume to be orthologous are not. Hybridization has been documented in *Cercocarpus* (Rieseberg 1989, 1994). Unlike the AFLP markers, the ETS dataset is a representation of a small non-coding area in the nrDNA. The phylogenetic history of this small non-coding region may be quite different than the species in which it is found. We have already cited examples above where ETS clones from a single individual did not unite in a monophyletic group, suggesting that hybridization and the maintenance of orthologous ETS sequences does occur in *Cercocarpus*, which may make the ETS an unsuitable nuclear marker for phylogenetic reconstruction of species relationships in *Cercocarpus*.

Taxon and population sampling in recently diverged groups

A major issue in phylogenetic inference is taxon sampling, which is usually defined as the selection of a small number of representative specimens for higher-level phylogenetic analysis used to infer evolutionary history. Insufficient taxon sampling is often cited as a major source of error in phylogenetic analysis (Murphy et al. 2001, Saunders and Edwards 2000, Soltis et al. 1999), and can occur because of constraints on time, money, or rarity (Rosenberg and Kumar

2001). The major argument for increased taxon sampling is that inclusion of more taxa counteracts long-branch attraction, where long, unrelated branches associate erroneously in the phylogenetic analysis. We argue here that in recently evolved groups, or groups with high rates of gene flow between taxa, increased population sampling within taxa is important for understanding the total variation present and properly delimitating species. If vaiation exists in the marker chosen for phylogenetic reconstruction that is shared between species, relationships between taxa can be reconstructed erroneously, because the individuals chosen to represent them (Wendel and Doyle 1998). These errors can be lessened if multiple populations are sampled.

In analyses based on both the ETS and AFLP datasets, we found that relationships between taxa and within a defined taxon depended on the individuals used to represent the taxon (Figure 3.6). MP trees created from different subsets of the taxa from the AFLP dataset (Figure 3.6) showed that we could reconstruct all three trees that represent the possible relationships of *C. betuloides*, *C. ledifolius*, and *C. montanus.* Even more striking, we produced two highly incongruent, but well supported MP trees based on different subsets of ETS sequences. These incongruencies can only be discovered because we sampled more than one individual per taxon, and highlight how increased population sampling within taxa is very important for correct inferences of both evolutionary history and taxon delimitations.

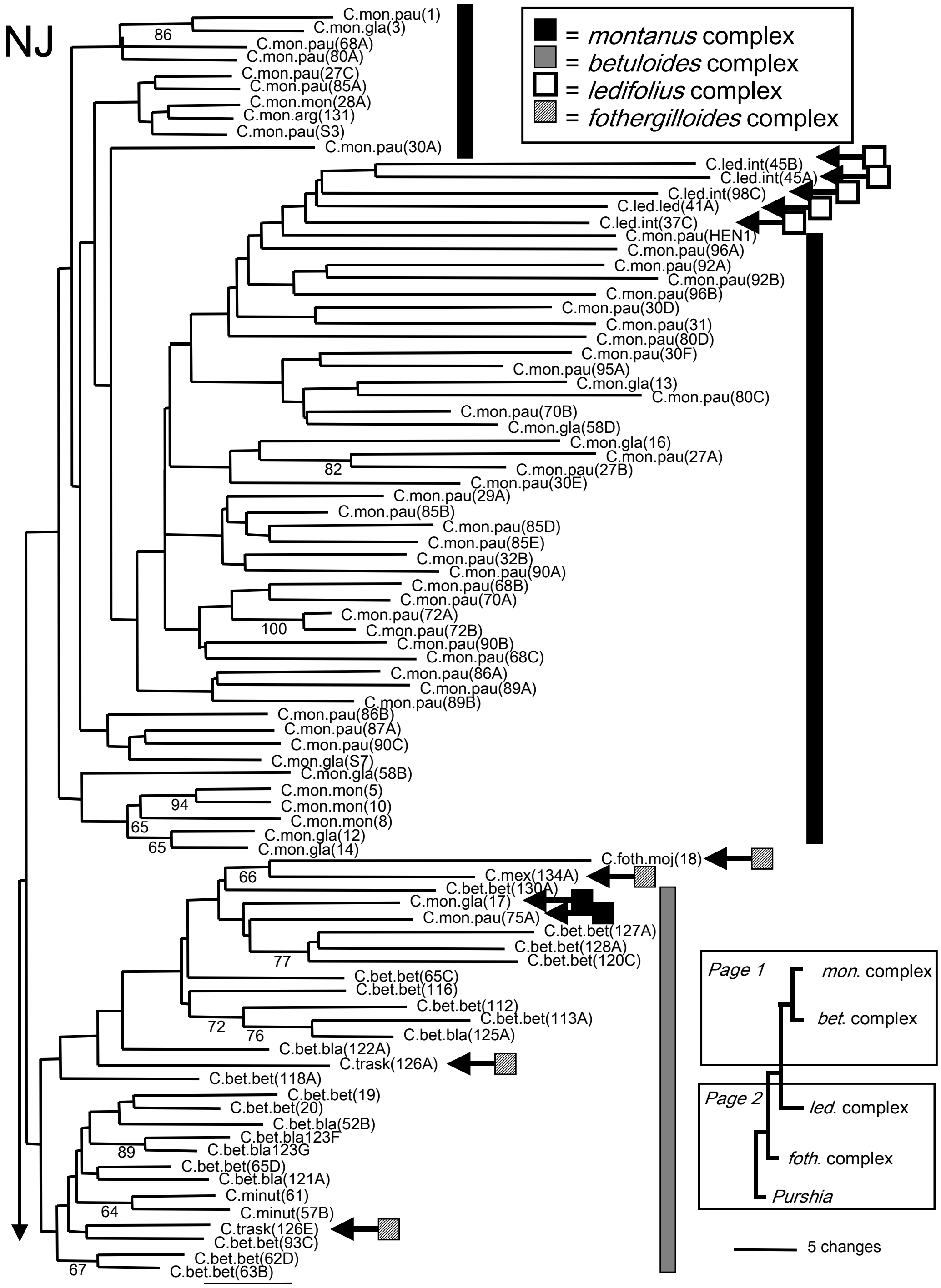


Figure 3.1. A neighbor joining (NJ) phenogram based on 376 presence/absence AFLP characters from three selective primer combinations for 153 individuals of *Cercocarpus*. The species names of the individuals have been abbreviated as in Figure 2.4. Numbers in parentheses refer to the collection number (Table 3.1). Numbers below the branches are a measure of bootstrap support. Only bootstrap values above 50% are reported.

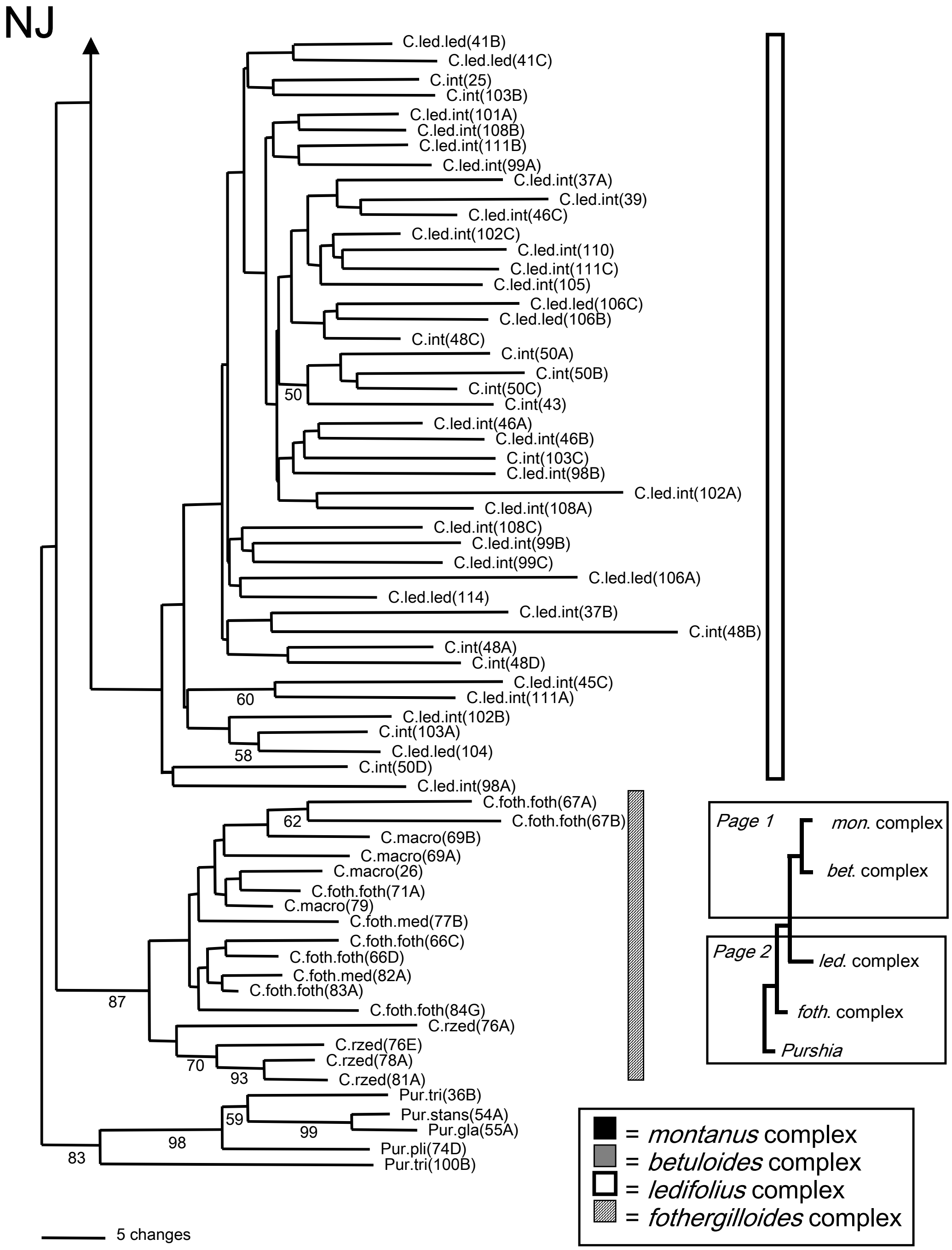


Figure 3.1: (continued)

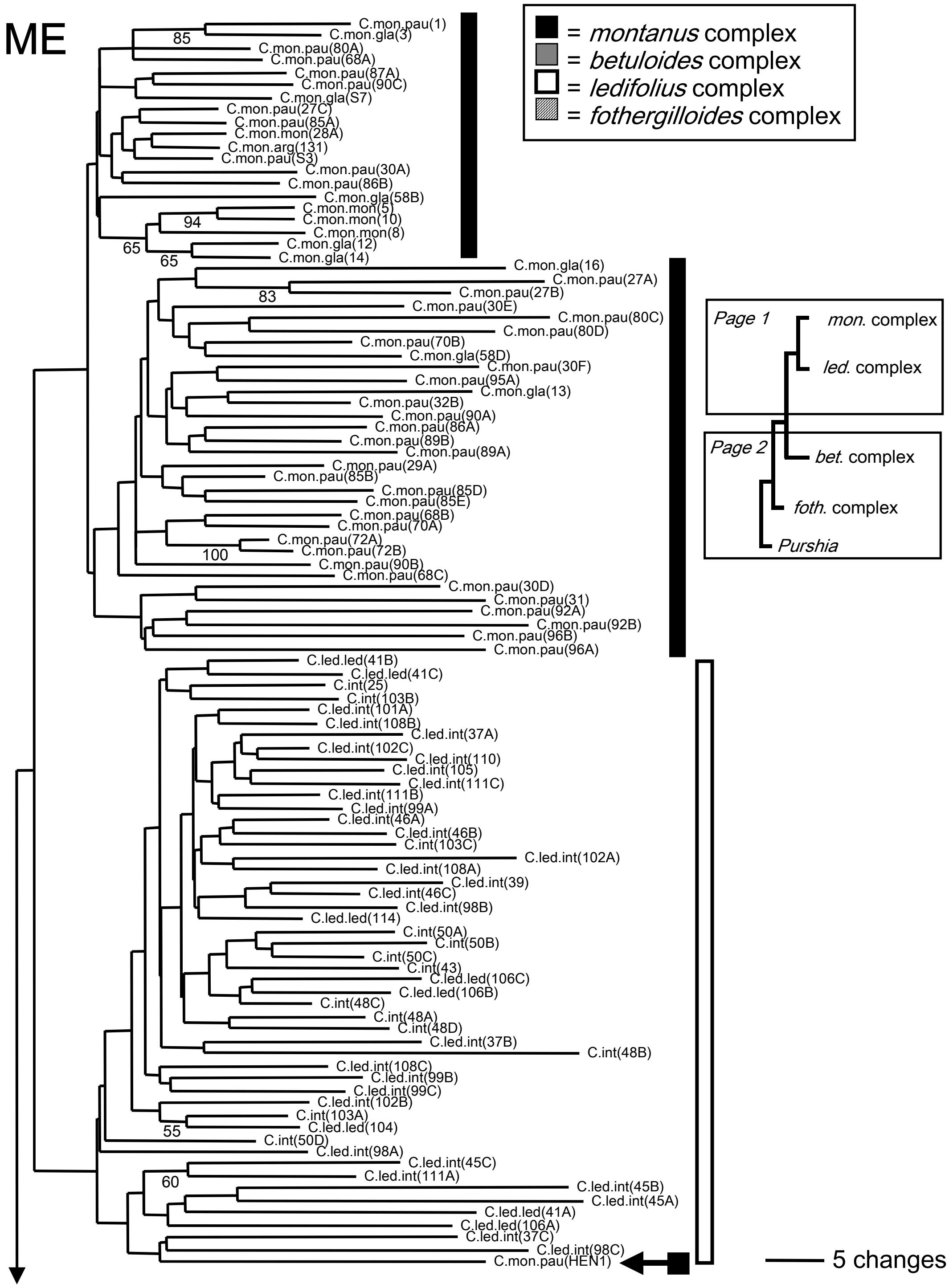


Figure 3.2. A minimum evolution (ME) phenogram based on 376 presence/absence AFLP characters from three selective primer combinations for 153 individuals of *Cercocarpus*. The species names of the individuals have been abbreviated as in Figure 2.4. Numbers in parentheses refer to the collection number (Table 3.1). Numbers below the branches are a measure of bootstrap support. Only bootstrap values above 50% are reported.

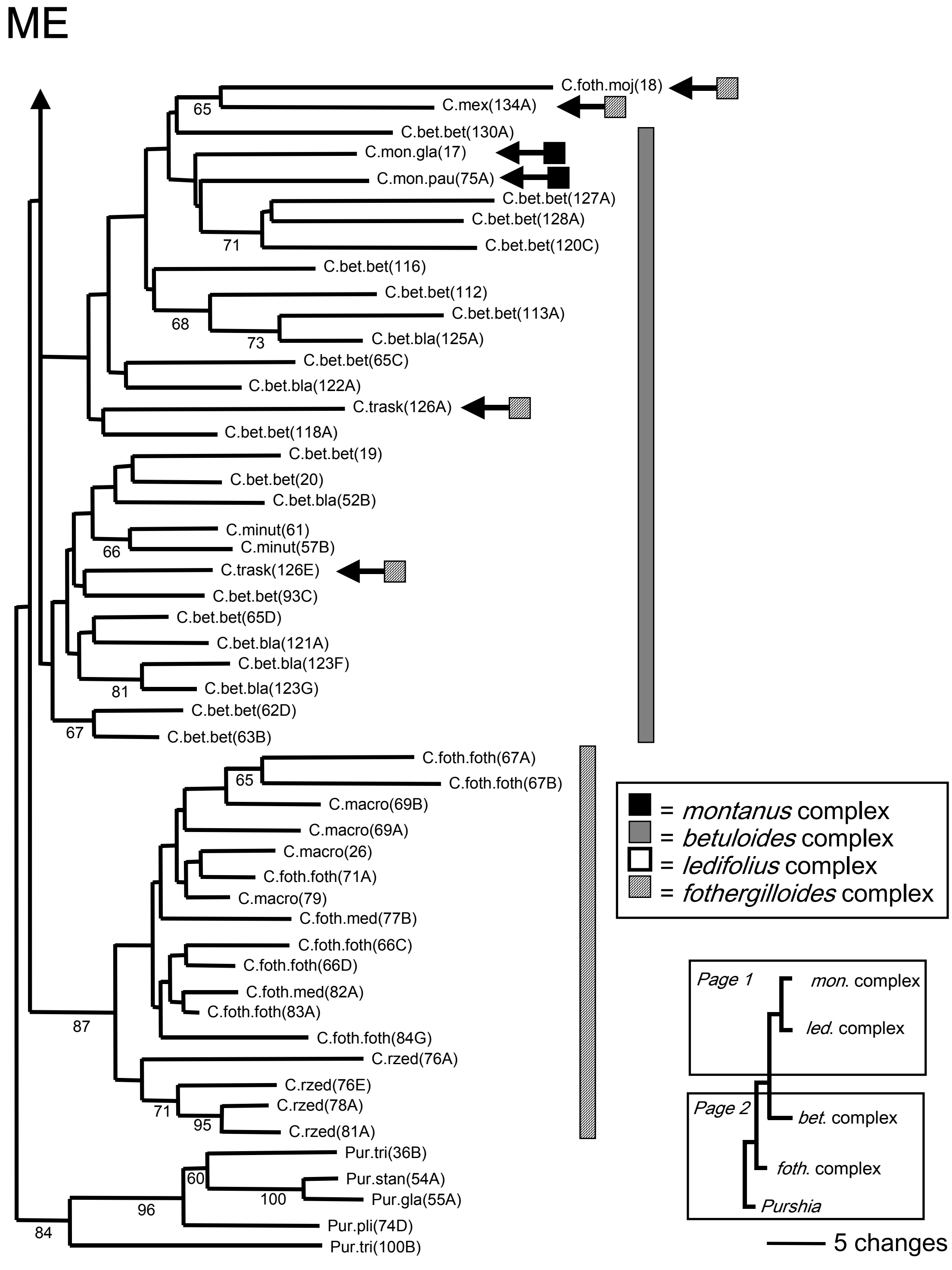


Figure 3.2: (continued)

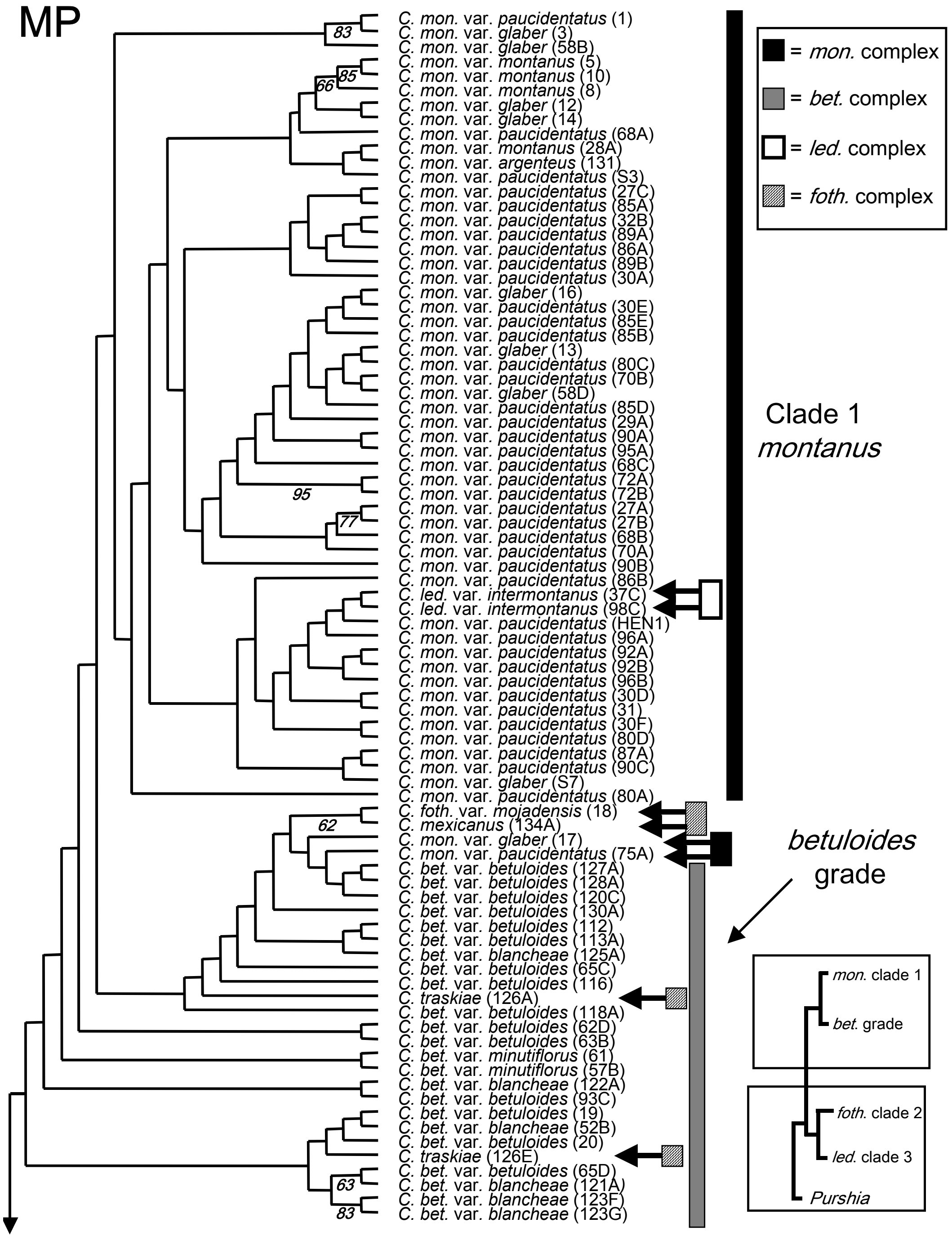


Figure 3.3. The single most parsimonious tree of length 2643 resulting from analysis of 376 presence/absence AFLP characters. Taxonomic identities have been abbreviated where C. = *Cercocarpus* and P. = *Purshia*. The three letters following are the first three letters of the species name. Numbers in parentheses refer to collection number (Table 3.1). The number of steps supporting the branch is shown above, the bootstrap support over 50 is shown below each branch.

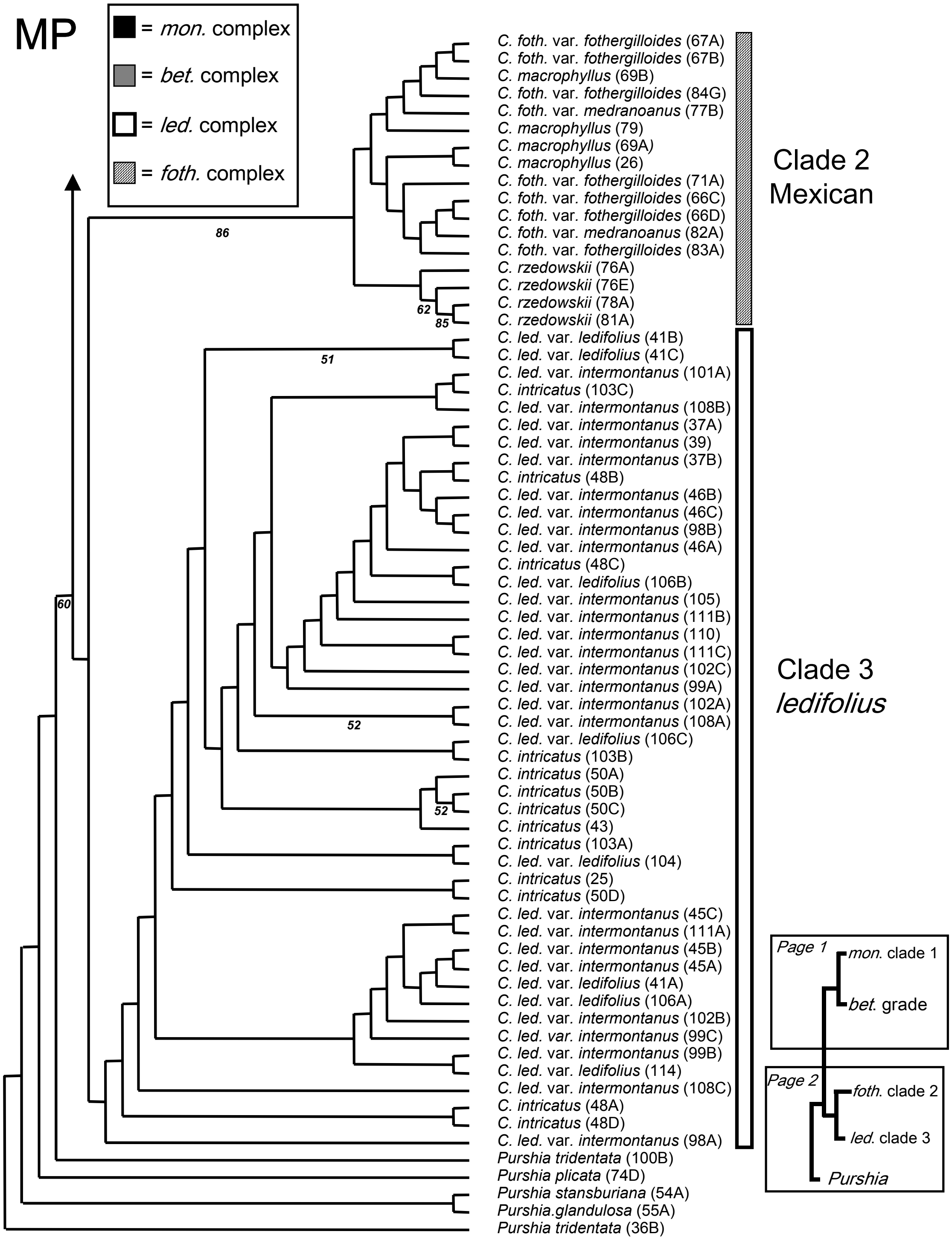


Figure 3.3: (continued)

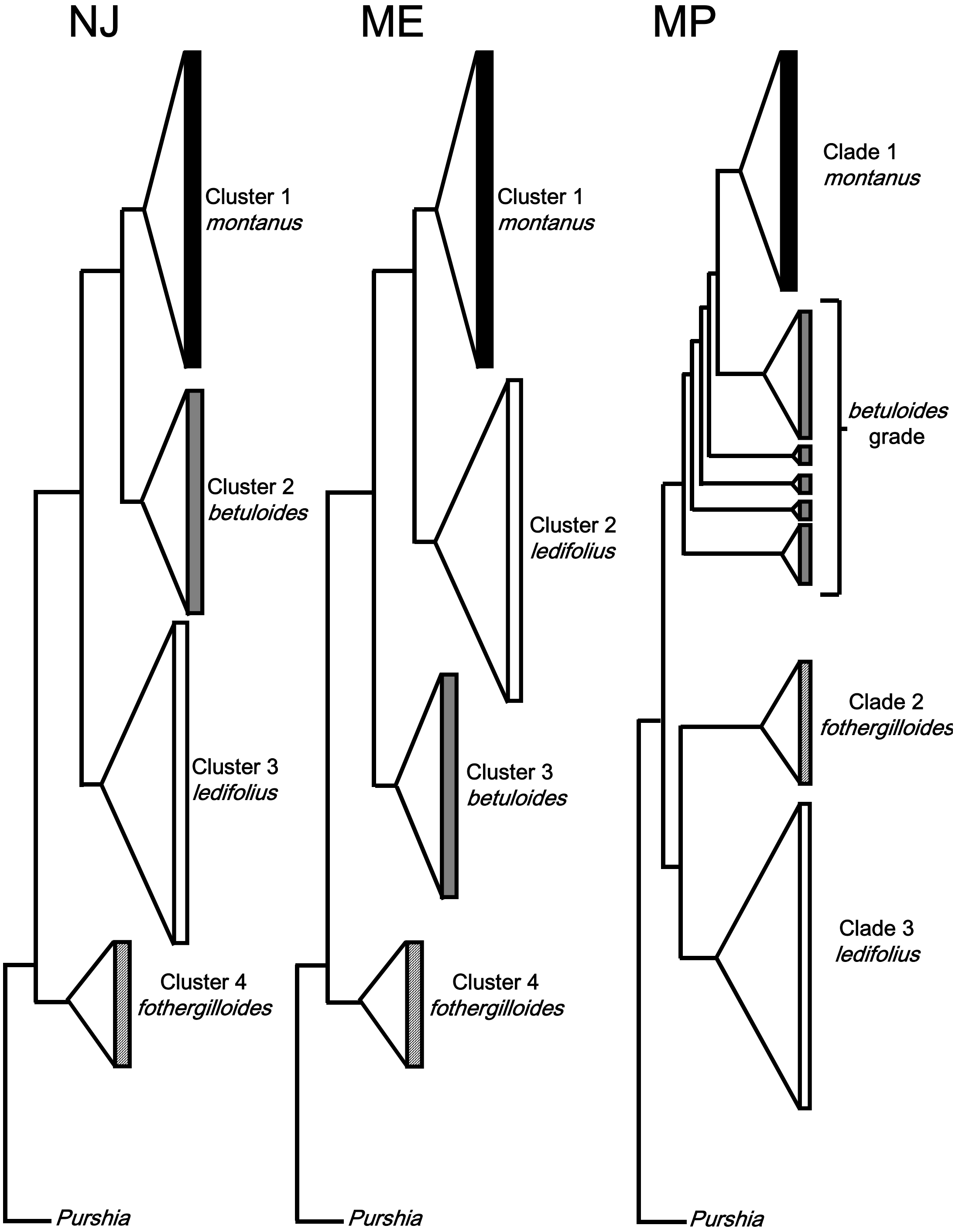


Figure 3.4. A comparison of the NJ phenogram, ME phenogram, and MP cladogram based on analyses of the AFLP characters. The three trees are stylized representations of Figures 3.1, 3.2, and 3.3. Colored bars are used to represent the 4 species complexes.

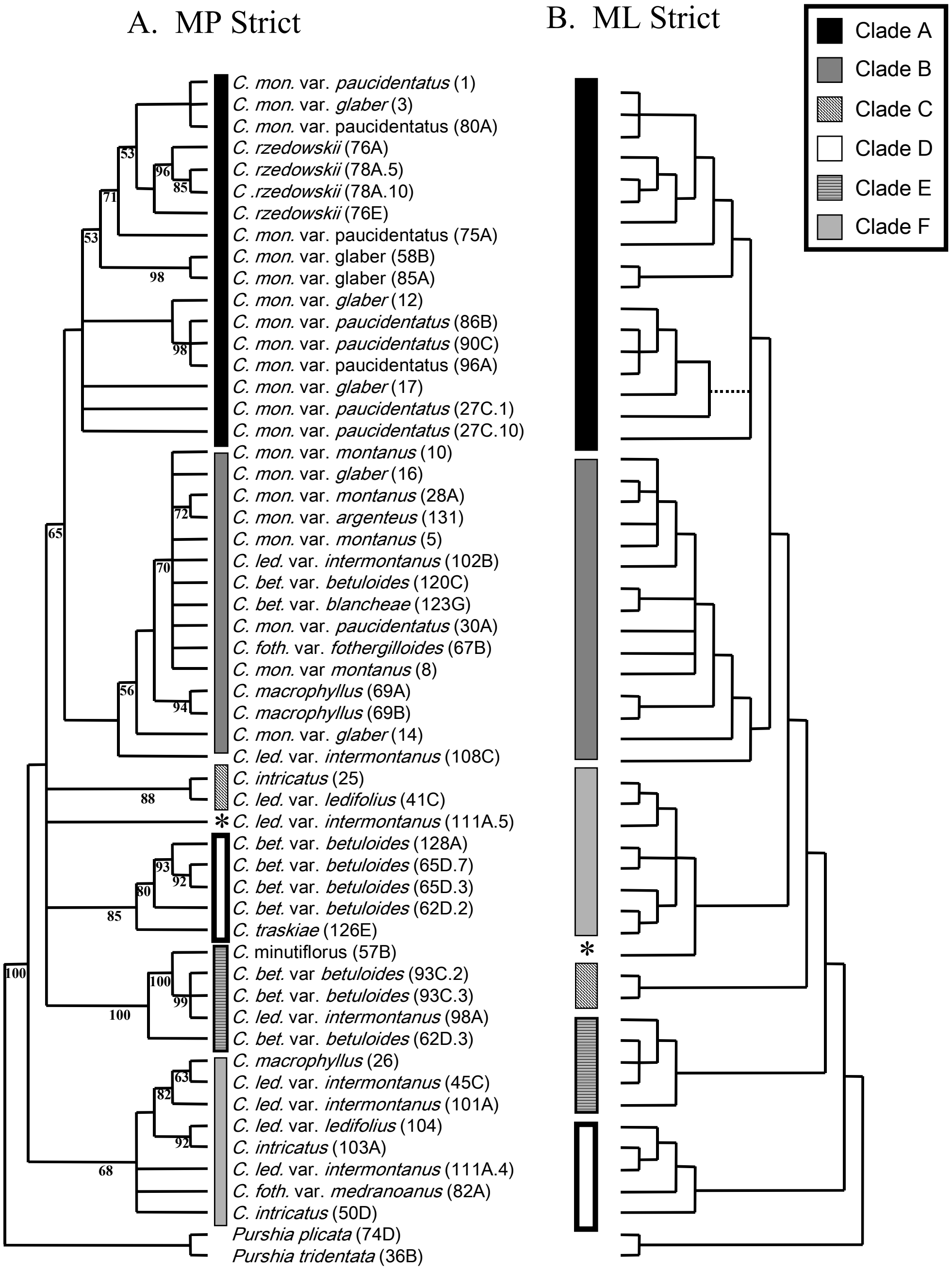


Figure 3.5. Strict consensus trees recovered from the MP and ML analyses conducted on the ETS dataset. A. The maximum parsimony (MP) strict consensus tree of over 15,000 most parsimonious trees of length 456. The numbers below the branches show bootstrap support above 50. The MP strict consensus tree has six major clades, labeled A through F, identified by the colored bars. The large asterisks denotes clone number 5 from *C. ledifolius* var. *intermontanus* 111A, which changes position between analyses. B. One of two most likely trees with a likelihood score of -Ln 4223.56020 recovered in the maximum likelihood analysis. The branch that collapses in the consensus tree is shown as a dotted line. Bootstrap support for the branches in the ML analysis were not estimated due to the computational time required.

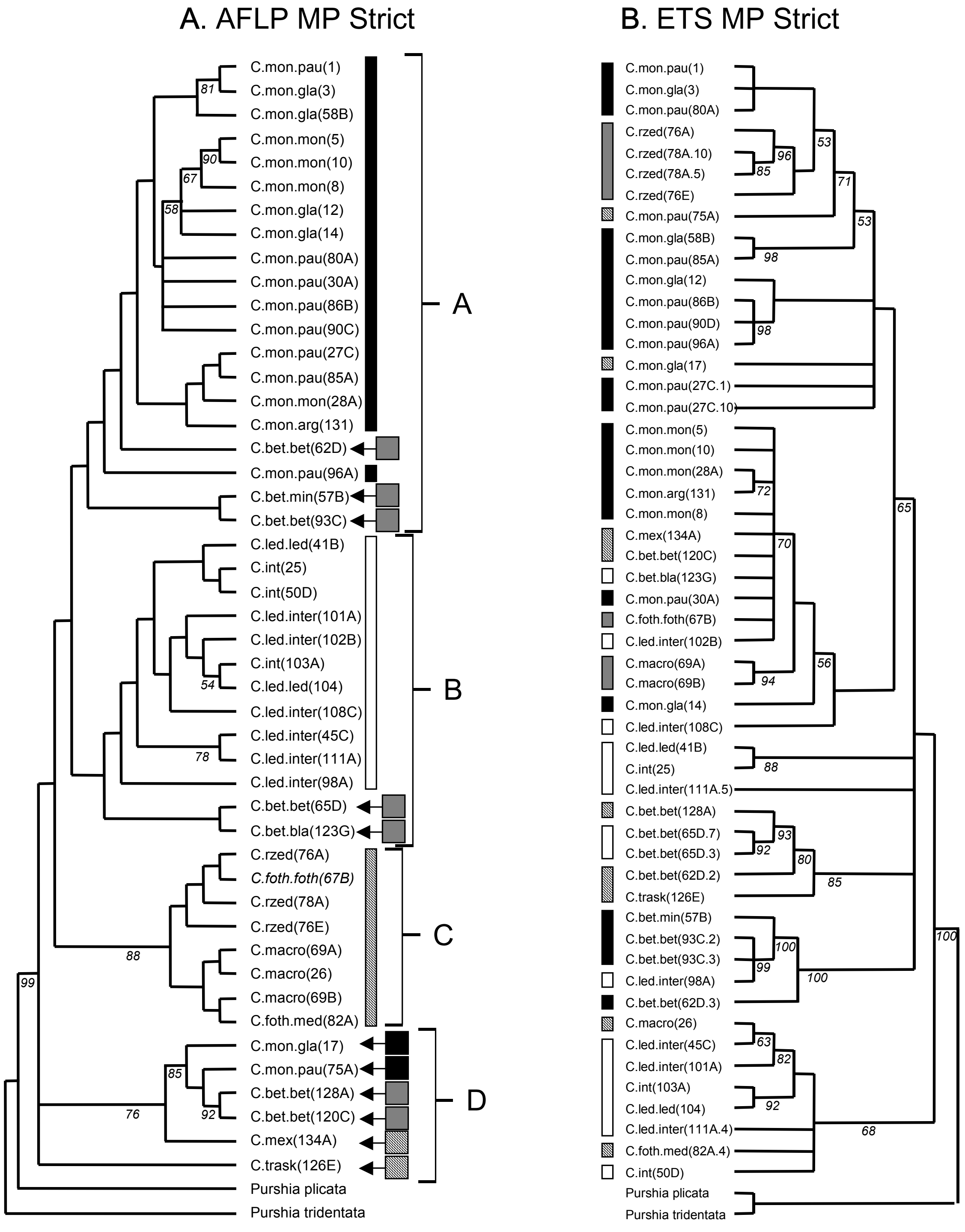


Figure 3.6. Comparison of the AFLP MP strict consensus tree and the ETS MP strict consensus tree. The names of the taxa have been abbreviated as in Figure 2.4. The number in parentheses at the end of the name refers to the collection number (Table 3.1). The colored bars indicate the species complex designations of Lis (1992).

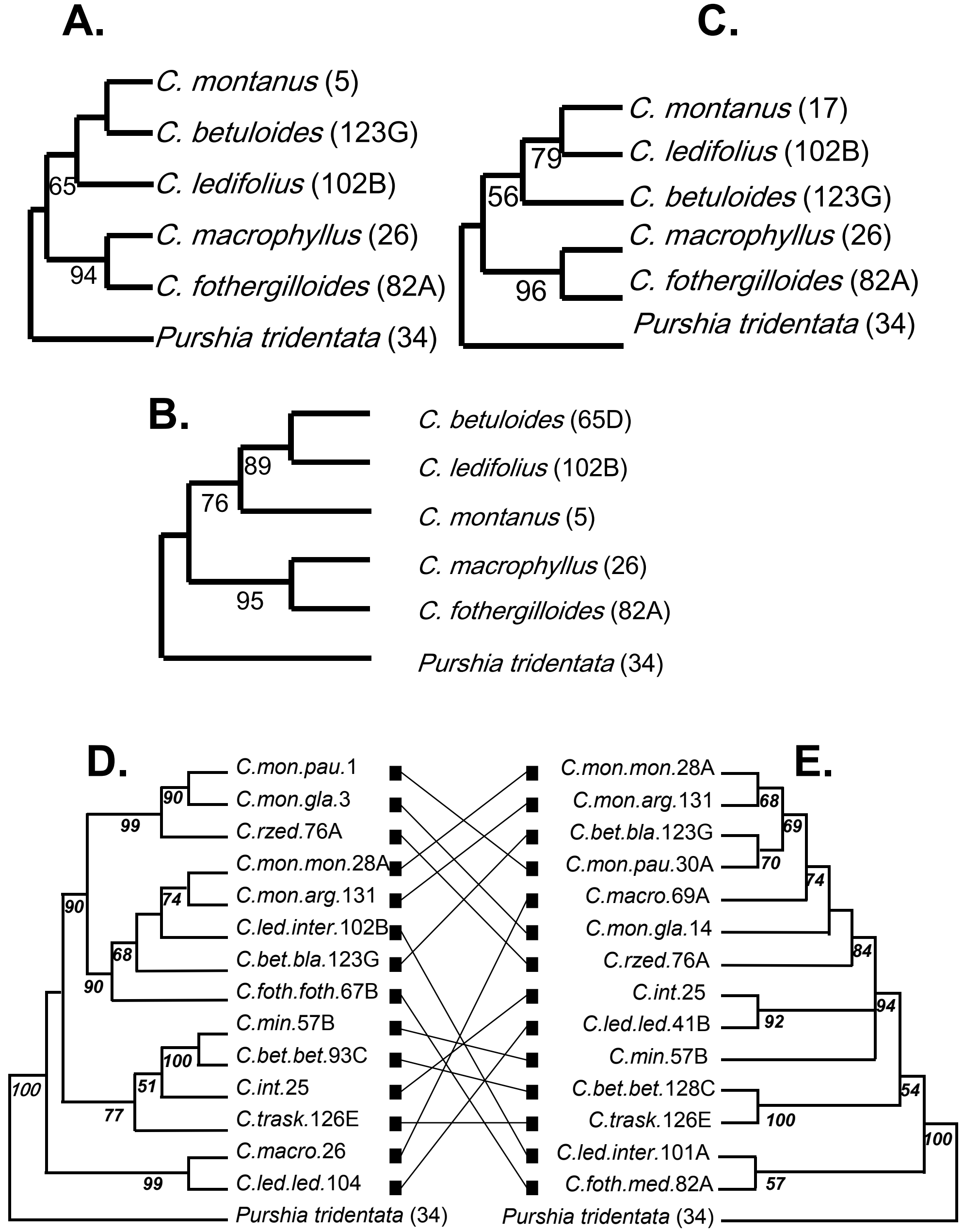


Figure 3.7. MP trees reconstructed using both AFLP and ETS characters showing the importance of taxon sampling in *Cercocarpus*. Figures A, B, and C are single most parsimonious trees based on exhaustive searches in PAUP, based on 376 presence/absence AFLP characters and individuals sampled in Table 3.3. Figure A, B, and C correspond to AFLP datasets A, B, and C in Table 3.3. Figure A had a length of 122. Figure B had a length of 114. Figure C had a length of 128. Numbers below the branches are a measure of bootstrap support based on 1000 replicates above 50. Figures D and E are trees that resulted from a heuristic MP search based on the aligned 977 ETS data set. Figures D and E correspond to ETS reduced datasets D and E in Table 3.4. Figure D is a strict consensus of 3 most parsimonious trees of length 201. Figure E is the most parsimonious tree of length 185. Numbers below the branches are measures of bootstrap support based on 100 replicates above 50.

Chapter 4: A phylogeny for *Cercocarpus* H.B.K (Rosaceae) based on cpDNA sequence data and the implications of congruence and incongruence with AFLP and ETS datasets

INTRODUCTION

The benefits of creating and comparing independent datasets of molecular characters for plant phylogenetic reconstruction are well discussed and follow two major lines of reasoning (Hillis 1987, Wendel and Doyle 1998). First, when phylogenetic topologies from independent molecular datasets are congruent, it is usually considered a reliable topology representing the evolutionary history of the taxa being studied (Hillis 1987). Under this condition, many researchers would support combining the datasets in order to improve support for particular branches or reconstruct groupings that were not present in the original, independent topologies (Bull et al. 1993, Kellogg et al. 1996). Second, multiple, independent datasets may show well-supported incongruence, suggesting different evolutionary histories for the molecular markers involved (Wendel and Doyle 1998). It is important to consider that most of the molecular characters used to reconstruct phylogenetic hypotheses come from DNA sequence data, representing small stretches of DNA. These small pieces of DNA may have different evolutionary histories from the species/genome within which they reside, as a result of past hybridization events or lineage sorting. Therefore, in both cases, multiple independent molecular datasets are desirable. This study presents the results of a chloroplast-based phylogeny and compares it with two independent nuclear phylogenies for the genus *Cercocarpus*. Congruencies and incongruencies between the three phylogenies are examined.

Recently, sequence data from the external transcribed spacer (ETS) of the nuclear ribosomal DNA (nrDNA) and presence/absence data generated from AFLP markers were used to reconstruct a nuclear phylogenetic hypothesis for *Cercocarpus* (Chapter 3). Unexpectedly, these phylogenies, based on two different nuclear datasets, were not congruent (Chapter 3). Three proposed reasons for the incongruence between the ETS and AFLP trees were presented, including; 1) an artifact of the data due to lack of support for branches, 2) excess homoplasy in the AFLP dataset, and 3) hybridization and a lack of concerted evolution, whereby ETS sequences assumed to be orthologous were not. Due to the incongruence between nuclear markers, a comparison of both the ETS and AFLP phylogenies to a third, truly independent molecular dataset is required. For the first case, after comparison to an independent phylogeny, we may identify a few individuals that were incorrectly reconstructed in both the ETS and AFLP trees due to a lack of phylogenetically informative variation, but were fully resolved in the independent tree. We might find the significant incongruence between the ETS and AFLP trees disappears after the individuals causing the incongruence are excluded.

For the second case, where the incongruence was produced from excess homoplasy in the AFLP dataset, we might find that ETS tree topology to be congruent to an independent phylogeny for *Cercocarpus*, but the AFLP tree will be incongruent due to its inherent homoplasy. For the third and last case, significant incongruencies between AFLP, ETS and an independent phylogeny may suggest different evolutionary histories for the molecular markers involved and yield insight into past evolutionary processes, including hybridization and introgression (Rieseberg et al. 1996).

Plastid sequence data is the choice most studies of plants make for a molecular marker that is independent of nuclear datasets. Recently, many non-coding chloroplast markers have been developed for widespread use in plant systematics (Soltis and Soltis 1998) and have been shown to be informative at inter and intra-specific levels (Hamilton 1999). Given three independent phylogenies, areas of topological congruence and incongruence may be identified (considering only well supported branches) that may allow the identification of the source of significant incongruence between the AFLP and ETS, and also past evolutionary processes.

A chloroplast phylogeny will also help to test hypotheses generated by previous authors about taxon relationships and evolutionary trends in *Cercocarpus* discussed in Chapter 3. Briefly, we would like to address whether or not *Cercocarpus betuloides* is indeed a separate taxon from *C. montanus* as the

AFLP and ETS datasets suggested. These results would indicate that the taxonomic changes made by Lis (1993) were warranted. A second question asks if *Cercocarpus traskiae* actually has closer affinities to *C. betuloides* than to *C. fothergilloides* as the AFLP and ETS datasets implied. Third, *Cercocarpus rzedowskii*, a newly described species from Tamaulipas, Mexico (Henrickson 1987) was found to cluster with *C. fothergilloides* in the AFLP trees, yet with *C. montanus* in the ETS dataset. *Cercocarpus mexicanus*, the other newly described species was found to cluster with *C. betuloides*. A cpDNA tree may help us determine the true relationships for these two taxa. Lastly, Stutz (1974) hypothesized that both *Cercocarpus montanus* and *C. intricatus* were formed by peripheral isolation from *C. ledifolius*. I did not find any support for this claim in either AFLP or ETS phylogenies, yet I will address this claim again with an independent cpDNA phylogeny.

MATERIALS AND METHODS

Taxon Sampling and DNA extraction

The same DNA accessions were used as in Chapter 3 with a few additions. Sampling for this cpDNA study attempted to overlap with both the ETS and AFLP datasets previously sampled (Table 4.1). One sample from each of two species of *Purshia*, *P. tridentata* and *P. plicata*, and one sample of *Chamaebatia foliolosa* were used as outgroups in this study. These outgroups were chosen based on the documentation of a close relationship between *Cercocarpus*, *Purshia*, and *Chamaebatia* in molecular phylogenies of the Rosaceae (Morgan et al. 1994, Evans et al. 2000). Fifty-four individuals, including the three outgroup individuals, are included in the cpDNA dataset. The 51 *Cercocarpus* individuals represent 46 populations of 17 taxa (Table 4.1).

