INFORMATION TO USERS

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

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

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

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

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

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

UMI®
800-521-0600

Butterfly Oviposition Behavior, Pika Biogeography, and Lentiviral Sequence Evolution

Approved by
Dissertation Committee:

James J. Bull, Co-Supervisor

Michael C. Singer, Co-Supervisor

David M. Hillis

Mark A. Kirkpatrick

Beryl B. Simpson

Butterfly Oviposition Behavior, Pika Biogeography, and Lentiviral Sequence Evolution

by

Kelly Kathleen Agnew, B.A.

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

December 1999

Dedication

For my parents, Brad and Sue Agnew.

Acknowledgements

I thank my supervisors, Jim Bull and Mike Singer, for their guidance and encouragement through my graduate career. They, and the other members of my committee, David Hillis, Mark Kirkpatrick, and Beryl Simpson, provided comments on this dissertation and each member provided valued advice.

Thanks to Mike Singer for teaching me how to read critically and for sharing his knowledge of *Euphydryas* and his beautiful field sites. V. Veit provided field assistance for the butterfly research, which was supported by National Science Foundation grant DEB-9307804. Work at Rabbit Meadow was courtesy of the US Forest Service (Sequoia National Forest, Hume Lake District).

Thanks to D. Hafner and the New Mexico Museum of Natural History for supporting the pika molecular work, field collections in New Mexico, and for providing access to tissues in the Museum's frozen collection. T. Lawlor provided a pika tissue sample from Nevada, and shared his knowledge of pika vocalizations. Discussions with R. Honeycutt helped structure my research questions. Thanks to R. Smith and the Rocky Mountain Biological Lab in Gothic, Colorado for introducing me to pikas and for supporting collections.

J. Bull, and his courses in Experimental Design and Infectious Diseases, helped me critically evaluate my own work and think about research directions. Discussions with D. Begun and H. Wichman provided valuable insight to my own research and to the scientific process in general--their thoughtful comments and willingness to listen are greatly appreciated. J. Bull and D. Hillis shared their lab space, and lab work was supported in part by the Zoology Department of the University of Texas at Austin. Special thanks go to M. Holder, M. Badgett, J. McGuire, and A. Graybeal for their patience and skill while teaching me the details of molecular lab work and phylogenetic analysis. Mark Holder also wrote several computer programs that helped in both the analyses of the pika biogeographic data and the lentiviral protein sequences—he deserves special recognition for his patience and generosity with his time and skill.

Finally, thanks to all the members of the Bull-Hillis-Cannatella labs for their support, advice and friendship--they were important sounding boards and are valued teachers and friends. They include J. Bull, D. Hillis, D. Cannatella, B. Mable, A. Graybeal, D. Brumbaugh, S. Messenger, J. McGuire, W. Crill, M. Badgett, L. Dries, M. Brauer, M. Holder, K. Kichler, T. Schlenke, and T. Wilcox.

Butterfly Oviposition Behavior, Pika Biogeography, and Lentiviral Sequence Evolution

Publication No._____

Kelly Kathleen Agnew, Ph.D.

The University of Texas at Austin, 1999

Supervisors: James J. Bull and Michael C. Singer

I present three different studies in three chapters. In Chapter 1, I investigate oviposition behavior in the checkerspot butterfly, *Euphydryas editha* (Nymphalidae). I conducted two field experiments to determine whether increasing internal egg load drives host plant discrimination in female insects searching for suitable oviposition sites. Supplemented by a series of dissections, the results indicate that internal egg load alone is not a causal factor in oviposition decisions. I also address the relationship between the strength of host preference and individual female age. Older butterflies are less discriminating among potential hosts and mean cluster size decreases with individual age in the butterfly population I studied.

In Chapter 2, I examine the biogeography of the American pika, *Ochotona princeps* (Lagomorpha). Pikas are small mammals restricted to montane habitat

In Chapter 2, I examine the biogeography of the American pika, *Ochotona princeps* (Lagomorpha). Pikas are small mammals restricted to montane habitat islands of mesic talus slopes throughout western North America. I sequenced 689 base pairs of mitochondrial DNA control region from 58 specimens representing every described *O. princeps* subspecies. The data were analyzed using parsimony, maximum likelihood and genetic algorithm methods. I identified five major clades of pikas, corresponding to the major mountain chains of western North America. Additionally, I used maximum likelihood to test specific biogeographic hypotheses about pika populations in western North America.

In Chapter 3, I examine the evolution of lentiviruses. Lentiviruses are complex retroviruses that infect mammals, usually causing immune dysfunction or nervous system disorders. I performed a maximum likelihood analysis of lentiviral relationships based on protein and nucleotide sequences of the *pol* polyprotein. *pol* is relatively conserved in lentiviruses and includes reverse