Table 4.1: A list of the taxonomic identity, code, location, and collector and voucher number of the individual leaf samples used in this study. BVH is the author and JH is James Henrickson. All vouchers are housed at the plant resources center (UTEX)* denotes those individuals that are in common with the AFLP dataset and were used in the cpDNA comparison. † denotes those individuals that are in common with the ETS type 1 dataset and were used in the cpDNA comparison.

taxon	code	location	collector and voucher
<u>*montanus* complex</u>			
C. montanus var. *montanus*	†*5	Castle Rock, Douglas Co., CO USA	BVH 5
	†*8	Hartville, Platte Co., WY USA	BVH 8
	†*10	Dewey, Custer Co., SD USA	BVH 10
	†*28A	Bailey Canyon, Otero Co., NM USA	BVH 28
C. montanus var. *glaber*	†*3	Jame, Coahuila, MX	JH 20439b
	†*12	Chisos Mts., Brewster Co., TX USA	BVH and JH 20662
	†*14	Chisos Mts., Brewster Co., TX USA	BVH and JH 20663
	†*15	Chisos Mts., Brewster Co., TX USA	BVH and JH 20665
	†*58B	Leaky, Real Co., TX USA	JH 22275
	*S.7	Guadalupe Mts., Culberson Co., TX USA	CJ 18482
C. montanus var. *paucidentatus*	†*1	La Peridida, Tamaulipas, MX	JH 19218
	†*27C	Guadalupe Mts., Eddy Co., NM USA	BVH 27
	†*30A	Pinos Altos, Grant Co., NM USA	BVH 30A
	†*75A	Inturbide, Coahuila, MX	BVH and JH 22429
	†*80A	Valle Hermosa, Tamaulipas, MX	BVH and JH 22457
	†*86B	Tunnel Springs, Bernalillo Co., NM USA	BVH 86
	*S.3	Guadaulupe Mts., Culberson Co., TX USA	CJ 24804

C. montanus var. *argenteus*	†*131	Guadaulupe Mts., Culberson Co., TX USA	McVaugh 7424
betuloides complex			
C. betuloides var. *betuloides*	†*62D	Porterville, Tulare Co., CA USA	BVH 127a
	†*65D	Camp Nelson, Tulare Co., CA USA	BVH 127b
	†*93C	Miami, Gila Co., AZ USA	BVH 93
	†*120C	American Canyon, San Luis Obispo Co., CA USA	BVH 120
	†*128A	Oakhurst, Madera Co., CA USA	BVH 128
C. betuloides var. *blancheae*	*121A	Parisima Hills, Santa Barbara Co., CA USA	BVH 121
	*123F	Santa Monica Mts., Los Angeles Co., USA	BVH 123
	†*123G	Santa Monica Mts., Los Angeles Co., USA	BVH 123
C. minutiflorus	†*57B	Del Mar, San Diego Co., CA USA	JH 22311
fothergilloides complex			
C. traskiae	†*126E	Santa Catalina Island, Los Angeles Co., CA USA	Thorne 36737
C. fothergilloides var. *fothergilloides*	†*67B	Los Lirios, Coahuila, MX	BVH and JH 22413
	*84G	San Roberto Junction, Nuevo Leon, MX	BVH and JH 22482
C. fothergilloides var. *medranoanus*	*77B	Valle Hermosa, Tamaulipas, MX	BVH and JH 22444
	†*82A	Valle Hermosa, Tamaulipas, MX	BVH and JH 22456b
C. pringlei	C. pri	Matatlan, Oaxaca, MX	Stevens 1318
C. macrophyllus	†*26	Durango, MX	Crozier n/a
	†*69A	Sierra de la Viga, Coahuila, MX	BVH and JH 22415
	†*69B	Sierra de la Viga, Coahuila, MX	BVH and JH 22415
New taxa			
C. mexicanus	*134A	San Antonio Pena Nevada, Nuevo Leon, MX	JH 19295
C. rzedowskii	†*76A	Miquihuana, Tamaulipas, MX	BVH and JH 22434
	†*76E	Miquihuana, Tamaulipas, MX	BVH and JH 22434
	*81A	Valle Hermosa, Tamaulipas, MX	BVH and JH 22460
ledifolius complex			
C. ledifolius var. intermontanus	†*45C	Mantua, Box Elder Co., UT USA	BVH 45

	†*98A	Verni, Sierra Co., CA USA	BVH 98
	†*101A	Toiyabe Range, Lander Co., NV USA	BVH 101
	†*102B	Ruby Mts., Elko Co., NV USA	BVH 102
	†*108C	Spring Mts., Clark Co., NV USA	BVH 108
	†*111A	West Walker River, Mono Co., CA USA	BVH 111
C. ledifolius var. ledifolius	†*41B	Emery, Emery Co., UT USA	BVH 41
	†*104	Ely, White Pine Co., NV USA	BVH 104
C. intricatus	†*25	Escalante, Garfield Co., UT USA	BVH 25
	†*50D	Vernal, Uintah Co., UT USA	BVH 50
	†*103A	Ely, White Pine Co., NV USA	BVH 103
Purshia tridentata	†*36B	Casto Canyon, Garfield Co., UT USA	BVH 36
Purshia plicata	†*74D	Interbide, Coahuila, MX	Nesom 6198
Chamebatia foliolosa	129	Wakalu Hep Yo Campground, Calaveras Co., CA USA	BVH 129

CpDNA Amplification and sequencing

I amplified and sequenced two non-coding cpDNA markers for the DNA accessions in Table 4.1. The first was the entire inter-genic spacer between the *trn*S and *trn*G coding regions using the forward primer (5'-GCCGCTTTAGTCCACTCAGC-3') and the reverse primer (5'-GAACGAATCACACTTTTACCAC-3') (Hamilton 1999). The other non-coding region was the entire inter-genic spacer between the *rpl*20 and *rps*12 coding regions using the forward primer (5'-TTTGTTCTACGTCTCCGAGC-3') and the reverse primer (5'-GTCGAGGAACATGTACTAGG-3') (Hamilton 1999). A negative control, which lacked template DNA, was included in all PCR runs. Fifty microliter PCRs were performed using 30mM Tricine, pH 8.4, 2mM $MgCl_2$, 50 mM KCl, 5% acetamide, 100 uM of each dNTP, 1.0 unit of Taq polymerase, and 30 nM of each amplification primer. Twenty nanograms of template DNA was added, and the reactions were overlaid with mineral oil and run on a MJ Research Thermal Cycler programmed for a hot start (95^{o} C, 5 min; 74 ^{o} C, 7 min) and 30 cycles of 94 ^{o} C for 30 sec, 50 ^{o} C for 45 sec, and 72 ^{o} C for 1 min. A final extension of 72 ^{o} C for 7 min was also included. Three microliters of the reaction products were run on a 1.5 % agarose gel in 0.5X TBE buffer. PCR products were visualized with ethidium bromide under UV light, and purified using Sephadex columns (Princeton Separations, Inc.). DNA sequencing used standard Big Dye reagents (Applied Biosystems) and was preformed on a MJ BaseStation (MJ Research). Each cpDNA region was proofread and then aligned using Clustal W in MegAlign (DNAstar, 1998), followed by adjustment by eye, yielding two aligned datasets (Appendix E.).

Phylogenetic analysis

Phylogenetic analyses were performed with maximum parsimony (MP) optimality criterion. I treated the characters as unordered. The analyses were performed in PAUP version 4.0b10 (Swofford 1999). I used a heuristic search strategy using accelerated transformation (ACCTRAN), with 10,000-sequence addition replicates and TBR branch swapping. Support for particular branches was evaluated using 100 bootstrap replicates (Felsenstein 1985), using the simple addition of sequences.

I performed this search strategy on each of the aligned cpDNA regions and a combined dataset (see below). Indel events of 2 bp or more were coded as presence/absence and added to the aligned sequence dataset. I also performed this search strategy on the 52 taxa common between the AFLP and cpDNA datasets, and on the 44 taxa common between the cpDNA and ETS datasets (Table 4.1). A description of the AFLP and ETS datasets can be found in Chapter 3. To compare the cpDNA dataset to the ETS dataset, I had to take into account the two cases of paraphyletic ETS clones from the individuals *Cercocarpus ledifolius* var. *intermontanus* 111A (clone 4 and 5) and *C. betuloides* var. *betuloides* 62DA (clones 2 and 3). Therefore, I constructed 4 ETS datasets which differed in the clones which represented the individuals 111A and 62D in all possible combinations (Table 4.2).

Table 4.2: A description of the clones included in each of the ETS datasets tested against the cpDNA dataset. 111A refers to the individual *Cercocarpus ledifolius* var. *intermontanus* collection 111A (table 1). 62D refers to *C. betuloides* var. *betuloides* collection 62D (table 1).

	111A clone 4	111A clone 5	62D clone 2	62D clone 3
ETS dataset 1	included	not included	included	not included
ETS dataset 2	included	not included	not included	included
ETS dataset 3	not included	included	included	not included
ETS dataset 4	not included	included	not included	included

Tests for incongruence

I tested for incongruence between the *trn*S-*trn*G inter-genic spacer region and the *rpl*20 and *rps*12 inter-genic spacer region using the partition homogeneity test (Farris et al. 1994, 1995) as implemented in PAUP (Swofford 1999). This test first repartitions the characters found in the two original datasets into two temporary data subsets, finding the shortest tree length for each and adds them together. This was done for 100 replicates. Then, the sum of the shortest tree lengths for the original datasets is compared to the distribution of the repartitioned tree length sums. If there is a low probability of finding shorter tree length sums from the repartitioned data than from the original data, the original datasets are interpreted as incongruent. The decision to combine the cpDNA non-coding regions into one dataset is based on the results of this test.

To assess the incongruence between the cpDNA dataset and the ETS datasets, and the incongruence between the cpDNA dataset and the AFLP dataset, I conducted the "compare-2" test (Faith 1991), the Templeton test (Templeton 1983), and the Kishino-Hasegawa test as implemented in PAUP (Swofford 1999). The methods for these tests are the same as in Chapter 3. The constraint trees for all three tests were the MP bootstrap consensus trees, which only included branches with bootstrap support over 50%.

RESULTS

Full cpDNA phylogenetic analysis

The length of the aligned sequence dataset for the *trn*S-*trn*G spacer was 580 bp, and for the *rps*12-*rpl*20 spacer was 757 bp. I recorded 6 indel events in the *trn*S-*trn*G spacer and recorded 8 indel events in the *rps*12-*rpl*20 spacer that were greater than 2 bp. The results of the partition homogeneity test for the congruence of the *trn*S-*trn*G spacer and the *rps*12-*rpl*20 spacer yielded a non-significant result (p=0.980) (Table 4.3). Therefore, I combined the two datasets and continued with a MP heuristic search. For the full cpDNA dataset of 54 individuals, our analysis recovered 6369 equally parsimonious trees of length 251 (Figure 4.1). The tree statistics for this search can be seen in Table 4.4.

Table 4.3: Results of the partition homogeneity test for the *trn*S- *trn*G inter-genic spacer versus the *rpl*20-*rps*12 inter-genic spacer. The asterisk denotes the sum of the tree lengths for the original, unpermuted partition.

sum of tree lengths	**number of replicates**
223	1
224	2
225	6
226	17
227	29
228	15
***229**	18
230	2

Table 4.4: A comparison of the tree statistics for the multiple MP heuristic searches conducted on all three datasets. CpDNA refers to the complete cpDNA dataset, including all 54 taxa. CpDNA vs. AFLP includes the 52 taxa in common between datasets. CpDNA vs. ETS includes the 44 taxa in common between datasets. ETS 1 – 4 identities refer to Table 4.2.

	CpDNA	**CPDNA vs AFLP**	**AFLP**	**CPDNA vs ETS**	**ETS 1**	**ETS 2**	**ETS 3**	**ETS 4**
# of ind.	54	52	52	44	44	44	44	44
# of equally parsimonious trees	6369	84	126	9342	1598	1520	2561	2584
tree length	251	210	827	168	395	400	404	409
# of characters	1352	1352	376	1352	977	977	977	977
# of parsimony informative characters	72	69	76	61	133	133	134	134
CI	0.847	0.838	0.314	0.851	0.767	0.769	0.757	0.768
RI	0.824	0.827	0.503	0.836	0.808	0.799	0.794	0.797

The MP phylogeny produced from the full cpDNA dataset lacks overall resolution, and supports four main clades (Figure 4.1). *Cercocarpus* is found to be monophyletic with good bootstrap support. Within *Cercocarpus*, clade A is weakly supported (bootstrap score below 50) and contained members of all four species complexes defined by Lis (1992). Clade B was weakly supported (bootstrap score of 62), in an unresolved position with clade A. Clade B contained all members of the *Cercocarpus betuloides* complex, except *C. betuloides* var. *betuloides* (93C) and *C. minutiflorus* (57B), which were present in clade A. Two individuals of *Cercocarpus ledifolius* var. *intermontanus* (collections 101A and 111A) were found nested within clade B. One individual of *Cercocarpus ledifolius* var. *intermontanus* (98A) was found in an unresolved position with clade B. Clades C and D were found in an unresolved position sister to clades A and B with moderate bootstrap support (score of 82). Clades C and D are shown as unresolved with respect to each other, yet 98% of the trees recovered in the MP search united them as sister to one another. Clades C and D are made up almost exclusively of members of the *Cercocarpus fothergilloides* complex and *C. rzedowskii*. The one exception, *C. montanus* var. *paucidentatus* (75A), was positioned in an unresolved position within clade C.

CpDNA vs. AFLP phylogenetic analysis

We pruned two individuals from the full cpDNA dataset for a total of 52 individuals to compare to the AFLP dataset (Table 4.1). Our MP heuristic search using the reduced cpDNA dataset resulted in 84 most parsimonious trees of length 210 with similar topology to the full cpDNA dataset (Figure 4.2). The MP heuristic search using the same 52 accessions from the AFLP dataset resulted in 126 most parsimonious trees of length 827 (Figure 4.2) (Table 4.4).

Although MP trees produced from the AFLP and cpDNA datasets do not support the same relationships for the 52 individuals in common between them, I found some well supported congruent topologies. First, both the cpDNA and AFLP trees support a *Cercocarpus fothergilloides* clade with high bootstrap support. Also, both the cpDNA and AFLP datasets support a close association between the *Cercocarpus fothergilloides* complex and *C. rzedowskii*. Finally, both trees support *Cercocarpus traskiae* as being more closely related to the *C. betuloides* complex than to the *C. fothergilloides* complex.

Due to the lack of resolution in the cpDNA tree, we are unable to make many claims about relationships between species. The tree based on the AFLP dataset supports the separation of the *Cercocarpus montanus* complex, the *C. ledifolius* complex, although these groupings have bootstrap support values lower than 50. These groups were not supported in the cpDNA tree, where the *Cercocarpus montanus* and *C. ledifolius* complexes were grouped together in clade A and B. The individual of *Cercocarpus montanus* var. *paucidentatus* (75A) is the only exception for the monophyly of the *C. montanus* complex in the AFLP tree. This is the same individual that was divergent in the cpDNA analysis as discussed above. The *Cercocarpus fothergilloides* complex is the only clade that has high bootstrap support in the AFLP tree (Figure 4.2). The *Cercocarpus fothergilloides* complex did group together in cpDNA tree in clades C and D. Members of the *Cercocarpus betuloides* complex are found in clades A and B in the cpDNA tree, yet are not grouped together in the AFLP tree, where they are found scattered and unresolved.

CpDNA vs ETS phylogenetic analysis

I pruned 10 individuals from the full cpDNA dataset for a total of 44 individuals to compare to the ETS datasets (Table 4.1). The tree statistics for the reduced cpDNA dataset and the four ETS datasets are illustrated in Table 4.4. The cpDNA phylogeny has the same structure as the previous two cpDNA phylogenies (Figure 4.3). The four ETS datasets differed in their topology due to the inclusion and exclusion of the different clones for individual 111A and 62D (Table 4.2). The MP strict consensus of the four ETS datasets is in Figure 4.3.

The ETS and cpDNA trees are similar in the placement of the *Cercocarpus montanus* complex as derived, although clade A is unresolved. Both trees also support the unity of the *Cercocarpus betuloides* complex (clades D and E in the ETS tree, clade B in the cpDNA tree). Overall, the cpDNA and ETS datasets do not support the same relationships for the 44 individuals. The ETS strict consensus tree contains 6 recognizable clades, A through F (Chapter 3). The most notable difference between these two trees is the placement of individuals of the *Cercocarpus fothergilloides* complex. The cpDNA tree supports a single clade with one exception (*Cercocarpus fothergilloides* var. *fothergilloides* 84G), yet members of the *C. fothergilloides* complex are present in three clades in the ETS tree (Clades A, B, and F)

Incongruence tests

The results of the compare-2, the Templeton, and the Kishino-Hasegawa tests can be seen in Table 4.5. For the compare-2 test, 100 permutations of each dataset were tested on each constraint tree. I found significant results with every dataset, where the difference in fit for the dataset on its own tree versus the constraint tree was significantly better than the difference with permuted data. The Templeton and Kishino-Hasegawa tests yielded highly significant results for each dataset, signaling significant incongruence between all datasets tested.

Table 4.5: Results for the tests of congruence between the cpDNA trees and the trees produced from ETS and AFLP datasets.

	COMPARE-2	**KISHINO-HASEGAWA**				**TEMPLETON**	
	p-value	diff. in tree length	s.d.	t	p-value	N	p-value
cpDNA vs AFLP							
cpDNA dataset	p=0.01	66	14.31537	4.6104	p=0.0001	27	p=0.0001
AFLP datset	p=0.01	319	38.5267	8.28	p=0.0001	109	p=0.0001
cpDNA vs ETS 1							
cpDNA dataset	p=0.01	65	18.71187	3.4737	p=0.0005	18	p=0.0002
ETS 1 dataset	p=0.01	296	40.12442	7.3771	p=0.0001	85	p=0.0001
cpDNA vs ETS 2							
cpDNA dataset	p=0.01	65	18.71187	3.4737	p=0.0005	18	p=0.0002
ETS 2 dataset	p=0.01	348	43.41185	8.0162	p=0.0001	103	p=0.0001
cpDNA vs ETS 3							
cpDNA dataset	p=0.01	61	18.01354	3.3863	p=0.0007	19	p=0.0002
ETS 3 datset	p=0.01	340	42.47163	8.0053	p=0.0001	86	p=0.0001
cpDNA vs ETS 4							
cpDNA dataset	p=0.01	66	19.31461	3.4171	p=0.0007	18	p=0.0002
ETS 4 dataset	p=0.01	301	40.86533	7.3657	p=0.0001	86	p=0.0001

DISCUSSION

My goal was to develop an independent molecular phylogeny for *Cercocarpus* to compare with the two incongruent nuclear phylogenies presented in Chapter 3. The *trn*S-*trn*G spacer and the *rps*12-*rpl*20 had congruent phylogenetic signal. The maximum parsimony phylogeny produced from the combined dataset yielded trees with 4 major clades, A through D. These trees strongly supported the monophyly of *Cercocarpus* using both *Purshia* and

Chamaebatia as outgroups. There was little resolution within the clades due to a lack of phylogenetically informative variation, and I was unable to resolve the relationships between the majority of *Cercocarpus montanus* and *C. ledifolius* samples in the dataset present in clade A. All but two members of the *Cercocarpus betuloides* complex were found in clade B, including *C. traskiae*. Clades C and D included mostly members of the *Cercocarpus fothergilloides* complex and the two newly described taxa, *C. rzedowskii* and *C. mexicana*. The chloroplast phylogeny was significantly incongruent to both the ETS and AFLP phylogenies for comparable individuals. Both the chloroplast and AFLP phylogenies suggest a well supported *Cercocarpus fothergilloides* complex and the close association between the *C. fothergilloides* complex and the two newly described taxa of *C. rzedowskii* and *C. mexicanus*. The cpDNA and ETS tree both support the *Cercocarpus betuloides* complex as unique from the C. *montanus* complex, with a few exceptions.

Taxonomic implications and testing previous hypotheses

I found all but two members of the *Cercocarpus betuloides* complex present in clade B, separate from the majority of *C. montanus* samples in clade A. This result was also found in the ETS and AFLP trees, and lends further support to the decision by Lis (1993) to recognize *C. betuloides* as a separate entity from *C. montanus*.

I also found *Cercocarpus traskiae* grouping with members of the *C. betuloides* complex. This result is consistent with both the AFLP and the ETS datasets. *Cercocarpus traskiae* had been hypothesized to be a part of the *C. fothergilloides* complex based of leaf characteristics (Lis 1991, Henrickson ined.). It appears based on all three datasets that *C. traskiae* belongs allied to the *C. betuloides* complex and not the *C. fothergilloides* complex.

The three individuals of *Cercocarpus rzedowskii* were found in the weakly supported clade C, along with one individual of *C. montanus* var. *paucidentatus* (75A) and *C. mexicanus* (134A). The relationship of clade C to clade D in the full cpDNA dataset, the *Cercocarpus fothergilloides* clade, is unresolved in the MP strict consensus tree, yet most (98%) of the equally parsimonious reconstructions supported a sister relationship. This implies that based on the cpDNA dataset, *Cercocarpus rzedowskii* and *C. mexicanus* both have affinities to the *C. fothergilloides* complex. This was an interesting result. Henrickson (1987) suggested in his description of these taxa that *C. rzedowskii* had similarities to *C. ledifolius* based on the revolute, entire margin and overall leaf shape. The ETS dataset strongly supported a close association between *Cercocarpus rzedowskii* and *C. montanus*, yet the AFLP suggested it was sister to the *C. fothergilloides* complex. Here I have more evidence for the inclusion of *Cercocarpus rzedowskii* into the *C. fothergilloides* complex. The similarities of leaf characters appear to be examples of convergent evolution.

In the description of *Cercocarpus mexicanus*, Henrickson (1987) hypothesized it was similar to the *C. montanus* complex, due to the serrate leaf margin and basic leaf shape displayed by this taxon. The AFLP dataset grouped the one individual of *Cercocarpus mexicanus* in this study deeply nested in the *C. betuloides* complex cluster. In the cpDNA tree, we see *Cercocarpus mexicanus* with *C. rzedowskii* in the *C. fothergilloides* cluster. More sampling of this taxon is required for a firm hypothesis on its affinities in *Cercocarpus*. Unfortunately, the one sample of *C. mexicanus* used in this study did not have an ETS type 1 to be included in the ETS analysis.

Stutz (1974) hypothesized that the origin of Cercocarpus montanus, with large, broad deciduous leaves arose as a peripheral isolate of C. ledifolius that became adapted to colder, higher altitudes in the Rocky Mountains. Stutz (1974) also hypothesized that Cercocarpus intricatus, with small, revolute leaves arose from adaptations by an ancestral C. ledifolius lineage that became adapted to drier, hotter conditions. I found limited support for this theory (Figure 4.1) with both Cercocarpus montanus and C. intricatus arising from a paraphyletic C. ledifolius clade. Although this is an interesting result, the lack of support for the cpDNA tree topology requires caution. Also, we have no support for these hypotheses with both the ETS and AFLP tree topologies.

Implications of incongruence and evidence for hybridization

I found significant incongruences for all comparisons between the cpDNA tree and the ETS and AFLP trees. This result might be explained as merely an artifact, i.e., a product of a lack of support for branches. As can be see in Figure 4.1, the cpDNA MP phylogeny is unresolved, and contains few branches with bootstrap support greater than 50%. I attempted to reduce the possibility of this explanation by using the bootstrap consensus trees that included only branches with at least 50% bootstrap support as the constraint trees in the incongruence tests. The significant result using these bootstrap constraint trees implies that the incongruence goes beyond simple lack of support for branches, and suggests that the markers used in creating the phylogenetic trees had different evolutionary histories in *Cercocarpus* and may have been subject to horizontal transmission between distinct lineages. Hybridization has long been discussed as a probable occurrence in the genus *Cercocarpus* (Blauer et al. 1975). Schneider (1905b) invoked hybridization to explain the intergradations in leaf morphology he saw in some populations of *Cercocarpus ledifolius*, although the individuals with intermediate morphology might also be explained by either plastic responses to water availability or selection for ecotypes on a rainfall gradient. The only clear genetic evidence for hybridization in *Cercocarpus* comes from Santa Catalina Island, where Rieseberg (1989, 1994) documented introgression between *C. betuloides* and *C. traskiae* using both allozymes and RAPD markers.

The use of hybridization as the cause for the incongruence between the chloroplast and AFLP or ETS trees can only be used if geographic proximity of the taxa involved would allow it, although we must also consider possible changes in range limits for hybridizing taxa. I identified four individuals that are located in clades that I would not expect based on their morphology. These four taxa are *Cercocarpus montanus* var. *paucidentatus* (75A) placed in clade C sister to *C. rzedowskii* and *C. macrophyllus*, two individuals of *C. ledifolius* var. *intermontanus* (101A and 111A) placed in clade B with individuals of the betuloides complex, and *C. betuloides* var. *betuloides* (93C) in clade A, with individuals identified as *C. montanus* and *C. ledifolius*. Arrows in Figure 4.2 and 4.3 identify these taxa. As might be expected, we find the individuals which are "out of place" in the chloroplast tree are found in geographical areas of overlap between ranges of species complexes. *Cercocarpus montanus* var. *paucidentatus* (75A) was collected in southwestern Coahuila, near the species range of *C. rzedowskii* and *C. mexicanus*, whose members it was grouped with in the chloroplast tree. The two individuals of *Cercocarpus ledifolius* var. *intermontanus* (101A and 111A) that were placed in the same clade as members of the *C. betuloides* complex, were collected in the western end of the *C. ledifolius* distribution near the eastern edge of the range of *C. betuloides*. Lastly, the individual of the *Cercocarpus betuloides* complex, *C. betuloides* var. *betuloides* (93C), that was placed in clade A was collected in central Arizona, well within the species range of the *C. montanus* complex. The anomalous placement of these individuals in the chloroplast tree based on their position in the AFLP tree, coupled with the geographic position of the collection sites supports hybridization and chloroplast capture as the explanation for the incongruences.

Molecular systematics of *Cercocarpus* illuminates the problems that may arise from attempting phylogenetic reconstruction in very closely related groups of plants. First, it can be difficult to obtain phylogenetically informative variation in DNA sequence regions, i.e., the chloroplast sequences presented here and the ITS region attempted in Chapter 3, due to a lack of sufficient time for substitutions to accumulate between taxa. Second, if you do indeed find DNA sequence variation, (for example the ETS region) issues of coalescent times can become paramount. If the coalescent time for the sequence variation obtained is older than the taxa trying to be reconstructed, many spurious relationships can be reconstructed (Chapter 3).

Solving the phylogenetic problems that accompany closely related groups of plants means finding a marker that is evolving at an appropriate rate, where homology is unambiguous, and is not subject to deep coalescence (Soltis and Soltis 1998). It may be that such a marker can be identified in *Cercocarpus*. Another possibility is that a robust phylogeny for *Cercocarpus* will never be developed because the history of the group contains an excess of reticulation, where identified taxa do not have bifurcating evolutionary histories. Plant groups with such a history might be better understood if relationships between species were viewed as networks, rather than phylogenies. With the inclusion of more data, I hope to uncover further relationships in *Cercocarpus* with the full understanding that the evolutionary history of taxa in this group may not conform to traditional, bifurcating tree topologies.

This work has laid a foundation for my work present and future work on the coevolutionary relationships between *Cercocarpus* and *Frankia*. Only by understanding the evolutionary history of both members in an ecological association can we study coevolution in a meaningful way. I hope to use what I have learned with *Cercocarpus* and apply it to other genera in the Rosaceae that associate with *Frankia*, expanding the study of coevolution to include all members of nitrogen-fixing members of the Rosaceae.

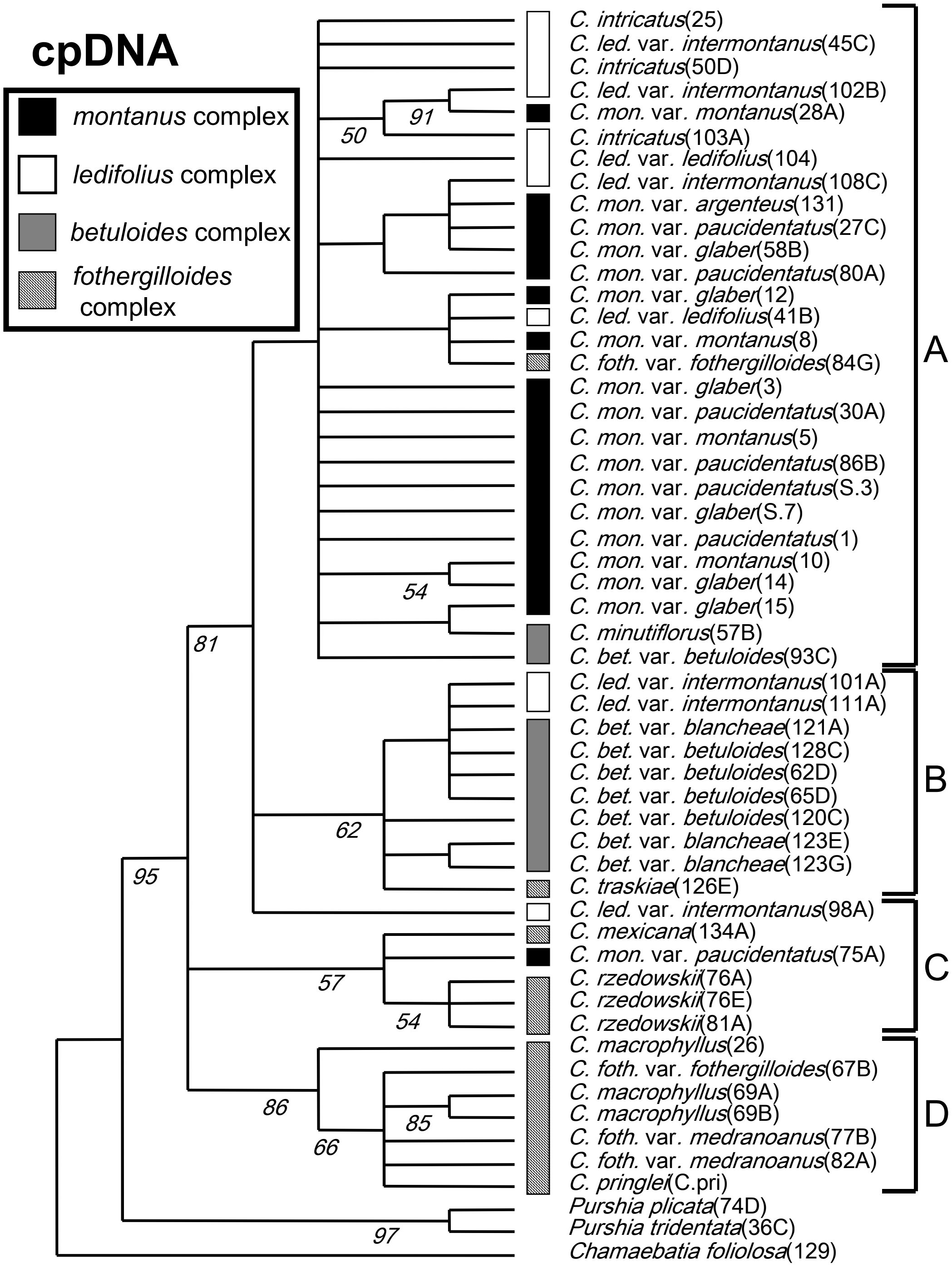


Figure 4.1: A strict consensus cladogram based on the combined cpDNA regions. It is composed of 6369 most parsimonious trees of length 251 steps. See table 4.4 for further tree statistics. Bars next to the taxon name designate species complex membership. Clades A-D are shown by the brackets.

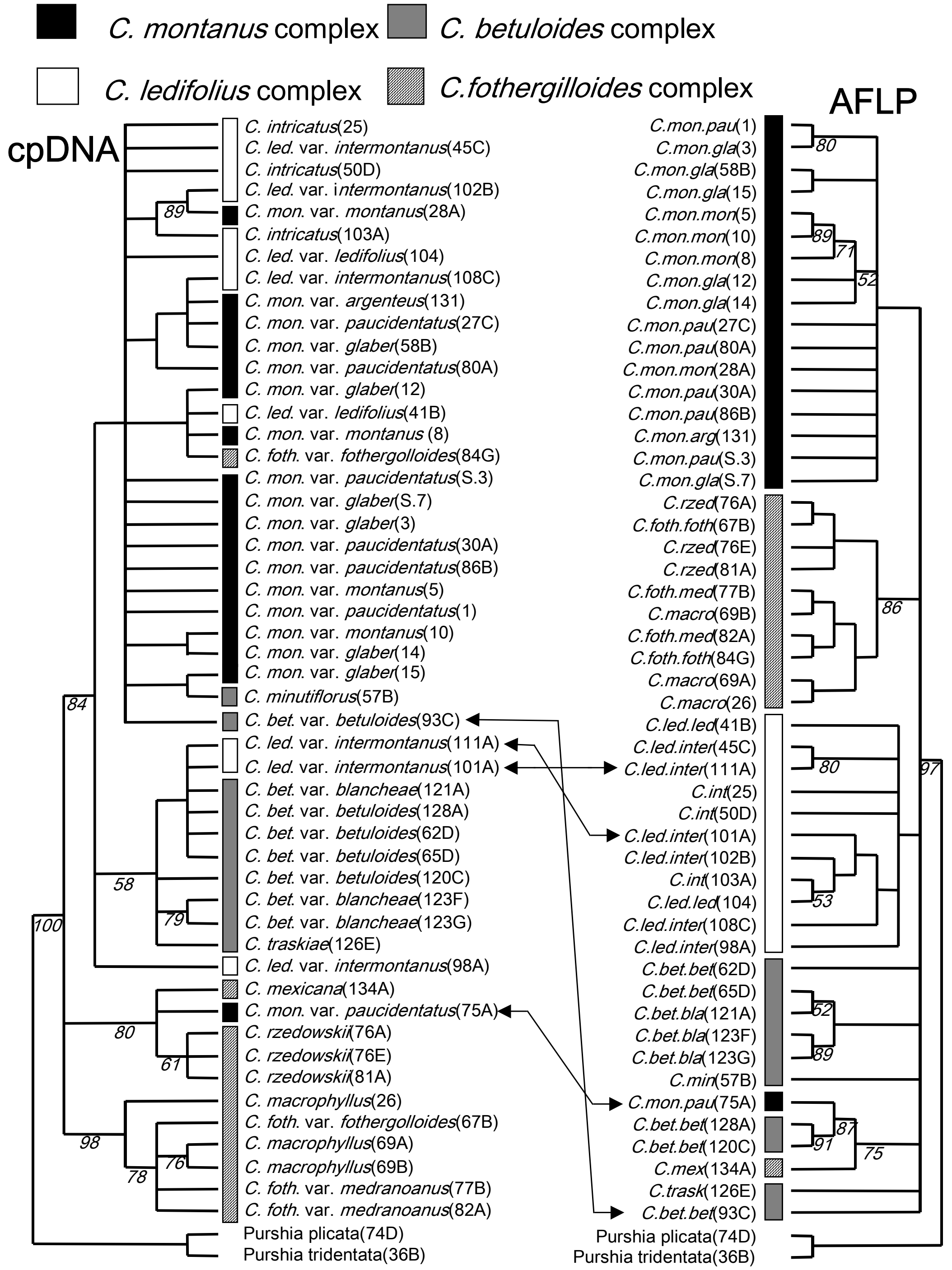


Figure 4.2: Comparison of the maximum parsimony strict consensus trees based on cpDNA sequences and AFLP data for 52 taxa. See Table 4.4 for further tree statistics. The cpDNA tree is the strict consensus of 84 most parsimonious trees of length 210 steps. The AFLP tree shown is the strict consensus of 126 most parsimonious trees of length 827 steps. Bars next to the taxon name designate species complex membership. Arrows refer to taxa in discussion.

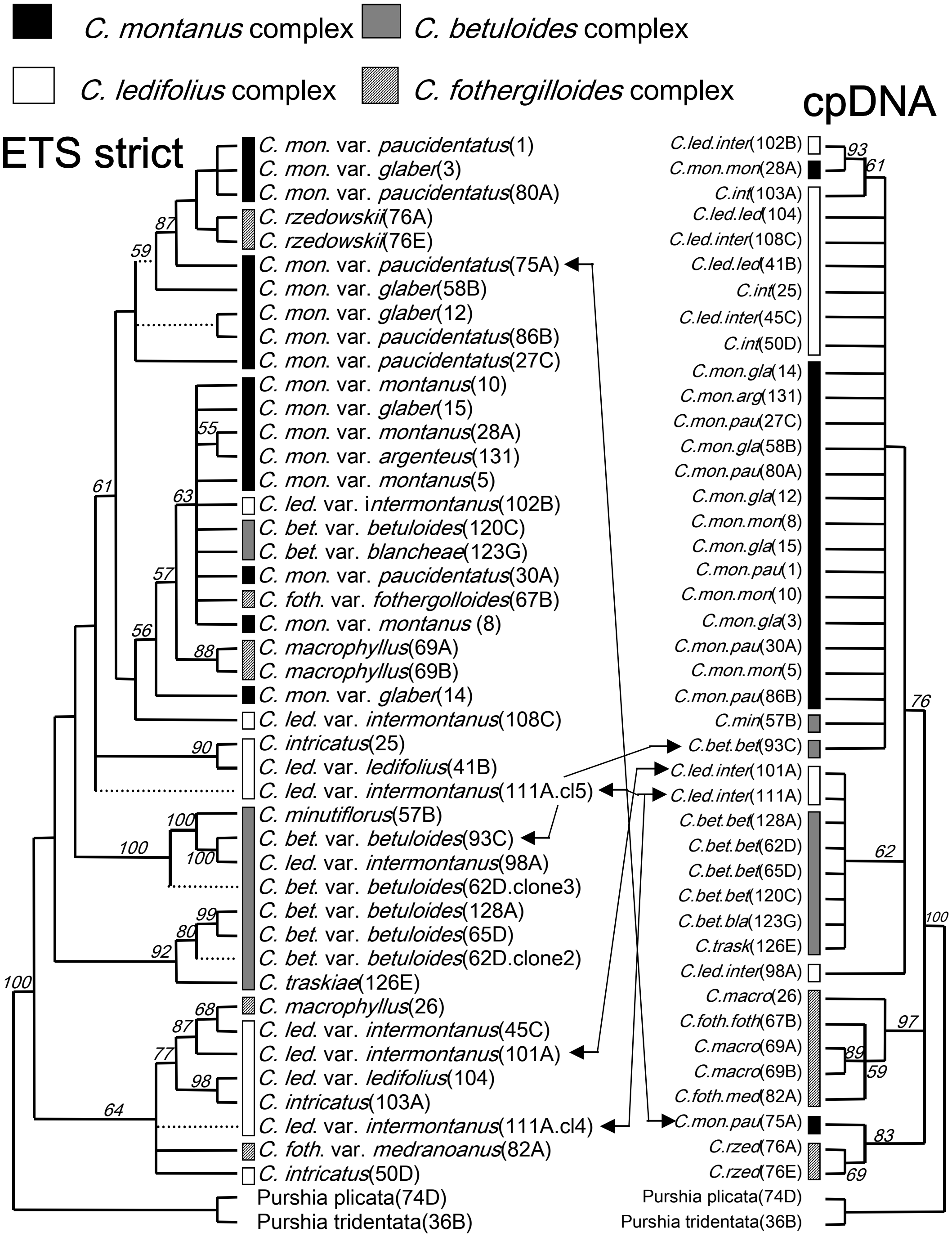


Figure 4.3: A comparison of maximum parsimony strict consensus trees based on cpDNA sequences and ETS sequence data for 44 taxa. See table 4.4 for tree statistics. The cpDNA tree shown is a strict consensus of 9342 most parsimonious trees of 168 steps. The ETS tree is the strict consensus of 2367 most parsimonious trees of 419 steps. Bars next to the taxon name designate species complex membership. Arrows refer to taxa in discussion.

Appendix A. ETS alignable DNA sequence data

This dataset contains the aligned sequences for the three alignable regions (5' alignable, 3' alignable, and the α and β regions) from the ETS types used in Chapter 2. The sequences for the three regions have been combined. The 5' alignable region is represented from positions 1-174. The 3' alignable region is represented from positions 175-382. The α and β regions are represented from positions 383-512. ID codes for the sequences correspond to those found in Table 1.1. ETS type 1, type 2, type 3, and type 4 have been separated and identified. Gaps in the sequence are indicated by a dash (-). Uncertainties are denoted by standard IUPAC/IUB codes as follows: N=unknown; R=A/G; Y=C/T; M=A/C; K=G/T; S=C/G; W=A/T.

Bases 1-60

5′ Common region →

ETS type 1

```
1       TCT-C-AATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
10      TST-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
14      TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
15      TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
12      TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
17      TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
25      TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
26      TCT-CCAATA TT---CTTAC GTGTTCTTGA GTGTTCATCT ACAGTCCTTG CA-TATTT-G
27C.1   TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
27C.10  -CT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
28A     TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
3       TCT-C-AATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
41B     TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
45      TCT-CCAATA TT---CTTAC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
5       TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
101A    TCT-CCAATA TT---CTTAC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
102B    TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
104     TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
108C    ---------- --------AC GTGTTCTTGA GTGTTCATCT ACAGTTCTCA CACTACTTTG
111A.4  ---------- ---------C GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
111A.5  ---------- ---------C GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
120C    ---------- ---------- -------TGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
123G    ---------- ---------- ---------- ---------- ---------- ----------
131     TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
30A     TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
57B     ---------- --------GC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
58B     TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
62D.2   TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
62D.3   TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
65D.7   TCT-GCA-TA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAATTCTTG CA-TATTT-G
65D.3   -CT-GCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAATTCTTG CA-TATTT-G
67B     TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
69A     TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
69B     ---------- ---------C GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
76A     TC---CAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
76E     TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
78A.10  TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
78A.5   TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
8       TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
80A     TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
82A     TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
85A     TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
86B     ---------- ---------C GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
90D     TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
93C.2   TCC----ATA TT---CTTGC GCGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
93C.3   TCT-CCAACA TT---CTTGC GCGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
98A     ---------- --------GC GCATTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
126E    ---------- ---------- ---------- -TGTTCATCT ACAATTCTTG CA-TATTT-G
103A    TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT GCAGTTCTTG CA-TATTT-G
50D     ---------- ---------- ------TTGA GTGTTCATCT ACAGTTCTTG CA-TATTTAG
68      ----CC-ATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
75A     TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
```

Bases 1-60

ETS type 2

```
121C     TCT-CCAACA TT---CTTAC GTGATCTTGA GTGTTTATCT GCAGTTCTCA CA-TATTT-G
18       TCT-CTAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
19       TCT-CCAACA TT---CTTAC GTGATCTTGA GTGTTTATCT GCAGTTCTCA CA-TATTT-G
20       TCT-CCAACA TT---CTTAC GTGATCTTGA GTGTTTATCT GCAGTTCTCA CA-TATTT-G
21       TCT-CCAACA TT---CTTAC GTGATCTTGA GTGTTTATCT GCAGTTCTCA CA-TATTT-G
122A     TCT-CCAACA TTG--CTTAC GTGATCTTGA GTGTTTATCT GCAGTTCTCA CA-TATTT-G
126.29   TCT-CCAACA TT---CTTAC GTGATCTTGA GTGTTTATCT GCAGTTCTCA CA-TATTT-G
26       TCT-CCAACA TT---CTTAC GTGATCTTGA GTGTTTATCT GCAGTTCTCA CA-TATTT-G
52B      TCT-CCAACA TT---CTTAC GTGATCTTGA GTGTTTATCT GCAGTTCTCA CA-TATTT-G
57B      TCT-CCAACA TT---CTTAC GTGATCTTGA GTGTTTATCT GCAGTTCTCA CA-TATTT-G
65C.3    TCT-CCCACT TT---CTTAC GTGTTCTTGA GTGTTTGTCT ACAGTTCTCA CA-TATTT-G
65C.4    TC---CCACT TT---CTTAC GTGTTCTTGA GTGTTTGTCT ACAGTTCTCA CA-TATTT-G
93C      TCT-CCAACA TCT--CTTAC GTGATCTTGA GTGTTTATCT GCAGTTCTCA CA-TATTT-G
118A     TCTCCCA-CA TCTCTCTTAC GTGTTCTTGA GTGTTTATCT ACAGTTCTAA CA-TATTT-G
126.BC1  TCTACCAACA TCTCTCTTAC GTGATCTTGA GTGTTTATYT GCAGTTCTCA CA-TATTT-G
126E.1   TCT-CCAAGA TT---CTTAC GTGATCTTGA GTGTTTATCT GCAGTCCTCA CA-TATTT-G
126E.3   TCT-CCAAAA TT---CTTAC GTGATCTTGA GTGTTTATCT GCAGTTCTCA CA-TATTT-G
126E.4   TCT-CCAAAA TT---CTTAC GTGATCTTGA GTGTTTATCT GCAGTTCTCA CA-TATTT-G
126A     TCT-CCAAAA TT---CTTAC GTGATCTTGA GTGTTTATCT GCAGTTCTCA CA-TATTT-G
61       TCT-CCAACA TT---CTTAC GTGATCTTGA GTGTTTATCT GCAGTTCTCA CA-TATTT-G
```

ETS type 3

```
121B     TCT-CCAAAA TT---CTTTC GTGTTCTTGA GTGTTCATCT ACAGTTCTCG CA-TATTT-G
66C      TCT-CCAACA TT---ATTTC ATGTTCTTGA GTGTTCATCT ACAGTTCTCG CA-TATTT-G
66D      TCT-CCAACA TT---CTTTC GTGTTCTTGA GTGTTCATCT ACAGTTCTCG CA-TATTT-G
67A      ---------- ---------C GTGTTCTTGA GTGTTCATCT ACAGTTCTCG CA-TATTT-G
69B      ---------- ---------C GTGTTCTTGA GTGTTCATCT ACAGTTCTCG CA-TATTT-G
71A      TCT-CCAACA TT---CTTTC GTGTTCTTGA GTGTTCATCT ACAGTTCTCG CA-TATTT-G
77B      ----CCAACA TT---CTTTC GTGTTCTTGA GTGTTCATCT ACAGTTCTCG CA-TATTT-G
82A      TCT-CCAACA TT---CTTTC ATGTTCTTAA GTGTTCATCT ACAGTTCTCG CA-TATTT-G
83B      ---------- -----CTTTC GTGTTCTTGA GTGTTCATCT ACAGTTCTCG CA-TATTT-G
86B.24   TCT-CCAACA TT---CTTTC GTGTTCTTGA GTGTTCATCT ACAGTTCTCG CA-TATTT-G
86B.29   TCT-CCAACA TT---CTTTC GTGTTCTTGA GTGTTCATCT ACAGTTCTCG CA-TATTT-G
86B.5    ---------- -----CTTTC GTGTTCTTGA GTGTTCATCT ACAGTTCTCG CA-TATTT-G
87B      -CT-CCAACA T--CTCTTTC GTGTTCTTGA GTGTTCATCT ACAGTTCTCG CA-TATTT-G
90D      TCT-CCAACA T----CTTTC GTGTTCTTGA GTGTTCATCT ACAGTTCTCG CA-TATTT-G
```

ETS type 4

```
118A     TCT-CCCACA TT---CTTAC GTGTTCTTGA GTGTTTATCT ACAGTTCTAA CA-TATTT-G
121B     TCT-CCCACA TT---CTTAC GTGTTCTTGA GTGTTTATCT ACAGTTCTAA CA-TATTT-G
123F     ---------- TT---CTTAC GTGTTCTTGA GTGTTTATCT ACAGTTCTCA CA-TATTT-G
126.29.1TCT-CCAACA TT---CTTAC GTGATCTTGA GTGTTTATCT GCAGTTCTCA CA-TATTT-G
126.29.3TCT-CCAACA TT---CTTAC GTGATCTTGA GTGTTTATCT GCAGTTCTCA CA-TATTT-G
63B      TCT-CCCACA TT---CTTAC GTGTTCTTGA GTGTTTGTCT ACAGTTCTCA CA-TATTT-G
81A      TNT-CCAGCA TT---CGTGN GTGNTCNTGA GTGTTCATCA ACAGTTCGTG CA-TATTT-A
76A      TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
76E      TCT-CCAATA TT---CTTGC GTGTTCTTGA GTGTTCATCT ACAGTTCTTG CA-TATTT-G
78B      TCT-CCAGCA TT---CTTGC GTGTTCTTGA GTGTTCATCA ACAGTTCTTG CA-TATTT-A
84A.1    TCA-CCCACA TT---CTTAC GTGTTCTTGA GTGTTTATCT ACAGTTCTAA CA-TATTT-G
84A.2    TCT-CCCACA TT---CTTAC GTGTTCTTGA GTGGTTATCT ACAGTTCTAA CA-TATTT-G
83B      TCT-CTAATA TT--TCTTGC GTGTTCTTGA GTGCTCATCT ACAGTTCTTG CA-TATTT-G
87A      TCT-CCAACA TT---CTTAC GTGTTCTTGA GTGTTCATCT GCAGTTCTCG CA-TATTT-G
126.WB3  TCT-CCCACA TT---CTTAC GTGTTCTTGA GTGTTTATCT ACAGTTCTCA CA-TATTT-G
Pur-34   ---------- ---------- ---------- ---------- ---------- ----------
Pur-132  ---------- ---------- ---------- ---------- ---------- ----------
```

Bases 61-120
ETS type 1

```
1        AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
10       AGCTG-TCTA ATCTACTT-T GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
14       AGCTG-TCTA ATCTACTT-T GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
15       AGCTG-TCTA ATCTACTT-T GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
12       AGCTG-TCTA ATCTACTT-T GGCAGGCR-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
17       AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
25       AGTTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTAACT TAGCTTGTTC TTGGAGGTGA
26       AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
27C.1    AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
27C.10   AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
28A      AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
3        AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGGTTGTTC TTGGAGGTGA
41B      AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTAACT TAGCTTGTTC TTGGAGGTGA
45C      AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
5Q       AGCTG-TCTA ATCTACTT-T GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
101A     AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
102B     AGCTG-GCTA ATCTACTT-T GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
104      AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
108C     AGCTACTCTG ATCTA-TT-T GGCAGGCACT TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
111A.4   AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGATT TAGCTTGTTC TTGGAGGTGA
111A.5   AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGATT TAGCTTGTTC TTGGAGGTGA
120C     AGCTG-TCTA ATCTACTT-T GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
123G     ---------- ---------- ---------- ---------- ---------- ----------
131      AGCTG-TCTA ATCTACTT-T GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
30A      AGCTG-TCTA ATCTACTT-T GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
57B      AGTTG-TCTA ATCTACTT-T GGCAGGCG-T TTAGCTTACT TAGCTTGTTC TTGGAGGTGA
58B      AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
62D.2    AGCTG-TCTA ATCTACTT-T GGCAAGCG-T TTAGCTAACT TAGCTTGTTT TTGGAGGTGA
62D.3    AGCTG-TCTA ATCTACTT-T GGCAAGCG-T TTAGCTAACT TAGCTTGTTC TTGGAGGTGA
65D.7    AGCTG-TCTA ATTTACTT-T GGCAGGCG-T TTGGATGACT TAGCTTGTTC TTGGAGGTGA
65D.3    AGCTG-TCTA ATTTACTT-T GGCAGGCG-T TTGGATGACT TAGCTTGTTC TTGGAGGTGA
67B      AGCTG-TCTA ATCTACTT-T GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
69A      AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
69B      AGCTG-TCTA ATCTACTT-T GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
76A      AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGCATGTTC TTGGAGGTGA
76E      AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
78A.10   AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGCATGTTC TTGGAGGTGA
78A.5    AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGCATGTTC TTGGAGGTGA
8        AGCTG-TCTA ATCTACTT-T GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
80A      AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGACT TARCTTGTTC TTGGAGGTGA
82A      AGCTG-TCTA ATCTACTT-T GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
85A      AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
86B      AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
90D      AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
93C.2    AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTAGCTAACT TAGCTTGTTC TTGGAGGTGA
93C.3    AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTAGCTAACT TAGCTTGTTC TTGGAGGTGA
98A      AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTAGCTAACT TAGCTTGTTC TTGGAGGTGA
126E     AGCTG-TCTA ATTTACTT-T GGCAGGCG-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
103A     AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
50D      AGCTG-GCTG ATCTACYTCT GGCAGGCG-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
68A      AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
75A      AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGSTTGTTC TTGGAGGTGA
```

Bases 61-120

ETS type 2

```
121C     AGATA-TCTG ATCTATTT-- AGCAGGCA-T TTGGCAGACT TGGCTTGTTC TTGGAGGTGA
18       GGCTG-TCTG ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
19       AGCTA-TCTG ATCTATTT-- AGCAGGCA-T TTGGCAGACT TAGCTTGTTC TTGGAGGTGA
20       AGCTA-TCTG ATCTATTT-- AGCAGGCA-T TTGGCAGACT TAGCTTGTTC TTGGAGGTGA
21       AGCTA-TCTG ATCTATTT-- AGCAGGCA-T TTGGCAGACT TAGCTTGTTC TTGGAGGTGA
122A     AGCTA-TCTG ATCTATTT-- AGCAGGCA-T TTGGCAGACT TAGCTTGTTC TTGGAGGTGA
126.29   AGCTA-TCTG ATCTATTT-- AGCAGGCA-T TTGGCAGACT TAGCTTGTTC TTGGAGGTGA
26       AGCTA-TCTG ATCTATTT-- AGCAGGCA-T TTGGCAGACT TAGCTTGTTC TTGGAGGTGA
52B      AGCTA-TCTG ATCTATTT-- AGCAGGCA-T TTGGCAGACT TAGCTTGTTC TTGGAGGTGA
57B      AGATA-TCTG ATCTATTT-- AGCAGGCA-T TTGGCAGACT TAGCTTGTTG TTGGAGGTGA
65C.3    AGCTA-TCTG ATCTATTT-- GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
65C.4    AGCTA-TCTG ATCTATTT-- GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
93C      AGCTA-TCTG ATCTATTT-- AGCAGGCA-T TTGGCAGACT TAGCTTGTTC TTGGAGGTGA
118A     AGATA-TCTG ATCTATTT-- AGCAGGCA-T TTGGCAGACT TGGCTTGTTC TTGGAGGTGA
126.BC1  AGCTA-TCTG ATCTATTT-- AGCAGGCA-T TTGGCAGACT TAGCTTGTTC TTGGAGGTGA
126E.1   AGCTA-TCTG ATCTATTT-- AGCAGGCA-T TTGGCAGACT TAGCCTGTTC TTGGAGGTGA
126E.3   AGCTA-TCTG ATCTATTT-- AGCAGGCA-T TTGGCAGACT TAGCTTGTTC TTGGAGGTGA
126E.4   AGCTA-TCTG ATCTATT--- AGCAGGCA-T TTGGCAGACT TAGCTTGTTC TTGGAGGTGA
126A     AGCTA-TCTG ATCTATTT-- AGCAGGCA-T TTGGCAGACT TAGCTTGTTC TTGGAGGTGA
61       AGCTA-TCTG ATCTATTT-- AGCAGGCA-T TTGGCAGACT TAGCTTGTTC TTGGAGGTGA
```

ETS type 3

```
121B     AGCTG-TCTG ATCTACTT-T GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGTGGTTA
66C      AGCTG-CCTG ATCTACTT-T GGCTGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTTA
66D      AGCTG-TCTG ATCTACTT-T GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTTA
67A      AGCTG-TCTG ATCTACTT-T GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTTA
69B      AGCTG-TCTG ATCTACTT-T GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTTA
71A      AGCTG-TCTG ATCTACTT-T GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTTA
77B      AGCTG-CCTG ATCTACTT-T GGCTGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTTA
82A      AGCTG-TCTG ATCTTCTT-T GGCAGACA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
83B      AGCTG-TCTG ATCTACTT-T GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTTA
86B.24   AGCTG-TCTG ATCTACTT-T GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTTA
86B.29   AGCTG-TCTG ATCTACTT-T GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTTA
86B.5    AGCTG-TCTG ATCTACTT-T GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTTA
87B      AGCTG-TCTG ATCTACTT-T GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTTA
90D      AGCTG-TCTG ATCTACTT-T GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTTA
```

ETS type 4

```
118A     AGCTA-TCTG ATCTATTT-- GGCAGGCA-T TCGGCTGACT TAGCTTGTTC TTGGAGGTGA
121B     AGCTA-TCTG ATCTATTT-- GGCAGGCA-T TCGGCTGACT TAGCTTGTTC TTGGAGGTGT
123F     AGCTA-TCTG ATCTATTT-- GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
126.29.1AGCTA-TCTG ATCTATTT-- GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
126.29.3---------- ---------- GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
63B      AGCTA-TCTG ATCTATTT-- GGCAGGCA-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
81A      AGCTG-TCTG ATATACTT-T GGCAGGCA-T TTGGCTGACT TAGCTAGTTC TTGGAGGTGA
76A      AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGCTAGTTC TTGGAGGTGA
76E      AGCTG-TCTA ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
78B      AGCTG-TCTG ATCTACTT-T GGCAGGCA-T TTGGCTGACT TAGCTAGTTC TTGGAGGTGA
84A.1    AGCTA-TCTG ATCTATTT-- GGCAGGCA-T TCGGCTGACT TAGCTTGTTC TTGGAGGTGA
84A.2    AGCTA-TCTG ATCTATTT-- GGCAGGCA-T TCGGCTGACT TAGCTTGTTC TTGGAGGTGA
83B      GGCTG-TCTG ATCTACTT-T GGCAGGCG-T TTGGCTGACT TAGCTTGTTC TTGGAGGTGA
87A      AGCTG-TCTG ATCT-TCTTT GGCAGGTG-T TTGGCTGACT TAGCTTGTTC TTGGAGGCGA
126.WB3  AGCTA-TCTG ATCTATTT-- GGCAGGCA-T TCGGCTGACT TAGCTTGTTC TTGGAGGTGA
Pur-34   ---------- ---------- ---------- ---------- ---------- ----------
Pur-132  ---------- ---------- ---------- ---------- ---------- ----------
```

Bases 121-174
ETS type 1

```
1        TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
10       TCAAGTCTAA GGTCATACAG GTAGAGGACT TCAAAACAAT CATCCCATGA CAGT
14       TCAAGTCTAA GGTCACACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
15       TCAAGTCTAA GGTCAYACAG GTAGAGGACT TCAAAACAAT CATCCCATGA CAGT
12       TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
17       TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
25       TTAAGTCTAA GGTCGCACGG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
26       TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
27C.1    TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
27C.10   TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCATAACAGT CATCCCATGA CAGT
28A      TCAAGTCTAA GGTCATACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
3        TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
41B      TTAAGTCTAA GGTCGCACGG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
45C      TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
5        TCAAGTCTAA GGTCATACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
101A     TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
102B     TCAAGTCTAA GGTCATACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
104      TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
108C     TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
111A.4   TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
111A.5   TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
120C     TCAAGTCTAA GGTCATACAG GTAGAGGACT TCAAAACAAT CATCCCATGA CAGT
123G     -----TCTAA GGTCATACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
131      TCAAGTCTAA GGTCATACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
30A      TCAAGTCTAA GGTCATACAG GTAGAGGACT TCAAAACAAT CATCCCATGA CAGT
57B      TCAAATCTAA GGTTGCACGG GTAGAGGACT TCAAAATAGT TATCCCATGA CAGT
58B      TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
62D.2    TCAAGTCTAA GATCGCATGG GTAGAGGACT TCAAAATAGT CATCCCATGA CAGT
62D.3    TCAAGTCTAA GGTCGCACGG GTAGAGGACT TCAAAATAGT TATCCCATGA CAGT
65D.7    TCAAGTCTAA GATCGCATGG GTAGAGGACT TCAAAATAGT CATCCCATGA CAGT
65D.3    TCAAGTCTAA GATCGCATGG GTAGAGGACT TCAAAATAGT CATCCCATGA CAGT
67B      TCAAGTCTAA GGTCATACAG GTAGAGGACT TCAAAATAAT CATCCCATGA CAGT
69A      TCAAGTCTAA GGTCATACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
69B      TCAAGTCTAA GGTCATACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
76A      TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
76E      TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
78A.10   TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCAAAACAGT TATCCCATGA CAGT
78A.5    TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCAAAACAGT TATCCCATGA CAGT
8        TCAAGTCTAA GGTCATACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
80A      TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
82A      TCAAGTCTAA GGTCATACAG GTAGAGGACT TCAAAACAAT CATCCCATGA CAGT
85A      TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
86B      TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCAAAATAGT CATCCCATGA CAGT
90D      TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCAAAATAGT CATCCCATGA CAGT
93C.2    TCAAATCTAA GGTCGCACGG GTAGAGGACT TCAAAATAGT TATCCCATGA CAGT
93C.3    TCAAATCTAA GGTCGCACGG GTAGAGGACT TCAAAATAGT TATCCCATGA CAGT
98A      TCAAATCTAA GGTCGCACGG GTAGAGGACT TCAAAATAGT TATCCCATGA CAGT
126E     TCAAGTCTAA GATCGCATGG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
103A     TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
50D      TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
68A      TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
75A      TCAAGTCTAA GGTCGCACAG GTAGAGGACT TCAAAACAGT CATCCCATGA CAGT
```

Bases 121-174

ETS type 2

```
121C    TCAAGTCTAA GGCCGCATGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
18      TCAAGTCTAA GGCCGCACGG GTAGAGGACT TCAAAGCAGT CATCCCGTGA CACT
19      TCAAGTCTAA GGCCGCATGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
20      TCAAGTCTAA GGCCGCATGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
21      TCAAGTCTAA GGCCGCATGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
122A    TCAAGTCTAA GGCCGCATGG GTAGAGGACT TCAAAACAGT TATCCCGTGA CACT
126.29  TCAAGTCTAA GGCCGCATGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
26      TCAAGTCTAA GGCCGCATGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
52B     TCAAGTCTAA GGCCGCATGG GTAGAGGACT TCAAAACAGT TATCCCGTGA CACT
57B     TCAAGTCTAA GGCCGCATGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
65C.3   TCAAGTCTAA GGCCGCACGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
65C.4   TCAAGTCTAA GGCCGCACGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
93C     TCAAGTCTAA GGCCGCATGG GTAGAGGACT TCAAAACGGT CATCCCGTGA CACT
118A    TCAAGTCTAA GGCCGCATGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
126.BC1 TCAAGTCTAA GGCCGCATGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
126E.1  TCAAGTCTAA GGCCGCATGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
126E.3  TCAAGTCTAA GGCCGCATGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
126E.4  TCAAGTCTAA GGCCGCATGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
126A    TCAAGTCTAA GGCCGCATGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
61      TCAAGTTTAA GGCCGCATGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
```

ETS type 3

```
121B    TCAAGTCTAA GGCCGCATGG GTAGAGGACT TCAAAATAGT CATCCCGTGA CACT
66C     TCAAGTCTAA GGCCGCATGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
66D     TCAAGTCTAA GGCCGCACGG GTAGAGGACT TCAAAACAGT TATCCCGTGA CACT
67A     TCAAGTCTAA GGCCGCACGG GTAGAGGACT TCAAAACAGT TATCCCGTGA CACT
69B     TCAAGTCTAA GGCCGCATGG GTAGAGGACT TCAAAATAGT CATCCCGTGA CACT
71A     TCAAGTCTAA GGCCGCATGG GTAGAGGACT TCAAAATAGT CATCCCGTGA CACT
77B     TCAAGTCTAA GGCCGCATGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
82A     TCAAGTCTAA GGCTGCACGG GTAGAGGACT TCAAAACAGT CATCCCGCGA CACT
83B     TCAAGTCTAA GGCCGCATGG GTAGAGGACT TCAAAATAGT CATCCCGTGA CACT
86B.24  TCAAGTCTAA GGCCACACGG GTAGAGGACT TCAAAACAGT TATCCCGTGA CACT
86B.29  TCAAGTCTAA GGCCACACGG GTAGAGGACT TCAAAACAGT TATCCCGTGA CACT
86B.5   TCAAGTCTAA GGCCACACGG GTAGAGGACT TCAAAACAGT TATCCCGTGA CACT
87B     TCAAGTCTAA GGCCGCACGG GTAGAGGACT TCAAAACAGT TATCCCATGA CACT
90D     TCAAGTCTAA GGCCGCACGG GTAGAGGACT TCAAAACAGT TATCCCATGA CACT
```

ETS type 4

```
118A    TCAAGTCTAA GGCCGCACGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
121B    TCAAGTCTAA GGCCGCACGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
123F    TCAAGTCTAA GGCCGCACGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
126.29.1TCAAGTCTAA GGCCGCACGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
126.29.3TCAAGTCTAA GGCCGCACGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
63B     TCAAGTCTAA GGCCGCACGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
81A     TCAAGTCTAA GGCCGCATGG GTAGAGGACT --AAAACAGT CATCCCTTGA CACT
76A     TCAAGTCTAA GGCCGCATGG GTAGAGGACT TGAAAACAGT CATCCCTTGA CACT
76E     TCAAGTCTAA GGCCGCATGG GTAGAGGACT TGAAAACAGT CATCCCTTGA CACT
78B     TCAAGTCTAA GGCCGCATGG GTAGAGGACT T--AAACAGT CATCCCTTGA CACT
84A.1   TTAAGTCTGA GGCCGCACGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
84A.2   TCAAGTCTGA GGCCGCACGG GTAGAGGACT TCAAAACAGT CATCCCGTGA CACT
83B     TCAAGTCTAA GGCCGCACGG GTAGAGGACT TCAAAGCAGT CATCCCGTGA CAGT
87A     TCAAGTCTA- GGCCGCACGA GTAGAGGACT TCAAAACAAT TATCCTGTAA CAGT
126.WB3 TCAAGTCTAA GGCCGCACGG GTAGAGGACT TCAAAACAAT CATCCCGTGA CACT
Pur-34  ---------- ---------- ---------- ---------- ---------- ----
Pur-132 ---------- ---------- ---------- ---------- ---------- ----
```

Bases 175-234

3′ common region→

ETS type 1

```
1        ATGACCCTCA CTCTCRGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
10       ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
14       ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
15       ATGACCCTCA CTCTCGGGTR CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
12       ATGACCCTCA STCTCGGGTG CTTCTGAAGG TGCAATGGGT GCTATTCGAT GTCGTATATG
17       ATGACCTTCA CTCTCGGGCA MTTGTGAAGG TGCATTGGKT GCTACTCGAT MTCTTCTATG
25       ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTCGTATG
26       ATGACCCTCA CTCTTGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
27C.1    ATGACCCTTA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTTGAT CTCTTCTATG
27C.10   ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
28A      ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
3        ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GGTAATCGGT CTCTTCTATG
41B      ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTGTATG
45C      ATGACCCTCA CTCTTGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
5        ATGACCCTCA CTGTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
101A     ATGACCCTCA CTCTTGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
102B     ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
104      ATGACCCTCA CTCTTGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
108C     ATGACCTTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
111A.4   ATGACCCTCA CTCTTGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATA
111A.5   ATGACCCTCA CTCTTGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTCATATG
120C     ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
123G     ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GGTACTCGAT CTTTTTTATG
131      ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
30A      ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
57B      ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTCGTATG
58B      ATGACCCTTA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
62D.2    ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT CCTACTCGAT CTCTCGTATG
62D.3    ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTCGTATG
65D.7    ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT CCTACTCGAT CTCTCGTATG
65D.3    ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT CCTACTCGAT CTCTCGTATG
67B      ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
69A      ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
69B      ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
76A      ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCCACTCGAT CTCTTCTATG
76E      ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTAGTT GCTACTCGAT CTCTTCTATG
78A.10   ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
78A.5    ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
8        ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
80A      ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTTTTCTATG
82A      ATGACCCTCA CTCTTGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTCGTATG
85A      ATGACCCTTA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
86B      ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
90D      ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTTCTATG
93C.2    ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTCGTATG
93C.3    ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTCGTATG
98A      ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTCGTATG
126E     ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT CCTACTCGAT CTCTCGTATG
103A     ATGACCCTCA CTCTTGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCAAT CTCTTCTATG
50D      ATGACCCTCA CTCTTGGGTG CTTGTGAAGG TGCATTGGTT GCTAYTGGAT CTCTTATATG
68A      ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCAATGATT GATATATGAT CTATTCTAWK
75A      ATGACCCTCA CTCTCGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGRT CTCTTCTATG
```

Bases 175-234

ETS type 2

```
121C    ATGACCCGTA CTCTCGGGTG CTTGTGAAGG TGCACTGGTT TCTACTCGAT CTCTTGGATG
18      ATGACCCGCA CTCTCGGGTG CTTGTGAAGG TGCACTGGTT GCTACTCGAT CTCTCGGATG
19      ATGACCCGTA CTCTCGGGTG CTTGTGAAGG TGCACTGGTT TCTACTCGAT CTCTTGGATA
20      ATGACCCGTA CTCTCGGGTG CTTGTGAAGG TGCACTGGTT TCTACTAGAT CTCTTGGATG
21      ATGACCCGTA CTCTCGGGTG CTTGTGAAGG TGCACTGGTT TCTACTAGAT CTCTTGGATG
122A    ATGACCCGTA CTCTCGGGTG CTTGTGAAGG TGCACTGGTT TCTACTCGAT CTCTTGGATG
126.29  ATGACCCGTA CTCTCAGGTG CTTGTGAAGG TGCACTGGTT TCTACTCGAT CTCTTGGATG
26      ATGACCCGTA CTCTCGGGTG CTTGTGAAGG TGCACTGGTT TCTACTCGAT CTCTTGGATG
52B     ATGACCCGTA CTCTCGGGTG CTTGTGAAGG TGCACTGGTT TCTACTCGAT CTCTTGGATG
57B     GTGACCCGTA CTCTCGGGTG CTTGTGAAGG TGCACTGGTT TCTACTCGAT CTCTTGGATG
65C.3   ATGACCCGTA CTCTCGGGTG CTTGTGAAGG TGCACTGGTT TCTACTCGAT CTCTTGGATT
65C.4   ATGACCCGTA CTCTCGGGTG CTTGTGAAGG TGCACTGGTT TCTACTCGAT CTCTTGGATT
93C     ATGACCCGTA CTCTCGGGTG CTTGTGAAGG TGCACTGGTT TCTACTCGAT CTCTTGGATG
118A    ATGACCCGTA CTCTTGGGTG CTTGTGAAGG TGCACTGGTT TCTACTCGAT CTCTTGGATT
126.BC1 ATGACCCGTA CTCTCGGGTG CTTGTGAAGG TGCACTGGTT TCTACTCGAT CTCTTGGATG
126E.1  ATGACCCGTA CTCTCGGGTG CTTGTGAAGG TGCACTGGTT TCTACTCGAT CTCCTGGATG
126E.3  ATGACCCGTA CTCTCGGGTG CTTGTGAAGG TGCACTGGTT TCTACTCGAT CTCTTGGATG
126E.4  ATGACCCGTA CTCTCGGGTG CTTGTGAAGG TGCACTGGTT TCTACTCGAT CTCTTGGATG
126A    ATGACCCGTA CTCTCGGGTG CTTGTGAAGG TGCACTGGTT TCTACTCGAT CTCTTGGATG
61      ATGACCCGTA CTCTCGGGTG CTTGTGAAGG TGCACTGGTT TCTACTCGAT CTCTTGGATG
```

ETS type 3

```
121B    ---------- ---------- ---------- ---------- ---------- ----------
66C     ---------- ---------- ---------- ---------- ---------- ----------
66D     ---------- ---------- ---------- ---------- ---------- ----------
67A     ---------- ---------- ---------- ---------- ---------- ----------
69B     ---------- ---------- ---------- ---------- ---------- ----------
71A     ---------- ---------- ---------- ---------- ---------- ----------
77B     ---------- ---------- ---------- ---------- ---------- ----------
82A     ---------- ---------- ---------- ---------- ---------- ----------
83B     ---------- ---------- ---------- ---------- ---------- ----------
86B.24  ---------- ---------- ---------- ---------- ---------- ----------
86B.29  ---------- ---------- ---------- ---------- ---------- ----------
86B.5   ---------- ---------- ---------- ---------- ---------- ----------
87B     ---------- ---------- ---------- ---------- ---------- ----------
90D     ---------- ---------- ---------- ---------- ---------- ----------
```

ETS type 4

```
118A    ATGACCCGTA CTCTCGGGTG CTTGTGAAGG TGCACTGGTT TCTACTCGAT CTCTTGGATT
121B    ATGACCCGTA CTCTTGGGTG CTTGTGAAGG TGCACTGGTT TCTACTCGAT CTCTTGGATT
123F    ATGACCCGTA CTCTCGGGTG CTTGTGAAGG TGCACTGGTT TCTACTCGAT CTCTTGGATT
126.29.1ATGACCCGTA CTCTCAGGTG CTTGTGAAGG TGCACTGGTT TCTACTCGAT CTCTTGGATG
126.29.3ATGACCCGTA CTCTCAGGTG CTTGTGAAGG TGCACTGGTT TCTACTCGAT CTCTTGGATG
63B     ATGACCCGTA CTCTCGGGTG CTTGTGAAGG TGCACTGGTT TCTACTCGAT CTCTTGGATT
81A     ATGACTCGTA CTCTTGGGTG CTTGTGAAGG TGCACTGGTT GCTACTCGAT CTCTTGGATG
76A     ATGACCCGTA CTCTTGGGTG CTTGTGAAGG TGCACTGGTT GCTACTCGAT CTCTTGGATG
76E     ATGACCCGTA CTCTTGGGTG CTTGTGAAGG TGCATTGGTT GCTACTCGAT CTCTCGGATG
78B     ATGACTCGTA CTCTTGGGTG CTTGTGAAGG TGCACTGGTT GCTACTCGAT CTCTTGGATG
84A.1   ATGACCCGTA CTCTTGGGTG CTTGTGAAGG TGCACTGGTT TCTACTCGAT CTCTTGGATT
84A.2   ATGACCCGTA CTCTTGGGTG CTTGTGTGGG TGCACTGGTT TCTACTCGAT CTCTTGGATT
83B     ---------A TTCTCGGGTG CTTGTGAAGG TGCACTGGTT GCTACTCGAT CTCTCGGATG
87A     ---------- ---------- ---------- ---------- ---------- ----------
126.WB3 ATGACCCGTA CTCTCGGGTG CTTGTGAAGG TGCACTGGTT TCTACTCGAT CTCTTGGATT
Pur-34  AAGACCTGCA TTGTTGGGTT TTTGTGTCGG TCCACGTGTT GTCACTCGAT CTCTCGGATG
Pur-132 AAGACTAGCA TTGTTAGGTT TTTGTGTCGG TCCATGTGTT GTCACTCGAT CTCTCGGATG
```

Bases 235-294
ETS type 1

```
1        CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-GAC GTC-AGATGC AAAA-GCATT
10       CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCY-GAM GTC-AGATGC AAAA-GCATT
14       CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-GAC GTC-AGATGC AAAA-GCATT
15       CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-GAC GTC-AGATGC AAAA-GCATT
12       CRGAACATAC AAAGGGCGCG GGGCTTTTCA GCCCCC-GAC GTC-AGATGC AAAW-GCATT
17       CGGAACATAC AAAGGGCGCG GGGGTCTTTA GCCCCC-GAC GTC-AGATGC AAAA-GCATT
25       CAGAACATAC AAAGGGTGCG GGGGTCTTCA GCCCTC-GAC GTC-CGATGC AAAA-GCATT
26       CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-GAC GTC-AGATGC AAAA-GCATT
27C.1    CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-GAC GTC-AGATGC AAAA-GCATT
27C.10   CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-GAC GTC-AGATGC AAAA-GCATT
28A      CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCT-GAM GTC-AGATGC AAAA-GCATT
3        CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-GAC GTC-AGATGC AAAA-GCATT
41B      TAGAACATAC AAAGGGTGCG GGGGTCTTCA GCCCCC-GAC GTC-CAATGC AAAA-GCATT
45C      CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-GAC GTC-AGATGC AAAA-GCATT
5        CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-GAC GTC-AGATGC AAAA-GCATT
101A     CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-GAC GTC-AGATGC AAAA-GCATT
102B     CGGAACATAC AAAGGGCGCG GGGGTCCTCA GCCCCC-GAC GTC-AGATGC AAAA-GCATT
104      CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCT-GAC GTC-AGATGC AAAA-GCATT
108C     CGGAACATAC AAAGGGCGCG GTGGTCTTCA GCCCCC-GAC GTC-AGAAGC AAAA-GCATT
111A.4   CGGAACATAC AAAGGGCGCG GGGGTCTTCC GCCCCC-GAC GTC-AGACGC AAAA-GCATT
111A.5   CAGAACATAC AAAGGGTACG GGGGTCTTCA GCCCCC-GAC GTC-CGATGC AAAA-GCATT
120C     CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-GAC GTC-AGATGC AAAA-GCATT
123G     CGGAACATAC AAAGGGSGGG GGGGT----- ---------- ---------- ----------
131      CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCT-GAA GTC-AGATGC AAAA-GCATT
30A      CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-GAC GTC-AGATGC AAAA-GCATT
57B      CGGAACATAC AAAGGGCGAG GGGGTCTTCA GCCCTC-GAT GTCC-GATGC AAAA-CCATT
58B      CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-GAC GTC-AGATGT AAAA-GCATT
62D.2    CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-GAC GTC-CGATGC AAAA-CCATT
62D.3    CCGAACATAC AAAGGGCGCG GGGGTTTTCA GCCCCC-GAC GTCC-GATGC AAAA-CCATT
65D.7    CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-GAC GTC-CGATGC AAAA-CCATT
65D.3    CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-GAC GTC-CGATGC GAAA-CCATT
67B      CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCA-GAC GTC-AGATGC AAAA-GCATT
69A      CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-GAC GTC-AGATGC AAAA-GCATT
69B      CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-GAC GTC-AGATGC AAAA-GCATT
76A      CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCN-TAC GTC-AGATGC AAAA-GCATT
76E      CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-GAC GTC-AGATGC AAAA-GCATT
78A.10   CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-TAC GTC-AGATGC AAAA-GCATT
78A.5    CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-TAC GTC-AGATGC AAAA-GCATT
8        CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-GAC GTC-AGAT-C AAAA-GCATT
80A      CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCTCC-GAC GTC-AGATGC AAAA-GCATT
82A      CAGAACATAC AAAGGGTGCG GGGGTCTTCA GCCCTC-GAC GTCC-GATGC AAAA-GCATT
85A      CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-GAC GTC-AGATGT AAAA-GCATT
86B      CAGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-GAC GTC-AGATGC AAAA-GCATT
90D      CAGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-GAC GTC-AAATGC AAAA-GCATT
93C.2    CGGAACATAC AAAGGGCGAG GGGGTCTTCA GCCCTC-GAT GTC-CGATGA GAA--CCATT
93C.3    CGGAACATAC AAAGGGCGAG GGGGTCTTCA GCCCTC-GAT GTCC-GATGC AAAA-CCATT
98A      CGGAACATAC AAAGGGCGAG GGGGTCTTCA GCCCTC-GAT GTCC-GATGC AAAA-CCATT
126E     CGGAACATAC AAAGGGCGCG AGGGTCTTCA GCCCCC-GAC GTCC-GATGC AAAAAGCATT
103A     CGGAACATAC AAAGGGCGCG GGAGTCTTCA GCCCCT-GAC GTCCAGATGC AAAA-GCATT
50D      CGGAACATAC AAAGGGTGCG GGGGTCTTCA GCCCCM-GAC GTC-AGAWGC AAGA-GCATT
68A      CGGAACAATC AAAGAGCKCG GGGGTCT--- ---------- ---------- ----------
75A      CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-GAC GTC-AGATGC AAAA-GCATT
```

```
Bases 235-294
ETS type 2
121C    CGGAACATAC AAAAGGCGCG GGGGTCTTCA GCCCCC-GAC GTCC-GATGC AAAAACCATT
18      CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCCC-GAC GTCC-GATGC AAAA-GCATT
19      CGGAACATAC AAAAGGCGCG GGGGTCTTCA GCCCCC-GAC GTCC-GATGC AAAAAGCATT
20      CGGAACATAC AAAAGGCGCG GGGGTCTTCA GCCCCC-GAC GTCC-GATGC AAAAAGCATT
21      CGGAACATAC AAAAGGCGCG GGGGTCTTCA GCCCCC-GAC GTCC-GATGC AAAAAGCATT
122A    CGGAACATAC AAAAGGCGCG GGGGTCTTCA GCCCCC-GAC GTCC-GATGC AAAAAGCATT
126.29  CGGAACATAC AAAAGGCGCG GGGGTCTTCA GCCCCC-GAC GTCC-GATGC AAAAAGCATT
26      CGGAACATAC AAAAGGCGCG GGGGTCTTCA GCCCCC-GAC GTCC-GATGC AAAAAGCATT
52B     CGGAACATAC AAAAGGCGCG GGGGTCTTCA GCCCCC-GAC GTCC-GATGC AAAAAGCATT
57B     CGGAACATAC AAAAGGCGCG GGGGTCTTCA GCCCCC-GAC GTCC-GATGC AAAAACCATT
65C.3   CGGAACATAC AAAAGGTGCG GGGGTTTTCA GCCCCC-GAC GTCC-GATGC AAAA-GCATT
65C.4   CGGAACATAC AAAAGGTGCG GGGGTTTTCA GCCCCC-GAC GTCC-GATGC AAAA-GCATT
93C     CGGAACATAC AAAAGGCGCG GGGGTCTTCA GCCCCC-GAC GTCC-TAAAA ACCATTCATT
118A    CGGAACATAC AAAAGGTGCG GGGGTCTTCA GCCCCC-GAC GTTC-GATGC AAA--GCATT
126.BC1 CGGAACATAC AAAAGGCGCG GGGGTCTTCA GCCCCC-GAC GTCC-GATGC AAAAASCATT
126E.1  CGGAACATAC AAAAGGCG-- GGGGTCTTCA GCCCCC-GAC GTCC-GATGC AAAAAGCATT
126E.3  CGGAACATAC AAAAGGCG-- GGGGTCTTCA GCCCCC-GAC GTCC-GATGC AAAAAGCATT
126E.4  CGGAACATAC AAAAGGCG-- GGGGTCTTCA GCCCCC-GAC GTCC-GATGC AAAAAGCATT
126A    CGGAACATAC AAAAGGCG-- GGGGTCTTCA GCCCCC-GAC GTCC-GATGC AAAAAGCATT
61      CAGAACATAC AAAAGGCGCG GGGGTCTTCA GCCCCC-GAC GTCC-GATGC AAAAAGCATT
ETS type 3
121B    ---------- ---------- ---------- ---------- ---------- ----------
66C     ---------- ---------- ---------- ---------- ---------- ----------
66D     ---------- ---------- ---------- ---------- ---------- ----------
67A     ---------- ---------- ---------- ---------- ---------- ----------
69B     ---------- ---------- ---------- ---------- ---------- ----------
71A     ---------- ---------- ---------- ---------- ---------- ----------
77B     ---------- ---------- ---------- ---------- ---------- ----------
82A     ---------- ---------- ---------- ---------- ---------- ----------
83B     ---------- ---------- ---------- ---------- ---------- ----------
86B.24  ---------- ---------- ---------- ---------- ---------- ----------
86B.29  ---------- ---------- ---------- ---------- ---------- ----------
86B.5   ---------- ---------- ---------- ---------- ---------- ----------
87B     ---------- ---------- ---------- ---------- ---------- ----------
90D     ---------- ---------- ---------- ---------- ---------- ----------
ETS type 4
118A    CGGAACATAC AAAAGGTGCG GGGGTCTTCA GCCCCC-GAC GTCC-GATGC AAAAACCATT
121B    CGGAACATAC AAAAGGTGCG GGGGTCTTCA ACCCCC-GAC GTTC-GATGC AAA--GCATT
123F    CGGAACATAC AAAAGGTGCG GGGGTCTTCA GCCCCC-GAC GTCC-GATGC AAAA-GCATT
126.29.1CGGAACATAC AAAAGGCGCG GGGGTCTTCA GCCCCC-GAC GTCC-GATGC AAAAACCATT
126.29.3CGGAACATAC AAAAGGCGCG GGGGTCTTCA GCCCCC-GAC GTCC-GATGC AAAAACCATT
63B     CGGAACATAC AAAAGGTGCG GGGGTCTTCA GCCCCC-GAC GTCC-GATGC AAAA-GCATT
81A     CGGAACATAC AAAGGGCGTG GGGGTCTTCA GCCCCC-GAC GTTC-GATGC AAAA-GTATT
76A     CGGAACATAC AAAGGGCGTG GGGGTCTTCA GCCCCC-GAC GTTC-GATGC AAAA-GTATT
76E     CGGAACATAC AAAGGGCGTG GGGGTCTTCA GCCCCC-GAC GTTC-GATAC AAAA-GTATT
78B     CGGAACATAC AAAGGGCGTG GGGGTCTTCA GCCCCC-GAC GTTC-GATGC AAAA-GTATT
84A.1   CGGAACATAC AAAAGGTGCG GGGGTCTTCA GCCCCC-GAC GTTC-GATGC AAA--GCATT
84A.2   CGGAACATAC GAAAGGTGCG GGGGTCTTCA GCCCCC-GAC GTTC-GATGC AAA--GCATT
83B     CGGAACATAC AAAGGGCGCG GGGGTCTTCA GCCCC--GAC GTCC-GATGC AAAA-GCATT
87A     ---------- ---------- ---------- ---------- ---------- ----------
126.WB3 CGGAACATAC AAAAGGTGCG GGGGTCTTCA GCCCCC-GAC GTCC-GATGC AAAA-GCATT
Pur-34  CAGAACATGT TGCGGGCGC- AGGGTCTTCG GCCCCTTGAC GTCCAAA--- ACAAAGCGTT
Pur-132 CGGAACATGT TGCGGGTGT- AGGTTCTTCG GCCCTTTGAC GTCCAAA--- ACAAAGCGTT
```

Bases 295-354
ETS type 1

```
1        T-GTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
10       TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
14       TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
15       TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
12       TTGTCCCTAK ACACGAACGA CAGTTGCGCG CTGTTGCACG GTTCTCGACC TT-CACGG-C
17       TTGTCCCTAT ACACGAACGG CAGTCGCGCG C-GTTGCAGG TT-CTKGACC TT-CACGG-C
25       TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTTGACC TT-CACGG-C
26       TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
27C.1    TTGTCCCTAT AC-------- ---------- ---------- ---------- ----------
27C.10   TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
28A      TTGTCCCTAT ACACGAACGA CAGTSGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
3        TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CATGG-C
41B      TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
45C      TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
5Q       TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
101A     TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CTT-TYGACC TT-CACGAGC
102B     TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
104      TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC WT-CACGG-C
108C     TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
111A.4   TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
111A.5   TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
120C     TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
123G     ---------- ---------- ---------- ---------- ---------- ----------
131      TTGTCCCTAT ACACGAACGA CAGTGGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
30A      TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
57B      TTGTCCCTAT ACACGAATGA TAGTTGTGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
58B      TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
62D.2    TTGTCCCTAC ACAAGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGGGC
62D.3    TTGTCCCTAT ACACGAATGA CAGTCGTGCG C-GCTGTAGG CT-CTCGACC TT-TACGG-C
65D.7    TTGTCCCTAC ACAAGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGGGC
65D.3    TTGTCCCTAC ACAAGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
67B      TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
69A      TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
69B      TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
76A      TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGGGC
76E      TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
78A.10   TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
78A.5    TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGGGC
8        TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
80A      TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CTTCTCGACC TT-CACGG-C
82A      TTGCCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGGGC
85A      TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CTCMTTGACC TTGCRCGG-C
86B      TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG TTACTCGACC TTGCACGG-C
90D      TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG TT-CTCGACC TT-CACGG-C
93C.2    TTGTCCCTAT ACACGAATGA CAGTTGTGCA C-GTTGCAGG CT-CTCGACC TT-CACGG-C
93C.3    TTGTCCCTAT ACACGAATGA CAGTTGTGCA C-GTTGCAGG CT-CTCGACC TT-CACGG-C
98A      TTGTCCCTAT ACACGAATGA CAGTTGTGCA C-GTTGCAGG CT-CTCGACC TT-CTCGG-C
126E     TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGGCC TT-CACGG-C
103A     TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-A------- ---------- ----------
50D      TTGTCCCTAT ACA--AAC-A TATT------ ---------- ---------- ----------
68A      ---------- ---------- ---------- ---------- ---------- ----------
75A      TTGTCCCTAT MCACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGACC TT-CACGG-C
```

```
Bases 295-354
ETS type 2
121C     TTGTCCCTAT ACACGAACAA CAGTCGCGCG C-GTTGCAGG CT-CTCGGCC TT-CACGG-C
18       TTGTCCCTAA ACAAGAACGA CAGTCGCGCG C-GTTGCAGT CT-ATCGACC TT-CAAGG-C
19       TTGTCCCTAT ACACGAATGA CAGTCGCGTG C-GTTGCAGG CT-CTCGGCC TT-CACGG-C
20       TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGGCC TT-CACGG-C
21       TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGGCC TT-CACGG-C
122A     TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGGCC TT-CACGG-C
126.29   TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGGCC TT-CACGG-C
26       TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGGCC TT-CACGGGC
52B      TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGGCC TT-CACGG-C
57B      TTGTCCCTAT ACACGAACGA CAGTCACGCG C-GTTGCAGG CT-CTCGGCC TT-CACGG-C
65C.3    TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGGCC TT-CACGG-C
65C.4    TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGGCC TT-CACGG-C
93C      TTGTCCCTAT ACACGGACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGGCC TT-CACGG-C
118A     TTGTCCCTAT ACAAGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGGCC TT-CACGG-C
126.BC1  TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGGCC TT-CACGG-C
126E.1   TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGGCC TT-CACGG-C
126E.3   TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGGCC TT-CACGG-C
126E.4   TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGGCC TT-CACGG-C
126A     TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGGCC TT-CACGG-C
61       TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CTTCTGGGCC TT-CACGGAC
ETS type 3
121B     ---------- ---------- ---------- ---------- ---------- ----------
66C      ---------- ---------- ---------- ---------- ---------- ----------
66D      ---------- ---------- ---------- ---------- ---------- ----------
67A      ---------- ---------- ---------- ---------- ---------- ----------
69B      ---------- ---------- ---------- ---------- ---------- ----------
71A      ---------- ---------- ---------- ---------- ---------- ----------
77B      ---------- ---------- ---------- ---------- ---------- ----------
82A      ---------- ---------- ---------- ---------- ---------- ----------
83B      ---------- ---------- ---------- ---------- ---------- ----------
86B.24   ---------- ---------- ---------- ---------- ---------- ----------
86B.29   ---------- ---------- ---------- ---------- ---------- ----------
86B.5    ---------- ---------- ---------- ---------- ---------- ----------
87B      ---------- ---------- ---------- ---------- ---------- ----------
90D      ---------- ---------- ---------- ---------- ---------- ----------
ETS type 4
118A     TTGTCCCTAT ACACGAACAA CAGTCGCGCG C-GTTGCAGG CT-CTCGGCC TT-CACGG-C
121B     TTGTCCCTAT ACAAGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGGCC TT-CACGG-C
123F     TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGGCC TT-CACGG-C
126.29.1TTGTCCCTAT ACACGAACTA CAGTCGCGCG T-GTTGCAGG CT-CTCGGCC TT-CACGG-C
126.29.3TTGTCCCTAT ACACGAACTA CAGTCGCGCG T-GTTGCAGG CT-CTCGGCC TT-CACGG-C
63B      TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGGCC TT-CACGG-C
81A      TTGTCCCTAT ACACGAACGA CAATCACCCG C-GTTGCAGG CT-MTYGACC TT-CACAG-C
76A      TTGTCCCTAT ACACGAACGG CAGTCACCCG C-GTTGCAGG CT-CTCGACC TT-CACAG-C
76E      TTGTCCCTAT ACACGAACGA CAGTCACCCG C-GTTGCAGG CT-CTCGACC TT-CACAG-C
78B      TTGTCCCTAT ACACGAACGA CAATCACCCG C-GTTGCAGG CT-CTCGACC TT-CACAG-C
84A.1    TTGTCCCTAT ACAAGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGGCC TT-CACGG-C
84A.2    TTGTCCCTAT ACAAGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGGCC TT-CACGG-C
83B      TTGTCCCTAA ACAAGAACGA CAGTCGCGCG C-GTTGCTGG CT-ATCGACC TT-CAAGG-C
87A      ---------- ---------- ---------- ---------- ---------- ----------
126.WB3  TTGTCCCTAT ACACGAACGA CAGTCGCGCG C-GTTGCAGG CT-CTCGGCC TT-CACGG-C
Pur-34   TT-YCGCTAG ACACGAACGA CAGTCGCGTA C-GTTGCGGG CTTTT-GACC TT-CGCGG-C
Pur-132  TT-TCGCTAG ACACGAATGA CTGTCGCGTA C-GTTGCGGG CTTGT-GACC TT-CGCGG-C
```

Bases 355-382
ETS type 1

1	CC-YTAGTC-	TTGCACCGCG	TCGA-GT-
10	CC-TTAGTC-	TTGCACCTCG	TCGA-GTC
14	CC-TTAGTC-	TTGCACCKCG	TCGACGTC
15	CC-TTAGTC-	TTGCACCTCG	TCGACGTC
12	CC-TTAGTC-	TTGCACCTCG	TAGA-GTC
17	CC-TTAGTC-	WTGCAC----	--------
25	CC-TTAGTC-	TTGCACCGCG	TTGACGTG
26	CC-TTAGTC-	TTGCACCTCG	TCGACGTC
27C.1	----------	----------	--------
27C.10	CC-TTAGTC-	TTGCA-----	--------
28A	CC-TTAGTC-	TTGCACCTCG	TCGACGTC
3	CC-TTAGTC-	TTGCACCGCG	TCGACGTC
41B	CC-TTAGTC-	TTGCACCGTG	TCGACYTC
45C	CC-TTAGTC-	TTGCACCGCG	TCGA-GTC
5	CC-TTAGTC-	TTGCACCTCG	TCGACGTC
101A	CC-TTAGTTC	TTG-------	--------
102B	CC-TTAGTCG	TTGCAC----	--------
104	CC-TTAGTTC	TT--------	--------
108C	CC-TTAGTC-	TTGCACC---	--------
111A.4	CC-TTAGTC-	TTGCA-----	--------
111A.5	CC-TTAGTC-	TTGCACCGCG	T-------
120C	CC-TTAGT--	----------	--------
123G	----------	----------	--------
131	CC-TTAGT--	----------	--------
30A	CC-TTAGTC-	TTGCACCTCG	TCGACGTC
57B	CCCTTAGTC-	TTGCACCGCG	TCAACGTC
58B	CC-TTAGT--	----------	--------
62D.2	CC-TTA-TTC	TTGCACCGCA	TCGATGTA
62D.3	CC-TTAGT-C	TTGCACCGCG	TCAACGT-
65D.7	CC-TTA-TTC	TTGCACCGCA	TCGATGTC
65D.3	CC-TTA-TTC	TTGCACCGCA	TCGATGTC
67B	CC-TTAGTC-	TTGCACCGCG	TCGACGTC
69A	C---------	----------	--------
69B	C---------	----------	--------
76A	CC-TTAGTC-	TT--------	--------
76E	CC-TTAGTC-	TTGCACCGCG	TCGACGTC
78A.10	CC-TTAGTC-	TTGCACCGCG	TCGACGTC
78A.5	CC-TTAGTC-	TTGCA-----	--------
8	CC-TTAGTC-	TTGCACC-TG	TCGACGTC
80A	CC-TTAGTTC	TTGCAAC---	--------
82A	CCCTTAGTC-	TTGCACCGCG	TT------
85A	CC-TTAGTTC	TTGC------	--------
86B	CC-TTAGTC-	TTGCA-C---	--------
90D	CC-TTAGTC-	TTGCACCG--	--------
93C.2	CC-TTAGTC-	TTACACCGCG	TCAACGTC
93C.3	CC-TTAGTCC	TTACACCGCG	TCAACGTC
98A	CC-TTAGTC-	TTACACCGCG	TCCA----
126E	CC-TTAGTCC	----------	--------
103A	----------	----------	--------
50D	----------	----------	--------
68A	----------	----------	--------
75A	CC-TTAGTCT	-TGCAAC---	--------

Bases 355-382

ETS type 2

```
121C    CC-TTAGTCC -TGCACCACG TCAACGTC
18      CC-TTAGCCC -TGCACCGCG TCGACGTC
19      CC-TTAGTCC -TGCACCGCG TCAACATC
20      CC-TTAGTCC -AGCACCGCG TCAACGTC
21      CC-TTAGTCC -AGCACCGCG TCAACGTC
122A    CC-TTAGTCC -TGCACCGCG T--ACAT-
126.29  CC-TTAGTCA -TGCACCGCG TCAACGTC
26      CCCTTAGTCC -AGCACCGCG TCAA----
52B     CC-TTAGTCC -TGCACCGCG TCAACGTC
57B     CC-TTAGTCC -TGCACCGCG TCAACGTC
65C.3   CCCTTAGTCC CTGCACCGCG TCGACGTC
65C.4   CC-TTAGTCC -TGCACCGCG TCGACGTC
93C     CC-TTAGTCC -TGCACCGCG TCAACGTC
118A    CC-TTAGTCC -TGCACCGCG T-------
126.BC1 CC-TTAGTCC -TGCACCGCG TCAACGTC
126E.1  CC-TTAGTCC -TGCACCGCG TCAACGTC
126E.3  CC-TTAGTCC -TGCACCGCG TCAACGTC
126E.4  CC-TTAGTCC -TGCACCGCG TCAACGTC
126A    CC-TTAGTCC -TGCACCGCG TCA---TA
61      CC-TTAGTCC CTGC------ --------
```

ETS type 3

```
121B    ---------- ---------- --------
66C     ---------- ---------- --------
66D     ---------- ---------- --------
67A     ---------- ---------- --------
69B     ---------- ---------- --------
71A     ---------- ---------- --------
77B     ---------- ---------- --------
82A     ---------- ---------- --------
83B     ---------- ---------- --------
86B.24  ---------- ---------- --------
86B.29  ---------- ---------- --------
86B.5   ---------- ---------- --------
87B     ---------- ---------- --------
90D     ---------- ---------- --------
```

ETS type 4

```
118A    CC-TTAGTCC -TGCACCGCG TCGACGTC
121B    CC-TTAGTCC -TGCACCGCG TCGACGTC
123F    CC-TTAGTC- ---------- --------
126.29.1CC-TTAGTCC -TGCACCGCG TCGACGTC
126.29.3CC-TTAGTCC -TGCACCGCG TCGACGTC
63B     CC-TTAGTCC -TGCACCGCG TCGACGTC
81A     CC-TTAGTCC -TGCACCGCA TCGACGTC
76A     CC-TTAGTCC -TGCACCGCA TCGACGTC
76E     CC-TTAGTCC -TGCATTGCA TCGACGTC
78B     CC-TTAGTCC -TGCACCGCA TCGACGTC
84A.1   CC-TTAGTCC -TGCACCGCG TCGACGTC
84A.2   CC-TTAGCCC -TGCACCGCG TCGACGTC
83B     CC-TTAGCCC -TGCACCGCG TCGACGTC
87A     ---------- ---------- --------
126.WB3 CC-TTAGTCC -TGCACCGCG TCGACGTC
Pur-34  CA-MTAGTCT -CGC------ --------
Pur-132 CA-TTAGTCC -CGT------ --------
```

Bases 383-442

α and β region→

ETS type 1

```
1       TT-TGTTGCC TTC-GAAATG GTGGATGAGA GACGAATTGG ATTGCCTATG CTTGCTTGTG
10      TT--GTTGCC TTT-GAAATG GTGGATGAGA GACGAATTGG ATTGCCTATG CTTGCTTGTG
14      TT--GTTGCC TTC-GAAATG GTGGATGAGA GACGAATTGG ATTGCCTATG CTTGCTTGTG
15      TT--GTTGCC TTC-GAAATG GTGGATGAGA GACGAATTGG ATTGCCTATG CTTGCTTGTG
12      TT--GAWGCC TTC-GAAATG GTGGATGAGA GACGAATTGG ATTGCCTATG CCTGMTTCTG
17      TT--GTTGCC TTC-GAAATG GTGGATGAGA GACGAATTGG ATTGCCTATG CTTGCTTGTG
25      TT--GTTGCC TTC-GAAATG GTGGATGAGA GACGAATTGG ATTGCCTATG CTTGCTTGTG
26      TT--GTTGCC TTC-AAAATG GTGGATAAGC GGCGAATTGG ATTGCCTATG CTTGCTTGTG
27C.1   ---------- -------ATG GTGGATGAGA GACGAATTGG ATTGCCTATG CTTGCTTGTG
27C.10  TT--GTTGCC TTC-GAAATG GTGGATGAGA GACGAATTGG ATTGCCTATG CTTGCTTGTG
28A     TT--GTTGCC TTT-GAAATG GTGGATGAGA GACGAATTGG ATTGCCTATG CTTGCTTGTG
3       TT--GTTGCC TTC-GAAATG GTGGATGAGA GACGAATTGG ATTGCCTATG CTTGCTTGTG
41B     TT--GTTGCC TTC-GAAATG GTGGATGAGA GACGAATTGG ATTGCCTATG CTTGCTTGTG
45C     TT--GTTGCC TTC-AAAATG GTGGATAAGC GGCGAATTGG ATTGCCTATG CTTGCTTGTG
5       TT--GTTGCC TTC-GAAATG GTGGATGAGA GACGAATTGG ATTGCCTATG CTTGCTTGTG
101A    TT--GTTGCC TTC-AAAATG GTGGATAAGC GGCGAATTGG ATTGCCTATG CTTGCTTGTG
102B    TT--GTTGCC TTC-GAAATG GTGGATGAGA GACGAATTGG ATTGCCTATG CTTGCTTGTG
104     TT--GTTGCC TTC-AAAATG GTGGATAAGC GGCGAATTGG ATTGCCTATG CTTGCTTGTG
108C    TT--GTTGCC TTC-GAAATG GTGGATGAGA GGCGAATTGG ATTGCCTATG CTTGCTTGTG
111A.4  TT--GTTGCC TTC-AAAATG GTGGATAAGC GGCGAATTGG ATTGCCTATG CTTGCTTGTG
111A.5  TT--GTTGCC TTC-GAAATG GTGGATTAGA GACGAATTGG ATTGCGTATG GTTGGTTGTG
120C    TT--GTTGCC TTC-GAAATG GTGGATGAGA GACGAATTGG ATTGCCTATG CTTGCTTGTG
123G    TT--GTTGCC TTC-GAAATG GTGGATGAGA GACGAATTGG ATTGCCTATG CTTGCTTGTG
131     TT--GTTGCC TTT-GAAATG GTGGATGAGA GACGAATTGG ATTGCCTATG CTTGCTTGTG
30A     TT--GTTGCC TTC-GAAATG GTGGATGAGA GACGAATTGG ATTGCCTATG CTTGCTTGTG
57B     TT-TGTTGCC TTC-GAAATG GTGGATGAGA GGCAAATTGG ATTGCCTATG CTTGCCTATG
58B     TT--GTTGCC TTC-GAAATG GTGGATGAGA GATGAATTGG ATTGCCTATG CTTGCTTGTG
62D.2   TT--CTTGCC TTC-AAAATG GTGGATGAGC GGCGAATTGG ATTGCCTATG CTTGCTTATG
62D.3   TT--GTTGCC TTC-GAAATG GTGGATGAGA GGCAAATTGG ATTGCCTATG CTTGCTTGTG
65D.7   TT--CTTGCC TTC-AAAATG GTGGATGAGC GGCGAATTGG ATTGCCTATG CTTGCTTGTG
65D.3   TT--CTTGCC TTC-AAAATG GTGGATGAGC GGCGAATTGG ATTGCCTATG CTTGCTTGTG
67B     TTGTCTCGCC TTCCGAAATG GTGCATGAGA GACGAATTGG ATTGCCTATG CTTGGTTGTG
69A     TT--GTTGCC TTC-GAAATG GTGGATGAGA GACGAATTGG ATTGCCTATG CTTGCTTGTG
69B     TT--GTTGCC TTC-GAAATG GTGGATGAGA GACGAATTGG ATTGCCTATG CTTGCTTGTG
76A     TT--GTTGCC TTC-GAAATG GTGGATGAGA GACGAATTGG ATTGCCTATG CTTGCTTGTA
76E     TT--GTTGCC TTC-GAAATG GTGGATGAGA GACGAATTGG ATTGCCTATG CTTGCTTGTG
78A.10  TT--GTTGCC TTC-GAAATG GTGGATGAGA GACGAATTGT ATTGCCTATG CTTGCTTGTA
78A.5   TT--GTTGCC TTC-GAAATG GTGGATGAGA GACGAATTGT ATTGCCTATG CTTGGTTGTA
8       TT--GTTGCC TTC-GAAATG GTGGATGAGA GACGAATTGG ATTGCCTATG CTTGCTTGTG
80A     TT--GTTGCC TTC-GAAATG GTGGATGAGA GACGAATTGG ATTGCCTATG CTTGCTTGTA
82A     TT--GTTGCC TTC-AAAATG GTGGATAAGC GGCGAATTGG ATTGCCTATG CTTGCTTGTG
85A     TT--GTTGCC TTC-GAAATG GTGGATGAGA GATGAATTGG ATTGCCTATG CTTGCTTGTG
86B     TT--GTTGCC TTC-GAAATG GTGGATGAGA GACCAATTGG ATTGCCTATG TTTGCTTGTG
90D     TT--GTTGCC TTC-GAAATG GTGGATGAGA GACCAATTGG ATTGCCTATG CTTGCTTGTG
93C.2   TT--GTTGCC TTC-GAAATG GTGGATGAGA GGCAAATTGG ATTGCCTATG CTTGCTTATG
93C.3   TT--GTTGCC TTC-GAAATG GTGGATGAGA GGCAAATTGG ATTGCCTATG CTTGCTTATG
98A     TT--GTTGCC TTC-GAAATG GTGGATGAGA GGCAAATTGG ATTGCCTATG CTTGCTTATG
126E    TT--CTTGCC TTC-AAAATG GTGGATGAGC GGCAAATTGG ATTGCCTATG CTTGCTTGTG
103A    TT--GTTGCC TTC-AAAATG GTGGATAAGC GGCGAATTGG ATTGCCTATG CTTGCTTGTG
50D     TT--GTTGCC TTC-AAAATG GTGGATAAGA GGCGAATTGG ATTGCCTATG CATGCTTGTG
68A     TT--GTTGCC TTC-GAAATG GTGGATGAGA GACGAATTGG ATTGCCTATG CTTGCTTGGG
75A     TT--GTTGCC TTC-GAAATG GTGGATGAGA GACGAATTGG ATTGCCTATG CTTGCTTGTG
```

```
Bases 383-442
ETS type 2
121C    ---------- ---------- ---------- ---------- -----CTATG CCTGCTTGTG
18      ---------- ---------- ---------- ---------- -----CTATG CTTGCTTGTG
19      ---------- ---------- ---------- ---------- -----CTATG CCTGCTTGTG
20      ---------- ---------- ---------- ---------- -----CTATG CCTGCTTGTG
21      ---------- ---------- ---------- ---------- -----CTATG CCTGCTTGTG
122A    ---------- ---------- ---------- ---------- -----CTATG CCTGCTTGTG
126.29  ---------- ---------- ---------- ---------- -----CTATG CCTGCTTGTA
26      ---------- ---------- ---------- ---------- -----CTATG CCTGCTTGTG
52B     ---------- ---------- ---------- ---------- -----CTATG CCTGCTTGTG
57B     ---------- ---------- ---------- ---------- -----CTATG CCTGCTTGTG
65C.3   ---------- ---------- ---------- ---------- -----CTATG CCTGCTTGTG
65C.4   ---------- ---------- ---------- ---------- -----CTATG CCTGCTTGTG
93C     ---------- ---------- ---------- ---------- -----CTATG CCTGCTTGTG
118A    ---------- ---------- ---------- ---------- -----CTATG CCTGCTTGTG
126.BC1 ---------- ---------- ---------- ---------- -----CTATG CCTGCTTGTG
126E.1  ---------- ---------- ---------- ---------- -----CTATG CCTGCTTGTG
126E.3  ---------- ---------- ---------- ---------- -----CTATG CCTGCTTGTG
126E.4  ---------- ---------- ---------- ---------- -----CTATG CCTGCTTGTG
126A    ---------- ---------- ---------- ---------- -----CTATG CCTGCTTGTG
61      ---------- ---------- ---------- ---------- -----CTATG CCTGCTTGTG
ETS type 3
121B    ---------- ---------- ---------- ---------- ---------- ----------
66C     ---------- ---------- ---------- ---------- ---------- ----------
66D     ---------- ---------- ---------- ---------- ---------- ----------
67A     ---------- ---------- ---------- ---------- ---------- ----------
69B     ---------- ---------- ---------- ---------- ---------- ----------
71A     ---------- ---------- ---------- ---------- ---------- ----------
77B     ---------- ---------- ---------- ---------- ---------- ----------
82A     ---------- ---------- ---------- ---------- ---------- ----------
83B     ---------- ---------- ---------- ---------- ---------- ----------
86B.24  ---------- ---------- ---------- ---------- ---------- ----------
86B.29  ---------- ---------- ---------- ---------- ---------- ----------
86B.5   ---------- ---------- ---------- ---------- ---------- ----------
87B     ---------- ---------- ---------- ---------- ---------- ----------
90D     ---------- ---------- ---------- ---------- ---------- ----------
ETS type 4
118A    ---------- ---------- ---------- ---------- ---------- ----------
121B    ---------- ---------- ---------- ---------- ---------- ----------
123F    ---------- ---------- ---------- ---------- ---------- ----------
126.29.1---------- ---------- ---------- ---------- ---------- ----------
126.29.3---------- ---------- ---------- ---------- ---------- ----------
63B     ---------- ---------- ---------- ---------- ---------- ----------
81A     ---------- ---------- ---------- ---------- ---------- ----------
76A     ---------- ---------- ---------- ---------- ---------- ----------
76E     ---------- ---------- ---------- ---------- ---------- ----------
78B     ---------- ---------- ---------- ---------- ---------- ----------
84A.1   ---------- ---------- ---------- ---------- ---------- ----------
84A.2   ---------- ---------- ---------- ---------- ---------- ----------
83B     ---------- ---------- ---------- ---------- ---------- ----------
87A     ---------- ---------- ---------- ---------- ---------- ----------
126.WB3 ---------- ---------- ---------- ---------- ---------- ----------
Pur-34  TTG-GTTGCC TTC-AGAATG GTGGATGTGC GGCGAATCGG ATTGCTTATG CTTGCTTGGG
Pur-132 ---------- ---------- ---------- ---------- ---------- ----------
```

Bases 443-502
ETS type 1

```
1        CGGAGCCCTT ACGGGCAGTG CACAA-GCTT GTCGACGATC GCACCCCGAG GAGCTGCGAT
10       CGGAGCCCTT ACGGGCAGTG CACAA-GCTT GTTGACGATG GCACCCCGAG GAGCTGCGAT
14       CGGAGCCCTT ACGGGCAGTG CACAA-GCTT GTTGACGATG GCACCCCGAG GAGCTGCGAT
15       CGGAGCCCTT ACGGGCAGTG CACAA-GCTT GTTGACGATG GCACCCCGAG GAGCTGCGAT
12       CGGAGCCCTT ACGGGCAGTG CACAA-GSTT GTCGACGATG GCACCCCGAG GAGTTGCGAT
17       CGGAGCCCTT ACGGGCAGTG CACAA-GCTT GTCGACGATG GCACCCCGAG GAGCTGCGAT
25       CGGAGCCCTT ACGAGCAATG CACAA-GCTT GTCGACGATG GCACCCCGAG GAGCTGCGAT
26       CGGAGCCCTT ACGGGCAATG CACAA-GCTT GTCGACAATG GCACCCCGAG GAGCTGCGAT
27C.1    CGAAGCCCTT GCGGGCAGTG CACAA-GCTT GTCGACGATG GCACCCCGAG GAGCTGCGAT
27C.10   CGGAGCCCTT ACGGGCAGTG CACAA-GCTT GTCGACGATG GCACCCCGAG GAGCTGCGAT
28A      CGGAGCCCTT ACGGGCAGTG CACAA-GCTT GTTGACGATG GCACCCCGAG GAGCTGCGAT
3        CGGAGCCCTT ACGGGCAGTG CACAA-GCTT GTCGACGATT GCACCCCGAG GAGCTGCGAT
41B      CGGAGCCCTT ACGGGCAATG CACAA-GCTT GTCGACGATG GCACCCCGAG GAGCTGCGAT
45C      CGGAGCCCTT ACGGGCAATG CACAA-GCTT GTCGACAATG GCACCCCGAG GAGCTGCGAT
5        CGGAGCCCTT ACGGGCAGTG CACAA-GCTT GTTGACGATG GCACCCCGAG GAGCTGCGAT
101A     CGGAGCCCTT ACGGGCAATG CACAA-GCTT GTCGACAATG GCACCCCGAG GAGCTGCGAT
102B     CGGAGCCCTT ACGGGCAGTG CACAA-GCTT GTTGACGATG GCACCCCGAG GAGCTGCGAT
104      CGGAGCCCTT ACGGGCAATG CACAA-GCTT GTCGACAATG GCACCCTGAG GAGCTGCGAT
108C     CGGAGCCCTT ACGGGCAGTG CACAA-GCTT GTCGACGATG GCACCCCGAG GAGCTGCGAT
111A.4   CGAAGCCCTT ACGGGCAATG CACAA-GCTT GTCGACAATG GCACCCCGAG GAGCTGCGAT
111A.5   CGGAGCCCTT ACGGGCAATG CACAA-GCTT GTCGACGATG GCACCCCGAG GAGCTGCGAC
120C     CGGAGCCCTT ACGGGCAGTG CACAA-GCTT GTTGCCCATG GCACCCCGAG GAGCTGCGAT
123G     CGGAGCCCTT ACGGGCAGTG CACAA-GCTT GTTGACGATG GCACCCCGAG GAGCTGCGAT
131      CGGAGCCCTT ACGGGCAGTG CACAA-GCTT GTTGACGATG GCACCCCGAG GAGCTGCGAT
30A      CGGAGCCCTT ATGGGCAGTG CACAA-GCTT GTTGACGATG GCACCCCGAG GAGCTGCGAT
57B      CGGAGCCCTT ACGGGCAATG CACAA-GCTT GTCGACGATG GCACCCCGAG GAGCTGTGAT
58B      CGGAGCCCTT ACGGGCAGTG CACAA-GCTT GTCGACGATG GCACCCCGAG GAGCTGCGAT
62D.2    CGGAGCCCTT ACGGGCAATG CACAA-GCTT GTCGACGATG GCACCCCGAG GAGCTGCGAT
62D.3    CGGAGCCCTT ACGGGCAATG CACAA-GCTT GTCGATGATG GCACCCCGAG GAGCTACGAT
65D.7    CGGAGCCCTT ACGGGCAATG CACAA-GCTT GTCGACGATG GCACCCCGAG GAGCTGCGAT
65D.3    CGGAGCCCTT ACGGGCAATG CACAA-GCTT GTCGACGATG GCACCCCGAG GAGCTGCGAT
67B      CGGAGCCATT ACGGGCAGTG CACAA-GCTT GTCGACGATG GCACCCCGAG GAGCTGCGAT
69A      CGGAGCCCTT ACGGGCAGTG TACAA-GCTT GTTGACGATG GCGCCCCGAG GAGCTGCGAT
69B      CGGAGCCCTT ACGGGCAGTG TACAA-GCTT GTTGACGATG GCGCCCCGAG GAGCTGCGAT
76A      CGGAGCCCTT ACGGGCAGTG CACAA-GCTT GTCGACGATC GCACCCCGAG GAGCTGCGAT
76E      CGGAGCCCTT ACGGGCAGTG CACAA-GCTT GTCGACGATC GCACCCCGAG GAGCTGCGAT
78A.10   CGGAGCCCTT ACGGGCAGTG CACAA-GCTT GTCGACGATC GCACCCCGAG GAGCTGCGAT
78A.5    CGGAGCCCTT ACTGGCAGTG CACAA-GCTT GTCGACGATC GCACCCCGAG GAGCTGCGAT
8        CGGAGCCCTT ACGGGCAGTG CACAA-GCTT GTTGACGATG GCACCCCGAG GAGCTGCGAT
80A      CGGAGCCCTT ACGGGCAGTG CACAA-GCTT GTCGACGATC GCACCCCGAG GAGCTGCGAT
82A      CGGAGCCCTT ACGGGCAACG CGCAA-GCTT GTCGACAATG GCACCCCGAG GAGCTGCGAT
85A      CGGAGCCCTT ACGGGCAGTG CACAA-GCTT GTCGACGATG GCACCCCGAG GAGCTGCGAT
86B      CGGAGCCCTT ACGGGCAGTG CACAA-GCTT GTCGACGATG GCACCCCGAA GAGCTGCGAT
90D      CGGAGCCCTT ACGGGCAGTG CACAA-GCTT GTCGACGATG GCACCCCGAA GAGCTGCGAT
93C.2    CAGAGCCCTT ATGGGTAATG CACAA-GCTT GTCGACGATG GCACCCCGAG GAGCTACGAT
93C.3    CAGAGCCCTT ACGGGTAATG CACAA-TGTC GACGGTGGCA CCCCGAGGAG CTACGATGAT
98A      CAGAGCCCTT ACGGGTAATG CACAA-GCTT GTCGACGATG GCACCCCGAG GAGCTACGAT
126E     CGGAGCCCTT ACGGGCAATG CACAA-GCTT GTCGACTATG GCACCCCGAG GAGCTGCGAT
103A     CGGAGCCCTT ACGGGCAATG CACAA-GCTT GTCGACAATG GCACCCTGAG GAGCTGCGAT
50D      CGGAGCCCTT ACGGGCAATG CACAA-GCTT GTCGACAATG GCACCCCAAG GAGCTGCGAT
68A      CGGAGCCCTT ACGGGCAGTG CACAA-GCTT GTCGACGATC GCACCCCGAG GAGCTGCGAT
75A      CGGAGCCCTT ACGGGCAGTG CACAA-GCTT GTCGACGATG GCACCCCGAG GAGCTGCGAT
```

Bases 443-502

ETS type 2

```
121C    CAAAGCCCAC CCATGCAATG CACAAAGCTT TTCGACGATG GCACTCCGAG GAGCTGCAGT
18      CAAAGCCCTC ACGGGCAGAG CACAA-GCTT GTCGAGGATG GCACTCCGAG GAGCTGCGAT
19      CAAAGCCCAC CCATGCAATG CACAAAGCTT TTCGACGATG GCACTCCGAG GAGCTGCAGT
20      CAAAGCCCAC CTATGCAATG CACAAAGCTT TTCGACGATG GCACTCCGAG GAGCTGCAGT
21      CAAAGCCCAC CTATGCAATG CACAAAGCTT TTCGACGATG GCACTCCGAG GAGCTGCAGT
122A    CAAAGCCCAC CCATGCAATG CACAAAGCTT TTCGACGATG GCACTCCGAG GAGCTGCAGT
126.29  CAAAGCCCAC CCATGCAATG CACAAAGCTT TTCGACGATG GCACTCCGAG GAGCTGCAGT
26      CAAAGCCCAC CTATGCAATG CACAAAGCTT TTCGACGATG GCACTCCGAG GAGCTGCAGT
52B     CAAAGCCCAC CCATGCAATG CACAAAGCTT TTCGACGATG GCACTCCGAG GAGCTGCAGT
57B     CAAAGCCCAC CCATGCAATG CACAAAGCTT TTCGACGATG GCACTCCGAG GAGCTGCAGT
65C.3   CAAAGCCCAC CCATGCAATG CACAAAGCTT GTCGACGATG GCACCCCGAG GAGCTTCAGT
65C.4   CAAAGCCCAC CCATGCAATG CACAAAGCTT GTCGACGATG GCACCCCGAG GAGCTTCAGT
93C     CAAAGCCCAC CCATGCAATG CACAAAGCTT TTCGACGATG GCACTCCGAG GAGCTGCAGT
118A    CAAAGCCCAC CCATGCAATG CACAAAGCTT TTCGACGATG GCACTCCGAG GAGCTGCAGT
126.BC1 CAAAGCCCAC C-ATGCAATG CACAAAGCTT TTCGACGATG GCACTCCGAG GAGCTGCAGT
126E.1  CAAAGCCCAC CCATGCAATG CACAAAGCTT TTTGACGATG GCACTCCGAG GAGCTGCAGT
126E.3  CAAAGCCCAC CCATGCAATG CACAAAGCTT TTTGACGATG GCACTCCGAG GAGCTGCAGT
126E.4  CAAAGCCCAC CCATGCAATG CACAAAGCTT TTTGACGATG GCACTCCGAG GAGCTGCAGT
126A    CAAAGCCCAC CCATGCAATG CACAAAGCTT TTTGACGATG GCACTCCGAG GAGCTGCAGT
61      CAAAGCCCAC CCATGCAATG CACAAAGCTT TTCGACGATG GCACTCCGAG GAGCTGCAGT
```

ETS type 3

```
121B    ---------- ---------- ---------- ---------- ---------- ----------
66C     ---------- ---------- ---------- ---------- ---------- ----------
66D     ---------- ---------- ---------- ---------- ---------- ----------
67A     ---------- ---------- ---------- ---------- ---------- ----------
69B     ---------- ---------- ---------- ---------- ---------- ----------
71A     ---------- ---------- ---------- ---------- ---------- ----------
77B     ---------- ---------- ---------- ---------- ---------- ----------
82A     ---------- ---------- ---------- ---------- ---------- ----------
83B     ---------- ---------- ---------- ---------- ---------- ----------
86B.24  ---------- ---------- ---------- ---------- ---------- ----------
86B.29  ---------- ---------- ---------- ---------- ---------- ----------
86B.5   ---------- ---------- ---------- ---------- ---------- ----------
87B     ---------- ---------- ---------- ---------- ---------- ----------
90D     ---------- ---------- ---------- ---------- ---------- ----------
```

ETS type 4

```
118A    ---------- ---------- ---------- ---------- ---------- ----------
121B    ---------- ---------- ---------- ---------- ---------- ----------
123F    ---------- ---------- ---------- ---------- ---------- ----------
126.29.1---------- ---------- ---------- ---------- ---------- ----------
126.29.3---------- ---------- ---------- ---------- ---------- ----------
63B     ---------- ---------- ---------- ---------- ---------- ----------
81A     ---------- ---------- ---------- ---------- ---------- ----------
76A     ---------- ---------- ---------- ---------- ---------- ----------
76E     ---------- ---------- ---------- ---------- ---------- ----------
78B     ---------- ---------- ---------- ---------- ---------- ----------
84A.1   ---------- ---------- ---------- ---------- ---------- ----------
84A.2   ---------- ---------- ---------- ---------- ---------- ----------
83B     ---------- ---------- ---------- ---------- ---------- ----------
87A     ---------- ---------- ---------- ---------- ---------- ----------
126.WB3 ---------- ---------- ---------- ---------- ---------- ----------
Pur-34  CGGAGCCTTT GTGGGCAATG CACAA-CATG TCGACGATGG CACCTCGAGC AGCCGCGATT
Pur-132 ---------- ---------- ---------- ---------- ---------- ----------
```

Bases 503-511
ETS type 1

1	GATTCCCAA
10	GATTCCCAA
14	GATTCCCAA
15	GATTCCCAA
12	GATTCCCAA
17	GATTCCCAA
25	GATTCCCAA
26	GATTCCTAA
27C.1	GATTCCCAA
27C.10	GATTCCCAA
28A	GATTCCCAA
3	GATTCCCAA
41B	GATTCCCAA
45C	GATTCCTAA
5	GATTCCCAA
101A	GATTCCTAA
102B	GATTCCCAA
104	GATTCCTAA
108C	GATTCCCAA
111A.4	GATTCCTAA
111A.5	GATTCCCAA
120C	GATTCCCAA
123G	GATTCCCAA
131	GATTCCCAA
30A	GATTCCCAA
57B	GATTCCCAA
58B	GATTCCCAA
62D.2	GATTCCCAA
62D.3	GATTCTCAA
65D.7	GATTCCCAA
65D.3	GATTCCCAA
67B	GATTCCCAA
69A	GATTCCCAA
69B	GATTCCCAA
76A	GATTCCCAA
76E	GATTCCCAA
78A.10	GATTCCCAA
78A.5	GATTCCCAA
8	GATTCCCAA
80A	GATTCCCAA
82A	GATTCCTAA
85A	GATTCCCAA
86B	GATTCCCAA
90D	GATTCCCAA
93C.2	GATTCCCAA
93C.3	GATTCCCAA
98A	GATTCCCAA
126E	GATTCCCAA
103A	GATTCCTAA
50D	GATTCCTAA
68A	GATTCCCAA
75A	GATTCCCAA

Bases 503-511

ETS type 2

121C	GATTCCCAA
18	GATTCCCAA
19	GATTCCCAA
20	GATTCCCAA
21	GATTCCCAA
122A	GATTCCCAA
126.29	GATTCCCAA
26	GATTCCCAA
52B	GATTCCCAA
57B	GATTCCCAA
65C.3	GATTCCCAA
65C.4	GATTCCCAA
93C	GATTCCCAA
118A	GATTCCCCA
126.BC1	GATTCCCAA
126E.1	GATTCCCAA
126E.3	GATTCCCAA
126E.4	GATTCCCAA
126A	GATTCCCAA
61	GATTCCCAA

ETS type 3

121B	---------
66C	---------
66D	---------
67A	---------
69B	---------
71A	---------
77B	---------
82A	---------
83B	---------
86B.24	---------
86B.29	---------
86B.5	---------
87B	---------
90D	---------

ETS type 4

118A	---------
121B	---------
123F	---------
126.29.1	---------
126.29.3	---------
63B	---------
81A	---------
76A	---------
76E	---------
78B	---------
84A.1	---------
84A.2	---------
83B	---------
87A	---------
126.WB3	---------
Pur-34	GACATCCCA
Pur-132	---------

Appendix B. AFLP presence/absence data matrix

ID codes correspond to those found in Table 3.1. Presence of the fragment is coded as ‘1’ and absence as a ‘0’. The AFLP selective primer pair that produced the fragment and the size of the individual fragments can be found in Appendix C.

```
Fragments 1-50
1       00000 00000 00000 00000 00001 10011 01000 00000 00000 00000
3       00000 00000 00000 00000 00001 10011 01000 00000 00000 00000
5       00000 00000 00010 00000 00001 10001 01000 00000 00000 00000
8       00000 00000 00000 00000 00001 10001 01000 00000 00000 00000
10      00000 00000 00010 00000 00001 10001 01000 00000 00000 00000
12      00000 00000 00000 00000 00001 10001 01000 00000 00000 00000
13      10000 00000 00010 00000 00001 10111 01000 10000 00000 00110
14      00000 00000 00000 00000 00001 10001 01000 00000 00000 00000
16      10000 00000 00000 00000 00001 10011 01011 00000 00000 00000
17      00000 00001 00000 01001 00001 10001 00000 00001 10000 00000
18      00000 00000 00000 00000 00001 10000 01011 00001 10001 00001
19      00000 00000 00000 00001 00001 10000 01000 00000 00000 00000
20      00000 00100 00000 00000 00001 10000 01000 00000 00000 00000
25      00000 00000 00000 00000 00001 00001 01000 00000 00000 00000
26      00010 00000 00000 00000 00001 10000 00000 00000 01000 00000
27A     10000 00000 00000 01001 00001 10011 01110 00000 00000 00000
27B     10000 00000 00000 01101 00001 10001 01110 00000 00000 00000
27C     00000 00000 00000 00000 00001 10001 01000 00000 00000 00000
28A     00000 00000 00000 00000 10001 10001 01000 00000 00000 00000
29A     00000 00000 00000 00000 00001 10001 01100 00000 00000 00000
30A     00000 00000 00000 00000 00001 10001 01000 00000 00000 00000
30D     10000 00000 00010 01001 00001 10001 01000 00000 00000 00100
30E     00000 00000 00000 00000 00001 10011 01011 00000 00000 00000
30F     10000 00100 00000 00001 00001 10001 01000 00000 00010 00000
31      10000 10101 00010 01001 00001 00101 01001 00000 00000 00100
32B     00000 00000 00000 00000 00001 10011 01000 00000 00000 00000
36B     00010 00000 00000 00000 00001 01001 00000 00000 00000 00001
37A     00000 00000 00000 00000 00001 10010 01011 00000 00000 00000
37B     10000 00000 00000 00000 00001 10011 01010 00000 00000 00000
37C     10001 01000 10100 01101 00001 00000 01000 00000 00000 00000
39      00000 00000 00000 00101 00001 00101 01011 00000 00000 00000
41A     10000 01000 00010 01001 00001 10000 01000 00000 00000 00000
41B     00000 00000 00000 00001 00001 10000 01000 00000 00000 00000
41C     00000 00000 00000 00001 00001 10000 01001 00000 00000 10000
43      00000 10000 00000 00000 00001 10000 01010 00000 00000 00000
45A     00000 10011 00100 01000 00011 00000 01000 00000 00000 10100
45B     10000 01001 01100 01101 00001 00010 01001 00000 00000 00000
45C     00000 00000 00000 00100 00011 10000 01000 00000 00000 00000
46A     00000 00000 00000 00010 00001 10101 01010 00000 00000 00000
46B     00000 00000 00000 00000 00001 10101 01000 00000 00000 00000
46C     00000 00000 00000 00000 00001 10101 01000 00000 00000 00000
48A     10000 00000 00000 00000 00001 10000 01011 00000 00000 00000
48B     10001 00001 00010 10010 00001 10011 01010 00000 00000 00000
48C     00000 00000 00000 00000 00001 10111 01010 00000 00000 00000
48D     00000 10000 00000 10000 00001 10000 01000 00000 00000 00000
50A     10000 00000 00000 00000 00001 10001 01010 11000 00000 00000
50B     00100 00000 00000 00000 00001 00000 00000 00000 00000 00000
```

```
Fragments 1-50
50C    00000 01000 00000 00000 00001 10001 01010 00000 10000 00000
50D    00000 00000 00000 00000 00001 00001 01000 00000 00000 00000
52B    00010 00000 00000 00000 00001 10000 01000 00000 00000 00000
54A    00010 00000 00000 00000 00001 01000 00000 00000 00000 00000
55A    00010 00000 00000 00000 00001 01000 00000 00000 00000 00000
57B    00010 00000 00000 10000 00001 10000 01000 00000 00000 00000
58B    00000 00000 00000 00000 00001 10101 01000 00000 00000 00000
58D    00010 00000 00000 00000 00001 10001 01010 00000 00000 00000
61     00010 00100 10000 00001 00001 10010 01000 00000 00000 00000
62D    00000 00000 00000 00010 00001 10000 01000 00000 00000 00000
63B    00000 00000 00000 00010 00001 10000 01000 00000 00000 00000
65C    00000 00000 01000 00000 00001 10000 01000 00001 00000 00000
65D    00000 00000 00000 00000 00001 10000 01000 00000 00000 00000
66C    00010 00000 00000 00000 00001 10000 01000 00000 01000 00000
66D    00010 00000 10000 00000 00001 10000 01000 00000 01000 00000
67A    00010 00000 00000 00000 00001 10000 00000 00000 01101 00000
67B    00000 00000 10000 00000 00001 10000 00000 00000 01001 01000
68A    00000 00000 00000 00000 00001 10001 01000 00000 00000 00000
68B    00000 00000 00000 00000 00001 10011 01000 00000 00000 00000
68C    00000 00000 00000 00000 00001 10011 01010 00000 00000 00000
69A    00010 00000 00000 00000 00001 10000 00000 00000 01000 00000
69B    00010 00000 00000 00000 00001 10000 00000 01000 01000 00000
70A    00000 00000 00000 00000 00001 10011 01000 00000 00000 00000
70B    00000 00000 00000 00000 00001 10011 01000 00000 00000 00000
71A    00010 00000 00000 00000 00001 10000 00000 00000 01000 00000
72A    10000 00000 00000 00000 00001 10001 01000 00000 00000 00000
72B    10000 00000 00000 00000 00001 10001 01000 00000 00000 00000
74D    00000 00000 00001 00000 01001 01000 01000 00000 00000 00000
75A    00000 00000 00010 00000 00001 10001 00000 00001 10000 00100
76A    00000 00000 00000 00000 00001 10101 01000 00000 01001 00000
76E    00000 00000 10000 00000 00001 10001 01000 00000 01000 00000
77B    00010 00000 00000 00000 00001 10001 00000 00000 01000 01000
78A    00010 00000 00100 00000 00001 10000 00000 00000 00000 01000
79     00010 00000 00000 00000 00001 10000 00000 00000 01000 00000
80A    00000 00000 00000 00000 00001 10001 01000 00000 00000 00000
80C    10000 00000 00001 00001 00001 10011 01000 10000 00000 00000
80D    10000 00100 00000 01001 00001 10001 01000 00000 00000 00000
81A    01010 00000 00000 00000 00001 10000 00000 00000 00000 01000
82A    00010 00000 00000 10000 00001 10000 00000 00000 01000 00000
83A    00010 00000 00000 00000 00001 10000 00000 00000 01000 00000
84G    00010 00000 00000 10000 00001 10000 00000 00000 01000 00000
85A    00000 00000 00000 00000 00001 10001 01000 00000 00000 00000
85B    00000 00000 00000 00000 00001 10001 01000 00000 00000 00000
85D    00000 00000 00000 00000 00001 10011 01000 00000 00000 00000
85E    00001 10000 00000 00000 00001 10011 01010 00000 00000 00000
86A    10000 00000 00000 00000 00001 10011 01010 00110 00000 00000
86B    00000 00000 00000 01000 00001 10001 01000 00000 00000 00000
```

```
Fragments 1-50
87A    00000 00000 00000 00000 00001 10001 01000 00000 00000 00000
89A    10000 00000 00000 00000 00001 10001 01000 00000 00000 00000
89B    10000 00000 00000 00000 00001 10011 01100 00000 00000 10000
90A    00000 00000 00000 00000 00001 10011 01000 00000 00000 00000
90B    10000 10000 00000 00000 00001 10001 01010 00000 00000 00010
90C    00000 00000 00000 01000 00001 10001 01000 00000 00000 00000
92A    10000 00101 00100 01000 00001 10001 01000 00011 00000 00000
92B    10000 00000 00100 01001 00001 10001 01000 00010 00000 00000
93C    00000 00000 00000 00010 00001 10000 01000 00000 00000 00000
95A    10000 00000 00000 00000 00001 10011 01000 00000 00000 00000
96A    10000 00001 00100 01001 00000 00010 01000 00100 00001 10000
96B    10001 00000 00100 01111 00001 10011 01000 00000 00000 00000
98A    00000 00000 00000 00000 00001 10001 01000 00000 00000 00000
98B    00000 00000 00000 00000 00001 10101 01000 00000 00000 00000
98C    10001 00011 11100 00001 00001 00000 01000 00001 10000 10100
99A    00000 00000 00000 00000 00001 10000 01011 00000 00000 00000
99B    10000 10000 00000 00000 00001 10010 01001 00000 00000 00000
99C    00000 10000 00000 00000 00001 00100 01000 00010 00000 00000
100B   10100 00000 00000 01001 00001 10000 01000 00000 00000 00000
101A   00000 01000 00000 00000 00001 10001 01010 00000 00000 00000
102A   00000 00000 00000 00000 00001 10100 01010 00001 00000 00000
102B   00100 01000 00000 00000 00001 00010 01000 00000 00000 00000
102C   00000 00000 00000 00000 00001 10000 01011 00000 00000 00000
103A   00000 01000 00000 00000 00001 00000 01000 00000 00000 00000
103B   00000 00000 00000 00000 00001 10101 01010 00000 00000 00000
103C   00000 00000 00000 00000 00001 10001 01000 01000 00000 00000
104    00000 00001 00000 00000 00001 00010 01000 00000 00000 00000
105    00000 10000 00000 00000 00001 10111 01011 00000 00000 00000
106A   10000 00100 00000 00001 00011 00100 01010 00001 00000 00100
106B   00000 00000 00000 00000 00001 00111 01010 00000 00000 00000
106C   00000 01000 00000 00000 00001 10000 01010 00001 00000 00000
108A   00000 00000 00000 00000 00001 10010 01000 00000 00000 00000
108B   00000 00000 00000 00000 00001 10000 01010 10000 00000 00000
108C   00000 00000 00000 00000 00001 00000 01000 00000 00000 00000
110    01000 10000 00000 00000 00001 10110 01011 00000 00000 00000
111A   00000 00000 00000 00000 00001 10100 01010 00000 00000 00000
111B   00000 10000 00000 00000 00001 10111 01011 00000 00000 00000
111C   00000 10000 00000 00000 00001 10110 01001 00000 00000 00000
112    00000 00000 00000 00010 00001 10000 01000 00000 00000 00100
113A   00000 00000 00000 01001 00001 10000 01000 00001 00000 00100
114    00000 00000 00000 00000 00001 10000 01010 00000 00000 00000
116    10000 00000 00010 00011 00001 10000 01000 00000 00000 00000
118A   00000 00000 00000 00000 00001 10000 01000 00000 00000 00000
120C   00000 00000 00000 00000 00111 10000 00000 00101 10000 00100
121A   00000 00000 00000 00010 00001 10000 01000 00000 00000 00000
122A   00000 00000 00000 00000 00001 10000 01000 00000 00000 00000
123F   00000 00000 00000 00000 00001 10000 01000 00000 00000 00000
```

```
Fragments 1-50
123G   00000 00000 00000 00000 00001 10000 01000 00000 00000 00000
125A   00000 00000 00010 01000 00001 00000 01000 00001 00000 00100
126A   00000 00000 00000 00000 00001 10010 01000 00000 00000 00000
126E   00000 00000 00000 00010 00001 10000 01000 00000 00000 00000
127A   00000 00000 00000 01000 01001 10000 00010 00001 10000 00001
128A   00000 00001 00010 01000 00001 10000 01000 00001 00000 00100
130A   00000 00000 00000 10001 00001 10100 01000 00001 00000 00000
131    00000 00000 00000 00000 00001 10001 01000 00000 00000 00000
134A   00000 00000 00000 10000 00001 10100 01010 00001 10001 00000
HEN1   10000 00101 11000 01101 00011 00100 00001 00000 00000 00000
S3     00000 00000 00100 00000 00001 10001 01000 00000 00000 00000
S7     00000 00100 00000 00000 00001 10001 01000 00000 00000 00000
```

```
Fragment 51-100
1       00000 00000 00000 00000 01000 00010 00000 00100 00000 00000
3       00000 00000 00000 00000 00000 00010 00000 00100 00000 00000
5       00010 00000 01000 00000 00000 00010 00000 00100 00000 00000
8       00000 00000 01000 00000 00000 00010 00000 00100 00000 00000
10      00000 00000 01000 00000 00000 00010 00000 00100 00000 00000
12      00010 00000 00000 00000 00000 00010 00000 00100 00000 00000
13      00100 00000 00000 00000 00000 00010 00000 00100 00000 00000
14      00010 00000 00000 00000 00000 00010 00000 00110 00000 00000
16      00000 00000 10000 00000 00000 00110 00000 01110 00000 00100
17      00000 00001 00000 00000 00000 00000 00000 00000 00010 00000
18      00000 00001 00000 00000 00001 01100 00000 00000 01100 00010
19      00000 00000 10000 00000 00000 00000 00010 00000 00000 00100
20      00000 00000 01000 00000 00000 00000 00000 00000 00000 00000
25      00000 00000 01000 00000 00000 00100 00000 00000 00000 00010
26      00000 00000 01000 00000 00000 01000 00000 00000 00000 00010
27A     00010 00000 00000 00000 10000 00010 00000 01110 00000 00100
27B     00010 00000 00000 00000 00000 00010 00000 00100 00000 00000
27C     00000 00000 00000 00000 00000 00010 00000 00100 00000 00000
28A     00000 00000 10000 00000 00000 00010 00000 00100 00000 00000
29A     00010 00000 00000 00000 00000 00010 00000 00100 00000 00100
30A     00010 00000 00000 00000 00000 00010 00000 00100 00000 00000
30D     00000 00000 00000 00000 00000 00010 00100 00100 00000 00000
30E     00000 00000 00000 00000 00000 00010 00000 00100 00000 00100
30F     00000 00000 00000 00000 00000 00010 00000 00100 00000 00100
31      10000 00000 00000 00000 00000 00010 00000 00100 00000 00000
32B     00000 00000 00000 00000 00000 00010 00000 00000 00000 00000
36B     00000 00000 00000 00000 00000 00000 00001 00000 00000 00000
37A     00000 00000 10000 00000 00000 00100 00000 00000 00001 10000
37B     00000 00000 10000 00000 00000 00100 00000 00000 00000 00000
37C     00000 00100 00000 00000 00000 00010 00000 01000 00000 00000
39      00000 00100 10000 00000 00000 00100 00000 00000 00000 00000
41A     00000 10000 00000 10000 00000 00000 00000 00000 00000 00100
41B     00000 00000 01000 00000 00000 00100 00000 00000 00000 00000
41C     00000 00000 10000 00000 00000 00100 00000 00000 00000 00000
43      00000 00000 01000 00000 00000 00110 00000 00000 00000 00000
45A     00000 00000 01000 00100 00000 00100 00000 00000 00000 00000
45B     00000 00010 00100 00001 10000 00100 00000 01000 00000 00000
45C     00000 00000 01000 00000 00100 00100 00000 00000 10000 00010
46A     00000 01000 10000 00000 00000 00100 00000 00000 00000 00000
46B     00000 01000 10000 00000 00000 00100 00000 00000 00000 00000
46C     00000 00000 10000 00000 00000 00100 00000 00000 00000 00000
48A     00000 00000 10000 00000 00000 00110 00000 00000 00000 00000
48B     00000 00000 10000 00000 00000 00100 00000 00000 00000 00010
48C     00000 00000 01000 00000 00000 00100 00000 00000 00000 00000
48D     00000 00000 01000 00000 00000 00110 00000 00000 00000 00010
50A     00000 00000 10000 00010 00000 00100 00000 00000 00000 00000
50B     00000 00000 10011 10000 00000 00110 00000 00000 00000 00000
```

```
Fragments 51-100
50C    00000 00000 10000 00000 00000 00100 00000 00000 00000 00000
50D    00000 00000 01000 10000 00000 00100 00000 01000 00000 00000
52B    00000 00000 10000 00000 10001 00000 00000 00000 00000 00000
54A    00100 00000 00000 00010 00000 00000 00001 00000 00000 00000
55A    00100 00000 00000 00000 00000 00000 00001 00000 00000 00000
57B    00000 00000 10000 00000 00000 00000 00001 00000 00000 00100
58B    00000 00000 00000 00000 00000 00010 00000 00100 00000 00000
58D    00000 00000 00000 00000 00000 00010 00000 00100 00000 00000
61     00000 00000 10000 00000 00000 00010 00000 00000 00000 00100
62D    00000 00000 10000 00000 10000 00000 00000 00000 00000 00000
63B    00000 00000 10000 00001 00000 00000 00000 00000 00000 00000
65C    00000 00000 10000 00001 00000 00000 00000 00000 00000 00100
65D    00000 00000 10000 00001 00000 00000 00000 00000 00000 00000
66C    00000 00000 10000 00001 00000 01000 00000 00000 00000 00010
66D    00000 00000 10000 00001 00000 01000 00000 00000 00000 00010
67A    00000 00000 00000 00000 00000 01000 00000 10001 00000 00010
67B    00000 00000 01000 00000 10000 01000 00000 00000 00100 00010
68A    00000 00000 00000 00000 11000 10010 00000 00100 00000 00000
68B    00000 00000 00000 00000 01000 00010 00000 00100 00000 00000
68C    00000 00000 00000 00000 01000 00010 00000 00100 00000 00000
69A    00000 00000 01000 00000 00000 01000 00000 00000 00000 00010
69B    00000 00000 01000 00000 00000 01000 00000 00000 00000 00010
70A    00000 00000 00000 00000 01100 00010 00000 00100 00000 00000
70B    00000 00000 00000 00000 01000 00010 00000 00100 00000 00000
71A    00000 00100 01000 00000 00000 01000 00000 00000 00000 00010
72A    00000 00000 00000 00000 01000 00010 01000 00100 00000 00000
72B    00000 00000 00000 00000 01000 00010 01000 00100 00000 00000
74D    00100 00000 00000 00000 00000 00001 00000 00000 00000 00000
75A    00000 00000 00000 00000 00000 00000 01000 00000 00000 00000
76A    00000 00000 00000 00000 00000 10000 00000 01000 00000 00010
76E    00000 00000 01000 00001 00000 10000 00000 01000 00000 00000
77B    00000 00000 01000 00001 00000 01000 00000 00000 00000 00000
78A    00000 00000 01000 00000 00000 10000 00000 01000 00000 00010
79     00000 00000 01000 00000 00000 01000 00000 00000 00000 00010
80A    00000 00000 00000 00000 01000 00010 00000 00100 00000 00000
80C    00000 00000 00000 00000 01100 00010 00000 00100 00000 00000
80D    00000 00000 00000 00000 01000 00010 00000 00100 00000 00000
81A    00000 00000 00000 00000 00000 10000 00000 01000 00000 00010
82A    00000 00000 10000 00001 00000 01000 00000 00000 00000 00010
83A    00000 00000 10000 00000 00000 01000 00000 00000 00000 00010
84G    00000 00000 10000 00000 00100 01000 00000 00000 00000 00010
85A    00000 00000 00000 00000 00000 00010 00000 00100 00000 00000
85B    00000 00000 00000 00000 00000 00010 00000 00100 00000 00000
85D    00000 00000 00000 00000 00000 00010 00000 00100 00000 00000
85E    00000 00000 00000 00000 00000 00010 00000 00100 00000 00000
86A    00000 00000 00000 00000 00000 00010 00000 00100 00000 00000
86B    00010 00000 00000 00000 00000 00010 00000 00100 00000 00000
```

```
Fragments 51-100
87A     00000 00000 01000 00000 00000 00010 00000 00100 00000 00100
89A     00010 00000 00000 00000 00000 00010 00000 00100 00000 00000
89B     00010 00000 00000 00000 00000 00010 00000 00100 00000 00000
90A     00000 00000 00000 00000 00000 00010 00000 00100 00000 00100
90B     00000 00000 00000 00000 00000 00000 00000 00100 00000 00000
90C     00010 00000 00000 00000 00000 00010 00000 00100 00000 00100
92A     00000 00000 00000 00000 00000 00011 00000 00000 00000 00010
92B     00000 00000 00000 00000 00001 00010 00000 00100 00000 00000
93C     00000 00000 01000 00000 10000 00010 00000 00000 00000 00000
95A     00000 00000 00000 00000 00000 00010 00000 00100 00000 00000
96A     00000 00000 00000 01000 00000 00000 10000 00000 00000 00000
96B     00000 01001 00000 00000 00000 00110 00010 00000 00000 00010
98A     00000 00001 01000 00000 00000 00100 00000 00000 00000 00010
98B     00000 00001 10000 00000 00000 00100 00000 00000 00000 00010
98C     00000 00000 10000 00000 00000 00000 01000 00000 00000 00001
99A     00000 00000 10000 00000 00000 00100 00000 00000 00000 00010
99B     00000 00000 01000 00000 00000 00100 00001 00000 00000 00010
99C     00000 00000 01000 00000 00000 00100 00000 00000 00000 00010
100B    00000 00000 00000 00010 00000 00000 00001 00000 00000 00000
101A    00000 00010 10000 00000 00000 00110 00000 00000 00000 00010
102A    00000 00000 10000 00000 00001 00100 00000 01000 00000 00000
102B    00000 00000 10000 00000 00000 00100 00000 00000 00000 00010
102C    00000 00000 10000 00000 00000 00100 00000 00000 00000 00010
103A    00000 00000 10000 00000 00000 00110 00000 00000 00000 00010
103B    00000 00000 10001 00000 00000 00100 00000 00000 00000 00010
103C    00000 01000 10000 00000 00000 00110 00000 00000 00000 00010
104     00000 00000 10000 00000 00000 00100 00000 00000 00000 00010
105     00000 00000 10000 00000 00000 00100 00001 00000 00000 00010
106A    00000 00000 01000 00000 00000 00100 00000 00000 00000 00010
106B    00000 00001 01000 00000 00000 00100 00000 00000 00000 00000
106C    00000 00000 10000 00000 00000 00100 00001 00000 00000 00000
108A    00001 00000 10000 00001 00000 00100 00000 00000 00000 00000
108B    00000 00010 10000 00000 00000 00100 00001 00000 00000 00010
108C    00000 10010 01000 00000 00000 00100 00001 00000 00000 00010
110     00000 00000 10010 00000 00000 00100 00000 00000 00000 00010
111A    00000 00000 01000 00000 00000 00100 00000 00000 00000 00010
111B    00000 00000 10000 00000 00000 00100 00000 00000 00000 00010
111C    00000 00000 00000 00000 00000 00100 00000 00000 00000 00010
112     10000 00000 00000 00000 10000 00000 00000 00010 10000 00000
113A    00000 10001 00000 10000 00000 00000 00010 00000 00100 01000
114     00000 00000 01000 00000 00000 00100 00000 00000 00000 00000
116     00000 00000 00000 00001 00000 00000 00000 00000 00000 00100
118A    00000 00000 10000 00001 00000 00000 00000 00100 00000 00100
120C    01001 00001 00000 00001 00010 00010 00100 00100 00010 00000
121A    01000 00000 10000 00000 00010 00000 00000 00000 00000 00000
122A    00000 00000 01000 00000 10000 00000 00000 00000 00000 00100
123F    00000 00000 01000 00000 00010 00000 00000 00000 00000 00000
```

```
Fragments 51-100
123G   00000 00000 01000 00000 00000 00000 00000 00000 00000 00000
125A   00000 01000 00000 10000 00000 00000 00010 00000 00100 00000
126A   00000 00000 00000 10000 00000 00000 00000 00100 00000 00010
126E   00000 00000 01000 00000 10000 00010 00000 00000 00000 00000
127A   00001 00010 00000 00000 00000 00000 01000 00000 10010 00000
128A   01000 00010 00000 00001 00000 00000 00100 00000 10001 00000
130A   01000 00000 01000 00000 10000 00100 00000 00000 00000 00000
131    00000 00000 01000 00000 00000 00010 00000 00100 00000 00000
134A   10000 00001 01000 10000 10000 00010 00100 00000 00100 01000
HEN1   10000 00000 00010 00000 00000 00000 00000 00000 00000 00000
S3     00010 00000 00000 00000 00000 00010 00000 00100 00000 00000
S7     00000 00000 00000 00000 00000 00010 00000 00100 00000 00000
```

```
Fragments 101-150
1       00000 00000 00000 00001 00000 00000 00000 00000 01010 00010
3       00000 00000 00000 00000 00000 00000 00000 00001 01000 10010
5       00000 10000 00000 00010 00000 00000 00000 00000 00100 10010
8       00000 10000 00000 00000 00000 00000 00000 00000 00100 10010
10      00000 10000 00000 00000 00000 00000 00000 00000 00100 10010
12      00000 11000 00000 00000 00000 00000 00000 00000 00100 10010
13      00000 00000 00000 00001 00000 00000 00010 00010 00010 01011
14      00000 10000 00000 00000 00000 00000 00000 00000 00100 10010
16      00000 01000 00000 00001 00000 00000 00000 00000 00010 01010
17      00000 00000 00000 00000 00000 00000 00000 00000 00000 00000
18      00000 00010 00000 00000 00000 00010 00000 00010 10110 00000
19      00001 00001 00000 00001 00100 00000 00000 00001 00110 00100
20      00001 00000 00000 00000 00100 00000 00000 00001 00010 10100
25      00010 00001 00000 00000 00001 01000 00000 00100 00011 11011
26      00000 00000 00000 00000 00000 00000 00000 00001 00010 10010
27A     00000 10000 00000 00001 00000 00000 00001 00000 01010 01010
27B     00000 00000 00000 00001 00000 00000 00001 00000 00010 01011
27C     00000 10000 00000 00000 00000 00000 00000 00000 00010 01011
28A     00000 10000 00000 00000 00000 00000 00000 00000 00010 11011
29A     00000 00000 00000 00000 00000 00100 00010 00000 00011 00010
30A     00000 00000 00000 00000 00000 00000 00000 00000 00010 01010
30D     00000 00000 00000 00000 00000 00000 01000 00000 00011 10011
30E     00000 00000 00000 00001 00000 00000 00000 01000 00010 01010
30F     00000 00000 00000 00000 00000 00000 01001 00000 00010 00011
31      00000 00000 00000 00001 00000 00000 01001 00000 00010 11011
32B     00000 01000 00000 00001 00000 00000 01010 00010 00010 01010
36B     00000 00000 00010 00000 00000 00000 00000 00000 00100 00100
37A     00000 00001 10000 10000 00000 01000 00000 00000 00011 00010
37B     00000 00001 00000 10100 01000 01000 00010 00000 00011 00010
37C     00000 00001 00000 00000 00000 00000 01000 00000 00011 00000
39      00000 00001 00000 10000 00001 10000 00001 00000 00011 01010
41A     11100 00000 00000 00000 00000 01000 01000 00000 00110 11010
41B     00000 00000 00000 00000 00000 01000 00001 00000 00111 10010
41C     00000 00001 00000 00000 00000 01000 00000 00100 00001 11010
43      00000 00001 00000 00100 00000 01001 00000 00100 00010 00010
45A     00000 00000 00000 00100 00000 00000 01000 10000 00011 01010
45B     10000 00001 00000 10100 01000 01000 01000 00000 00001 01010
45C     10000 00001 00000 11000 00001 01000 00000 00000 00100 10010
46A     00000 00001 10000 00100 00000 01000 00000 00000 00010 01010
46B     00000 00001 00000 10100 00001 01000 01000 00000 00011 00010
46C     00000 00001 00000 10100 00000 01000 01000 00000 00011 01010
48A     00000 00001 00000 10100 01001 10000 00000 00000 00011 11010
48B     10000 00001 00000 10100 01001 00000 00000 00000 00001 01011
48C     00000 00001 00000 00001 00001 00000 00000 00000 00001 01010
48D     10000 00001 00000 10100 00001 00000 00000 00000 00010 01010
50A     00000 00001 00000 10000 00000 01000 00000 00000 00011 01010
50B     00000 00001 00000 00000 00001 01000 00000 00100 00011 01010
```

```
Fragments 101-150
50C    00000 00000 00000 00000 00001 01000 00000 00100 00011 00010
50D    00000 00001 00000 00000 00001 01000 00000 00000 00010 00000
52B    00001 00000 00000 00010 00100 00000 00000 10001 00011 00010
54A    00000 00000 00000 00000 00000 00000 00000 00000 00010 00100
55A    00000 00000 00000 00000 00000 00000 00000 00000 00000 00111
57B    00001 00000 00000 00001 00100 00000 00000 00000 00010 10010
58B    00000 01000 00000 00000 00000 00000 00000 00000 00100 00100
58D    00000 01000 00000 00000 00000 00000 00000 00000 00010 10011
61     00001 00000 00000 00000 00100 00000 00000 00000 00011 10010
62D    00000 00000 00000 00000 00000 00000 00000 00000 00010 01010
63B    00000 00000 00000 00000 00010 00000 00000 00000 00010 00010
65C    00000 00000 00000 00000 00010 00000 01000 00000 00110 10010
65D    00000 00000 00000 00000 00100 00000 00000 00000 00110 10010
66C    00000 00000 00000 00000 00100 00000 00000 10000 10010 10010
66D    00000 00000 00000 00000 00000 00000 00000 00000 00010 10010
67A    00000 00000 00000 00000 00000 00000 00000 00000 00000 11010
67B    00000 00000 00000 00001 00000 00000 00000 00000 10110 11010
68A    00000 00000 00000 00001 00000 00000 00000 00000 00011 11010
68B    00000 00000 00000 00001 00000 00100 00011 00000 00010 01011
68C    00000 00000 00000 00000 00000 00000 00000 00000 00010 00010
69A    00000 00000 00000 00000 00000 00000 00000 00001 10000 10010
69B    00000 00000 00000 00001 00000 00000 00000 00000 10010 10010
70A    00000 00000 00000 00000 00000 00000 00000 00000 00010 01011
70B    00000 00000 00000 00000 00000 00000 00000 00000 00010 00011
71A    00000 00000 00000 00000 00000 00000 00000 00000 00010 10010
72A    00000 00000 00000 00000 00000 00000 00000 00000 00011 01010
72B    00000 00000 00000 00001 00000 00000 00000 00000 00011 01010
74D    00000 00000 00100 00000 00000 00100 00000 00000 00000 00110
75A    00000 00000 00000 00000 00000 00000 01000 00000 00000 00011
76A    00000 00000 00000 00001 00000 00000 00000 00000 00000 00000
76E    00000 00000 00000 00000 00000 00000 00000 00000 00111 11011
77B    00000 00000 00000 00001 00000 00000 00000 00001 00011 01011
78A    00000 00000 00000 00000 00000 00000 00000 00000 00100 11010
79     00000 00000 00000 00000 00000 00000 00000 00000 00010 10010
80A    00000 00000 00000 00001 00000 00000 00000 00000 00010 01010
80C    00000 00000 00000 00001 00000 00000 00010 00000 00010 10011
80D    00000 00000 00000 00001 00000 00000 01000 00000 00010 01011
81A    00000 00000 00000 00000 00000 00000 00000 00000 00100 11010
82A    00000 00000 00000 00001 00000 00000 00000 00000 00010 10010
83A    00000 00000 00000 00000 00000 00000 00000 00000 00010 10010
84G    00000 00000 00000 00000 00000 00000 00000 10000 10010 10010
85A    00000 10000 00000 00000 00000 00000 00000 00000 00010 01001
85B    00000 01000 00000 00000 00000 00000 00000 00000 00010 00010
85D    00000 00100 00000 00000 00000 00000 00010 00000 00010 00010
85E    00000 10000 00000 00000 00000 00000 00010 00000 00010 00010
86A    00000 10000 00000 00000 00000 00000 01010 00000 00010 01010
86B    00000 00000 00000 00000 00000 00000 00000 00000 00010 01011
```

```
Fragments 101-150
87A    00000 10000 00000 00000 00000 00000 00000 00000 00000 10011
89A    00000 10000 00000 00000 00000 00000 01010 00000 00010 01010
89B    00000 10000 00000 00000 00000 00000 01000 00000 00010 01011
90A    00000 10000 00000 00000 00000 00000 01010 00000 00010 00010
90B    00000 00000 00000 00000 00000 00000 00000 00000 00010 01010
90C    00000 00000 00000 00000 00000 00000 00000 00000 00010 01010
92A    00000 00000 00000 00000 00000 00000 01010 00000 00010 00011
92B    00000 10000 00000 00000 00100 00000 01010 00000 00010 10011
93C    00001 00000 00000 00001 00010 00000 00000 10001 00010 11010
95A    00000 00000 00000 00000 00000 00000 01011 10000 00010 00011
96A    00001 00000 00000 00000 00000 00000 01010 00000 00010 01110
96B    00000 00001 00000 00001 00000 01000 01010 00000 10010 00011
98A    00000 00001 00000 10000 00000 00000 01010 00000 00010 10000
98B    00000 00001 00000 10000 00000 01000 01000 00000 00010 01010
98C    00000 00001 00001 00000 00000 01000 01010 00000 00010 11000
99A    00000 00001 00000 00000 00000 01000 00000 00000 00010 00010
99B    00000 00001 00000 00000 00000 01000 00001 00000 00001 00110
99C    00000 00001 00000 00000 00000 01000 00000 00000 01010 01010
100B   00000 00000 00010 00000 00000 00000 00000 00001 00000 10010
101A   00000 00001 00000 00000 00000 11000 00000 00000 00010 01010
102A   00000 00001 00010 00000 00000 01000 00000 00000 00010 01010
102B   00000 00001 00000 00000 00000 10000 00000 00000 00010 01010
102C   00000 00001 00000 00000 00000 01000 00010 00000 00011 01010
103A   00000 00001 00000 00000 00001 00000 00000 00000 00010 01010
103B   00000 00001 00000 00000 00000 11000 00000 00000 00010 11011
103C   00000 00001 00000 00000 00000 01000 01000 00000 00010 01010
104    00000 00001 00000 10001 00000 10000 00000 00000 00010 01010
105    00000 00001 10000 10000 00000 01000 00000 00000 00011 01010
106A   00010 00001 10000 00000 00000 01000 01001 00000 00000 01010
106B   00000 00001 00000 00100 00001 01000 00000 00000 00011 01010
106C   00000 00001 00000 00000 00001 01000 00000 00000 00001 01010
108A   00000 00001 00000 00000 00000 01000 00000 00000 00010 01010
108B   00000 00001 00000 00000 00000 01000 00000 00100 00010 01010
108C   00000 00101 00000 10000 00000 01000 00000 00000 00000 01010
110    00000 00001 00000 00000 00000 01000 00010 00100 00011 01010
111A   00000 00001 00000 00000 00000 01000 00100 00000 00010 00010
111B   00000 00001 00000 00000 00000 01000 01000 00000 00010 11010
111C   00000 00001 00000 00000 00000 01000 00000 00000 00011 01010
112    00000 00000 00000 00001 00100 00000 00000 00001 00110 00010
113A   00000 00000 00000 00000 00000 00000 10000 00001 00110 00010
114    00000 00001 00000 10000 00000 01000 00000 00000 00010 00010
116    00001 00000 00000 00001 00000 00000 00000 00000 00000 00010
118A   00000 00000 00000 00001 00000 00000 00000 00000 00010 11010
120C   00000 00000 00000 00000 00000 00000 00000 00000 00010 10010
121A   00001 00000 00000 00000 00100 00000 01000 00000 00110 10010
122A   00001 00000 00000 00000 00010 00000 00000 00001 00010 11010
123F   00001 00000 00000 00001 00100 00000 00000 00001 00010 00010
```

```
Fragments 101-150
123G   00001 00000 00000 00000 00100 00000 00000 00001 00011 10010
125A   00000 00000 00000 00000 00000 00000 10000 00001 00010 00010
126A   00001 00000 00000 00001 00100 00000 00000 00001 00110 00010
126E   00001 00000 00000 00000 00100 00000 00000 00000 00000 00000
127A   00000 00000 00000 00000 00000 00000 10000 00000 00000 00001
128A   00000 00000 00000 00100 00010 00000 00010 00000 00110 10000
130A   00000 00000 00000 00001 00010 00000 00000 00000 00000 00001
131    00000 10000 00000 00000 00000 00000 00000 00000 00011 01010
134A   00010 00000 00000 00000 00000 00000 10000 00000 00000 00000
HEN1   00000 00000 00000 00000 00000 00000 01010 00000 00111 01000
S3     00000 00000 00000 00000 00000 00000 00000 00000 00010 11010
S7     00000 10000 00000 00000 00000 00000 00000 00000 00010 01010
```

```
Fragments 151-200
1       00100 10000 00000 01010 00000 00011 00000 10100 00001 00100
3       00100 00000 00000 00010 00000 00011 00000 10100 00000 00000
5       00000 00000 00000 00000 10000 00011 00000 00000 00000 00100
8       00100 00000 00000 00000 10000 00001 00000 00101 00000 00000
10      00000 00000 00000 00000 10000 00001 00100 10000 00000 00000
12      00000 00000 00000 00000 10000 00001 00000 00000 00010 00100
13      00011 00000 10000 00010 10010 01001 11000 10100 00010 10000
14      00000 00000 00000 00000 10000 00001 00000 00000 00000 10000
16      11001 01000 00010 00110 10010 00011 01000 00101 00000 00010
17      00100 00000 00000 00000 00000 00000 00100 00000 00000 00000
18      01000 10000 00000 00000 00000 00011 10010 00010 00000 00000
19      00100 00010 00000 00000 10000 00110 01000 00100 00000 00000
20      00100 00000 00000 00010 10000 00011 01000 00000 00000 00000
25      00100 00000 00000 00100 10010 10011 01000 00000 00000 00110
26      00000 00000 00000 00010 10100 00011 00000 00100 00000 00010
27A     10111 00000 00000 00010 10010 00001 00000 10100 00000 00010
27B     00101 00000 00000 00010 10010 00001 00000 10100 00000 00010
27C     00101 00000 00000 00010 10000 00001 00000 10000 00000 00010
28A     00000 00000 00000 00000 10000 00001 00000 10000 00000 00000
29A     10001 00000 00000 00010 10010 10001 00100 00100 00010 00010
30A     00000 00110 00000 00010 10000 00001 00100 00001 00000 10010
30D     00100 01000 01000 00010 10010 00001 01000 01000 10010 10000
30E     00111 01010 00000 00010 10000 01001 00000 00101 00010 10000
30F     10101 00100 10000 00010 10000 00001 00110 00100 00010 10010
31      00000 01100 01000 01010 10000 01001 00100 00000 00000 00010
32B     00001 00000 00000 00010 10000 01001 01000 00100 00000 10000
36B     00000 11001 00000 00000 00000 00001 10010 00000 00000 00000
37A     00000 00000 00000 00100 10010 00000 00010 01000 00000 00110
37B     10000 00100 00000 00100 10010 00001 00100 01001 00000 00110
37C     00100 00100 00010 00100 11010 00011 00000 01001 00000 00010
39      00100 00000 00000 00100 10010 10001 00000 01001 00010 00110
41A     00000 01000 00100 00100 10010 00001 10000 00000 00000 00110
41B     00100 00000 00000 00100 10000 00011 00000 00000 00000 00100
41C     00100 00000 00000 00100 10010 00011 01000 00000 00000 00100
43      00100 00000 00000 00100 10010 00011 00000 01001 00000 00100
45A     00000 00000 00000 10100 11010 00101 10010 01001 00010 00100
45B     10000 00010 00000 00100 10010 00010 11010 01000 10000 00100
45C     00100 00000 00000 00000 10010 00011 00000 00001 00000 00100
46A     00100 00010 00000 00100 10010 00011 00000 00001 00000 00110
46B     00100 00000 00000 00100 10010 00011 00000 01001 00000 00110
46C     00000 00000 00000 00100 10010 10001 00000 01001 00000 00010
48A     00000 01000 00000 00100 10010 10011 00000 01000 00000 00110
48B     11101 01000 00000 00100 10010 10011 01010 00001 10100 10110
48C     00100 00000 00000 00100 10010 00011 00000 00000 00000 00110
48D     00000 00000 01000 00100 10010 01001 00010 00000 00010 00100
50A     00100 00010 00000 00100 10010 00011 00000 00000 00000 00100
50B     00100 00010 00000 00100 10010 00011 00000 00001 00000 00100
```

```
Fragments 151-200
50C     00100 00000 00000 00100 10010 00011 00000 00000 00000 00100
50D     00100 00000 00000 00000 10010 00011 00000 00000 00000 00000
52B     00100 00010 00000 00000 10000 00010 01000 00000 00000 00000
54A     00100 00001 00000 00000 10010 00011 10010 00000 00000 00000
55A     00100 10001 00001 00000 10010 00011 10010 00000 00000 00000
57B     00000 00010 00000 00010 10000 00010 11000 00000 00000 00000
58B     10100 00100 00000 00010 10000 00011 01000 00100 00010 00000
58D     00111 00110 00000 00010 10000 11001 11100 00101 00010 00000
61      00100 00000 00000 00000 10000 00011 01000 00000 00000 00000
62D     00000 00000 00000 00000 10000 00011 01000 00001 00000 00000
63B     00000 00000 00000 00000 10000 00011 01000 00000 00000 00000
65C     00100 00000 00000 00000 10000 00011 01000 00000 00000 00000
65D     00100 00000 00000 00000 10000 00011 01000 00000 00000 00000
66C     00000 10001 00000 00010 10100 00010 00010 00000 00000 00010
66D     00000 00000 00000 00010 10100 00011 00000 00000 00000 00010
67A     00000 00000 00000 00010 00100 00011 00001 00010 00000 01010
67B     10101 00000 00000 00011 00100 00011 00000 00010 00000 00010
68A     00100 00000 00000 00000 10000 00011 00000 00001 00001 00000
68B     00000 00100 00000 00010 10000 00001 00000 10000 00000 00010
68C     00111 00000 10000 00000 10000 01001 10000 10100 00000 10000
69A     00100 00000 00000 00010 10000 00011 00000 00000 00000 00000
69B     10001 00000 00000 00010 10100 00011 00000 00010 00000 01000
70A     00000 00000 10000 00010 10000 11001 00100 10000 00000 00010
70B     01111 01000 00000 00010 10000 11011 00000 00101 00010 00010
71A     00000 00000 00000 00010 10100 00011 00000 00000 00000 00000
72A     00001 00000 10000 00010 10000 00001 10000 10000 00000 00010
72B     00001 00000 10000 00010 10000 00001 10000 10000 00000 00010
74D     00000 00001 00000 00010 10010 00011 10010 00000 00000 00000
75A     00111 00000 10100 00010 10010 00011 10100 00000 00000 10000
76A     00100 00000 00000 00000 00000 00000 00000 00010 00000 00000
76E     00000 00000 00000 00001 00000 00011 00000 00010 00000 00000
77B     00000 00000 00000 00010 10100 00011 00000 00010 00000 00010
78A     00100 00000 00000 00001 00000 00011 00000 00010 00000 00000
79      00000 00000 00000 00010 10000 00011 00000 00000 00000 00000
80A     00000 00000 00000 00010 10000 00001 00000 00000 00000 00000
80C     00111 01100 00000 00010 10000 11001 11000 10100 00010 10110
80D     10111 01100 10000 00010 11001 11001 01000 10001 00000 00000
81A     00000 00000 00000 00001 00000 00001 00001 00010 00000 00000
82A     00000 00000 00000 00010 10000 00011 00000 00000 00000 00000
83A     00000 00000 00000 00010 10000 00011 00000 00000 00000 00000
84G     00000 10000 00000 00010 10100 00011 00000 00010 00000 00000
85A     00001 00000 00000 00000 10000 00001 00000 00001 00000 00010
85B     00001 00000 00000 00010 10010 01001 01000 00001 00000 00010
85D     10111 01000 00000 00010 10010 11001 01000 00101 00000 00010
85E     00001 01000 00000 00010 10010 01001 01000 00101 00000 00010
86A     00001 00100 00000 00010 10010 00001 00000 00100 00000 00010
86B     00000 00100 00000 00000 10000 00001 00000 10000 00000 00010
```

Fragments 151-200

87A	00100	00000	00000	00010	10000	00011	00000	00001	00000	00010
89A	00101	00000	00000	00010	10010	01001	00000	10100	00000	10000
89B	00000	00100	00000	00010	10000	00001	00000	10000	00000	00010
90A	00000	00000	00000	00010	10010	01001	11000	10100	00000	00010
90B	00000	00000	10000	00010	10000	00001	01000	00000	00000	00010
90C	00100	00000	00000	00000	10000	00001	00100	00001	00000	00000
92A	00000	00100	10000	00100	00001	00001	01010	10000	00000	10000
92B	11000	00100	00000	00010	10001	11011	10000	10000	00010	10000
93C	00100	10000	00000	00000	10000	00011	00000	00000	00000	00010
95A	00101	00100	00000	00010	10010	01001	10000	10000	00000	00010
96A	10100	00100	00000	00010	00000	00001	00000	00000	00000	00000
96B	01100	00000	00000	00010	10000	00001	01000	00001	00010	00010
98A	00000	00000	00000	00000	10100	00011	01000	00000	00100	00110
98B	00000	00000	00000	00100	10010	10001	00100	00001	00010	00010
98C	00100	01010	00000	00000	11010	00011	01000	00000	00100	00110
99A	01100	01000	00000	00100	10010	00011	01000	00001	00000	00111
99B	00100	00010	00000	00100	10010	01011	00100	00000	00000	00010
99C	00000	00000	01000	00100	10010	00001	00000	00000	00000	00010
100B	00100	01000	00000	00000	10010	10011	10001	00000	00000	00000
101A	00100	00000	00000	00100	10010	00011	01000	00000	00000	00110
102A	00100	00000	10000	00101	10010	00011	11010	00001	00000	00010
102B	00100	00000	00000	00100	10010	00001	00000	00000	00000	00010
102C	00100	00000	00000	00100	10010	00011	01000	01000	00000	00110
103A	00100	00000	00000	00100	10010	00011	00000	00000	00000	00100
103B	00000	00000	00000	00100	10010	00011	01000	00000	00000	00100
103C	00100	00000	00000	00100	10010	10011	01000	00001	00000	00110
104	00000	00000	00000	00010	10010	00011	00000	00000	00000	00100
105	00100	00000	00000	00000	10010	10010	00000	00000	00000	00111
106A	00101	00000	00000	00100	10010	01001	00110	00000	00000	00100
106B	00000	00000	01000	00100	10010	00011	00000	00000	00000	00110
106C	00000	01000	00000	00100	10010	00011	11001	00000	00000	00100
108A	00100	00000	00000	00001	10010	00001	01010	01001	00000	00110
108B	00100	00000	00000	00100	10010	00001	01000	00000	00000	00110
108C	00000	00000	00000	00100	10100	00011	00000	00001	00000	00110
110	00000	00000	00000	00100	10010	00001	01000	01001	00000	00110
111A	00100	00010	00000	00100	10100	00011	11000	00000	00010	00110
111B	00100	00000	00000	00100	10010	00001	01000	00000	00000	00010
111C	00100	00000	00000	00000	10010	00010	01000	01001	00000	00111
112	00100	00010	01010	00010	10000	00011	01000	00000	00000	00110
113A	00100	00010	01010	00010	10000	00011	01000	00000	00000	00010
114	00101	00000	01000	00100	10010	00001	00100	00001	00000	00010
116	00100	00010	00000	00000	00000	00011	01000	00000	00000	00000
118A	00000	00010	01000	00010	00000	00011	00000	00000	00000	00000
120C	00110	00010	01010	00010	10000	00111	01100	00000	00000	10000
121A	00100	00000	00000	10000	10000	00011	01000	00000	00000	00000
122A	00000	00000	00000	00000	10000	00011	01000	00001	00000	00000
123F	00000	00010	00000	00000	10000	00011	01000	00000	00000	00000

```
Fragments 151-200
123G   00000 00010 00000 00000 10000 00011 01000 00000 00000 00000
125A   00000 00010 00000 00000 10000 00011 01000 00000 00000 00000
126A   00000 00000 10000 00010 10000 00010 00000 00000 00000 00000
126E   00000 00000 00000 00000 00000 00000 00000 00000 00000 00000
127A   00110 00110 01010 10010 11000 00111 00100 00000 00000 00000
128A   00110 00000 01010 00010 11000 00011 00100 00000 00000 10000
130A   00100 00000 01010 00000 00000 00011 01100 00000 00000 10000
131    00000 00000 00000 00000 10000 00011 00000 10001 00000 00000
134A   10101 00000 00000 00000 10100 00011 00000 00000 00000 00000
HEN1   00100 00100 00100 00000 10000 00000 00000 00000 00000 00011
S3     00000 00000 00000 00000 10000 00001 00000 10000 00000 00000
S7     00100 00000 00000 00000 10000 00001 01000 00100 00000 00100
```

```
Fragments 201-250
1       00000 01010 00000 00100 00010 00010 00000 00001 00000 00001
3       00000 01010 00000 00000 00010 00010 00000 00001 00000 00001
5       00000 10010 01000 00000 00000 00010 00000 00101 00100 00001
8       00000 10010 01000 00000 00000 01010 00000 00011 00001 00000
10      00000 10010 01000 10000 00000 00010 00001 00001 00100 00001
12      00000 01010 00000 00000 00011 00010 00000 00001 00000 00001
13      00000 01010 00000 10100 00010 00010 00000 10001 00000 00000
14      00000 01010 00000 00000 00011 00010 00000 00001 00000 00001
16      00010 01010 00000 10100 00011 00010 00000 10001 00000 00011
17      00000 00010 00000 00000 00011 00000 00000 00000 00000 00000
18      00001 00100 00000 00000 00000 00000 00000 00000 00000 00000
19      00010 00100 00000 00000 00000 00010 00000 00000 00000 00000
20      00010 10100 00000 00000 00000 00010 00000 00000 00000 00000
25      00000 10010 00000 00000 00000 00010 00000 10000 00010 00000
26      00001 00100 00000 00000 00000 00100 00000 01001 00000 00000
27A     00000 01000 00000 10100 10010 00010 00000 10011 00000 00001
27B     00000 01010 00000 10100 10010 00010 00000 10001 00000 00100
27C     00000 01010 00000 00000 00010 00010 00000 00001 00100 00000
28A     00000 01000 01000 00000 00000 00010 00000 00001 00000 00001
29A     00000 01010 00100 00000 00010 00010 00000 10001 00000 00000
30A     00000 01010 00000 00000 00010 00010 00000 10001 00100 00001
30D     00000 01010 00000 00000 00010 00011 00000 10001 00000 00000
30E     00000 01000 00010 00000 00010 00010 00000 10011 00000 00011
30F     00000 01010 01000 00010 00010 00010 00001 10001 00100 00011
31      00000 01000 01000 00000 00010 00010 00000 10001 00100 00000
32B     00000 01010 00000 10000 00011 00010 00000 00001 00100 00000
36B     00000 00000 00000 00000 00000 00100 00100 00000 00000 01000
37A     00000 10110 00000 01000 00000 00010 00000 10000 00010 00000
37B     00000 10110 00000 01000 00000 00010 00000 10000 00010 00000
37C     00000 10000 00000 01000 00000 00010 00000 10000 00010 00000
39      00000 10110 00000 01001 00000 00010 00000 10000 00010 00000
41A     10000 10110 00000 01011 00000 00010 10000 10000 00010 00000
41B     00000 10100 00000 00000 00000 00010 00000 10000 00010 00001
41C     00000 10010 00000 00000 00000 00010 00000 10000 00010 00001
43      00000 10110 00000 01000 00000 00010 00000 10000 00010 00010
45A     00000 10100 10000 01000 01000 00010 10000 10000 01000 00011
45B     00000 10110 00000 01000 00000 00010 00000 10000 00010 00000
45C     00000 10110 00000 00000 00000 00010 00000 10000 00010 00001
46A     00001 10100 00000 00000 00000 00010 00000 10000 00010 00000
46B     00000 10100 00000 01000 00000 00010 00010 10000 00010 00001
46C     00000 10010 10000 01001 00000 00010 00000 10000 00010 00000
48A     00000 10010 00000 00000 00000 00010 00000 10000 00010 00011
48B     00000 10010 00000 01001 00000 00010 00000 10000 00010 00001
48C     00000 10110 00000 00000 00000 00010 00000 10000 00010 00000
48D     00000 10010 00000 01000 00000 00010 00000 10000 00010 00011
50A     00000 10110 00000 01000 00000 00010 00010 10000 00010 00010
50B     00000 10010 00000 01000 00000 00010 00010 10000 00010 00010
```

```
Fragments 201-250
50C    00000 10010 00000 00000 00000 00010 00010 10000 00010 00010
50D    00000 10010 00000 00000 00000 00000 00000 10000 00000 00000
52B    00000 10100 00001 00000 00000 00010 00000 00001 00000 10000
54A    00000 10100 00000 00000 00001 00100 00100 00000 00000 00000
55A    00000 10000 00000 00000 00001 00100 00100 00000 00000 00000
57B    00000 01010 00000 00000 00000 00010 00000 00001 00000 00000
58B    00000 10100 00000 00000 00011 00010 00000 00011 00000 00001
58D    00100 01010 00000 10000 00010 00010 00000 10001 00000 00001
61     00000 01010 00000 00000 00000 00010 00000 00001 00000 00000
62D    00000 01010 00000 00000 00000 00010 00000 00001 00000 00010
63B    00000 01010 00000 00000 00000 00010 00000 00001 00000 00010
65C    00000 10100 00000 00000 00000 00010 00000 00000 00000 00010
65D    00000 10100 00000 00000 00000 00010 00000 00000 00000 00000
66C    00001 00100 00000 00000 00000 00100 00000 00001 00000 00000
66D    00001 00100 00000 00000 00010 00100 00000 00001 00000 00000
67A    00011 00100 00000 00000 00000 00100 01000 01001 00000 00000
67B    00011 00100 00000 00000 00010 00100 01000 00001 00000 00000
68A    00000 01010 00000 00000 00010 00010 00000 00001 00000 00001
68B    00000 01000 00000 10100 00010 00010 00000 10001 00000 00001
68C    00000 00000 00000 00000 00000 00000 00000 00000 00000 00000
69A    00001 00100 00000 00000 00000 00000 00000 00001 00000 00000
69B    00001 00100 00000 00000 00000 00100 00000 00001 00000 00000
70A    00000 01000 00000 10000 00010 00010 00000 00001 00000 00001
70B    00000 01010 00000 10000 00010 00010 00000 10001 00100 00001
71A    00001 00100 00000 00000 00010 00100 00000 00001 00000 00000
72A    00000 01010 00000 00000 00010 00010 00000 10001 00000 00001
72B    00000 01010 00000 00000 00010 00010 00000 00000 00000 00001
74D    00000 00000 00000 00000 00000 00000 00000 00000 00000 00000
75A    00000 00000 00000 00000 00001 00000 00000 00000 00000 00000
76A    00011 00100 00000 00000 00010 00100 01000 01001 00000 00000
76E    00001 00100 00000 00000 00000 00100 00000 00001 00000 00000
77B    00001 00100 00000 00000 00000 00100 00000 00001 00000 00000
78A    00001 00100 00000 00000 00000 00100 00000 00001 00000 00000
79     00001 00100 00000 00000 00000 00100 00000 00001 00000 00000
80A    00000 01000 00000 00000 00011 00010 00000 00001 01000 00001
80C    00001 01000 00000 10000 01010 00010 00010 10001 00000 00001
80D    00001 01000 00000 00000 00010 00010 00000 00001 01000 00001
81A    00001 00000 00000 00000 00000 00100 00000 00001 00000 00000
82A    00001 00100 00000 00000 00010 00100 00000 00001 00000 00000
83A    00001 00100 00000 00000 00010 00100 00000 00001 00000 00000
84G    00001 00100 00000 00000 00010 00100 00000 00001 00000 00000
85A    00000 01010 00000 00000 00001 00000 00000 00001 00000 00000
85B    00000 01010 00000 00000 00001 00010 00001 10001 00000 00000
85D    00001 01010 01000 10000 00011 00010 00001 00001 10000 00000
85E    00000 01010 00000 00000 00011 00010 00000 00001 00000 00001
86A    00000 01010 00000 00000 00010 00010 00000 00001 00100 00001
86B    00000 01010 00000 00000 00001 00010 00000 00001 00000 00001
```

Fragments 201-250

87A	00000	01010	00000	00000	00010	00010	00000	10001	00000	00001
89A	00000	01010	00000	10000	00000	00010	00010	00000	00100	00001
89B	00000	01010	00000	00000	00010	00010	00000	00001	00100	00001
90A	00000	01000	00000	00000	00011	00010	00001	00001	00101	00000
90B	00000	01000	00000	00000	00011	00010	00001	00001	00000	00001
90C	00000	01000	00000	00000	00010	00010	00000	10001	00101	00001
92A	00100	01010	00000	00000	00000	00000	00000	10000	10101	00001
92B	00000	01000	00000	00010	00000	01010	00000	00001	00100	00000
93C	00000	01010	00000	00000	00000	00010	00000	00001	00000	00000
95A	00000	01000	10000	00000	00011	00010	00000	10001	00100	00000
96A	00000	01010	00010	00000	00010	00000	00000	00011	00000	00000
96B	00001	01011	00000	10000	00000	00011	00010	10010	00010	00001
98A	00000	10110	00010	00000	00000	00010	00010	10000	00000	00000
98B	00001	10111	10000	01000	00000	00010	00010	10000	00010	00000
98C	00000	10110	00010	00000	00000	00010	00010	10000	00000	00000
99A	11000	10010	10000	00000	00000	00010	10000	10000	00010	00000
99B	00000	10010	10000	00000	00000	00010	00000	10000	00110	00000
99C	00000	10110	00000	00000	00000	00010	00000	10000	00010	00000
100B	00000	01010	00000	00000	00001	00010	10100	10000	00000	01000
101A	00001	10110	00000	00000	00100	00010	00000	10000	00010	00001
102A	00000	10100	00000	00001	00101	00010	10011	10011	00010	00100
102B	00000	10100	00000	00000	00000	00010	00000	10000	00010	00000
102C	00000	10110	00000	01000	00000	00010	00000	10000	00010	00000
103A	00000	10110	00000	00000	00000	00010	00010	10001	00010	00000
103B	00000	10010	00000	00000	00000	00010	00000	10000	00010	00000
103C	10000	10110	00000	00000	00000	00010	00010	10000	00010	00011
104	00000	10110	00000	00000	00000	00010	00000	10000	00010	00000
105	00000	10110	00000	00000	00000	00010	10000	10000	00010	00011
106A	00000	10110	10000	01001	00000	00010	00000	10000	00010	00000
106B	00000	10110	00000	01000	00000	00010	00000	10000	00010	00000
106C	00000	10110	00000	00000	00000	00010	00000	10000	00010	00000
108A	00000	10110	00000	00000	00000	00010	00000	10000	00010	00010
108B	00000	10110	00000	00000	00000	00010	00000	10000	00010	00000
108C	00000	10100	00000	00000	00000	00010	00010	10000	00010	00001
110	00000	10110	00000	01000	00000	00010	00000	10000	00010	00000
111A	00000	10010	00000	00000	00000	00010	00000	10000	00010	00001
111B	00000	10010	10000	00000	00000	00010	00000	10000	00010	00000
111C	00011	10110	00000	01000	00000	00010	00000	10000	00010	00000
112	00000	10010	10100	00000	00000	10010	01000	00001	00000	00000
113A	00000	10010	10100	00000	00000	10010	01000	00001	00000	00000
114	00000	10010	10000	01000	00000	00010	00000	10000	00010	00000
116	00000	01010	00000	00000	00000	00010	00000	10000	00000	00000
118A	00000	01010	00000	00000	00000	00010	00000	00001	00000	00000
120C	00000	00000	00000	00000	00001	00000	00000	00000	00000	00000
121A	00000	10010	00000	00000	00000	00000	00000	00001	00000	00000
122A	00000	10010	00000	00000	00000	00010	00000	10001	00000	00000
123F	00000	10010	00001	00000	00000	00010	01000	10000	00000	00000

```
Fragments 201-250
123G   00010 10010 00000 00000 00000 00010 01000 00000 00000 00000
125A   00010 10010 00000 00000 00000 10010 00000 00001 00000 00000
126A   00000 01010 00000 00000 00000 00010 00000 00001 00000 00000
126E   00000 00000 00000 00000 00000 00000 00000 00000 00000 00000
127A   00000 00000 00000 00000 00000 00000 00000 00000 00000 00000
128A   00000 10010 00000 00000 00001 00000 00000 00000 00000 00000
130A   00000 10010 00000 00000 00000 00000 00000 00001 00000 00000
131    00000 01010 01000 00000 00000 00010 00000 00001 00100 00000
134A   00000 00000 00000 00000 00000 00000 00000 00000 00000 00000
HEN1   00000 00000 00000 00000 00000 00000 00000 00001 00000 00000
S3     00000 01000 00000 00000 00010 00000 00000 00001 00000 00000
S7     00000 01010 00000 00000 00011 00010 00000 00001 00000 00001
```

```
Fragments 251-300
1       00000 00000 00000 01010 00000 00000 00000 00100 00101 00000
3       00000 00000 00000 00010 00000 00010 10000 00010 00001 00010
5       00000 00000 00000 00010 00000 00000 00000 00010 00101 00000
8       01000 00000 00000 00000 00000 00000 00000 00010 00100 00000
10      00000 00000 00000 00010 00000 00000 00000 00010 00101 00000
12      01000 00000 00000 00010 00000 00000 00000 00010 00101 00000
13      01000 00000 00000 00010 00000 00000 00000 00010 00110 00000
14      00000 00000 00000 00010 00000 00000 10000 00010 00110 00000
16      01001 00000 00000 01010 00010 00000 00000 00100 00100 00000
17      00000 00000 00000 00010 00000 00000 10000 00010 00110 00001
18      00000 00000 00010 00010 00000 00010 10000 00011 00111 00001
19      00000 00000 00000 00010 00000 00000 00000 00010 00101 00000
20      00000 00000 00000 00010 00000 00000 00000 00010 00111 00000
25      01100 00000 00000 00000 00000 00000 00000 00010 00100 00000
26      00000 00000 00000 00010 00000 00000 00000 00010 00101 00000
27A     00000 00000 00000 00010 00000 00010 00000 00011 00101 00000
27B     01000 00000 00000 00010 00001 00001 00000 00011 10010 00010
27C     00000 00000 00000 00010 00000 00000 00000 00010 00100 00000
28A     00000 00000 00000 00010 00000 00000 00000 00010 00100 00000
29A     01000 00000 00000 00010 00000 00000 00000 00010 00110 00000
30A     00100 00000 00000 00010 00000 00000 00000 00010 00100 00000
30D     00001 00000 00000 01000 00000 00001 01000 00011 00110 00000
30E     01110 00100 00000 00010 00000 00000 10000 00010 00101 00001
30F     01000 00000 00000 00000 00000 00000 00000 00010 00100 00001
31      01000 01000 00000 01010 00000 00000 10000 00011 00110 01001
32B     01000 00000 00000 00010 00001 00000 00000 00011 00110 00000
36B     00000 00000 10000 00001 00000 00000 00000 00010 00101 00000
37A     01000 00000 00000 00000 00000 00000 00000 00010 00110 00000
37B     01000 00000 00000 00010 00100 00000 00000 00011 00110 00000
37C     00000 00000 00000 00010 10000 00000 01000 00011 10110 00001
39      01000 00000 00000 00000 00000 00000 00000 00100 00110 00000
41A     01000 00000 00000 00010 00000 00000 00000 00011 00110 00001
41B     01000 00000 00000 00000 00000 00000 00000 00010 00101 00000
41C     01000 00000 00000 00000 00000 00000 00000 00000 00001 00000
43      10000 00000 00000 00000 00010 00000 00000 00100 10100 00000
45A     01000 01000 00000 01010 00000 00000 10000 00011 01100 00000
45B     01000 00000 00000 00000 00000 00000 10000 00011 00110 00000
45C     01000 00000 00000 00010 00000 00010 10000 00011 00111 00001
46A     01000 00000 00000 00000 00000 00000 00000 00010 00110 00000
46B     01000 00000 00000 00000 00000 00000 00000 00010 00110 00000
46C     01000 00000 00000 00000 00000 00000 00000 00010 00100 00000
48A     00000 00000 00000 00000 00000 00000 00000 00010 00100 00000
48B     10000 00000 00000 01010 00000 00010 11000 00010 00111 00000
48C     01000 00000 00000 00000 00000 00000 00000 00100 00100 00000
48D     00000 00000 00000 00000 00000 00000 00000 00010 00100 00000
50A     10000 00000 00000 00000 00000 00000 00000 00110 00100 00000
50B     10000 00000 00000 00000 00000 00000 00000 00100 00101 00000
```

```
Fragments 251-300
50C    10000 00000 00000 00000 00000 00000 00000 00100 00101 00000
50D    00000 00000 00000 00010 00000 00000 00000 00010 00100 00000
52B    00000 00000 00000 00010 00000 00000 00000 00010 00101 00000
54A    00000 10000 10000 00001 00000 00001 00000 00100 00100 00000
55A    00000 10010 10000 00001 00000 00001 00000 00100 00100 00000
57B    00000 00000 00000 00010 00000 00000 00000 00010 00111 10000
58B    00000 00000 00000 00010 00000 00000 00000 00010 00100 00000
58D    00000 00000 00000 00010 00000 00000 00000 00010 00101 01000
61     00000 00000 00000 00010 00000 00000 00000 00010 00111 00100
62D    00000 00000 00000 00010 00000 00000 00000 00010 00100 00000
63B    00000 00000 00000 00000 00000 00000 00000 00010 00101 00000
65C    00000 00000 00000 00010 00000 00000 10000 00010 00111 00001
65D    00000 00000 00000 00010 00000 00000 00000 00010 00100 00001
66C    00000 00000 00000 00010 00000 00000 00000 00010 00100 00000
66D    00000 00000 00000 00010 00000 00000 00000 00010 00100 00000
67A    00000 00000 00000 00010 00000 00000 00000 00010 00110 00100
67B    00000 00000 00000 00010 00000 00100 00000 00010 00100 00100
68A    00000 00000 00000 00010 00000 00000 10000 00010 00100 00010
68B    00000 00000 00001 00010 00000 00000 00000 00010 00111 00000
68C    00000 00000 00001 00110 00000 00000 00000 00110 00110 00000
69A    00000 00000 00000 00010 00000 00000 00000 00010 00101 00100
69B    00000 00000 00000 00010 00000 00000 10000 00010 00110 00100
70A    01000 00000 00000 00000 00000 00000 00000 00110 00100 01000
70B    01000 00000 00000 00010 00000 00000 00000 00110 00111 00000
71A    00000 00000 00000 00010 00000 00000 00000 00010 00100 00000
72A    01000 00000 00000 00000 00000 00000 00000 00010 00110 00000
72B    01000 00000 00000 00000 00000 00000 00000 00010 00111 00000
74D    00000 10010 10000 00001 00000 00000 00000 00100 10100 00000
75A    00000 00000 00000 00010 00000 00000 00000 00011 10110 00001
76A    00000 00000 00000 00010 00000 00100 00000 00110 00111 00100
76E    00000 00000 00000 00010 00000 00100 00000 00010 00100 00000
77B    00000 00000 00000 00010 00000 00000 10000 00010 00110 00000
78A    00000 00000 00000 00000 00000 00100 00000 00010 00100 00000
79     00000 00000 00000 00010 00000 00000 10000 00010 00110 00000
80A    00000 00000 00000 00010 00000 00000 00000 00010 00101 00000
80C    01000 00000 00000 00010 00000 00000 10000 00111 00110 01000
80D    01000 00000 00000 00000 00000 00000 10000 00011 10110 01010
81A    00000 00000 00000 00010 00000 00100 00000 00010 00100 00000
82A    00000 00000 00000 00010 00000 00000 00000 00010 00110 00000
83A    00000 00000 00000 00010 00000 00000 00000 00010 00100 00000
84G    00000 00000 00000 00010 00000 00000 00000 00010 00110 00001
85A    00000 00000 00000 00010 00000 00000 00000 00010 00100 00000
85B    01000 00000 00000 00010 00000 00000 00000 00110 00110 00000
85D    01000 00000 00000 00000 00000 00000 00000 00010 00110 00000
85E    01000 00000 00000 00010 00000 00000 00000 00110 00110 00000
86A    01100 00000 00000 00010 00000 00000 10000 00111 00110 00000
86B    00000 00000 00000 00010 00000 00000 00000 00011 00110 00001
```

```
Fragments 251-300
87A     00000 00000 00000 00010 00000 00000 10000 00010 00110 00001
89A     01000 00000 00000 00100 00000 00010 00000 00010 00100 00001
89B     01000 00000 00000 00000 00000 00000 00010 00010 00100 00000
90A     00010 00000 00000 00000 00000 00000 00000 00010 00110 01000
90B     00010 00000 00000 00010 00000 10000 10000 00110 00110 00000
90C     00000 00000 00000 00010 00000 00000 10000 00010 00010 00001
92A     00000 00000 00000 01110 00000 00000 00000 00011 00100 00000
92B     01100 00000 00000 00000 00010 00010 00000 00011 01010 00011
93C     00000 00000 00000 00010 00000 00000 00000 00010 00101 00000
95A     01000 00000 00000 00010 00000 00000 00000 00110 00100 00000
96A     00000 00100 00000 00000 00010 00000 01000 00011 01100 00001
96B     01000 00000 00000 00000 00000 00010 00000 00011 00110 00001
98A     00000 00000 00000 00000 00000 00000 00000 00000 00000 00000
98B     01000 00000 00000 00000 00000 00000 00000 00111 00110 00000
98C     00000 00001 00000 00000 00000 00000 10000 00011 10110 00000
99A     01000 00000 00000 00000 00000 00000 00000 00110 00110 00000
99B     01000 00000 00100 00010 00000 00000 00000 00011 00110 00000
99C     01000 00000 00000 00010 00000 01000 00000 00111 01111 00000
100B    00000 10000 10000 00000 00000 00010 01000 00110 00110 00000
101A    00000 00000 00000 00010 00000 00000 00000 00010 00110 00000
102A    01000 00000 00000 00000 00000 00000 00000 00110 00110 00000
102B    00000 00000 00000 00010 00000 00000 10000 00010 00110 00001
102C    01000 00000 00000 00000 00000 00000 00000 00110 00110 00000
103A    01000 00000 00000 00010 00000 00000 10000 00010 00100 00000
103B    11000 00000 00000 00100 00000 00000 00000 00010 00110 00000
103C    00100 00001 00000 00000 00000 00000 00000 00000 00010 00000
104     01000 00000 00000 00010 00000 00000 10000 00010 00110 00000
105     00000 00000 00000 00000 00000 00000 00000 00110 00110 00000
106A    10000 00000 00000 00000 00000 00000 00000 00010 00100 00001
106B    10000 00001 00000 00000 00000 00000 00000 00100 00110 00000
106C    01000 00000 00000 00000 00000 00000 00000 00100 00110 00000
108A    01000 00000 00000 00000 00000 00000 00000 00100 00110 00000
108B    00000 00000 00000 00000 00000 00000 00000 00110 00110 00000
108C    00000 00000 00000 00010 00000 00000 00000 00010 00100 00000
110     10000 00000 00000 00100 00000 10000 00000 00110 00110 00000
111A    01000 00000 00000 00010 00000 00000 00000 00011 00110 00001
111B    01000 00000 00000 00000 00000 00000 00000 00110 00111 00000
111C    01000 00000 00000 00000 00000 00000 00100 10100 00100 00000
112     00000 00000 00000 00010 00000 00000 10000 00010 00101 00001
113A    00000 00000 00010 00010 00000 00000 00000 00011 00110 00001
114     01000 00000 00000 00000 00000 00000 10000 00010 00110 00000
116     00000 00000 00000 00010 00000 00000 00000 00011 00111 00001
118A    00000 00000 00000 00010 00000 00000 10000 00010 00101 00000
120C    00000 00000 00000 00110 00000 00000 00000 00011 00110 00001
121A    00000 00000 00000 00010 00000 00000 00000 00010 00100 00000
122A    00000 00000 00000 00010 00000 00010 10000 00010 00111 00001
123F    00000 00000 00000 00010 00000 00000 00000 00010 00100 00000
```

```
Fragments 251-300
123G   00000 00000 00000 00010 00000 00000 00000 00010 00100 00000
125A   00000 00000 00000 00010 00000 00000 10000 00011 00100 00001
126A   00000 00000 00000 00010 00000 00010 10000 01011 10111 10001
126E   00000 00000 00000 00010 00000 00000 00000 00010 00101 10000
127A   00000 00000 00000 01010 00000 00000 10000 00011 00100 00001
128A   00000 00000 00000 00010 00000 00000 10000 00011 00110 00001
130A   00100 00000 00000 01010 00000 00000 10000 00011 00110 00001
131    00000 00000 00000 00010 00000 00000 00000 00010 00100 00000
134A   00000 00000 00000 00010 00000 00000 10000 00010 00111 00001
HEN1   00000 00000 00000 00010 00000 00010 10000 00011 00110 00001
S3     00000 00000 00000 00010 00000 00000 00000 00010 00100 00000
S7     00000 00000 00000 00010 00000 00000 00000 00011 00110 00001
```

```
Fragments 301-350
1       00100 00010 00000 00000 10000 00010 00001 00000 00000 00000
3       00100 00010 00000 00000 11000 00010 00000 00001 10000 00000
5       00000 00100 00000 00000 10000 00011 01000 00000 10000 00010
8       00100 00100 00100 00000 10000 00011 00000 00000 10010 00000
10      00000 00100 00000 00000 10000 00011 00000 00000 10000 00010
12      00000 00100 00000 00000 10000 00010 00000 00000 10010 00000
13      00000 00100 00000 01000 10000 01011 00001 11010 10000 00000
14      00100 00100 00000 00000 10000 00010 00001 00000 10010 00000
16      00001 00100 00000 00000 10000 00010 00001 10000 10000 00000
17      01100 00100 00001 00100 00000 01101 00001 00001 00000 00000
18      01100 00010 00001 00100 00000 01100 00001 10001 00000 00001
19      00000 00110 00000 00000 10000 00000 00000 00000 00000 00000
20      00000 00010 00000 00000 10000 00010 00000 00000 01000 00000
25      00000 00100 00000 00000 10000 00010 00000 00000 01000 00000
26      00000 00000 00000 00000 10000 00010 00000 01000 01000 00001
27A     00001 00100 00101 01010 10000 00011 00000 10000 00000 00000
27B     00100 00100 00000 01000 10010 00010 00000 00000 10000 00000
27C     00000 00000 00000 00000 10000 00010 00000 00000 10000 00000
28A     00100 00100 00000 00000 10000 00010 00000 00000 10000 00000
29A     00000 00100 00000 00000 10000 00011 01000 00000 10000 00000
30A     00000 10100 00000 00000 10000 00010 00000 00000 10011 00100
30D     00100 00100 01000 10100 10100 00000 00001 00100 10000 00000
30E     00000 00100 00000 00000 10000 00010 00000 00000 00000 00000
30F     00100 00100 00000 00000 10001 01000 01001 11000 10000 00000
31      01100 00100 00000 01100 10000 00000 00000 00000 00000 00000
32B     00000 00100 00001 01000 10000 00000 00000 00000 10000 00000
36B     00000 00000 00000 00000 10000 00000 00010 00000 10000 00000
37A     00000 00000 00000 00000 10000 00011 10000 00000 10011 10000
37B     00100 00100 00000 01010 10001 00011 01000 00000 10001 00000
37C     00110 00100 00000 01000 10000 00000 00000 10011 00000 00000
39      00000 00000 00000 00000 10000 00011 10000 00000 10000 10000
41A     01100 00101 00001 00110 10000 00100 00000 00010 00000 00000
41B     00000 00000 00000 00000 10000 00011 00000 00000 01000 10000
41C     00000 00000 00000 00000 10000 00011 00000 00000 10000 10100
43      00000 00100 10000 00000 10000 00011 00001 00001 10000 00000
45A     01100 00100 01000 10010 00000 00100 00000 00001 00000 00000
45B     10100 01100 01000 00011 10000 00100 00000 00000 00000 00000
45C     00100 00100 00001 00110 10001 00100 00000 00001 00000 00000
46A     00000 00000 00000 00000 10000 00011 00000 00000 10000 10000
46B     00000 00100 01000 00010 10100 00011 00000 00000 10000 00000
46C     00000 00000 00000 00000 10000 00011 10000 00000 10000 10000
48A     00000 00000 10000 00000 10000 00011 00000 00000 10000 00000
48B     10100 00100 00000 01000 10000 00010 00001 00001 00000 00000
48C     00000 00000 00000 00000 10000 00011 10000 00000 10000 00000
48D     00000 00000 10000 00001 10000 00011 00000 00000 10000 00000
50A     00000 00000 00000 00000 10000 00011 00000 00000 10110 00000
50B     00000 00000 00000 00000 10100 00010 00000 00000 10000 00000
```

Fragments 301-350

50C	00000	00000	00000	00000	10100	00011	10000	00000	10000	00000
50D	00000	00100	00000	00000	10000	00011	00000	00000	01000	00000
52B	00000	00100	00000	00000	10000	00011	00000	00000	01000	00000
54A	00000	00000	00000	00000	10000	00000	00010	00000	10000	00000
55A	00000	00000	00000	00000	10000	00000	00010	00000	10000	00000
57B	00000	00100	00000	00000	10000	00000	00000	00000	00000	00000
58B	00000	00100	00000	00000	10000	00010	00000	00000	01000	00001
58D	00100	00000	00000	00100	10000	00011	00001	10100	10000	00000
61	00000	00100	00000	00000	10000	00000	00000	00000	00000	00000
62D	00100	00010	00000	00000	10000	00011	01000	00000	01000	00000
63B	00000	00100	00000	00000	10000	00010	00000	00000	00000	00000
65C	01100	00101	00001	00100	10000	00100	00001	00001	00000	00000
65D	00100	10100	00000	00000	10000	00011	00000	00000	10000	00000
66C	00000	00010	00000	00000	10000	00010	01000	00000	00000	00000
66D	00100	00010	00000	00000	10000	00010	00000	00000	10000	00001
67A	00000	00010	00000	00000	10000	00000	00001	11000	01000	00001
67B	00000	00010	00000	00000	10000	00000	00001	01000	00000	00001
68A	00100	00100	00000	00000	10000	00010	00000	00000	10000	00000
68B	00000	00100	00000	00000	10000	00011	00001	00000	10000	00000
68C	00000	00000	00000	00100	10000	00011	01000	00000	10001	00000
69A	00000	00010	00000	00000	10000	00010	00001	11000	00000	00100
69B	00000	00010	00000	00000	10000	00010	00001	00000	00000	00001
70A	00000	00100	00100	00000	10000	00011	00001	00000	10000	00000
70B	00100	00100	00000	00000	10000	00011	11001	00000	10000	00000
71A	00000	00010	00000	00000	10000	00010	00000	00000	00000	00001
72A	00000	00100	00000	00000	10000	00011	00000	00001	10000	00000
72B	00000	00100	00000	00100	10000	00011	00100	00000	10000	00000
74D	00000	00100	00000	00000	10000	00000	00010	00000	10000	00000
75A	01100	00100	00001	00100	00000	00000	00000	00001	00000	00000
76A	00000	00010	00000	00000	00000	00000	00000	01000	00000	00001
76E	00000	00010	00000	00000	10000	00000	00000	00000	00000	00001
77B	00100	00010	00000	00000	10000	00010	00000	00000	00000	00001
78A	00000	00000	00000	00000	10000	00000	00000	00000	00000	00000
79	00100	00010	00000	00000	10000	00000	00000	00000	00000	00001
80A	00100	00101	00000	00000	10000	00110	00001	00000	00000	00000
80C	00000	00100	00000	01000	10001	00011	01001	00001	00000	00000
80D	00100	00000	00001	00000	11000	00010	00100	00000	00000	00000
81A	00000	00000	00000	00000	10000	00000	00000	00000	00000	00001
82A	00000	00010	00000	00000	10000	00011	00000	00000	10000	00001
83A	00000	00010	00000	00000	10000	00010	00000	00000	10000	00000
84G	00100	00100	00000	00100	10000	00000	00001	00000	00000	00001
85A	00000	00000	00000	00000	00000	00000	00001	00000	10000	00000
85B	00000	00100	00000	00000	10000	00011	00101	00000	10000	00000
85D	00000	00100	00000	00000	10000	00011	00000	10010	10000	00000
85E	00100	00100	10000	00000	10001	00111	11000	00000	10000	00000
86A	00000	00000	00000	00000	10000	01010	00000	00000	10000	00000
86B	00100	00100	01010	00000	10000	00110	01000	00000	00000	00000

Fragments 301-350

87A	01000	00100	00000	00100	10000	00000	01001	10001	10000	00000
89A	01100	00000	00000	00000	10000	00000	00000	00001	00000	00010
89B	00000	00000	00000	00000	10000	00011	00000	00000	10000	10000
90A	00000	00100	00000	00000	10000	01011	01000	00001	10000	00000
90B	00000	00000	00000	00010	10000	00011	00000	00000	10000	00000
90C	01100	00100	00000	00110	10000	00000	00001	00001	00000	00000
92A	00100	00100	00000	00110	10000	01000	00000	00010	00000	00000
92B	01100	00100	00000	00100	10000	00100	00001	00010	00100	00000
93C	00000	00100	00000	00000	10000	00000	00000	00000	00000	00000
95A	01000	00100	00000	00100	10000	01111	01001	01000	10000	00010
96A	00100	00100	00010	10000	00000	00000	00000	00000	00000	00000
96B	00100	00100	00000	00100	10000	00000	00000	00000	00000	00000
98A	00000	00000	00000	00000	00000	00000	00000	00000	00000	00000
98B	00000	00100	01000	00000	10000	00011	00000	00000	11000	00000
98C	01100	00001	00000	00010	00000	00000	00000	00000	00000	00000
99A	00000	00100	00000	00000	10000	00011	00000	00000	10000	00000
99B	00000	00100	00000	00000	10000	00011	00000	00000	10000	00000
99C	00000	00100	00000	00010	10000	10011	00000	00001	10000	00000
100B	00000	00100	00100	00000	10000	00000	00011	00000	10000	00000
101A	00000	00100	00000	00000	10000	00011	00000	00000	10000	00000
102A	00000	00001	00100	00000	10000	00011	01000	00000	10011	00100
102B	00100	00100	00010	00100	10000	00111	00000	00001	10000	00000
102C	00000	00100	00000	00000	10100	00011	00000	00000	10000	00000
103A	00000	00100	00000	00011	10000	00111	00000	00000	10000	00000
103B	00000	00000	10000	00000	10001	00010	10000	00000	10000	00000
103C	00000	00000	00000	00000	10000	00011	00000	00000	10001	10000
104	00000	00100	00000	00011	10000	00010	00000	00000	00000	01000
105	00000	00000	00000	00000	10000	00011	00000	00000	10000	00000
106A	00000	01100	01000	00000	10000	00100	00000	00000	00001	00000
106B	00000	10000	00000	00000	10000	00011	00000	00000	10000	00000
106C	00000	10000	00000	00000	10000	00011	10010	00000	10000	00000
108A	00000	00000	00000	00000	10100	00011	00000	00000	11000	00100
108B	00001	00100	00000	00000	10000	00011	10000	00000	10000	00000
108C	00000	00100	00000	00000	10000	00011	00000	00000	10000	00000
110	00000	00000	00000	00000	10000	00010	10000	00000	10000	00000
111A	01100	00100	00001	00111	10000	00100	00001	10001	00000	00000
111B	00000	00000	00000	00000	10000	00111	00000	00000	10000	00000
111C	00000	00000	00000	00010	10000	00011	00000	00000	10000	00000
112	01100	00100	00000	00100	00000	00100	00000	00001	00000	00000
113A	01100	00100	00001	00110	00000	00100	00001	10001	00000	00000
114	00000	00000	00000	00000	10000	00011	00000	00000	10000	00000
116	01100	00100	00000	00110	00000	00100	00001	10001	00000	00000
118A	00100	00100	00000	00000	10000	00110	00001	00000	00000	00000
120C	01100	00101	00000	00100	00000	00100	00001	10001	00000	00000
121A	00000	00100	00000	00000	10000	00010	00000	00000	10000	01000
122A	00100	00100	00000	00100	10000	00100	00000	00000	00000	00000
123F	00100	10100	01000	00000	10000	00010	00001	10000	10000	00000

```
Fragments 301-350
123G   00000 00100 00000 00000 10000 00010 00101 00000 10000 00000
125A   01100 00100 00001 00110 10000 00100 00001 10001 00000 00000
126A   01100 00100 00010 00100 10000 00111 00001 00001 00000 01000
126E   00000 00100 00000 00000 10000 00010 00000 00000 00000 00000
127A   00100 00100 00010 00110 00000 01100 00001 00001 00000 00000
128A   01100 00100 00000 00110 00000 01100 00001 00001 00000 00000
130A   01000 00100 00001 00100 10000 00100 00001 10001 00000 00000
131    00100 00100 00000 00000 10000 00011 00000 00000 10000 00000
134A   01100 00101 00001 00100 10000 00100 00001 10001 00000 00000
HEN1   01000 00101 00000 00000 00001 00100 00001 01001 00000 00000
S3     00000 00100 00000 00000 10000 00010 00000 00000 10000 00000
S7     01100 00100 00001 00100 10000 00000 00001 00000 00000 00000
```

```
Fragments 351-376
1      10000 00000 00000 00000 11000 1
3      10000 00000 00000 00000 11000 0
5      00000 00000 00000 00000 01010 0
8      00000 00000 01000 00000 01010 1
10     00010 00000 00000 00000 01000 1
12     10000 00000 00000 00000 01010 1
13     00100 00000 00000 00000 01000 1
14     00010 00000 00000 00000 01000 1
16     10000 00000 00100 00010 01010 0
17     00000 00000 00000 00000 01000 0
18     00000 00000 00000 00000 00000 0
19     00000 00000 00000 00000 01000 0
20     00000 00000 00000 00000 01000 0
25     00000 00000 00000 00010 01000 0
26     00000 00000 00000 00000 11000 0
27A    11000 00010 00000 00000 01010 1
27B    11000 00000 00000 00000 01000 1
27C    10000 00000 00000 00000 01000 1
28A    10000 00000 00000 00000 01000 0
29A    00000 00000 00010 00000 01000 0
30A    00010 00000 00000 00000 01000 1
30D    00000 00000 00000 00000 01000 1
30E    00000 00000 00000 00000 01000 1
30F    00000 00000 00100 00000 01000 1
31     00000 00000 00000 00000 01000 1
32B    00100 00000 00000 00000 01000 1
36B    00000 10000 00000 00000 00000 0
37A    00000 00000 00000 00000 01100 0
37B    00010 00000 00000 00000 01000 0
37C    00000 00000 00000 00000 01000 0
39     00000 00000 00100 00000 01000 0
41A    00000 00000 00000 00000 01000 0
41B    00000 00000 00000 00010 01000 0
41C    00000 00000 00000 00010 01000 0
43     00000 00000 00000 00010 01000 0
45A    00000 00000 00000 00000 01000 0
45B    00000 00000 00000 00000 01000 0
45C    00000 00000 00000 00000 01000 0
46A    00000 00000 00000 00000 01000 0
46B    00000 00000 00000 00010 01100 0
46C    01000 00000 00000 00000 01000 0
48A    00000 00000 00000 00000 01000 0
48B    00000 00000 00000 00010 01000 0
48C    00000 00000 00000 00000 01000 0
48D    01000 00000 00000 00000 01000 0
50A    00000 00000 00000 00010 01100 0
50B    00000 00000 00000 00010 01100 0
```

```
Fragments 351-376
50C    00000 00000 00000 00000 01010 0
50D    00000 00100 00000 00010 01000 0
52B    00000 00000 00000 00000 01000 0
54A    00000 10000 00001 00000 00000 0
55A    00000 10000 00001 00000 00000 0
57B    00000 00000 00000 00000 01000 0
58B    10000 00000 00000 00000 01100 0
58D    10000 00000 00000 00000 01000 0
61     00000 00000 00000 00000 01000 0
62D    00000 00000 00000 00000 01000 0
63B    00000 00000 00000 00000 01000 0
65C    00000 00000 00000 00000 01000 0
65D    00000 00000 00000 00000 01000 0
66C    00000 00000 00000 00000 01000 0
66D    00000 00000 00000 00000 01000 0
67A    00001 00000 00000 00000 11000 0
67B    00000 00000 00000 10001 11000 0
68A    01000 00000 00000 00000 01000 1
68B    11000 00000 00000 00000 11000 1
68C    00100 00000 00000 00000 01000 1
69A    00000 00000 00000 10001 01010 0
69B    00001 00000 00000 00000 01000 0
70A    10000 00000 00000 00000 01000 1
70B    00000 00000 00000 00000 01000 1
71A    00000 00000 00000 00000 11001 0
72A    00000 00000 00000 00000 01000 1
72B    00000 00000 00000 00000 01000 1
74D    00000 10000 00001 00000 01000 0
75A    00000 00000 00000 00000 01000 0
76A    00000 00000 00000 10000 01000 0
76E    00000 00000 00000 00000 01000 0
77B    00000 00000 00000 00000 01001 0
78A    00000 00000 00000 00000 01000 0
79     00000 00000 00000 00000 11000 0
80A    10000 00000 00000 00000 01000 1
80C    10000 10010 00000 00000 11000 1
80D    00000 00000 00100 00000 11000 0
81A    00000 00000 00000 00000 01000 0
82A    00000 00000 00000 00000 01000 0
83A    00000 00000 00000 00000 01000 0
84G    00000 00000 00000 00000 01001 0
85A    00000 00000 00000 00000 01000 0
85B    00000 00000 00000 00000 01010 0
85D    00000 01000 00000 00000 01000 0
85E    00000 00000 00000 00000 01010 0
86A    10100 00000 00000 00000 01010 1
86B    00000 00000 00000 00000 01010 1
```

```
Fragments 351-376
87A    10000 00000 00000 00000 01000 1
89A    10100 00000 00000 00000 01010 1
89B    00100 00000 10000 00000 01000 1
90A    00101 00000 00000 00000 01000 1
90B    00100 00000 00000 00000 01000 1
90C    00000 00000 00000 00000 01000 0
92A    00000 00100 00000 00000 01000 0
92B    00000 00000 00000 00000 01000 0
93C    00000 00000 00000 00000 01000 0
95A    00000 00000 00100 00000 01000 1
96A    00000 00000 00000 00000 11000 0
96B    00000 00000 00000 00000 01000 0
98A    00000 00000 00000 00000 01000 0
98B    01000 00000 00000 00000 01000 0
98C    00000 00100 00000 00000 01000 0
99A    00000 00000 00000 00000 01000 0
99B    00001 00000 00000 00000 01000 0
99C    00000 00000 00000 00000 01000 0
100B   00000 00010 00000 00000 01000 0
101A   00000 00000 00100 00000 01000 0
102A   00000 00000 01000 00010 01000 0
102B   00000 00100 00000 00000 01000 0
102C   00000 00000 00000 00000 01000 0
103A   00000 00000 00000 00010 01000 0
103B   00000 00000 00000 00010 01000 0
103C   00000 00000 00000 00010 01000 0
104    00000 00000 00000 00010 01000 0
105    00000 00000 00000 00000 01100 0
106A   00000 00000 00000 00000 01000 0
106B   00000 00010 00000 00010 01000 0
106C   00000 00010 00101 00010 01000 0
108A   00010 00000 00000 00010 01000 0
108B   00000 00000 00000 00000 01000 0
108C   00001 00000 00000 00000 01000 0
110    00000 00000 00000 00000 01000 0
111A   00000 00000 00000 00000 01000 0
111B   00000 00000 00000 00000 01000 0
111C   00000 00000 00100 00000 01000 0
112    00000 00000 00000 00000 01000 0
113A   00000 00000 00000 00000 01000 0
114    00000 00000 00000 00000 01000 0
116    00000 00000 00000 00000 01000 0
118A   00000 00000 00000 00000 01000 0
120C   00000 00000 00000 00000 01000 0
121A   00000 00000 00000 00000 01000 0
122A   00000 00000 00000 00000 01000 0
123F   00000 00000 00000 00000 01000 0
```

```
Fragments 351-376
123G   00000 00000 00000 00000 01000 0
125A   00000 00000 00000 00000 01000 0
126A   00000 00000 00000 00000 01000 0
126E   00000 00000 00000 00000 01000 0
127A   00000 00000 00000 00000 01000 0
128A   00000 00000 00000 00000 01000 0
130A   00000 00000 00000 00000 01000 0
131    10000 00000 00000 00000 01000 0
134A   00000 00000 00000 00000 01000 0
HEN1   00000 00000 00000 00000 01000 0
S3     10000 00000 00000 00000 01000 1
S7     10000 00000 00000 00000 01010 0
```

Appendix C. AFLP fragment sizes

This appendix gives the size in base pairs (bp) for the AFLP fragments listed in Appendix B. AFLP fragments 1 -135 were generated with the selective primer combination having extensions -CCC/-AGC, 136-269 were generated with the selective primer combination having extensions -CCC/-ATA, and 270-376 were generated with the selective primer combination having the extensions -CGG/-AGC. The extensions refer to the EcoRI and MseI adapters, respectively.

AFLP Fragment	Size (bp)	AFLP Fragment	Size (bp)	AFLP Fragment	Size (bp)
1	38	46	112	91	217
2	42	47	113	92	221
3	47	48	115	93	223
4	49	49	117	94	224
5	52	50	120	95	231
6	53	51	122	96	239
7	54	52	123	97	242
8	55	53	125	98	245
9	56	54	126	99	247
10	61	55	128	100	261
11	64	56	129	101	265
12	65	57	130	102	269
13	67	58	131	103	271
14	68	59	132	104	285
15	69	60	133	105	292
16	70	61	135	106	297
17	71	62	136	107	299
18	72	63	138	108	302
19	73	64	139	109	305
20	74	65	141	110	307
21	76	66	143	111	309
22	77	67	145	112	313
23	78	68	148	113	321
24	81	69	152	114	332
25	84	70	155	115	363
26	86	71	156	116	367
27	87	72	161	117	370
28	88	73	165	118	372
29	89	74	166	119	373
30	91	75	167	120	374
31	92	76	170	121	375
32	93	77	173	122	389
33	94	78	174	123	396
34	95	79	176	124	403
35	96	80	178	125	409
36	97	81	182	126	413
37	98	82	188	127	416
38	100	83	190	128	420
39	101	84	192	129	431
40	103	85	196	130	444
41	105	86	199	131	496
42	107	87	204	132	40
43	108	88	208	133	43
44	109	89	210	134	45
45	111	90	214	135	47

AFLP Fragment	**Size (bp)**	**AFLP Fragment**	**Size (bp)**	**AFLP Fragment**	**Size (bp)**
136	50	181	137	226	225
137	52	182	140	227	227
138	55	183	144	228	233
139	56	184	148	229	234
140	60	185	150	230	238
141	63	186	151	231	245
142	64	187	153	232	248
143	65	188	154	233	251
144	67	189	155	234	254
14	69	190	156	235	258
146	71	191	159	236	261
147	73	192	160	237	262
148	74	193	162	238	263
149	75	194	164	239	270
150	77	195	167	240	276
151	80	196	169	241	280
152	82	197	170	242	282
153	84	198	171	243	286
154	88	199	173	244	290
155	90	200	175	245	327
156	94	201	177	246	330
157	95	202	179	247	332
158	96	203	180	248	342
159	99	204	182	249	345
160	100	205	183	250	348
161	101	206	184	251	349
162	103	207	185	252	350
163	104	208	187	253	352
164	105	209	188	254	359
165	106	210	190	255	361
166	108	211	194	256	368
167	110	212	195	257	380
168	113	213	198	258	382
169	114	214	200	259	411
170	116	215	203	260	487
171	117	216	205	261	490
172	120	217	206	262	38
173	121	218	208	263	39
174	122	219	211	264	40
175	125	220	213	265	42
176	127	221	214	266	43
177	128	222	215	267	46
178	130	223	217	268	52
179	132	224	220	269	56
180	135	225	223	270	57

AFLP Fragment	Size (bp)
271	58
272	59
273	60
274	64
275	66
276	67
277	68
278	69
279	71
280	72
281	73
282	74
283	75
284	76
285	77
286	78
287	79
288	81
289	83
290	86
291	88
292	91
293	93
294	95
295	99
296	101
297	103
298	104
299	105
300	107
301	108
302	109
303	112
304	113
305	114
306	116
307	117
308	118
309	119
310	120
311	122
312	124
313	126
314	131
315	132
316	135
317	136
318	138
319	142
320	144
321	146
322	147
323	148
324	152
325	156
326	158
327	161
328	168
329	172
330	175
331	177
332	180
333	184
334	185
335	188
336	190
337	192
338	195
339	198
340	199
341	202
342	203
343	209
344	210
345	212
346	214
347	216
348	218
349	219
350	221
351	222
352	223
353	225
354	227
355	228
356	234
357	236
358	240
359	242
360	248
361	259
362	261
363	266
364	268
365	284
366	286
367	332
368	335
369	343
370	349
371	357
372	360
373	362
374	363
375	364
376	448

Appendix D. ETS type 1 sequence data

This dataset contains 977 aligned base pairs for 55 external transcribed spacer (ETS) sequences used to reconstruct phylogenetic relationships in Chapter 3. ID codes for the sequences correspond to those found in Table 3.1. Gaps in the sequence are indicated by a dash (-). Uncertainties are denoted by standard IUPAC/IUB codes as follows: N=unknown; R=A/G; Y=C/T; M=A/C; K=G/T; S=C/G; W=A/T.

Bases 1-60

```
1       TCTC-AATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
10      TSTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
14      TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
15      TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
12      TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
17      TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
25.2    TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGTT
26.1    TCTCCAATAT TCTTACGTGT TCTTGAGTGT TCATCTACAG TCCTTGCA-T A-TTTGAGCT
27C.1   TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
27C.10  -CTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
28A     TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
3       TCTC-AATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
41B     TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
45C     TCTCCAATAT TCTTACGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
5.11    TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
101A    TCTCCAATAT TCTTACGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
102B.39 TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
104     TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
108C    ---------- ----ACGTGT TCTTGAGTGT TCATCTACAG TTCTCACACT ACTTTGAGCT
111A.4  ---------- -----CGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
111A.5  ---------- -----CGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
120C    ---------- ---------- ---TGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
123G    ---------- ---------- ---------- ---------- ---------- ----------
128C    ---------- ---------- ---------- ---------- ---------T A-TT-GAGCT
131     TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
30A     TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
57B.2   ---------- ----GCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGTT
58B     TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
62D.2   TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
62D.3   TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
65D.7   TCTGCA-TAT TCTTGCGTGT TCTTGAGTGT TCATCTACAA TTCTTGCA-T A-TTTGAGCT
65D.3   -CTGCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAA TTCTTGCA-T A-TTTGAGCT
67B.3   TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
69A.2   TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
69B.6   ---------- -----CGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
76A.1   TC--CAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
76E.2   TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
78A.10  TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
78A.5   TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
8       TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
80A     TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
82A.4   TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
85A     TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
86B     ---------- -----CGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
90D.11  TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
93C.2   TC-C--ATAT TCTTGCGCGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
93C.3   TCTCCAACAT TCTTGCGCGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
98A.12  ---------- ----GCGCAT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
126E.5  ---------- ---------- -------TGT TCATCTACAA TTCTTGCA-T A-TTTGAGCT
103A    TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTGCAG TTCTTGCA-T A-TTTGAGCT
50D     T--------- ---------- ---TGAGTGT TCATCTACAG TTCTTGCA-T ATTTAGAGCT
75A     TCTCCAATAT TCTTGCGTGT TCTTGAGTGT TCATCTACAG TTCTTGCA-T A-TTTGAGCT
96A     ---------- ---------- ---------- ---------- ---------- ----------
Pur-132 ---------- ---------- ---------- ---------- ---------- ----------
Pur-34  ---------- ---------- ---------- ---------- ---------- ----------
```

Bases 61-120

```
1       G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
10      G-TCTAATCT ACTT-TGGCA GGCA-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
14      G-TCTAATCT ACTT-TGGCA GGCA-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
15      G-TCTAATCT ACTT-TGGCA GGCA-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
12      G-TCTAATCT ACTT-TGGCA GGCR-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
17      G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
25      G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTAACTTAGC TTGTTCTTGG AGGTGATTAA
26      G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
27C.1   G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
27C.10  G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
28A     G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
3       G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTGACTTAGG TTGTTCTTGG AGGTGATCAA
41B     G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTAACTTAGC TTGTTCTTGG AGGTGATTAA
45C     G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
5       G-TCTAATCT ACTT-TGGCA GGCA-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
101A    G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
102B    G-GCTAATCT ACTT-TGGCA GGCA-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
104     G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
108C    ACTCTGATCT A-TT-TGGCA GGCACTTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
111A.4  G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTGATTTAGC TTGTTCTTGG AGGTGATCAA
111A.5  G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTGATTTAGC TTGTTCTTGG AGGTGATCAA
120C    G-TCTAATCT ACTT-TGGCA GGCA-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
123G    ---------- ---------- ---------- ---------- ---------- ----------
128C    G-TCTAATTT ACTT-TGGCA GGCG-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
131     G-TCTAATCT ACTT-TGGCA GGCA-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
30A     G-TCTAATCT ACTT-TGGCA GGCA-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
57B     G-TCTAATCT ACTT-TGGCA GGCG-TTTAG CTTACTTAGC TTGTTCTTGG AGGTGATCAA
58B     G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
62D.2   G-TCTAATCT ACTT-TGGCA AGCG-TTTAG CTAACTTAGC TTGTTTTTGG AGGTGATCAA
62D.3   G-TCTAATCT ACTT-TGGCA AGCG-TTTAG CTAACTTAGC TTGTTCTTGG AGGTGATCAA
65D.7   G-TCTAATTT ACTT-TGGCA GGCG-TTTGG ATGACTTAGC TTGTTCTTGG AGGTGATCAA
65D.3   G-TCTAATTT ACTT-TGGCA GGCG-TTTGG ATGACTTAGC TTGTTCTTGG AGGTGATCAA
67B     G-TCTAATCT ACTT-TGGCA GGCA-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
69A     G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
69B     G-TCTAATCT ACTT-TGGCA GGCA-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
76A     G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTGACTTAGC ATGTTCTTGG AGGTGATCAA
76E     G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
78A.10  G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTGACTTAGC ATGTTCTTGG AGGTGATCAA
78A.5   G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTGACTTAGC ATGTTCTTGG AGGTGATCAA
8       G-TCTAATCT ACTT-TGGCA GGCA-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
80A     G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTGACTTARC TTGTTCTTGG AGGTGATCAA
82A     G-TCTAATCT ACTT-TGGCA GGCA-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
85A     G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
86B     G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
90D     G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
93C.2   G-TCTAATCT ACTT-TGGCA GGCG-TTTAG CTAACTTAGC TTGTTCTTGG AGGTGATCAA
93C.3   G-TCTAATCT ACTT-TGGCA GGCG-TTTAG CTAACTTAGC TTGTTCTTGG AGGTGATCAA
98A     G-TCTAATCT ACTT-TGGCA GGCG-TTTAG CTAACTTAGC TTGTTCTTGG AGGTGATCAA
126E    G-TCTAATTT ACTT-TGGCA GGCG-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
103A    G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
50D     G-GCTGATCT ACYTCTGGCA GGCG-TTTGG CTGACTTAGC TTGTTCTTGG AGGTGATCAA
75A     G-TCTAATCT ACTT-TGGCA GGCG-TTTGG CTGACTTAGS TTGTTCTTGG AGGTGATCAA
96A     ---------- ---------- ---------- ---------- ---------- ----------
Pur-132 ---------- ---------- ---------- ---------- ---------- ----------
Pur-34  ---------- ---------- ---------- ---------- ---------- ----------
```

Bases 121-180

```
1       GTCTAAGGTC GCACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
10      GTCTAAGGTC ATACAGGTAG AGGACTTCAA AACAATCATC CCATGACAGT TGGTATCAG-
14      GTCTAAGGTC ACACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
15      GTCTAAGGTC AYACAGGTAG AGGACTTCAA AACAATCATC CCATGACAGT TGGTATCAG-
12      GTCTAAGGTC GCACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
17      GTCTAAGGTC GCACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAS-
25      GTCTAAGGTC GCACGGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATTAG-
26      GTCTAAGGTC GCACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
27C.1   GTCTAAGGTC GCACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
27C.10  GTCTAAGGTC GCACAGGTAG AGGACTTCAT AACAGTCATC CCATGACAGT TGGTATCAG-
28A     GTCTAAGGTC ATACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
3       GTCTAAGGTC GCACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
41B     GTCTAAGGTC GCACGGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
45C     GTCTAAGGTC GCACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
5       GTCTAAGGTC ATACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
101A    GTCTAAGGTC GCACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
102B    GTCTAAGGTC ATACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
104     GTCTAAGGTC GCACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
108C    GTCTAAGGTC GCACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
111A.4  GTCTAAGGTC GCACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
111A.5  GTCTAAGGTC GCACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
120C    GTCTAAGGTC ATACAGGTAG AGGACTTCAA AACAATCATC CCATGACAGT TGGTATCAG-
123G    -TCTAAGGTC ATACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
128C    GTCTAAGATC GCATGGGTAG AGGACTTCAA AATAGTCATC CCATGACAGT TGGTATCAGT
131     GTCTAAGGTC ATACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
30A     GTCTAAGGTC ATACAGGTAG AGGACTTCAA AACAATCATC CCATGACAGT TGGTATCAG-
57B     ATCTAAGGTT GCACGGGTAG AGGACTTCAA AATAGTTATC CCATGACAGT TGGTATCAG-
58B     GTCTAAGGTC GCACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
62D.2   GTCTAAGATC GCATGGGTAG AGGACTTCAA AATAGTCATC CCATGACAGT TGGTATCAG-
62D.3   GTCTAAGGTC GCACGGGTAG AGGACTTCAA AATAGTTATC CCATGACAGT TGATATCAG-
65D.7   GTCTAAGATC GCATGGGTAG AGGACTTCAA AATAGTCATC CCATGACAGT TGGTATCAG-
65D.3   GTCTAAGATC GCATGGGTAG AGGACTTCAA AATAGTCATC CCATGACAGT TGGTATCAG-
67B     GTCTAAGGTC ATACAGGTAG AGGACTTCAA AATAATCATC CCATGACAGT TGGTATCAG-
69A     GTCTAAGGTC ATACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
69B     GTCTAAGGTC ATACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
76A     GTCTAAGGTC GCACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
76E     GTCTAAGGTC GCACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
78A.10  GTCTAAGGTC GCACAGGTAG AGGACTTCAA AACAGTTATC CCATGACAGT TGGTATCAG-
78A.5   GTCTAAGGTC GCACAGGTAG AGGACTTCAA AACAGTTATC CCATGACAGT TGGTATCAG-
8       GTCTAAGGTC ATACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
80A     GTCTAAGGTC GCACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAA-
82A     GTCTAAGGTC ATACAGGTAG AGGACTTCAA AACAATCATC CCATGACAGT TGGTATCAG-
85A     GTCTAAGGTC GCACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
86B     GTCTAAGGTC GCACAGGTAG AGGACTTCAA AATAGTCATC CCATGACAGT TGGTATCAG-
90D     GTCTAAGGTC GCACAGGTAG AGGACTTCAA AATAGTCATC CCATGACAGT TGGTATCAG-
93C.2   ATCTAAGGTC GCACGGGTAG AGGACTTCAA AATAGTTATC CCATGACAGT TGGTATCAG-
93C.3   ATCTAAGGTC GCACGGGTAG AGGACTTCAA AATAGTTATC CCATGACAGT TGGTATCAG-
98A     ATCTAAGGTC GCACGGGTAG AGGACTTCAA AATAGTTATC CCATGACAGT TGGTATCAG-
126E    GTCTAAGATC GCATGGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATTAG-
103A    GTCTAAGGTC GCACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
50D     GTCTAAGGTC GCACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
75A     GTCTAAGGTC GCACAGGTAG AGGACTTCAA AACAGTCATC CCATGACAGT TGGTATCAG-
96A     ---------- ---------- ---------- -----TCATC CCATGACAGT TGGTCTCAC-
Pur-132 ---------- ---------- ---------- ---------- ---------- ----------
Pur-34  ---------- ---------- ---------- ---------- ---------- ----------
```

Bases 181-240

```
1       AGCAGGATCG ATATGTCAGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
10      AGCAGGATCG ATATGTCGGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
14      AGCAGGATCG ATATGTCAGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
15      AGCAGGATCG ATATGTCAGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
12      AGCAGGATCG ATATGTCAGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
17      AGCAGGATCG ATATGTCAGA CGAGGTGGTA ACTATAGTGG ACATGACAAA TGTCACCGTG
25      AGCAGGATTG ATATGTCGGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
26      AGCAGGATCG ATATGTCGGA CAAGGTGGTA ACTATGGTGG ACATAACAAA TGTCACCGTG
27C.1   AACAG--TCG ATATGTCAGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
27C.10  AGTAGGATCG ATATGTCAGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
28A     AGCAGGATCG ATATGTCGGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
3       AGCAGGATCG ATATGTCAGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
41B     AGCAGGATTG ATATGTCGGA CGAGGTGGTA ACTATGGTAG ACATGACAAA TGTCACCGTG
45C     AGCAGGATCG ATATGTCGGA CAAGGTGGTA ACTATGGTGG ACATAACAAA TGTCACCGTG
5       AGCAGGATCG ATATGTCGGA CGAGGTGGTA ACTATGGCGG ACATGACAAA TGTCACCGTG
101A    AGCAGGATCG ATATGTCGGA CAAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
102B    AGCAGGATCG ATATGTCGGA CGAGGTGGTA ACTATGGTGG ACATGACAAA CGTCACCGTG
104     AGCAGGATCG ATATGTCGGA CAAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
108C    AGCAGGATCG ATATGTCGGA CAAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
111A.4  AGCAGGATCG ATATGTCGGA CAAGGTGGTA ACCATGGTGG ACATGACAAA TGTCACCGTG
111A.5  AGCAGGATCG ATATGTCGGA CAAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACTGTG
120C    AGCAGGATCG ATATGTCGGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
123G    AGCAGGATCG ATATGTCGGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
128C    AGTAGGATCG ATATGTCGGA CGAGGTGGTA ACTATGGTGG ACATGAAAAA TGTCACCGTG
131     AGCAGGATCG ATATGTCGGA CGAGGGGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
30A     AGCAGGATCG ATATGTCGGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
57B     AGCAGGATCG ATATGTCAGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
58B     AGTAGGATCG ATATGTCAGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
62D.2   AGTAGGATCG ATATGTCGGA CGAGGTGGTA ACTATGGTGG ACATGAAAAA TGTCACCGTG
62D.3   AGCAGGATCG ATATGTCGGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
65D.7   AGTAGGATCG ATATGTCGGA CGAGGTGGTA ACTATGGTGG ACATGAAAAA TGTCACCGTG
65D.3   AGTAGGATCG ATATGTCGGA CGAGGTGGTA ACTATGGTGG ACATGAAAAA TGTCACCGTG
67B     AGCAGGATCG ATATGTCGGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
69A     AGCAGGATCG ATATGTCGGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
69B     AGCAGGATCG ATATGTCGGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
76A     AGCAGGATCG ATATGTCAGA CGAGGTGGTA ACTATGCTGG ACATGACAAA TGTCACCGTG
76E     AGCAGGATCG ATATGTCAGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
78A.10  AGCAGGATCG ATATGTCAGA CGAGGTGGTA ACTATGCTGG ACATGACAAA TGTCACCGTG
78A.5   AGCAGGATCG ATATGTCAGA CGAGGTGGTA ACTATGCTGG ACATGACAAA TGTCACCGTG
8       AGCAGGATCG ATATGTCGGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
80A     AGCAGGATCG ATATGTCAGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
82A     AGCAGGATCG ATATGTCGGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
85A     AGTAGGATCG ATATGTCAGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
86B     AGCAGGGTCG ATATGTGAGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
90D     AGCAGGATCG ATATGTGAGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
93C.2   AGCAGGATCG ATATGTCGGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
93C.3   AGCAGGATCG ATATGTCGGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
98A     AGCAGGATCG ATATGTCGGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
126E    AGTAGGATCG ATATGTCGGA CGAGGTGGTA ACTATGGTGG ACATGAAAAA TGTCACCGTG
103A    AGCAGGATCG ATATGTCGGA CAAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
50D     AGCAGGATCG ATATGTCGGA CAAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
75A     AGCAGGATCG ATATGTCAGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
96A     AGCAGGATCG ATATGTGAGA CGAGGTGGTA ACTATGGTGG ACATGACAAA TGTCACCGTG
Pur-132 ---------- ---------- ---------- ---------- ---------- ----------
Pur-34  ---------- ---------- ---------- ---------- ---------- ----------
```

Bases 241-300

```
1       GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCCG TGCGATGAAA
10      GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCAG TGCGATGAAA
14      GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TGCGATGAAA
15      GAGCGTGGTG GCAAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCAG TGCGATGAAA
12      GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TGCGATGAAA
17      GAGCGTGGTG GCAAAAAGTC CAAGAGGGGA AGGTCAAAGA ATATTGTCGG TGCGATGAAA
25      GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TGCGATGAAA
26      GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TACGATGAAA
27C.1   GAGCGTGGTG GCGAAAAGTC CAAGAGAGGA AGGTCAAAGG ATATTGTCGG TGCGATGAAA
27C.10  GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TGCGATGAAA
28A     GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TGCGATGAAA
3       GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCCG TGCGATGAAA
41B     GAGCGTGGTG GYGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TGCGATGAAA
45C     GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TACGATGAAA
5       GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TGCGATGAAA
101A    GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TACGATGAAA
102B    GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA GGGTCAAAGG ATATTGTCGG TGCGATGAAA
104     GAGCGTGGTG GCGAAAAGTC CAAGAGTGGA AGGTCAAAGG ATATTGTCGG TGCGATGAAA
108C    GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TACGATGAAA
111A.4  GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TGCGATGAAA
111A.5  GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TGTGATGAAA
120C    GAGCGTGGTG GCAAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCAG TGCGATGAAA
123G    GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCAG TGCGATGAAA
128C    GAGCATGGTG GCGAAAAGTT CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TGCGATGAAA
131     GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TGCGATGAAA
30A     GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCAG TGCGATGAAA
57B     GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG AAATTGTCGG TGCGATGAAA
58B     GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCCG TGCGATGAAA
62D.2   GAGCATGGTG GCGAAAAGTT CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TGCGATGAAA
62D.3   GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TGCGATGAAA
65D.7   GAGCATGGTG GCGAAAAGTT CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TGCGATGAAA
65D.3   GAGCATGGTG GCGAAAAGTT CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TGCGATGAAA
67B     GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCAG TGCGATGAAA
69A     GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TGCGATGAAA
69B     GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TGCGATGAAA
76A     GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCCG TGCGATGAAA
76E     GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCCG TGCGATGAAA
78A.10  GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCCG TGCGATGAAA
78A.5   GAGCGTGGTG GCGAAAAGTC CTAGAGGGGA AGGTCAAAGG ATATTGTCCG TGCGATGAAA
8       GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TGCGATGAAA
80A     GASCGTGGTG GCTAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCCG TGCGATGAAA
82A     GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TGCGATGAAA
85A     GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCCG TGCGATGAAA
86B     GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TGCGATGAAA
90D     GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCAG TGCGATGAAA
93C.2   GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG AAATTGTCGG TGCGATGAAA
93C.3   GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG AAATTGTCGG TGCGATGAAA
98A     GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG AAATTGTCGG TGCGATGAAA
126E    GAGCATGGTG GCGAAAAGTT CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TGCGATGAAA
103A    GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TGCGATGAAA
50D     GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TGCGATGAAA
75A     GAGCGTGGTG GCAAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCCG TGCGATGAAA
96A     GAGCGTGGTG GCGAAAAGTC CAAGAGGGGA AGGTCAAAGG ATATTGTCGG TGCGATGAAA
Pur-132 ---------- ---------- ---------- ---------- ---------- ----------
Pur-34  ---------- ---------- ---------- ---------- ---------- ----------
```

Bases 301-360

```
1        GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
10       GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
14       GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
15       GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
12       GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
17       GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TAAGGAAGAT
25       GGACGGCTTT TGAGGATGGA GAAGGCCATG GCCTATTTTG AGACTCAGCT TGAGGAAGAT
26       GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
27C.1    GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
27C.10   GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
28A      GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
3        GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
41B      GGACGACTTT TGAGGATGGA GAAGGCCATG GCCTATTTTG AGACTCAGCT TGAGGAAGAT
45C      GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
5        GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
101A     GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
102B     GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
104      GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
108C     GGGTGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
111A.4   GGACGGCTTT TGAGGATGGA GAAGGCCATG GCCTATTTTG AGACTCAGTT TGAGGAAGAT
111A.5   GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGAAAGAT
120C     GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
123G     GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
128C     GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TTAGGAAGAT
131      GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
30A      GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
57B      GGATGACTTT CGAGGATGAA GAAGGCCATG GCCGATTTTG AGACTCAGCT TAAGAAAGAT
58B      GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
62D.2    GGACGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TTAGGAAGAT
62D.3    GGACGGCTTT CAAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TAAGAAAGAT
65D.7    GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TTAGGAAGAT
65D.3    GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TTAGGAAGAT
67B      GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
69A      GGATTGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
69B      GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
76A      GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
76E      GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
78A.10   GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
78A.5    GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
8        GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
80A      GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
82A      GGACGGCTTT TGAGGATGGA GAAGGCCATG GCCTATTTTG AGACTCAGCT TGAGGAAGAT
85A      GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
86B      GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCTATTTTG AGACTCAGCT TGAGGAAGAT
90D      GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
93C.2    GGACGACTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TAAGAAAGAT
93C.3    GGACGACTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TAAGAAAGAT
98A      GGACGACTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TAAGAAAGAT
126E     GGACGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAACT TTAGGAAGAT
103A     GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
50D      GGACGGCTTT TGAGGATGGA GAAGGCCATG GCCTATTTTG AGACTCAGCT TGAGGAAGAT
75A      GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
96A      GGATGGCTTT CGAGGATGGA GAAGGCCATG GCCGATTTTG AGACTCAGCT TGAGGAAGAT
Pur-132  ---------- ---------- ---------- ---------- ---------- ----------
Pur-34   ---------- ---------- ---------- ---------- ---------- ----------
```

Bases 361-420

```
1       GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
10      GTTCCTGAGA TGGAGACCCT TAAGGCAATG GTAGAAGATC TTAAGGGAGA GTTCCAAGGA
14      GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
15      GTTCCTGAGA TGGAGACCCT TAAGGCAATG GTAGAAGATC TTAAGGGAGA GTTCCAAGGA
12      GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
17      GTTCCTGAGA TGGGGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCTAAGGA
25      GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
26      GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
27C.1   GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAAAAGATC TTAAGGGAGA GTTCCAAGGA
27C.10  GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
28A     GTTCCTGAGA TGGAGACCCT TAAGGCAATG GYAGAAGATC TTAAGGGAGA GTTCCAAGGA
3       GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
41B     GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
45C     GTTCCTGAGA TGGACACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
5       GTTCCTGAGA TGGAGACCCT TAAGGCAATG GTAGAAGATC TTAAGGGAGA GTTCCAAGGA
101A    GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
102B    GTTCCTGAGA TGGAGACCCT TAAGGCAATG GTAGAAGATC TTAAGGGAGA GTTCCAAGGA
104     GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
108C    GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
111A.4  GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
111A.5  GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTC------
120C    GTTCCTGAGA TGGAGACCCT TAAGGCAATG GTAGAAGATC TTAAGGGAGA GTTCCAAGGA
123G    GTTCCTGAGA TGGAGACCCT TAAGGCAATG GTAGAAGATC TTAAGGGAGA GTTCCAAGGA
128C    GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
131     GTTCCTGAGA TGGAGACCCT TAAGGCAATG GYAGAAGATC TTAAGGGAGA GTTCCAAGGA
30A     GTTCCTGAGA TGGAGACCCT TAAGGCAATG GTAGAAGATC TTAAGGGAGA GTTCCAAGGA
57B     GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
58B     GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
62D.2   GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
62D.3   GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
65D.7   GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
65D.3   GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
67B     GTTCCTGAGA TGGAGACCCT TAAGGCAATG GTAGAAGATC TTAAGGGAGA GTTCCAAGGA
69A     GTTCCTGAGA TGGAGACCCT TAAGGCAATG GTAGAAGATC TTAAGGGAGA GTTCCAAGGA
69B     GTTCCTGAGA TGGAGACCCT TAAGGCAATG GTAGATGATC TTAAGGGAGA GTTCCAAGGA
76A     GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
76E     GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
78A.10  GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
78A.5   GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
8       GTTCCTGAGA TGGAGACCCT TAAGGCAATG GTAGAAGATC TTAAGGGAGA GTTCCAAGGA
80A     GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
82A     GTTCCTGAGA TGGAGACCCT TAAGGCAATG GTAGAAGATC TTAAGGGAGA GTTCCAAGGA
85A     GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
86B     GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
90D     GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
93C.2   GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
93C.3   GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
98A     GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
126E    GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
103A    GTTCCTGAGA TGGTGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
50D     GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
75A     GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
96A     GTTCCTGAGA TGGAGACCCT TAAGGCAATG GCAGAAGATC TTAAGGGAGA GTTCCAAGGA
Pur-132 ---------- ---------- ---------- ---------- ---------- ----------
Pur-34  ---------- ---------- ---------- ---------- ---------- ----------
```

Bases 421-480

```
1       GCACTCACCT TTCTGGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
10      GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
14      GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
15      GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
12      GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
17      GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
25      GCACTCACCT CTCTCGTAGA CACAATCCAA ATGGTAGTTA AGGAGTAGCT TGACCAGTAC
26      GTACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTG AGGAGCAGCT TGACCAGTAC
27C.1   GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
27C.10  GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
28A     GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
3       GCACTCACCT TTCTGGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
41B     GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
45C     GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
5       GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
101A    GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
102B    GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
104     GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
108C    GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
111A.4  GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
111A.5  ---------- ---------- ---------- --------TA AGGAGCAGCT TGACCAGTAC
120C    GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
123G    GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
128C    GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
131     GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
30A     GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
57B     GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGTT TGACCAGTAC
58B     GCACTCACCT TTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
62D.2   GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
62D.3   GCACTCACCA CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGTT TGACCAGTAC
65D.7   GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
65D.3   GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
67B     GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
69A     GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGGGCAGCT TGACCAGTAC
69B     GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
76A     GCACTCACCT TTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
76E     GCACTCACCT TTCTGGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
78A.10  GCACTCACCT TTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
78A.5   GCACTCACCT TTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
8       GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
80A     GCACTCACCT TTCTGGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
82A     GCACTCGCCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
85A     GCACTCACCT TTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
86B     GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
90D     GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
93C.2   GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGTT TGACCAGTAC
93C.3   GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGTT TGATCAGTAC
98A     GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGTT TGACCAGTAC
126E    GCACTCACCT CTGTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
103A    GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
50D     GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
75A     GCACTCACCT TTCTGGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
96A     GCACTCACCT CTCTCGCAGA CACAATCCAA ATGGAAGTTA AGGAGCAGCT TGACCAGTAC
Pur-132 ---------- ---------- ---------- ---------- ---------- ----------
Pur-34  ---------- ---------- ---------- ---------- ---------- ----------
```

Bases 481-540

```
1       CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
10      CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
14      CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
15      CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
12      CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
17      CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
25      CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
26      CTTGCTGAGT TCACCATAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
27C.1   CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
27C.10  CTTGCTGAGT TCACCACAAT TCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
28A     CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
3       CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGG GATTGGGCCT
41B     CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
45C     CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
5       CTTGCTGAGT TCACCACAAT CCGTTGAGGA GATTGAGTGA CATGAAAGGA GATTGGGCCT
101A    CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGAGCCT
102B    CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
104     CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
108C    CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
111A.4  CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
111A.5  CTTTCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGAGCCT
120C    CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
123G    CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
128C    CTTGTTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
131     CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
30A     CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
57B     CTTGCTGAGT TCACCATAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
58B     CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
62D.2   CTTGTTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
62D.3   CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
65D.7   CTTGTTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
65D.3   CTTGTTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
67B     CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
69A     CTTGTTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
69B     CTTGTTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
76A     CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
76E     CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
78A.10  CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
78A.5   CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
8       CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
80A     CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
82A     CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
85A     CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
86B     CTTGCTGAGT TCACCACAAT CTGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
90D     CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
93C.2   CTTGCTGAGT TCACCATAAT CCGT-GAGGA GAT-GAGTGA CATAAAAGGA GATTGGGCCT
93C.3   CTTGCTGAGT TCACCATAAT CCGT-GAGGA GAT-GAGTGA CATAAAAGGA GATTGGGCCT
98A     CTTGCTGAGT TCACCATAAT CCGT-GAGGA GAT-GAGTGA CATAAAAGGA GATTGGGCCT
126E    CTTGTTGAGT TCACCACAAT CAGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
103A    CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
50D     CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
75A     CTTGCTGAGT TCACCACAAT YCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
96A     CTTGCTGAGT TCACCACAAT CCGT-GAGGA GAT-GAGTGA CATGAAAGGA GATTGGGCCT
Pur-132 ---------- ---------- ---------- ---------- ---------- ----------
Pur-34  ---------- ---------- ---------- ---------- ---------- ----------
```

Bases 541-600

```
1       TATGCAAAAA GGCAGTGCTC AACAA-TTCC CATGGGC--- ----GC--GG AGATGGATGC
10      TATGCAAAAA GGCAGTGCTC AA----TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
14      TATGCAAAAA GGCAGTGCTC AACATCTTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
15      TATGCAAAAA GGCAGTGCTC AA----TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
12      TATGCAAAAA GGCAGTGCTC AACA-CTTCC CATGGGCCTA AGGAGCC-GA AGATGGATGC
17      TATGCAAAAA GGCAGTGCTC AACAA-TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
25      TATGCAAAAA GGCAGTGCTC AACAA-TTCC CATGGACCTA GGGAGCC-GA AGATGGATGC
26      CATGCAAAAA GGCAGTGCTC GACAA-TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
27C.1   TATGCAAAAA GGCAGTGCTC AAAAA-TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
27C.10  TATGCAAAAA GGCAGTGCTC AACAA-TTCC CATGGGCCTA GAGAGCC-GA AGATGGATGC
28A     TATGCAAAAA GGCAGTGCTC AA----TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
3       TATGCAAAAA GGCAGTGAT- AACAA-TTCC CATGGGC--- ----GC--GG AGATGGATGC
41B     TATGCAAAAA GGCAGTGCTC AACAA-TTCC CATGG-CCTA GGGAACC-GA AGATGGATGC
45C     CATGCAAAAA GGCAGTGCTC AACAA-TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
5       TATGCAAAAA GGCAGTGCTC AA---ATTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
101A    CATGCAAAAA GGCAGTGCTC AACAA-TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
102B    TATGCAAAAA GGCAGTGCTC AA----TTCC CATGGGCCTA AGGAGCT-GA AGATGGATGC
104     TATGCAAAAA GGCAGTGCTC AACAA-TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
108C    CATGCAAAAA GGCAGTGCTC AACAA-TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
111A.4  TATGCAAAAA GGCAGTGCTC AACAA-TTCC CATGGGCCTA GGGAGCC-GA AGATAGATGC
111A.5  TATGCAAAAA GGCAGTGCTC AACAA-TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
120C    TATGCAAAAA GGCAGTGCTC AA----TTCC CATGGGCCTA AGGAGCT-GA AGATGGATGC
123G    TATGCAAAAA GGCAGTGCTC AA----TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
128C    TATGCAAAAC GGCAGTGCTC AACAA-TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
131     TATGCAAAAA GGCAGTGCTC AA----TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
30A     TATGCAAAAA GGCAGTGCTC AA----TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
57B     TATGCAAAAA GGCAGTGCTC AACAA-TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
58B     TATGCAAAAA GGCAGTGCTC AACAA-TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
62D.2   TATGCAAAAC GGCAGTGCTC AACAA-TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
62D.3   TATGCAAAAA GGCAGTGCTC AACAA-TTCC CATGGGCCTA GGGAGCT-GA AGATGGATAC
65D.7   TATGCAAAAC GGCAGTGCTC AACAA-TTCC CATGGGCCTA GGGAGCCCGA AGATGGATGC
65D.3   TATGCAAAAC GGCAGTGCTC AACAA-TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
67B     TATGCAAAAA GGCAGTGCTC AA----TTCC CATGGGCCTA GAGAGCC-GA AGATGGATGC
69A     TATGCAAAAA GGCAGTGCTC AA----TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
69B     TATGCAAAAA GGCAGTGCTC AA----TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
76A     TATGCAAAAA GGCAGTGCTC AACAA-TTCC CATGGGC--- ----GC--GG AGATGGATGC
76E     TATGTAAAAA GGCAGTGCTC AACAA-TTCC CATGGGC--- ----GC--GG AGATGGATGC
78A.10  TATGCAAAAA GGCAGTGCTC AACAA-TTCC CATGGGC--- ----GC--GG AGATGGATGC
78A.5   TATGCAAAAA GGCAGTGCTC AACAA-TTCC CATGGGC--- ----GC--GG AGATGGATGC
8       TATGCAAAAA GGCAGTGCTC AA----TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
80A     TATGCAAAAA GGCAGTGCTC AACAA-TTCC CATGGGC--- ----GC--GG AGATGGATGC
82A     TATGCAAAAA GGCGGTGCTC AACAA-TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
85A     TATGCAAAAA GGCAGTGCTC AACAA-TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
86B     TATGCAAAAA GGCAGTGCTC ACCAA-TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
90D     TATGCAAAAA GGCAGTGCTC ACCAA-TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
93C.2   TATGCAAAAA GGCAGTGCTC AACAA-TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
93C.3   TATGCAAAAA GGCAGTGCTC AACAA-TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
98A     TATGCAAAAA GGCAGTGCTC AACAA-TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
126E    TATGCAAAAC GGCAGTGTTC AACAA-TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
103A    TATGCAAAAA GGCAGTGCTC AACAA-TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
50D     TATGCAAAAA GGCAGTGCTC AACAA-TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
75A     TATGCAAAAA GGCAGTGCTC AACAA-TTCC CATGGGC--- -----C---- AGATGGATGC
96A     TATGCAAAAA GGCAGTGCTC ACCAA-TTCC CATGGGCCTA GGGAGCC-GA AGATGGATGC
Pur-132 ---------- ---------- ---------- ---------- ---------- ----------
Pur-34  ---------- ---------- ---------- ---------- ---------- ----------
```

Bases 601-660

```
1       TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGYTG
10      TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
14      TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
15      TTT--CAAGC CYA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
12      TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
17      TTT--CAAGC CTA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
25      TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
26      TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
27C.1   TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
27C.10  TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
28A     TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
3       TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGKGGAGTTG
41B     TTTTTCAAGC CCAACCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
45C     TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
5       TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
101A    TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
102B    TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CAAGGAGCTG
104     TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
108C    TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
111A.4  TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
111A.5  TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAAGAGCTG
120C    TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CAAGGAGCTG
123G    TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGGC GATGGCACCT CGAGGAGCTG
128C    TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGACACCC CGAGGAGCTG
131     TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
30A     TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
57B     TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
58B     TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATAGCACCC CGAGGAGCTG
62D.2   TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGACACCC CGAGGAGCTG
62D.3   TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
65D.7   TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGACACCC CGAGGAGCTG
65D.3   TCT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGACACCC CGAGGAGCTG
67B     TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGCC GATGGCACCC CGAGGAGTTG
69A     TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
69B     TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GACGGCACCC CGAGGAGCTG
76A     TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
76E     TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
78A.10  TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
78A.5   TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
8       TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
80A     TTT--CAAGC CCA-CCCATG CAATGCACAA ACTTGTCGAC GATGGCACCC CGAGGAGCTG
82A     TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAT GATGGCACCC CGAGGAGCTG
85A     TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATAGCACCC CGAGGAGCTG
86B     TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGTTG
90D     TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
93C.2   TTT--CAAGC CCA-CCTATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
93C.3   TTT--CAAGC CCA-CCTATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
98A     TTT--CAAGC CCA-CCTATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
126E    TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGTTG
103A    TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
50D     TTT--CAAGC CCA-CCCATG CAATGTACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTA
75A     TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
96A     TTT--CAAGC CCA-CCCATG CAATGCACAA GCTTGTCGAC GATGGCACCC CGAGGAGCTG
Pur-132 ---------- ---------- ---------- ---------- ---------- ----------
Pur-34  ---------- ---------- ---------- ---------- ---------- ----------
```

Bases 661-720

```
1        CGATGATTTC CAAATGTTGC CTTCGAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
10       CAATGATTCC CAAATGTTGC CTTCGAATTT GTT--GCCTT T-GAAATGGT GGATGAGAGA
14       CGATGATTCC CAAATGTTGC CTTTGAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
15       CAATGATTCC CAAATGTTGC CTTCGAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
12       CGATGATTTC CAAATGTTGC CTTTGMATTT GAW--GCCTT C-GAAATGGT GGATGAGAGA
17       CGATGATTCC CAAATGTTGC CTTCGAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
25       CGATGATTCC CAAATGTTGC CTTCGAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
26       CGATGATTCC CAAATGTTGC CTTCGGATTT GTT--GCCTT C-AAAATGGT GGATAAGCGG
27C.1    CGATGATTCC CAAATGTTGC CTTCGAA--- ---------- -----ATGGT GGATGAGAGA
27C.10   CGATGATTCC CAAATGTTGC CTTCGAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
28A      CAATGATTCC CAAATGTTGC CTTCGAATTT GTT--GCCTT T-GAAATGGT GGATGAGAGA
3        CGATGATTTC CAAATGTTGC CTTAGAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
41B      TGATGATTCC CAAATGTTGC CTTCGAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
45C      CGATGATTCC CAAATGTTGC CTTCGGATTT GTT--GCCTT C-AAAATGGT GGATAAGCGG
5.11     CAATGATTCC CAAATGTTGC CTTCGAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
101A     CGATGATTCC CAAATGTTGC CTTCGGATTT GTT--GCCTT C-AAAATGGT GGATAAGCGG
102B     CAATGATTCC CAAATGTTGC CTTCGAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
104      CAATGATTCC CAAATGTTGC CTTCGGATTT GTT--GCCTT C-AAAATGGT GGATAAGCGG
108C     CGATGATTCC CAAATGTTGC CTTCGAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGG
111A.4   CGATGATTCC CAAATGTTGC CTTCGGATTT GTT--GCCTT C-AAAATGGT GGATAAGCGG
111A.5   CGATGATTCC CAAATGTTGC CTTCGAATTT GTT--GCCTT C-GAAATGGT GGATTAGAGA
120C     CAATGATTCC CAAATGTTGC CTTCGAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
123G     CAATGATTCC CAAATGTTGC CTTCGAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
128C     CGATGATTCC TAAATGTTTC CTTCGGATTT CTT--GCCTT C-AAAATGGT GGATGAGCGG
131      CAATGATTCC CAAATGTTGC CTTCGAATTT GTT--GCCTT T-GAAATGGT GGATGAGAGA
30A      CAATGATTCC CAAATGTTGC CTTCGAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
57B      TGATGATTCC CAAATGTTGC CTTCAGATTT GTT--GCCTT C-GAAATGGT GGATGAGAGG
58B      CGATGATTCC CAAATGTTGC CTTCGAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
62D.2    CGATGTTTCC CAAATGTTTC CTTCAGATTT CTT--GCCTT C-AAAATGGT GGATGAGCGG
62D.3    CGATGATTCC CAAATGTGGC CTTCAGATTT GTT--GCCTT C-GAAATGGT GGATGAGAGG
65D.7    CGATGATTCC TAAATATTTC CTTCGGATTT CTT--GCCTT C-AAAATGGT GGATGAGCGG
65D.3    CGATGATTCC TAAATATTTC CTTCGGATTT CTT--GCCTT C-AAAATGGT GGATGAGCGG
67B      CGATGATTCC CAAATGTTGC CTTCGAATTT GTCTCGCCTT CCGAAATGGT GCATGAGAGA
69A      CGATGATTCC CAAATGTTGC CTTCGAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
69B      CGATGATTCC CAAATGTTGC CTTCGAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
76A      CGATGATTCC CAAATGTTGC CTTCAAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
76E      CGATGATTCC CAAATGTTGC CTTCAAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
78A.10   CGATGATTCC CAAATGTTGC CTTCAAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
78A.5    CGATGATTCC CAAATGTTGC CTTCAAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
8        CAATGATTCC CAAATGTTGC CTTCGAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
80A      CGATGATTTC CAAATGTTGC CTTCGAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
82A      CGATGATTCC CAAATGTTGC CTTCGGATTT GTT--GCCTT C-AAAATGGT GGATAAGCGG
85A      CGATGATTCC CAAATGTTGC CTTCGAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
86B      CGATGATTCC CAAATGTTGC CTTTGAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
90D      CGATGATTCC CAAATGTTGC CTTTGAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
93C.2    TGATGATTCC CAAATGTTGC CTTCAGATTT GTT--GCCTT C-GAAATGGT GGATGAGAGG
93C.3    TGATGATTCC CAAATGTTGC CTTCAGATTT GTT--GCCTT C-GAAATGGT GGATGAGAGG
98A      TGATGATTCC CAAATGTTGC CTTCAGATTT GTT--GCCTT C-GAAATGGT GGATGAGAGG
126E     CGATGATTCC CAAATGTTGC CTTCGGATTT CTT--GCCTT C-AAAATGGT GGATGAGCGG
103A     CAATGATTCC CAAATGTTGC CTTCGGATTT GTT--GCCTT C-AAAATGGT GGATAAGCGG
50D      CGATGATTCC CAAATGTTGC CTTCGGATTT GTT--GCCTT C-AAAATGGT GGATAAGAGG
75A      CGATGATTYC CAAATGTTGC CTTCGAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
96A      CGATGATTGC CAAATGTTGC CTTTGAATTT GTT--GCCTT C-GAAATGGT GGATGAGAGA
Pur-132  ---------- ---------- ---------- ---------- ---------- ----------
Pur-34   ---------- ---------- -------TG- GTT--GCCTT C-AGAATGGT GGATGTGCGG
```

Bases 721-780

```
1       CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAGTGCA CAAGCTTGTC
10      CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAGTGCA CAAGCTTGTT
14      CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAGTGCA CAAGCTTGTT
15      CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAGTGCA CAAGCTTGTT
12      CGAATTGGAT TGCCTATGCC TGMTTCTGCG GAGCCCTTAC GGGCAGTGCA CAAGSTTGTC
17      CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAGTGCA CAAGCTTGTC
25      CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GAGCAATGCA CAAGCTTGTC
26      CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAATGCA CAAGCTTGTC
27C.1   CGAATTGGAT TGCCTATGCT TGCTTGTGCG AAGCCCTTGC GGGCAGTGCA CAAGCTTGTC
27C.10  CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAGTGCA CAAGCTTGTC
28A     CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAGTGCA CAAGCTTGTT
3       CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAGTGCA CAAGCTTGTC
41B     CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAATGCA CAAGCTTGTC
45C     CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAATGCA CAAGCTTGTC
5       CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAGTGCA CAAGCTTGTT
101A    CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAATGCA CAAGCTTGTC
102B    CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAGTGCA CAAGCTTGTT
104     CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAATGCA CAAGCTTGTC
108C    CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAGTGCA CAAGCTTGTC
111A.4  CGAATTGGAT TGCCTATGCT TGCTTGTGCG AAGCCCTTAC GGGCAATGCA CAAGCTTGTC
111A.5  CGAATTGGAT TGCGTATGGT TGGTTGTGCG GAGCCCTTAC GGGCAATGCA CAAGCTTGTC
120C    CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAGTGCA CAAGCTTGTT
123G    CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAGTGCA CAAGCTTGTT
128C    CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAATGCA CAAGCTTGTC
131     CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAGTGCA CAAGCTTGTT
30A     CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAT GGGCAGTGCA CAAGCTTGTT
57B     CAAATTGGAT TGCCTATGCT TGCCTATGCG GAGCCCTTAC GGGCAATGCA CAAGCTTGTC
58B     TGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAGTGCA CAAGCTTGTC
62D.2   CGAATTGGAT TGCCTATGCT TGCTTATGCG GAGCCCTTAC GGGCAATGCA CAAGCTTGTC
62D.3   CAAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAATGCA CAAGCTTGTC
65D.7   CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAATGCA CAAGCTTGTC
65D.3   CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAATGCA CAAGCTTGTC
67B     CGAATTGGAT TGCCTATGCT TGGTTGTGCG GAGCCATTAC GGGCAGTGCA CAAGCTTGTC
69A     CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAGTGTA CAAGCTTGTT
69B     CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAGTGTA CAAGCTTGTT
76A     CGAATTGGAT TGCCTATGCT TGCTTGTACG GAGCCCTTAC GGGCAGTGCA CAAGCTTGTC
76E     CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAGTGCA CAAGCTTGTC
78A.10  CGAATTGTAT TGCCTATGCT TGCTTGTACG GAGCCCTTAC GGGCAGTGCA CAAGCTTGTC
78A.5   CGAATTGTAT TGCCTATGCT TGGTTGTACG GAGCCCTTAC TGGCAGTGCA CAAGCTTGTC
8       CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAGTGCA CAAGCTTGTT
80A     CGAATTGGAT TGCCTATGCT TGCTTGTACG GAGCCCTTAC GGGCAGTGCA CAAGCTTGTC
82A     CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAACGCG CAAGCTTGTC
85A     TGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAGTGCA CAAGCTTGTC
86B     CCAATTGGAT TGCCTATGTT TGCTTGTGCG GAGCCCTTAC GGGCAGTGCA CAAGCTTGTC
90D     CCAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAGTGCA CAAGCTTGTC
93C.2   CAAATTGGAT TGCCTATGCT TGCTTATGCA GAGCCCTTAT GGGTAATGCA CAAGCTTGTC
93C.3   CAAATTGGAT TGCCTATGCT TGCTTATGCA GAGCCCTTAC GGGTAATGCA CAAGCTTGTC
98A     CAAATTGGAT TGCCTATGCT TGCTTATGCA GAGCCCTTAC GGGTAATGCA CAAGCTTGTC
126E    CAAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAATGCA CAAGCTTGTC
103A    CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAATGCA CAAGCTTGTC
50D     CGAATTGGAT TGCCTATGCA TGCTTGTGCG GAGCCCTTAC GGGCAATGCA CAAGCTTGTC
75A     CGAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAGTGCA CAAGCTTGTC
96A     CCAATTGGAT TGCCTATGCT TGCTTGTGCG GAGCCCTTAC GGGCAGTGCA CAAGCTTGTC
Pur-132 ---------- ---------- ---------- ---------- ---------- ----------
Pur-34  CGAATCGGAT TGCTTATGCT TGCTTGGGCG GAGCCTTTGT GGGCAATGCA CAAGCATGTC
```

Bases 781-840

```
1        GACGATCGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CRGGTGCTTG
10       GACGATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
14       GACGATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
15       GACGATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTRCTTG
12       GACGATGGCA CCCCGAGGAG TTGCGATGAT TCCCAAATGA CCCTCASTCT CGGGTGCTTC
17       GACGATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCTTCACTCT CGGGCAMTTG
25       GACGATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
26       GACAATGGCA CCCCGAGGAG CTGCGATGAT TCCTAAATGA CCCTCACTCT TGGGTGCTTG
27C.1    GACGATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTTACTCT CGGGTGCTTG
27C.10   GACGATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
28A      GACGATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
3        GACGATTGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
41B      GACGATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
45C      GACAATGGCA CCCCGAGGAG CTGCGATGAT TCCTAAATGA CCCTCACTCT TGGGTGCTTG
5        GACGATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTGT CGGGTGCTTG
101A     GACAATGGCA CCCCGAGGAG CTGCGATGAT TCCTAAATGA CCCTCACTCT TGGGTGCTTG
102B     GACGATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
104      GACAATGGCA CCCTGAGGAG CTGCGATGAT TCCTAAATGA CCCTCACTCT TGGGTGCTTG
108C     GACGATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCTTCACTCT CGGGTGCTTG
111A.4   GACAATGGCA CCCCGAGGAG CTGCGATGAT TCCTAAATGA CCCTCACTCT TGGGTGCTTG
111A.5   GACGATGGCA CCCCGAGGAG CTGCGACGAT TCCCAAATGA CCCTCACTCT TGGGTGCTTG
120C     GCCCATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
123G     GACGATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
128C     GACGATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
131      GACGATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
30A      GACGATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
57B      GACGATGGCA CCCCGAGGAG CTGTGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
58B      GACGATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTTACTCT CGGGTGCTTG
62D.2    GACGATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
62D.3    GATGATGGCA CCCCGAGGAG CTACGATGAT TCTCAAATGA CCCTCACTCT CGGGTGCTTG
65D.7    GACGATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
65D.3    GACGATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
67B      GACGATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
69A      GACGATGGCG CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
69B      GACGATGGCG CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
76A      GACGATCGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
76E      GACGATCGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
78A.10   GACGATCGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
78A.5    GACGATCGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
8        GACGATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
80A      GACGATCGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
82A      GACAATGGCA CCCCGAGGAG CTGCGATGAT TCCTAAATGA CCCTCACTCT TGGGTGCTTG
85A      GACGATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTTACTCT CGGGTGCTTG
86B      GACGATGGCA CCCCGAAGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
90D      GACGATGGCA CCCCGAAGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
93C.2    GACGATGGCA CCCCGAGGAG CTACGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
93C.3    GACGGTGGCA CCCCGAGGAG CTACGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
98A      GACGATGGCA CCCCGAGGAG CTACGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
126E     GACTATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
103A     GACAATGGCA CCCTGAGGAG CTGCGATGAT TCCTAAATGA CCCTCACTCT TGGGTGCTTG
50D      GACAATGGCA CCCCAAGGAG CTGCGATGAT TCCTAAATGA CCCTCACTCT TGGGTGCTTG
75A      GACGATGGCA CCCCGAGGAG CTGCGATGAT TCCCAAATGA CCCTCACTCT CGGGTGCTTG
96A      GACGATGGCA CCCCGAAGAG ---------- ---------- ---------- ----------
Pur-132 ---------- ---------- ---------- ----A---GA CTAGCATTGT TAGGTTTTTG
Pur-34   GACGATGGCA CCTCGAGCAG CCGCGATGAC ATCCA---GA CCTGCATTGT TGGGTTTTTG
```

Bases 841-900

```
1        TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
10       TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
14       TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
15       TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
12       TGAAGGTGCA ATGGGTGCTA TTCGATGTCG TATATGCRGA ACATACAAAG GGCGCGGGGC
17       TGAAGGTGCA TTGGKTGCTA CTCGATMTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
25       TGAAGGTGCA TTGGTTGCTA CTCGATCTCT CGTATGCAGA ACATACAAAG GGTGCGGGGG
26       TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
27C.1    TGAAGGTGCA TTGGTTGCTA CTTGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
27C.10   TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
28A      TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
3        TGAAGGTGCA TTGGTTGGTA ATCGGTCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
41B      TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TGTATGTAGA ACATACAAAG GGTGCGGGGG
45C      TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
5        TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
101A     TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
102B     TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
104      TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
108C     TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGTGG
111A.4   TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATACGGA ACATACAAAG GGCGCGGGGG
111A.5   TGAAGGTGCA TTGGTTGCTA CTCGATCTCT CATATGCAGA ACATACAAAG GGTACGGGGG
120C     TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
123G     TGAAGGTGCA TTGGTTGGTA CTCGATCTTT TTTATGCGGA ACATACAAAG GGSGGGGGGG
128C     TGAAGGTGCA TTGGTTCCTA CTCGATCTCT CGTATGCGGA ACATACA--- ----------
131      TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
30A      TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
57B      TGAAGGTGCA TTGGTTGCTA CTCGATCTCT CGTATGCGGA ACATACAAAG GGCGAGGGGG
58B      TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
62D.2    TGAAGGTGCA TTGGTTCCTA CTCGATCTCT CGTATGCGGA ACATACAAAG GGCGCGGGGG
62D.3    TGAAGGTGCA TTGGTTGCTA CTCGATCTCT CGTATGCCGA ACATACAAAG GGCGCGGGGG
65D.7    TGAAGGTGCA TTGGTTCCTA CTCGATCTCT CGTATGCGGA ACATACAAAG GGCGCGGGGG
65D.3    TGAAGGTGCA TTGGTTCCTA CTCGATCTCT CGTATGCGGA ACATACAAAG GGCGCGGGGG
67B      TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
69A      TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
69B      TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
76A      TGAAGGTGCA TTGGTTGCCA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
76E      TGAAGGTGCA TTAGTTGCTA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
78A.10   TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
78A.5    TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
8        TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
80A      TGAAGGTGCA TTGGTTGCTA CTCGATCTTT TCTATGCGGA ACATACAAAG GGCGCGGGGG
82A      TGAAGGTGCA TTGGTTGCTA CTCGATCTCT CGTATGCAGA ACATACAAAG GGTGCGGGGG
85A      TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
86B      TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCAGA ACATACAAAG GGCGCGGGGG
90D      TGAAGGTGCA TTGGTTGCTA CTCGATCTCT TCTATGCAGA ACATACAAAG GGCGCGGGGG
93C.2    TGAAGGTGCA TTGGTTGCTA CTCGATCTCT CGTATGCGGA ACATACAAAG GGCGAGGGGG
93C.3    TGAAGGTGCA TTGGTTGCTA CTCGATCTCT CGTATGCGGA ACATACAAAG GGCGAGGGGG
98A      TGAAGGTGCA TTGGTTGCTA CTCGATCTCT CGTATGCGGA ACATACAAAG GGCGAGGGGG
126E     TGAAGGTGCA TTGGTTCCTA CTCGATCTCT CGTATGCGGA ACATACAAAG GGCGCGAGGG
103A     TGAAGGTGCA TTGGTTGCTA CTCAATCTCT TCTATGCGGA ACATACAAAG GGCGCGGGAG
50D      TGAAGGTGCA TTGGTTGCTA YTGGATCTCT TATATGCGGA ACATACAAAG GGTGCGGGGG
75A      TGAAGGTGCA TTGGTTGCTA CTCGRTCTCT TCTATGCGGA ACATACAAAG GGCGCGGGGG
96A      ---------- ---------- ---------- ---------- ---------- ----------
Pur-132  TGTCGGTCCA TGTGTTGTCA CTCGATCTCT CGGATGCGGA ACATGTTGCG GGTGT-AGGT
Pur-34   TGTCGGTCCA CGTGTTGTCA CTCGATCTCT CGGATGCAGA ACATGTTGCG GGCGC-AGGG
```

Bases 901-960

```
1       TCTTCAGCCC CC-GACGTC- AGATGCAAAA -GCATTT-GT CCCTATACAC GAACGACAGT
10      TCTTCAGCCC CY-GAMGTC- AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
14      TCTTCAGCCC CC-GACGTC- AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
15      TCTTCAGCCC CC-GACGTC- AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
12      TTTTCAGCCC CC-GACGTC- AGATGCAAAW -GCATTTTGT CCCTAKACAC GAACGACAGT
17      TCTTTAGCCC CC-GACGTC- AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGGCAGT
25      TCTTCAGCCC TC-GACGTCC -GATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
26      TCTTCAGCCC CC-GACGTC- AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
27C.1   TCTTCAGCCC CC-GACGTC- AGATGCAAAA -GCATTTTGT CCCTATAC-- ----------
27C.10  TCTTCAGCCC CC-GACGTC- AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
28A     TCTTCAGCCC CT-GAMGTC- AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
3       TCTTCAGCCC CC-GACGTC- AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
41B     TCTTCAGCCC CC-GACGTCC -AATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
45C     TCTTCAGCCC CC-GACGTC- AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
5       TCTTCAGCCC CC-GACGTC- AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
101A    TCTTCAGCCC CC-GACGTC- AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
102B    TCCTCAGCCC CC-GACGTC- AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
104     TCTTCAGCCC CT-GACGTC- AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
108C    TCTTCAGCCC CC-GACGTC- AGAAGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
111A.4  TCTTCCGCCC CC-GACGTC- AGACGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
111A.5  TCTTCAGCCC CC-GACGTCC -GATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
120C    TCTTCAGCCC CC-GACGTC- AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
123G    T--------- ---------- ---------- ---------- ---------- ----------
128C    ---------- ---------- ---------- ---------- ---------- ----------
131     TCTTCAGCCC CT-GAAGTC- AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
30A     TCTTCAGCCC CC-GACGTC- AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
57B     TCTTCAGCCC TC-GATGTCC -GATGCAAAA C-CATTTTGT CCCTATACAC GAATGATAGT
58B     TCTTCAGCCC CC-GACGTC- AGATGTAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
62D.2   TCTTCAGCCC CC-GACGTCC -GATGCAAAA C-CATTTTGT CCCTACACAA GAACGACAGT
62D.3   TTTTCAGCCC CC-GACGTCC -GATGCAAAA C-CATTTTGT CCCTATACAC GAATGACAGT
65D.7   TCTTCAGCCC CC-GACGTCC -GATGCAAAA C-CATTTTGT CCCTACACAA GAACGACAGT
65D.3   TCTTCAGCCC CC-GACGTCC -GATGCGAAA C-CATTTTGT CCCTACACAA GAACGACAGT
67B     TCTTCAGCCC CA-GACGTC- AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
69A     TCTTCAGCCC CC-GACGTC- AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
69B     TCTTCAGCCC CC-GACGTC- AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
76A     TCTTCAGCCC CN-TACGTC- AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
76E     TCTTCAGCCC CC-GACGTC- AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
78A.10  TCTTCAGCCC CC-TACGTC- AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
78A.5   TCTTCAGCCC CC-TACGTC- AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
8       TCTTCAGCCC CC-GACGTC- AGAT-CAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
80A     TCTTCAGCCT CC-GACGTC- AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
82A     TCTTCAGCCC TC-GACGTCC -GATGCAAAA -GCATTTTGC CCCTATACAC GAACGACAGT
85A     TCTTCAGCCC CC-GACGTC- AGATGTAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
86B     TCTTCAGCCC CC-GACGTC- AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
90D     TCTTCAGCCC CC-GACGTC- AAATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
93C.2   TCTTCAGCCC TC-GATGTCC -GATGCAGAA C-CATTTTGT CCCTATACAC GAATGACAGT
93C.3   TCTTCAGCCC TC-GATGTCC -GATGCAAAA C-CATTTTGT CCCTATACAC GAATGACAGT
98A     TCTTCAGCCC TC-GATGTCC -GATGCAAAA C-CATTTTGT CCCTATACAC GAATGACAGT
126E    TCTTCAGCCC CC-GACGTCC -GATGCAAAA AGCATTTTGT CCCTATACAC GAACGACAGT
103A    TCTTCAGCCC CT-GACGTCC AGATGCAAAA -GCATTTTGT CCCTATACAC GAACGACAGT
50D     TCTTCAGCCC CM-GACGTC- AGAWGCAAGA -GCATTTTGT CCCTATACA- -AACA----T
75A     TCTTCAGCCC CC-GACGTC- AGATGCAAAA -GCATTTTGT CCCTATMCAC GAACGACAGT
96A     ---------- ---------- ---------- ---------- ---------- ----------
Pur-132 TCTTCGGCCC TTTGACGTCC AAA-ACAAAG CGTTTTTCG- --CTAGACAC GAATGACTGT
Pur-34  TCTTCGGCCC CTTGACGTCC AAA-ACAAAG CGTTTTYCG- --CTAGACAC GAACGACAGT
```

Bases 961-977

```
1       CGCGCGC-GT TGCAGGC
10      CGCGCGC-GT TGCAGGC
14      CGCGCGC-GT TGCAGGC
15      CGCGCGC-GT TGCAGGC
12      TGCGCGCTGT TGCACGG
17      CGCGCGC-GT TGCAGGC
25      CGCGCGC-GT TGCAGGC
26      CGCGCGC-GT TGCAGGC
27C.1   ---------- -------
27C.10  CGCGCGC-GT TGCAGGC
28A     SGCGCGC-GT TGCAGGC
3       CGCGCGC-GT TGCAGGC
41B     CGCGCGC-GT TGCAGGC
45C     CGCGCGC-GT TGCAGGC
5       CGCGCGC-GT TGCAGGC
101A    CGCGCGC-GT TGCAGGC
102B    CGCGCGC-GT TGCAGGC
104     CGCGCGC-GT TGCAGGC
108C    CGCGCGC-GT TGCAGGC
111A.4  CGCGCGC-GT TGCAGGC
111A.5  CGCGCGC-GT TGCAGGC
120C    CGCGCGC-GT TGCAGGC
123G    ---------- -------
128C    ---------- -------
131     GGCGCGC-GT TGCAGGC
30A     CGCGCGC-GT TGCAGGC
57B     TGTGCGC-GT TGCAGGC
58B     CGCGCGC-GT TGCAGGC
62D.2   CGCGCGC-GT TGCAGGC
62D.3   CGTGCGC-GC TGTAGGC
65D.7   CGCGCGC-GT TGCAGGC
65D.3   CGCGCGC-GT TGCAGGC
67B     CGCGCGC-GT TGCAGGC
69A     CGCGCGC-GT TGCAGGC
69B     CGCGCGC-GT TGCAGGC
76A     CGCGCGC-GT TGCAGGC
76E     CGCGCGC-GT TGCAGGC
78A.10  CGCGCGC-GT TGCAGGC
78A.5   CGCGCGC-GT TGCAGGC
8       CGCGCGC-GT TGCAGGC
80A     CGCGCGC-GT TGCAGGC
82A     CGCGCGC-GT TGCAGGC
85A     CGCGCGC-GT TGCAGGC
86B     CGCGCGC-GT TGCAGGT
90D     CGCGCGC-GT TGCAGGT
93C.2   TGTGCAC-GT TGCAGGC
93C.3   TGTGCAC-GT TGCAGGC
98A     TGTGCAC-GT TGCAGGC
126E    CGCGCGC-GT TGCAGGC
103A    CGCGCGC-A- -------
50D     -A-------T T------
75A     CGCGCGC-GT TGCAGGC
96A     ---------- -------
Pur-132 CGCGTAC-GT TGCGGGC
Pur-34  CGCGTAC-GT TGCGGGC
```

Appendix E. Chloroplast DNA sequence data

This dataset contains the aligned sequences for the *trn*S and *trn*G intergenic spacer and the *rpl*20 and *rps*12 intergenis spacer of the chloroplast genome used in Chapter 4. The sequences of the two regions have been combined for a total of 1337 bp. Sequence positions 1 - 580 bp represent the *trn*S and *trn*G intergenic spacer and positions 581 - 1337 represents the *rps*12-*rpl*20 intergenic spacer. 14 indels of 2 bp or more are coded in poitions 1338-1351. ID codes for the sequences correspond to those found in Table 4.1. Gaps in the sequence are indicated by a dash (-). Uncertainties are denoted by standard IUPAC/IUB codes as follows: N=unknown; R=A/G; Y=C/T; M=A/C; K=G/T; S=C/G; W=A/T.

```
Bases 1-60
1       ---------- ---------- ---------- ------CATT CATTTCATTT T--AAAAACT
10      TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
101A    TGG-AACTTT TGTCGAACAA GGAGATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
102B    TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
103A    -------TTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
104     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
108C    TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
111A    TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT TT-AAAAACT
12      TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
120C    TGG-AACTTT TGTCGAACAA GGAGATAGGA AAAGA-CATT CATTTCATTT T--AAAAACT
121A    TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGGACATT CATTTCATTT ---AAAAACT
123E    TGGCAACTTT TGTCGAACAA GAA-ATAGGA AAAGA-CATT CATTTCATTT T--AAAAACT
123G    TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
126E    TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
128C    TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
129     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTAATTT A--AATAACT
131     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
132     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTAATTT A--AAAAACT
134A    T--------- ---------- ------AGGA AAAGAGCATT CATTTCATTT T--AAAAACT
14      TGGCAACTTT TGTCGAACAA GAA-ATAGGG AAAGAGCATT CATTTCATTT T--AAAAACT
15      TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
25      TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTAATTT T--AAAAACT
26      TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AATAACT
27C     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
28A     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT A--AAAAACT
3       TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T-AAAAAACT
30A     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
36C     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTAATTT A--AAAAACT
41B     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
45C     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT TTAAAAAACT
5       -------TTT TGTCGAACAA GAA-ATAGGA AA-GAGCAT- CATTTCATTT T--AAAAACT
50D     -------TTT TGTCGAACAA GAG-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
57B     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T-AAAAAACT
58B     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT TTAAAAAACT
62D     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
65D     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
67B     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
69A     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T-AAAAAACT
69B     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
75A     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
76A     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
76E     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AATAACT
77B     -------TTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
8       TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
80A     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
81A     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
82A     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
84F     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
86B     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
93C     TGG-AACTTT TGTCGAACAA GAA--TAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
98A     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
C.pri   TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
S.3     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
S.7     TGG-AACTTT TGTCGAACAA GAA-ATAGGA AAAGAGCATT CATTTCATTT T--AAAAACT
```

```
Bases 61-120
1       AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
10      AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
101A    AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
102B    AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
103A    AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
104     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
108C    AAT--AC--- AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
111A    AAT--AC--- AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
12      AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
120C    AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
121A    AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
123E    AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
123G    AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
126E    AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
128C    AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAA-CTATTT
129     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTTGAT AAAACGATTT
131     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGTTCTCGAT ACAACTATTT
132     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACGATTT
134A    AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACGATTT
14      AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
15      AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGAC-TATA CGT-CTCGAT AAAACTATTT
25      AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
26      AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTTACCTATA CGT-CTCGAT AAAACGATTT
27C     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
28A     AAATAAC--- AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
3       AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
30A     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAA-CTATTT
36C     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACGATTT
41B     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
45C     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
5       AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
50D     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
57B     AAATAAACAA AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
58B     AAATAAC--- AGTTTTTTTG CTGCAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
62D     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
65D     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
67B     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACGATTT
69A     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACGATTT
69B     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACGATTT
75A     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACGATTT
76A     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACGATTT
76E     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACGATTT
77B     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACGATTT
8       AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
80A     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
81A     AAATAAACAA -GTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACGATTT
82A     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACGATTT
84F     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
86B     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
93C     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
98A     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
C.pri   AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACGATTT
S.3     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
S.7     AAATAAC--A AGTTTTTTTG CTGGAGGTTT TTGACCTATA CGT-CTCGAT AAAACTATTT
```

```
Bases 121-180
1       AAG-CCATAA C-ATAAAAAT GGATGGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
10      AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
101A    AAG-CCATAA C-ATAAAAGT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
102B    AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
103A    AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
104     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
108C    CAG-CCATCA C-ATAAAAAT GGATTGTGCA AGCATCCACA GTTCCTTTTT TT-CT---TA
111A    AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
12      AAG-CCATCA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TTTCT---TA
120C    AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
121A    AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
123E    AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
123G    AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
126E    AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TTTCT---TA
128C    AAGGCCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
129     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TTTTT--TTA
131     AAG-CCATCA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
132     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TTCTTATTTA
134A    AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
14      AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
15      AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
25      AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
26      AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT T--CT---TA
27C     AAGCCCATCA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
28A     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TTTCT---TA
3       AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
30A     AAG-CCATAA CCATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
36C     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TTCTTATTTA
41B     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
45C     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
5       AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
50D     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
57B     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
58B     AAG-CCATCA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
62D     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
65D     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
67B     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
69A     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
69B     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
75A     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
76A     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
76E     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
77B     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
8       AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
80A     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
81A     AAG-CCATAA C-ATAAAAAG GGATGGGGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
82A     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
84F     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
86B     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
93C     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
98A     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
C.pri   AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
S.3     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
S.7     AAG-CCATAA C-ATAAAAAT GGATTGTGCA AGAATCCACA GTTCCTTTTT TT-CT---TA
```

```
Bases 181-240
1       TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
10      TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
101A    TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
102B    TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
103A    TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
104     TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
108C    TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
111A    TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
12      TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
120C    TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
121A    TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
123E    TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
123G    TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
126E    TTTA----TT TACTTT---- --------AA CTGGAATAAA AAAGAATAAA CTAAATAAAT
128C    TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
129     TT-ACTTTTT TACTTT---- ---------A CTGGAATAAA AAA-AATAAA CTAAATTAAT
131     TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATATA CTAAATAAAT
132     TTTACTTTTT -ACTTC---- ---------A CTGGAATAAA AA-GAATAAA CTAAATAAAT
134A    TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAA--
14      TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
15      TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
25      TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
26      TTTA----TT TACTTT---- ---------A CTGGAATAAA AA-GAATAAA CTAAATAAAT
27C     TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
28A     TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
3       TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
30A     TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
36C     TTTACTTTTT TACTTTTTAT TTCACTGGAA TAGGAATAAA AA-GAATAAA TTAAATAAAT
41B     TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
45C     TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
5       TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
50D     TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
57B     TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
58B     TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
62D     TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
65D     TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
67B     TTTA----TT TACTTT---- --------A- CTGGAATAAA AA-GAATAAA CTAAATAAAT
69A     TTTA----TT TACTTT---- --------A- CTGGAATAAA AA-GAATAAA CTAAATAAAT
69B     TTTA----TT TACTTT---- --------A- CTGGAATAAA AA-GAATAAA CTAAATAAAT
75A     TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
76A     TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAA--
76E     TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAA--
77B     TTTA----TT TACTTT---- --------A- CTGGAATAAA AA-GAATAAA CTAAATAAAT
8       TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
80A     TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
81A     TTTA----TT TACTTT---- --------AA CTGGGAATAA AAAGAATAAA CTAAATAA--
82A     TTTA----TT TACTTT---- --------A- CTGGAATAAA AA-GAATAAA CTAAATAAAT
84F     TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
86B     TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
93C     TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
98A     TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
C.pri   TTTA----TT TACTTT---- --------A- CTGGAATAAA AA-GAATAAA CTAAATAAAT
S.3     TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
S.7     TTTA----TT TACTTT---- --------AA CTGGAATAAA AA-GAATAAA CTAAATAAAT
```

```
Bases 241-300
1        AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
10       AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
101A     AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGA TTCGATATCA TTTGATTCTC
102B     AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
103A     AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
104      TAGAAATTAA GATAAGAAA- GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
108C     AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
111A     AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGA TTCGATATCA TTTGATTCTC
12       AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
120C     AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGA TTCGATATCA TTTGATTCTA
121A     AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGA TTCGATATCA TTTGATTCTC
123E     AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGA TTCGATATCA TTTGATTCTA
123G     AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGA TTCGATATCA TTTGATTCTA
126E     AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGA TTCGATATCA TTTGATTCTA
128C     AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGA TTCGATATCA TTTGATTCTC
129      AATAAATTAA AATAAAAAAA AATAAAAAAT GACACTTAAA TTCGATATCA TTTGATTTTA
131      AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
132      AATAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGA TTCGATATCA TTTGATTCTA
134A     --GAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGA TTCGATATCA TTTGATTCTA
14       AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
15       AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
25       AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
26       AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGA TTCGATATCA TTTGATTCTA
27C      AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
28A      AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
3        AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
30A      AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
36C      AATAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGA TTCGATATCA TTTGATTCTA
41B      AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
45C      AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
5        AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
50D      AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
57B      AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
58B      AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
62D      AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGA TTCGATATCA TTTGATTCTC
65D      AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGA TTCGATATCA TTTGATTCTC
67B      AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGA TTCGATATCA TTTGATTCTA
69A      AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGA TTCGATATCA TTTGATTCTA
69B      AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGA TTCGATATCA TTTGATTCTA
75A      AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGA TTCGATATCA TTTGATTCTA
76A      --GAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGA TTCGATATCA TTTGATTCTA
76E      --GAAATTAA GATTAGAAAA GATAAGAAAT GACACTTTGA TTCGATATCA TTTGATTCTA
77B      AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGA TTCGATATCA TTTGATTCTA
8        AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
80A      AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
81A      --GAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGA TTCGATATCA TTTGATTCTA
82A      AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGA TTCGATATCA TTTGATTCTA
84F      AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
86B      AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
93C      AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
98A      AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGA TTCGATATCA TTTGATTCTA
C.pri    AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGA TTCGATATCA TTTGATTCTA
S.3      AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
S.7      AAGAAATTAA GATAAGAAAA GATAAGAAAT GACACTTTGG TTCGATATCA TTTGATTCTA
```

```
Bases 301-360
1       TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
10      TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
101A    TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
102B    TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
103A    TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
104     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
108C    TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
111A    TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
12      TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
120C    TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
121A    TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
123E    TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
123G    TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
126E    TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
128C    TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
129     TATCATAAAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCAAGTATTT
131     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
132     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCAAGTATTT
134A    TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
14      TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
15      TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
25      TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
26      TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
27C     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
28A     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
3       TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
30A     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
36C     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCAAGTATTT
41B     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
45C     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
5       TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
50D     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
57B     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
58B     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
62D     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
65D     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
67B     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
69A     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
69B     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
75A     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
76A     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
76E     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
77B     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
8       TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
80A     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
81A     TATCATAGAA ATAGTTCTTT ATTATTTTTT CCAATCGCAA ----TAACAA GCACGTATTT
82A     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
84F     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
86B     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
93C     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
98A     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
C.pri   TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAAC GCAATAACAA GCACGTATTT
S.3     TATCATAGGA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ---TAACCAA GCACGTATT-
S.7     TATCATAGAA ATAGTTCTTT ATTATTTTTT CAATCGCAA- ----TAACAA GCACGTATTT
```

```
Bases 361-420
1       GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
10      GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
101A    GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
102B    GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
103A    GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
104     GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
108C    GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
111A    GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
12      GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGC---A
120C    GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
121A    GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
123E    GAGTGTTCAA GTTAAATTAT TAAATTAAAT AAATCATTTG ATTTACGATT CGTTGG---A
123G    GAGTGTTCAA GTTAAATTAT TAAATTAAAT AAATCATTTG ATTTACGATT CGTTGG---A
126E    GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
128C    GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT C-TTGG---A
129     TAGTTTTCAA -----ATTA- -------AAT AAATCATTTA ATTTATGATT CGTTGGAACA
131     GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CCTTGG---A
132     TAGTGTTCAA ATTAAATTA- -------AAT AAATCATTTT ATTTACGATT CGTTGG---A
134A    GAGTGTTCAA GTTAAATTAA -ATATTAAAT AAATCATTTT ATTTATGATT CGTTGG---A
14      GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
15      GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CCTTGG---A
25      GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
26      GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTT ATTTACGATT CGTTGG---A
27C     GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGC--CA
28A     GAGTGTTCAA GTTAAATT-- ------AAAT AAATCATTTG ATTTACGATT CGTTGG---A
3       GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
30A     GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
36C     TAGTGTTCAA ATTAAATTA- -------AAA AAATCATTTA ATTTACGATT CGTTGG---A
41B     GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
45C     GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
5       GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
50D     GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
57B     GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CCCTCC---A
58B     GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
62D     GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
65D     GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
67B     GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTT ATTTACGATT CGTTGG---A
69A     GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTT ATTTACGATT CGTTGG---A
69B     GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTT ATTTACGATT CGTTGG---A
75A     GAGTGTTCAA GTTAAATTAA -ATATTAAAT AAATCATTTT ATTTACGATT CGTTGG---A
76A     GAGTGTTCAA GTTAAATTAA -ATATTAAAT AAATCATTTT ATTTACGATT CGTTGG---A
76E     GAGTGTTCAA GTTAAATTAA -ATATTAAAT AAATCATTTT ATTTACGATT CGTTGG---A
77B     GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTT ATTTACGATT CGTTGG---A
8       GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
80A     GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
81A     GAGTGTTCAA GTTAAATTAA -ATATTAAAT AAATCATTTT ATTTACGATT CGTTGG---A
82A     GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTT ATTTACGATT CGTTGG---A
84F     GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
86B     GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
93C     GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
98A     GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
C.pri   GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTT ATTTACGATT CGTTGG---A
S.3     GAGTGTTCAA GGTTAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
S.7     GAGTGTTCAA GTTAAATTA- -------AAT AAATCATTTG ATTTACGATT CGTTGG---A
```

```
Bases 421-480
1       ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
10      ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
101A    ACAAAAAGGG C-GG-TCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
102B    ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
103A    ACAAAAAGGG C-GG-GCCCG GCC-TGGTCC AGTACTTAGG CCGGGCCGTG GAACTAAAAA
104     ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
108C    ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
111A    ACAAAAAGGG C-GG-TCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
12      ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
120C    ACAAAAAGGG C-GG-TCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
121A    ACAAAAAAGG C-GG-TCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
123E    ACAAAAAGGG C-GG-TCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
123G    ACAAAAAGGG C-GG-TCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
126E    ACAAAAAGGG CCGGTCCCCG GCC-TGGTCC AGTATTTAG- -CCGGCCGTG GAACTAAAAA
128C    ACAAAAAGGG C-GG-TCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
129     AAAAAAAAAG GGCGGGCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
131     ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG CAAGTAAAAA
132     ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTATTTAG- CCGGGCCGTG GAACTAAAAA
134A    ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
14      ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
15      ACAAAAAGGG C-GG-GCCCG CCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
25      ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
26      ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
27C     ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
28A     ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
3       ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
30A     ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
36C     ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTATTTAG- CCGGGCCGTG GAACTAAAAA
41B     ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
45C     ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
5       ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
50D     ACAAAAA-GG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAG
57B     ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
58B     ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
62D     ACAAAAAGGG C-GG-TCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
65D     ACAAAAAGGG C-GG-TCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
67B     ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
69A     ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
69B     ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
75A     ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
76A     ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
76E     ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
77B     ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
8       ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
80A     ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
81A     ACAAAAAGGG CCGGGCCCCG GCCCTGGTCC AGTATTTAGC CCGGGCCGTG GAACTAAAAA
82A     ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
84F     ACAAAAAGGC G-GG--CCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
86B     ACAAAAAGGG C-GG--CCCG GCC-TGGTC- AGTACTTAG- CCCGGCCGTG GAACTAAAAA
93C     ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
98A     ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
C.pri   ACAAAAAGGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
S.3     ACAAAAAGGG C-GG--CCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
S.7     ACAAAAAAGG C-GG-GCCCG GCC-TGGTC- AGTACTTAG- CCGGGCCGTG GAACTAAAAA
```

```
Bases 481-540
1       GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
10      GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTG TCAAGCCTTA
101A    GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
102B    GCCCCTTCCG GACGAAATCC AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
103A    GCCCCTTCCG GACGAAATCC AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
104     GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
108C    GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
111A    GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
12      GCCCCTTCCG GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
120C    GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
121A    GCCCCTTC-G -ACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
123E    GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
123G    GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
126E    GCCCCTTC-G GACGAAATCC AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
128C    GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
129     GCCCCTTC-G GACGAAATCA AAAAA----G AGTGTTATGA AGGG-CTTTT TCAAGCCTTA
131     GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
132     GCCCCTTC-G GACGAAGTCA AAAAAGAATG AATATTATGA AGGG-CTTTT TCAAGCCTTA
134A    GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
14      GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTG TCAAGCCTTA
15      GCCCCTTCCG GACGAAATC- AAAAAA---G AATATTATGA AGGGGCTTTT TCCACCCTTA
25      GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAG-CTTA
26      GCCC-TTC-G GGCGAAATCA AAAAA----G AATATTAT-A AGGG-CTTTT TCAAGC-TTA
27C     GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
28A     GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
3       GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
30A     GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
36C     GCCCCTTC-G GACGAAGTCA AAAAATAATG AATATTATGA AGGG-CTTTT TCAAGCCTTA
41B     GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTT- TCAAGCCTTA
45C     GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTT- TCAAGCCTTA
5       GCCC-TTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGC-TTA
50D     GCCCCTTC-G -ACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
57B     GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
58B     GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
62D     GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
65D     GCCC-TTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
67B     GTCCCTTC-G GGCGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
69A     GTCCCTTC-G GGCGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
69B     GTCC-TTC-G GGCGAAATC- ATAAAA---G AATATTATGA AGGG-CTTTT TCCACCCTTA
75A     GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
76A     GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
76E     GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
77B     GTCCC-TTCG GGCGAAATC- AAAAAA---G AATATTATGA AGGG-CTTGT TCAAGCCTTA
8       GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
80A     GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
81A     GCCCCTTCCG GACGAAATCC AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
82A     GTCCCCTTCG GGCGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
84F     GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
86B     GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
93C     GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
98A     GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
C.pri   GTCCCTTC-G GGCGAAATC- AAAAAA---T AATATTATGA AGGG-CTTTT TCAAGCCTTA
S.3     GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
S.7     GCCCCTTC-G GACGAAATC- AAAAAA---G AATATTATGA AGGG-CTTTT TCAAGCCTTA
```

```
Bases 541-580
1        TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
10       TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
101A     TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTAT---
102B     TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
103A     TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
104      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
108C     TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATTTA
111A     TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATTTA
12       TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
120C     TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTAT---
121A     TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
123E     TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
123G     TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
126E     TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
128C     TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
129      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAG------
131      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
132      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTCTT--
134A     TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
14       TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
15       TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
25       TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
26       TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
27C      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
28A      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
3        TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
30A      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
36C      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTCTT--
41B      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
45C      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
5        TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
50D      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
57B      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTAT---
58B      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
62D      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
65D      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
67B      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
69A      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
69B      TTTTTTATAA TGTAACTT-A TAATGTAATC ATAG------
75A      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
76A      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
76E      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
77B      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
8        TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
80A      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
81A      TTTTTTATAA TGTAACTT-A TAATGTAA-C AT--------
82A      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
84F      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
86B      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
93C      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
98A      TTTTTTATAA TGTAACTTTA TAATGTAA-C ATAGTATT--
C.pri    TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAG------
S.3      TTTTTTATAA TGTAACTT-A TAATGTAA-C ATAGTATT--
S.7      TTTTTTATAA TGTAA--TTA TAATGTAA-C ATAGTATT--
```

```
Bases 581-640
1       AAAACAATTG -ATCCAGTAT CAACG-ATCA G-TACCAG-- GAATAGGA-T AGAATGGAAG
10      AAAACAATTG -ATCCAGTAT CAACG-ATCA G-TACCAGT- GAATAGGA-T AGAATGGAAG
101A    AAAACAATTA -ATCCA-TAT CAACG-ATCA G-TACCAGT- GAATAGGA-T AGAATGGAAG
102B    AAAACAATTG -ATCCAGTAT CAACG-ATCA G-TACCAGT- GAATAGGA-T AGAATGGAAG
103A    ---------- -------TAT CAA-GGATCA G-TACCAGT- GAATAGGA-T AGAA-GGAAG
104     ---------- -------TAT CAA-GGATCA G-TACCAGT- GAATAGGA-T AGAA-GGAAG
108C    AAAACAATTG -ATCCAGTAT CAACGGATCA G-TACCAGT- GAATAGGA-T AGAA-GGAAG
111A    AAAACAATTA -ATCCA-TAT CAA-GGATCA G-TACCAGT- GAATAGGA-T AGAA-GGAAG
12      AAAACAATTG -ATCCAGTAT CAACG-ATCA G-TACCAGT- GAATAGGA-T AGAATGGAAG
120C    AAAACAATTA -ATCCAGTAT CAACGGATCA G-TACCAGT- GAATAGGA-T AGAATGGAAG
121A    AAAACAATTA -ATCCAGTAT CAA-GGATCA G-TACCAGT- GAATAGGA-T AGAA-GGAAG
123E    AAAACAATTA -ATCCAGTAT CAACG-ATCA G-TACCAG-- GAATAGGA-T AGAATGGAAG
123G    AAAACAATTA -ATCCA-TAT CAA-GGATCA G-TACCAG-- GAATAGGA-T AGAA-GGAAG
126E    AAAACA-TTA -ATCCAGTAT CAA-GGATCA G-TACCAGT- GAATAGGA-T AGAA-GGAAG
128C    AAAACAATTA -ATCCAGTAT CAACGGATCA G-TACCAG-- GAATAGGA-T AGAA-GGAAG
129     AAAACAATTG -ATCCAGTAT CAA-GGATCA G-TACCAG-- GAATAGGA-T AGAA-GGAAG
131     AAAACAATTG TATCCAGTAT CAACGGATCA G-TACCAGTG GAATAGGA-T AGAA-GGAAG
132     AAAACAATTG -ATCCAGTAT CAA-GGATCA G-TACCAG-- GAATAGGA-T AGAA-GGAAG
134A    ----CAATTG -ATCCAGTAT CAACG-ATCA G-TACCAGT- GAATAGGA-T AGAATGGAAG
14      AAAACAATTG -ATCCAGTAT CAACG-ATCA G-TACCAGT- GAATAGGA-T AGAATGGAAG
15      AAAACAATTG -ATCCAGTAT CAACG-ATCA G-TACCAGT- GAATAGGAAT AGAATGGAAG
25      AAAACAATTG -ATCCAGTAT CAACG-ATCA G-TACCAGT- GAATAGGA-T AGAATGGAAG
26      AAAACAATTG -ATCCAGTAT CAACG-ATCA G-TACCAG-- GAATAGGA-T AGAATGGAAG
27C     ---------- ---------- ---------- ---------- ---------- ----------
28A     AAAACAATTG -ATCCA-TAT CAA-GGATCA G-TACCAG-- -AATAGGA-T AGAA-GGAAG
3       ---------- ---------- ---------- ---------- ---------- ----------
30A     AAAACAATTG -ATCCAGTAT CAA-GGATCA G-TACCAG-- GAATAGGA-T AGAA-GGAAG
36C     AAAACAATTG -ATCCAGTAT CAACG-ATCA G-TACCAGT- GAATAGGA-T AGAATGGAAG
41B     AAAACAATTG -ATCCAGTAT CAACG-ATCA G-TACCAGT- GAATAGGA-T AGAATGGAAG
45C     AAAACAATTG -ATCCAGTAT CAA-GGATCA G-TACCAGT- GAATAGGA-T AGAA-GGAAG
5       AAAACAATTG -ATCCAGTAT CAACG-ATCA G-TACCAG-- GAATAGGA-T AGAATGGAAG
50D     AAAACAATTG -ATCCAGTAT CAACG-ATCA G-TACCAGT- GAATAGGA-T AGAATGGAAG
57B     AAAACAATGG -ATCCA-TAT CAA-GGATCA G-TACCA--- GAATAGGA-T AGAA-GGAAG
58B     AAAACAATTG -ATCCA-TAT CAA-GGATCA GCTACCAGT- GAATAGGA-T AGAAGGGAAG
62D     AAAACAATTA -ATCCAGTAT CAACGGATCA G-TACCAGT- GAATAGGA-T AGAATGGAAG
65D     AAAACAATTA -ATCCAGTAT CAACGGATCA G-TACCAG-- GAATAGGA-T AGAAGGGAAG
67B     AAAACAATTG -ATCCAGTAT CAA-GGATCA G-TACCA-T- GAATAGGA-T AGAATGGAAG
69A     AAAACAATTG -ATCCAGTAT CAACG-ATCA G-TACCAGT- GAATAGGA-T AGAATGGAAG
69B     AAAACAATTG TATCCAGTAT CAA--GATCA G-TACCAG-- -AATAGGA-T AGAA-GGAAG
75A     AAAACA-TTG -ATCCAGTAT CAACG-ATCA G-TACCAGT- GAATAGGA-T AGAATGGAAG
76A     AAAACAATTG -ATCCAGTAT CAACG-ATCA G-TACCAGT- GAATAGGA-T AGAATGGAAG
76E     ---------- ---------- ---------- ---------- ---------- ----------
77B     AAAACAATTG -ATCCAGTAT CAACG-ATCA G-TACCAG-- GAATAGGA-T AGAATGGAAG
8       AAAACAATTG -ATCCAGTAT CAACG-ATCA G-TACCAGT- GAATAGGA-T AGAATGGAAG
80A     AAAACAATTG -ATCCAGTAT CAACGGATCA G-TACCAGTG GAATAGGA-T AGAAGGGAAG
81A     AAAACAATTG -ATCCAGTAT CAACG-ATCA G-TACCAGT- GAATAGGA-T AGAATGGAAG
82A     AAAACAATTG -ATCCAGTAT CAA-GGATCA G-TACCAGT- GAATAGGA-T AGAA-GGAAG
84F     AAAACAATTG -ATCCA-TAT CAACG-ATCA G-TACCAGT- GAATAG--AT AGAATGGAAG
86B     AAAACAATTG -ATCCAGTAT CAA-G-ATCA G-TACCAG-- -AATAGGA-T AGAA-GGAAG
93C     AAAACAATTG -ATCCAGTAT CAA-GGATCA G-TACCAG-- GAATAGGA-T AGAA-GGAAG
98A     AAAACAATTG -ATCCAGTAT CAACG-ATCA G-TACCAGT- GAATAGGA-T AGAATGGAAG
C.pri   AAAACAATTG -ATCCAGTAT CAACG-ATCA G-TACCAGT- GAATAGGA-T AGAATGGAAG
S.3     AAAACAATTG -ATCCAGTAT CAAC-GATCA G-TACCAGT- GAATAGGA-T AGAATGGAAG
S.7     AAAACAATTG -ATCCAGTAT CAA-G-ATCA G-TACCAG-- GAATAGGA-T AGAA-GGAAG
```

```
Bases 641-700
1       AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
10      AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
101A    AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
102B    AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
103A    AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
104     AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
108C    AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
111A    AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
12      AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
120C    AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
121A    AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
123E    AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
123G    AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
126E    AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
128C    AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
129     AACTCCACTC AATATCTAAA AAT-AAAAAA GATCTATCCT TCTATTGTGG TATCAAATGT
131     AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
132     AACTCCACTC AATATCTAAA AAT-AAAAAA GATCTATCCT TCTATTGTGG TATCAAATGT
134A    G-CTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTATTGTGG TATCAAATGT
14      AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
15      AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
25      AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
26      AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTATTGTGG TATCAAATGT
27C     --CTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
28A     AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
3       --CTCCACTC AATATCTAAA AATCAAAAAG GATCTATCCT GCTGTTGTGG TATCAAATGT
30A     AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
36C     AACTCCACTC AATATCTAAA AATAAAAAAG -ATCTATCCT TCTATTGTGG TATCAAATGT
41B     AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTGTCCT GCTGTTGTGG TATCAAATGT
45C     AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
5       AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
50D     AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
57B     AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
58B     AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
62D     AACTCCACTA AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
65D     AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
67B     AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTATTGTGG TATCAAATGT
69A     AACTCCACTC AATATCTAAA AATCAAAAAG GATCTATCCT GCTATTGTGG TATCAAATGT
69B     AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTATTGTGG TATCAAATGT
75A     AACTCCACTC AATATCTAAA AATCAAAAAA GATCTATCCT GCGATTGTGG TATCAAATGT
76A     AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTATTGTGG TATCAAATGT
76E     --CTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTATTGTGG TATCAAATGT
77B     AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTATTGTGG TATCAAATGT
8       AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
80A     AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
81A     AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTATTGTGG TATCAAATGT
82A     AACTCCACTC AATATCTAAA AATCAAAA-G -ATCTATCCT GCTATTGTGG TATCAAATGT
84F     AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
86B     AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
93C     AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
98A     AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
C.pri   AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTATTGTGG TATCAAATGT
S.3     AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
S.7     AACTCCACTC AATATCTAAA AATCAAAAAG -ATCTATCCT GCTGTTGTGG TATCAAATGT
```

```
Bases 701-760
1        ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
10       ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
101A     ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
102B     ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
103A     ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
104      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
108C     ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
111A     ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
12       ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
120C     ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
121A     ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
123E     ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
123G     ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
126E     ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
128C     ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
129      ATGGTTTCCG TTGGTGCAAA TCCA-ATCAT CTCAATTTTA AATTTAAG-A TGAGAAAGAA
131      ATGGTTTCCA TTGGTGCAGA TCCAGATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
132      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
134A     ATGGGTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATGTAAGTA TGAGAAAGAA
14       ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
15       ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
25       ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
26       ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
27C      ATGGTTTCCA TTGGTGCAAA TGCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
28A      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
3        ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
30A      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
36C      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
41B      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
45C      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
5        ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
50D      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
57B      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
58B      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
62D      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
65D      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
67B      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
69A      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
69B      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAGGA TGAGAAAGAA
75A      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
76A      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
76E      ATGGTTCCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
77B      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAGGA TGAGAAAGAA
8        ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
80A      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
81A      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
82A      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
84F      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
86B      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
93C      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
98A      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
C.pri    ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
S.3      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
S.7      ATGGTTTCCA TTGGTGCAAA TCCA-ATCAC CTCAATTTTA AATTTAAG-A TGAGAAAGAA
```

```
Bases 761-820
1        TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
10       TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
101A     TTCTCCATTG ATAGAAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
102B     TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
103A     TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
104      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
108C     TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
111A     TTCTCCATTG ATAGAAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
12       TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
120C     TTCTCCATTG ATAGAAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
121A     TTCTCCATTG ATAGAAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
123E     TTCTCCATTG ATAGAAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
123G     TTCTCCATTG ATAGAAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
126E     TTCTCCATTG ATAGAAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
128C     TTCTCCATTG ATAGAAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
129      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTTTTTT AAAAAAGCAG
131      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
132      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAT
134A     TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
14       TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGG-AA ATCCTATTTA AAAAAAGCAG
15       TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
25       TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
26       TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
27C      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
28A      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
3        TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
30A      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
36C      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
41B      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
45C      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
5        TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
50D      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
57B      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
58B      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
62D      TTCTCCATTG ATAGAAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
65D      TTCTCCATTG ATAGAAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
67B      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGG-AA ATCCTATTTA AAAAAAGCAG
69A      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAAGGGAA ATCCTATTTA AAAAAAGCAG
69B      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAAGGGAA ATCCTATTTA AAAAAAGCAG
75A      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
76A      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAAGGGAA ATCCTATTTA AAAAAAGCAG
76E      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
77B      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
8        TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
80A      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
81A      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
82A      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
84F      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
86B      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
93C      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
98A      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
C.pri    TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
S.3      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
S.7      TTCTCCATTG ATAGCAAATG CTATCCATTA AGCAGGGGAA ATCCTATTTA AAAAAAGCAG
```

```
Bases 821-880
1        AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
10       AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
101A     AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
102B     AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
103A     AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
104      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
108C     AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
111A     AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
12       AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
120C     AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
121A     AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
123E     AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
123G     AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
126E     AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
128C     AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
129      AAAAATTATG CCTTACTCTT G-----AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
131      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
132      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
134A     AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
14       AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
15       AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
25       AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
26       AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
27C      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
28A      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
3        AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
30A      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
36C      AAAAATTATG CCTTACTCTT GGAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
41B      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
45C      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
5        AAAAATTATG CCTTACTCTT GCAAGGAACG GACTAACAGG GTCAGCTACT CAGCTAATCT
50D      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
57B      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
58B      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
62D      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
65D      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
67B      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
69A      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
69B      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG -TCAGCTACT CAGCTAATCT
75A      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
76A      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
76E      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
77B      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
8        AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
80A      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
81A      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
82A      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
84F      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
86B      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
93C      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
98A      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
C.pri    AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
S.3      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
S.7      AAAAATTATG CCTTACTCTT GCAAG-AACG GACTAACAGG GTCAGCTACT CAGCTAATCT
```

```
Bases 881-940
1        TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
10       TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
101A     TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
102B     TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
103A     TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
104      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
108C     TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
111A     TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
12       TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG TTGAAAGACC TAGTACTGGA
120C     TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
121A     TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
123E     TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
123G     TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
126E     TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
128C     TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
129      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
131      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
132      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
134A     TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
14       TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
15       TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
25       TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
26       TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
27C      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
28A      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
3        TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
30A      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
36C      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
41B      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
45C      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
5        TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
50D      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
57B      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
58B      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
62D      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
65D      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
67B      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
69A      TCATAATTAA ATACCGGTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
69B      TCATAATTAA ATACCGGTAC TGTATAAGTA GATCTCGCTG T-GAAAGACC TAATACTGGA
75A      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
76A      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTAGTTG T-GAAAGACC TAGTACTGGA
76E      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTAGTTG T-GAAAGACC TAGTACTGGA
77B      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
8        TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
80A      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
81A      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTAGTTG T-GAAAGACC TAGTACTGGA
82A      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
84F      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
86B      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
93C      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
98A      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
C.pri    TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
S.3      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
S.7      TCATAATTAA ATACCGTTAC TGTATAAGTA GATCTCGTTG T-GAAAGACC TAGTACTGGA
```

```
Bases 941-1000
1        TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
10       TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
101A     TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
102B     TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
103A     TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
104      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
108C     TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
111A     TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
12       TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
120C     TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
121A     TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
123E     TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
123G     TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
126E     TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
128C     TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
129      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
131      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
132      TAATTATGGG TAGAGCCAAA GAGTGTGAAC GGTACAAGTT AACAATAACA TTGATTAAAT
134A     TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
14       TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAA-T
15       TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
25       TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
26       TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
27C      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
28A      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
3        TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
30A      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
36C      TAATTATGGG TAGAGCCAAA GAGTGTGAAC GGTACAAGTT AACAATAACA TTGATTAAAT
41B      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
45C      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
5        TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
50D      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
57B      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
58B      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
62D      TAATAATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
65D      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
67B      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
69A      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
69B      TAATTATGGG TAGAGCCAAA GAG-GTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
75A      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
76A      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
76E      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
77B      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
8        TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
80A      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
81A      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
82A      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
84F      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
86B      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
93C      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
98A      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
C.pri    TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
S.3      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
S.7      TAATTATGGG TAGAGCCAAA GAGTGTGAAC TGTACAAGTT AACAATAACA TTGATTAAAT
```

```
Bases 1001-1060
1        GAAGGAAGGG CTCCGGGGTA TAGAGAGGAC CTCACCGTT- AAGAAGTAAC CATAGAAA-C
10       GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
101A     GAAGGAAGG- CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
102B     GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCCCCGTTT AAGAAGTAAC CATAGAAA-C
103A     GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
104      GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
108C     GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
111A     GAAGAAGGG- CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
12       GAAAGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
120C     GAAGGAAGG- CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
121A     GAAG-AAGG- CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
123E     GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
123G     GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
126E     GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
128C     GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
129      GAAGGAAGG- CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATATAAA-C
131      GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
132      GAAGGAAGG- CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
134A     GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
14       GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGGTT AAGAAGTAAC CATAGAAA-C
15       GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
25       GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
26       GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
27C      GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
28A      GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCCCCGTTT AAGAAGTAAC CATAGAAA-C
3        GAAG-AAGG- CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
30A      GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
36C      GAAAGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
41B      -AAAGGAAGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
45C      GAAG-AAGG- CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
5        GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
50D      GAAGGAAGG- CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
57B      GAAG-AAGG- CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
58B      GAAG-AAGG- CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
62D      GAAGGAAGG- CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAC-C
65D      GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
67B      GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
69A      GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
69B      GAAG-AA-GG CTC-GGTGTA TAGAGAGGAC -TCACCGTTT AAGAAGTAAC CATAGAAA-C
75A      GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
76A      GAAG-AA-GG CTCCG-TGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
76E      GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
77B      GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
8        GAAAGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTACC CATAGAAA-C
80A      GAAGGAAGG- CTCCGGTGTA TAGAGAGGAC CTCACGGTTT AAGAAGTAAC -ATAGAAA-C
81A      GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
82A      GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
84F      GAAAGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGGTT AAGAAGTAAC CATAGAAA-C
86B      GAAGAAGGG- CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
93C      GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
98A      GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAAAC
C.pri    GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
S.3      GAAG-AAGG- CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
S.7      GAAGGAAGGG CTCCGGTGTA TAGAGAGGAC CTCACCGTTT AAGAAGTAAC CATAGAAA-C
```

```
Bases 1061-1120
1       GATGGAACCC ACTATATCCA TT-ATATTAA CTTTACTACT TTTTAATTTA CT-AT-----
10      GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
101A    GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
102B    GATGGAACCC CCTATATCCC TTTATATTTA CTTTTCTACT TTTT-ATTTA CT-AT-----
103A    GATGGAACCC CCTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
104     GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
108C    GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
111A    GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
12      GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
120C    GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
121A    GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
123E    GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
123G    GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
126E    GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
128C    GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
129     GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTGTTATTTA TT-AT-----
131     GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
132     GATGTAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
134A    GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTTTATTTA CT-AT-----
14      GATGGAACCC ACTATATCCA TTTATATATA CTTTACTACT TTTT-ATTTA CT-AT-----
15      GATGGAACCC ACTATATCCA TT-ATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
25      GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
26      GATGGAACCC ACTATATCCA TTTATATTTA CTGTACTACT TTTTTATTTA TTTATTTACT
27C     GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
28A     GATGGAACCC CCTATATCCC TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
3       GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
30A     GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
36C     GATGTAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTTTATTTA CT-AT-----
41B     GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
45C     GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
5       GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
50D     GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
57B     GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
58B     GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
62D     GATGGAACCC ACTATATCCA TTTATATT-A CCTTACTACT TTTT-ATTTA ACTAT-----
65D     GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
67B     GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTTTATTTA TTTATTTACT
69A     GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTTTATTTA TTTATTTACT
69B     GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTTTATTTA TTTATTTACT
75A     GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTTTATTTA CT-AT-----
76A     GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTTTATTTA CT-AT-----
76E     GATGGAACCC ACTATATGCA TTTATATTTA CTTTACTACT TTTTTATTTA CT-AT-----
77B     GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTTTATTTA TTTATTTACT
8       GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
80A     GATGGAACC- ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
81A     GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTTTATTTA CT-AT-----
82A     GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTTTATTTA TTTATTTACT
84F     GATGGA-CCC ACTATATCCA TTTATATTTA CTT-ACTACT TTTT-ATTTA CT-AT-----
86B     GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
93C     GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
98A     GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
C.pri   GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTTTATTTA TTTATTTACT
S.3     GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
S.7     GATGGAACCC ACTATATCCA TTTATATTTA CTTTACTACT TTTT-ATTTA CT-AT-----
```

```
Bases 1121-1180
1        GGGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
10       GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT TGAGAGGTGA AATGCCTAGA
101A     GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAAA
102B     GTGT-TTTT- GATACCTAGT ATCCATA-CC ATTTTTGATT T-AGAGGTGA AATGCCTAGA
103A     GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
104      GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
108C     GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAAGTGA AATGCCTAGA
111A     GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
12       GTGT-TTTTT GATACCTAGT ATCCATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
120C     GTGT-TTTT- GATACCTAGT ATCAATA-CA -TTTTTGATT --AGAGGTGA AATGCCTAGA
121A     GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGGGA AATGCCTAGA
123E     GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
123G     GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA A-TGCCTAGA
126E     GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
128C     GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
129      GTAT-TTTT- GATACCTAGT ATCAATA-CA ATTTTT-ATT T-AGAGGTGA AATGCCTAGA
131      GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
132      GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTGTATT T-AGAGGTGA AATACCTATC
134A     GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
14       GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
15       GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
25       GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
26       ATGTGTTTTT GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
27C      GTGT-TTTT- GATACCTAGT ATCAATA-CC ATTTTTGATT T-AGAGGTGA AATGCCTAGA
28A      GTGT-TTTT- GATACCTAGT ATCAATA-CC ATTTTTGATT T-AGAGGTGA AATGCCTAGA
3        GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
30A      GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
36C      GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTGTATT T-AGAGGTGA AATGCCTAGA
41B      GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
45C      GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
5        GTGT-TCTT- GATACCTAGT ATCAATA-CA ATTTTTGATT TGAGAGGTGA AATGCCTAGA
50D      GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
57B      GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
58B      GTGT-TTTT- GATACCTAGT ATCAATT-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
62D      GTGT-TTTT- GATATCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
65D      GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
67B      ATGTGTTTTT GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
69A      ATGTGTTTTT GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
69B      ATGTGTTTTT GATACCTAGT ATCAATA-CA TTTTTTGATT T-AGAGGTGA AATGC-TAGA
75A      GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
76A      GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAAGTGA AATGCCTAGA
76E      GTGT-TTTT- GATACCTAGT ATCAATAGCA ATTTTTGATT T-AGAGGAGA AATGCCTAGA
77B      ATGTGTTTTT GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
8        GTGT-TTTTT GATACCTAGT ATCAATA-CA ATTTTTGTTT T-AGAGGTGG AATGCCTAGA
80A      GTGG-TTTT- GATAC-TAG- ATCA-TA-C- ATTTTTGATT --AAAGGTGA A-TGCCTAGA
81A      GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
82A      ATGTGTTTTT GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
84F      GTGT-TTTT- GATACCTAGT ATCAATA-CC ATTTTTGATT T-AGAG-TGA AATGCCTAGA
86B      GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
93C      GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
98A      GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
C.pri    ATGTGTTTTT GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
S.3      GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
S.7      GTGT-TTTT- GATACCTAGT ATCAATA-CA ATTTTTGATT T-AGAGGTGA AATGCCTAGA
```

```
Bases 1181-1240
1        AAAAAAAAAA A-------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
10       AAAAAAAAAA --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
101A     AAAAAAAAAA A-------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
102B     AAAAAAAAAA --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
103A     AAAAAAAAA- --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
104      AAAAAAAAA- --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
108C     AAAAAAAAAA --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
111A     AAAAAAAAAA AAA-----GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
12       AAAAAAAAAA --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
120C     AAAAAAAAAA A-------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
121A     AAAAAAAAAA AAAA----GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
123E     AAAAAAAAAA AAA-----GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
123G     AAAAAAAAAA AAAAA---GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
126E     AAAAAAAA-- --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
128C     AAAAAAAAAA A-------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
129      AAAAAAAAAA AAAA---CTA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
131      AAAAAAAAAA --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
132      AAAAAAAAAA AAAAAAAAGA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
134A     AAAAAAAAAA AAAA----GA AAAAATCGTG GTTGGTAAGG TTATAGTAGC AAAAGCCATT
14       AAAAAAAAAA --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
15       AAAAAAAAAA --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
25       AAAAAAAAAA --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
26       AAAAAAAA-- --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
27C      AAAAAAAAAA --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
28A      AAAAAAAAAA --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
3        AAAAAAAAAA AAA-----GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
30A      AAAAAAAAAA A-------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
36C      AAAAAAAAAA --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
41B      AAAAAAAAAA AA------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
45C      AAAAAAAAAA AAA-----GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
5        AAAAAAAAAA --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
50D      AAAAAAAAAA --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
57B      AAAAAAAAA- --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
58B      AAAAAAAAAA --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
62D      AAAAAAAAAA A-------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
65D      AAAAAAAAAA --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
67B      AAAAAAAA-- --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
69A      AAAAAAAA-- --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
69B      AAAAAAAA-- --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
75A      AAAAAAAAAA AAA-----GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
76A      AAAAAAAAAA AAAAAA--GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
76E      AAAAAAAAAA AA------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
77B      AAAAAAAA-- --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
8        AAAAAAAAAA --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
80A      AAAAAAAAAA A-------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
81A      AAAAAAAAAA A-------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
82A      AAAAAAAA-- --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
84F      AAAAAAAAAA --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
86B      AAAAAAAAAA --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
93C      AAAAAAAAAA --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
98A      AAAAAAAAAA AAA-----GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
C.pri    AAAAAAAA-- --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
S.3      AAAAAAAAAA AA------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
S.7      AAAAAAAAAA --------GA AAAAATCGTG GTTGGGAAGG TTATAGTAGC AAAAGCCATT
```

Bases 1241-1300

```
1        GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
10       GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
101A     GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
102B     GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
103A     GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
104      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
108C     GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
111A     GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
12       GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
120C     GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
121A     GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGA-T AGGACACGGG
123E     GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
123G     GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGA-T AGGACACGGG
126E     GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
128C     GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
129      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
131      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
132      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
134A     GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
14       GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGA-T AGGACACGGG
15       GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
25       GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGA-T AGGACACGGG
26       GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGAGT AGGACACGGG
27C      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
28A      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
3        GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
30A      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
36C      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
41B      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGA-T AGGACACGGG
45C      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
5        GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGA-T AGGACACGGG
50D      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
57B      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
58B      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
62D      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
65D      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
67B      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
69A      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
69B      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
75A      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
76A      GGCATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
76E      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
77B      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
8        GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
80A      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
81A      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
82A      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
84F      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
86B      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
93C      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
98A      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
C.pri    GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
S.3      GGAATTTTTA TTTTATCCAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
S.7      GGAATTTTTA TTTTATACAT AGGAAGAATC CGTTTTGTTA TTAATAGACT AGGACACGGG
```

```
Bases 1301-1337 with the 14 presence/absence indel characters
1       AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
10      AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
101A    AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
102B    AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
103A    AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
104     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
108C    AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     1111101011 0010
111A    AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     1111101011 0010
12      AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
120C    AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
121A    AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
123E    AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101111 1010
123G    AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101111 1010
126E    AAGGGAATAA CTGAAAGAAA GGAAATCA-T TAGTTATT     0111101011 1010
128C    AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
129     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0110101001 1100
131     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
132     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0100101010 1010
134A    AACGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111111111 1010
14      AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
15      AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
25      AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
26      AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1011
27C     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
28A     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
3       AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
30A     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
36C     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0100001010 1010
41B     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
45C     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
5       AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
50D     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
57B     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0011101011 1010
58B     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
62D     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
65D     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
67B     AAGGGAATAA CTGAAAGAAA GGAA-TCAAT TAGTTATT     0111101011 1011
69A     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1011
69B     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1111
75A     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111111111 1010
76A     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111111111 1010
76E     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111111111 1010
77B     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1011
8       AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
80A     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
81A     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0011111111 1110
82A     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1011
84F     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
86B     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
93C     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
98A     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
C.pri   AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111100011 1111
S.3     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
S.7     AAGGGAATAA CTGAAAGAAA GGAAATCAAT TAGTTATT     0111101011 1010
```

References

Abrams, L. R. 1910. Studies on the Flora of Southern California III. Bulletin of the Torrey Botanical Club 37:149-153.

Appels, R. and R. L. Honeycutt. 1986. rDNA: evolution over a billion years. *In* DNA Systematics, Vol. II, Plants, ed. S. K. Dutta, pp. 81-135. CRC Press, Boca Raton, Florida.

Arnheim N., M. Krystal, R. Schmickel, G. Wilson, O. Ryder, and E. Zimmer. 1980. Molecular evidence for genetic exchanges among ribosomal genes on nonhomologous chromosomes in man and apes. Proceedings of the National Academy of Sciences U.S.A. 77:7323-7327.

Arnheim, N. 1983. Concerted evolution of multigene families. *In* Evolution of Genes and Gene Proteins, eds. M. Nei and R. K. Koehn, pp. 38-61. Sinauer Associates, Sunderland, Massachusetts.

Baker, D.D. and C.R. Schwintzer. 1990. Introduction *In* The Biology of Frankia and Actinorhizal Plants, eds. C.R. Schwintzer and J.D. Tjepkema, pp. 1-31. Academic Press, New York, New York.

Baldwin, B. and S. Markos. 1998. Phylogenetic utility of the external transcribed spacer (ETS) of 18S-26S rDNA: Congruence of ETS and ITS trees of Calycadenia (Compositae). Molecular Phylogenetics and Evolution 10:449-463.

Baldwin, B.G., M.J. Sanderson, J.M. Porter, M.F. Wojciechowski, C.S. Campbell, and M.J. Donoghue. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. Annals of the Missouri Botanical Garden 82:247-277.

Bena, G., Jubier, M.-F., Olivieri, I., and B. LeJueune. 1998. Ribosomal external and internal transcribed spacers: Combined use in phylogenetic analysis of *Medicago* (Leguminosae). Journal of Molecular Evolution 46:299-306.

Blauer, A.C., A.P. Plummer, E.D. McArthur, R. Stevens, and B.C. Giunta. 1975. Characteristics and hybridization of important Intermountain shrubs. I. Rose family. USDA Forest Service Research Paper INT-169. 36 pgs.

Brayton, R. and H.A. Mooney. 1966. Population variability of *Cercocarpus* in the White Mountains of California as related to habitat. Evolution 20:383-391.

Buckler, E. S. IV, A. Ippolito, and T. P. Holtsford. 1997. The evolution of Ribosomal DNA: Divergent Paralogues and phylogenetic implication. Genetics 145:821-832.

Buckler, E. S., IV, and T. P. Holtsford. 1996. *Zea* Systematics: ribosomal ITS evidence. Molecular Biology and Evolution 13:612-622.

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

Correll, D. S., and M. C. Johnston. 1970. Manual of the Vascular Plants of Texas. Texas Research Foundation, Renner, Texas.

Campbell, C. S., M. F. Wojciechowski, B. G. Baldwin, L. A. Alice, and M. J. Donoghue. 1997. Persistent nuclear ribosomal DNA sequence polymorphism in *Amelanchier* agamic complex (Rosaceae). Molecular Biology and Evolution 14:81-90.

Davis, J.N. and J.D. Brotherson. 1991. Ecological Relationships of Curlleaf Mountain Mahogany (*Cercocarpus ledifolius* Nutt.) Communities in Utah and implications for Management. Great Basin Naturalist 51:153-166.

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

Doyle, J. J., and S. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemistry Bulletin 19:11-15.

Dvorak, J. 1990. Evolution of multigene families: the ribosomal RNA loci of wheat and related species. *In* Plant Populational Genetics, Breeding, and Genetic Resources, eds. A. H. D. Brown and M. T. Clegg, pp. 83-97. Sinauer Associates, Sunderland, Massachusetts.

Eastwood, A. 1898. Proceedings of the California Academy of Sciences 3:136.

Evans, R. C., L. A. Alice, C. S. Campbell, E. A. Kellogg, and T. A. Dickson. 2000. The granule-bound starch synthase (GBSSI) gene in the Rosaceae:

Multiple loci and pylogenetic utility. Molecular Phylogenetics and Evolution 17:388-400.

Faith, D. P. 1991. Cladistic permutation test for monophyly and nonmonophyly. Systematic Zoology 40:366-375.

Farris, J. S., M. Kallersjo, A. G., Kluge, and C. Bult. 1994. Testing significance of incongruence. Cladistics 10:315-319.

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

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

Gray, A. 1852. Plantae wrightianae part 1, pp. 54. Smithsonian Institute, G. P. Putnam, New York, New York.

Hamby, K. R. and E. A. Zimmer. 1992. Ribosomal RNA as a phylogenetic tool in plant systematics. *In* Molecular Systematics of Plants, eds. P. Soltis, D. Soltis, and J. Doyle, pp. 50-91. Chapman and Hall, New York, New York.

Hamilton, M.B. 1999. Four primer pairs for the amplification of chloroplast intergenic regions with interspecific variation. Molecular Ecology 8:521-523.

Hedren, M., M. F. Fay, and M. W. Chase. 2001. Amplified fragment length polymorphisms (AFLP) reveal details of polyploid evolution in *Dactylorhiza* (Orchidicaeae). American Journal of Botany 88:1868-1880.

Henrickson, J. 1987. Two new species of *Cercocarpus* (Rosaceae) from Mexico. Systematic Botany 12:293-298.

Hillis, D. M. 1987. Molecular versus morphological approaches to systematics. Annual review of Ecology and Systematics 18:23-42.

Hillis, D. M. and M. T. Dixon. 1991. Ribosomal DNA: Molecular evolution and phylogenetic inference. Quarterly Review Biology 66:411-452.

Hitchcock, C. L., and A. Cronquist. 1961. *Cercocarpus In* Vascular Plants of the Pacific Northwest, Part 3: Saxifragaceae to Ericaceae, pp. University of Washington Press, Seattle, Washington.

Holmgren, N. H. 1987. *Cercocarpus ledifolius* var. *intermontanus* (Rosaceae), a new varietal name for the intermountain curl-leaf mountain-mahogany. Brittonia 39:423-427.

Hooker, W. J. 1840. Icones plantarum series 1, vol. 4, pp. 332. Royal Botanic Society, Kew Gardens, London.

Hooker, W. J., and G. A. W. Arnott. 1839. The Botany of Captain Beechey's Voyage, part 8, pp. 337. H. G. Bohn, London.

Humboldt, F. W. H. A., A. Bonpland, and C. S. Kunth. 1832. *Cercocarpus In* Nova Genera et Species Plantarum, vol. 6, pp. 232. reprinted by Weldon, Wesely and Hafner, New York, New York.

Jansen, R. K. and J. D. Palmer. 1987. Chloroplast DNA from lettuce and *Barnadesia* (Asteraceae): structure, gene location, and characterization of a large inversion. Current Genetics 11:553-564.

Jones, M. E. 1891a. New plants from Arizona, Utah, and Nevada. Zoe 2:12-17.

Jones, M. E. 1891b. New species and notes of Utah Plants. Zoe 2:236-252.

Jourgensen, R. A., R. E. Cuellar, W. F. Thompson, and T. A. Kavanagh. 1987. Structure and Variation in Ribosomal RNA genes of pea: characterization of a cloned ribosomal DNA repeat and chromosomal ribosomal DNA variants. Plant Molecular Biology 8:3-12.

Kearney, T. and R. Peebles. 1951. *Cercocarpus In* Arizona Flora, pp. 388. University of Arizona Press, Tucson, Arizona.

Kellogg, E. A., R. Appels, and R. J. Mason-Gamer. 1996. When genes tell different stories: the diploid genera of Triticeae (Gramineae). Systematic Botany 21:321-347.

Klemmedson, J.O. 1979. Ecological importance of actinomycete-nodulated plants in the western U.S. Botanical Gazette 140(suppl.):S91-S96.

Koopman, W. J. M., M. J. Zevenbergen, and R. G. Van den Berg. 2001. Species relationships in *Lactuca* s.l. (Lactuceae, Asteraceae) inferred from AFLP fingerprints. American Journal of Botany 88:1881-1887.

Linder, C. R., L. R. Goertzen, B. Vanden Heuvel, J. Francisco-Ortega, and R. K. Jansen. 2000. The complete External Transcribed Spacer of 18S - 26S rDNA: Amplification and Phylogenetic utility at low taxonomic levels in the Asteraceae and closely allied families. Molecular Phylogenetics and Evolution 19:167-176.

Lis, R. A. 1992. Leaf Architectural Survey of *Cercocarpus* (Rosaceae) and its systematic significance. International Journal of Plant Science 153:258-272.

Lis, R. A. 1993. *Cercocarpus*. *In* The Jepson Manual: Higher Plants of California, ed. J. C. Hickman, pp. 948-949. University of California Press, Berkeley, California.

Little, E. L. 1953. Five varietal transfers of United States trees. Phytologia 4:305-310.

MacAurthur, E. D., and S. C. Anderson. 1985. A Cytotaxonomic Contribution to the western North American Rosaceous flora. Madrono 32:24-28.

Martin, F. L. 1950. A revision of *Cercocarpus*. Brittonia 7:91-111.

Mayol, M., and J. A. Rossello. 2001. Why Nuclear Ribosomal DNA spacers (ITS) tell different stories in *Quercus*. Molecular Phylogenetics and Evolution 19:167-176.

McVaugh, R. 1952. Remarks of the genus *Cercocarpus* in Texas. Field and Laboratory 10:35-40.

Mindell, D. P. and R. L. Honeycutt. 1990. Ribosomal RNA in vertebrates: evolution and phylogenetic applications. Annual Review of Ecology and Ststematics 21:541-566.

Morgan, D. R., D. E. Soltis, and K. Robertson. 1994. Systematic and Evolutionary Implications of rbcL sequence variation in Rosaceae. American Journal of Botany 81:890-903.

Morley, T. 1949. *In* Documented chromosome number of plants. Madrono 10:95.

Mortenson, T.H. 1973. Ecological variation in the leaf anatomy of selected species of *Cercocarpus*. Aliso 8:19-48.

Mueller, U. G., and L. L. Wolfenbarger. 1999. AFLP genotyping and fingerprinting. Trends in Ecology and Evolution 14:389-394.

Munz, P. A. 1959. A California Flora. University of California Press, Berkeley, California.

Murphy, W. J., E. Eizirik, W. E. Johnson, Y. P. Zhang, O. A. Ryder, and S. J. O'Brien. 2001. Molecular phylogenetics and the origins of placental mammals. Nature 409:614-618.

Nickrent, D. L. and D. E. Soltis. 1995. A comparison of angiosperm phylogenies from nuclear 18S rDNA and *rbcL* sequences. Annals of the Missouri Botanical Garden 82:208-234.

O'Kane, S. L., Schaal, B. A., and I. A. Al-Shehbuz. 1996. The origins of *Arabadopsis suecica* (Brassica

Palmer, J. D. 1986. Isolation and structural analysis of chloroplast DNA. Methods in Enzymology 118:167-186.

Rafinesque, C. S. 1832. Twenty new genera of plants from the Oregon Mountains. Atlantic Journal 1:146.

Rieseberg, L. H., J. Whitton, and C. R. Linder. 1996. Molecular marker incongruence in plant hybrid zones and phylogenetic trees. Acta Botanica Neerlandica 45:243-262.

Rieseberg, L. H., and O. Gerber. 1994. Hybridization in the Catalina Island Mountain Mahogany (*Cercocarpus traskiae*):RAPD evidence. Conservation Biology 9:199-203.

Rieseberg, L. H., S. Beckstromsternberg, and K. Doank. 1990. *Helianthus annuus* ssp *texanus* has chloroplast DNA and nuclear ribosomal RNA genes of *Helianthus debilis* ssp *cucumerifolius*. Proceedings of the National Academy of Science USA 87:593-597.

Rieseberg, L. H., S. Zona, L. Aberbom, and T. D. Martin. 1989. Hybridization in the island endemic, Catalina Mahogany. Conservation Biology 3:52-58.

Rosenberg, M. S., and S. D. Kumar. 2001. Incomplete taxon sampling is not a problem for phylogenetic inference. Proceedings of the National Academy of Sciences USA 98: 10751-10756.

Russell S. K., and E. W. Schupp. 1998. Effects of microhabitat patchiness on patterns of seed dispersal and seed predation of *Cercocarpus ledifolius* (Rosaceae). Oikos 81:434-443.

Russell, S. K. 1997. Thesis. Utah State University.

Rydberg, P. A. 1913. *Cercocarpus*. North American Flora 22:418-424.

Sanderson, M. J. and J. J. Doyle. 1992. Reconstruction of organismal gene phylogenies from data on multigene families: concerted evolution, homoplasy, and confidence. Systematic Biology 41:4-17.

Sang, T. Crawford, D. J., and Stuessy, T. F. 1995. Documentation of reticulate evolution in peonies (*Paeonia*) using sequences of internal transcribed spacer of nuclear ribosomal DNA: Implications for biogeography and concerted evolution. Proceedings of the National Academy of Sciences USA. 92:6813-6817.

Sang, T., D. J. Crawford, T. F. Stuessy, and O. Silva. 1995. ITS sequences and the phylogeny of the genus *Robinsonia* (Asteraceae). Systematic Botany 20:55-64.

Saunders, M. A., and S. V. Edwards. 2001. Dynamics and phylogenetic implications of mtDNA control region sequences in new world jays (Aves:Corvidae). Journal of Molecular Evolution 51:97-109.

Searcy, K. B. 1969. Variation in *Cercocarpus* in southern California. New Phytologist 68:829-839.

Schneider, C. K. 1905a. Illustrietes Handbusch der Laubholzkunde, vol. 1, pp. 530. Verlag von Gustav, Fisher, Jena.

Schneider, C. K. 1905b. Beitrag zur Kenntnis der Arten und Formen der Guttung *Cercocarpus* Kunth. Mitteilungen der Deutschen Dendrologischen Gesellschaft 14:125-129.

Soltis, D.E. and P.S. Soltis. 1998. Introduction *In* Molecular Systematics of Plants II: DNA sequencing. Soltis, P.S., D.E. Soltis, and J.J. Doyle (eds). Kluwer Academic publishers, Boston, Massachusetts.

Soltis, P. S., and D. E. Soltis. 1991. Multiple origins of the allotetraploid *Tragapogon mirus* (Compositae)-rDNA evidence. Systematic Botany 16:407-413.

Soltis, P. S., D. E. Soltis, P. G. Wolf, D. L. Nickrent, S. –M. Chaw, and R. L. Chapman. 1999. The phylogeny of land plants inferred from 18S rDNA sequences: Pushing the limits of rDNA signal? Molecular Biology and Evolution 16:1774-1784.

Steane D. A., G. E. McKinnon, R. E. Vaillancourt, and B. M. Potts. 1999. ITS sequence data resolve higher level relationships among the eucalypts. Molecular Phylogenetics and Evolution 12:215-223.

Stutz, H.C. 1974. Rapid Evolution in Western Shrubs. Utah Science 35:16-20.

Suh, Y., L. B. Thien, H. E. Reeve, and E. A. Zimmer. 1993. Molecular evolution and phylogenetic implications of internal transcribed spacer sequences of ribosomal DNA in Winteraceae. American Journal of Botany 80:1042-1055.

Swofford, D. L. 1999. PAUP*: Phylogenetic Analysis Using Parsimony beta test version 4.08b. Sinauer, Sunderland, Massachusetts

Templeton, A. R. 1983. Convergent evolution and non-parametric inference from restricted fragment and DNA sequence data *In* Statistical Analysis of DNA sequence Data, ed. B. Wier, pps. 151-179. Marcel Dekker, New York, New York.

Thompson, W. F. and R. B. Flavell. 1988. DnaseI sensitivity of ribosomal RNA genes in chromatin and nucleolar dominance in wheat. Journal of Molecular Biology 204:535-548.

Thorne, R. F. 1978. New Subspecific Combinations for Southern California Plants. Aliso 9:189-196.

Torrey, J., and A. Grey. 1840. A Flora of North America. Vol. 1, part 3. Wiley and Putnam, New York, New York.

Vargas P., B. G. Baldwin, and L. Constance. 1998. Nuclear ribosomal DNA evidence for a western North American origin of Hawaiian and South American species of *Sanicula (*Apiaceae). Proceedings of the National Academy of Sciences USA 95:235-240.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. Van De Lee, M. Hornes, A. Freijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP, a new technique for DNA fingerprinting. Nucleic Acid Research 23:4407-4414.

Waters, E. R. and B. A. Schaal. 1996. Biased gene conversion is not occurring among rDNA repeats in the *Brassica* triangle. Genome 39:150-154.

Watson, S. 1882. Proceedings of the American academy of Arts and Sciences 17:353.

Watson, S. 1876. Botany of California, vol. 1, pp. 175. Welch, Bigelow and Co., Cambridge, Massachusetts.

Watson, S. 1875. Revision of the Genus *Ceanothus*, and Descriptions of new plants. Proceedings of the American Academy of Arts and Sciences 10:333-350.

Wendel, J. F. and J. J. Doyle. 1998. Phylogenetic incongruence: window into genome history and molecular evolution. In Molecular Systematics of Plants II: DNA sequencing, eds. D. E. Soltis, P. S. Soltis and J. J. Doyle, pp 265-296. Kluwer Academic Publishing, Boston, Massachusetts.

Wendel, J. F., A. Schnabel, and T. Seelanan. 1995. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proceedings of the National Academy of Sciences USA 92:280-284.

Wendel, J. F., A. Schnubel, and T. Seelanan. 1995. An unusual ribosomal DNA sequence from *Gossypium gossypioides* reveals ancient, cryptic intergenomic introgression. Molecular Phylogenetics and Evolution 4:298:313

Zhang, L. B., H. P. Comes, and J. W. Kadereit. 2001. Phylogeny and Quaternary history of the European Montane/Alpine endemic *Soldanella* (Primulaceae) based on ITS and AFLP variation. American Journal of Botany 88:2331-2345.

Zimmer, E. A., S. L. Martin, S. M. Beverly, Y. M. Kan, and A. C. Wilson. 1980. Rapid duplication and loss of genes coding for the chains of hemoglobin. Proceedings of the National Academy of Sciences U.S.A. 77:2158-2162.

Vita

Brian David Vanden Heuvel was born in Greeley, Colorado on September 4, 1975. Joseph and Raedene Purma adopted him soon after his birth. After the death of Joseph Purma, his mother remarried Steve Vanden Heuvel, and both mother and son changed their last name to Vanden Heuvel. He grew up in Castle Rock, Colorado, and graduated from Douglas County High School in 1993. He attended Fort Lewis College in Durango, Colorado from 1993 to 1994. In 1994, he transferred to Colorado State University to pursue a Bachelor of Science degree, majoring in Botany. He graduated Phi Beta Kappa and cum laude in 1997. In the fall of 1997, he began his graduate career at The University of Texas at Austin.

Permanent address: 2511 Enfield Apartment D, Austin, Texas 78703
This dissertation was typed by Brian David Vanden Heuvel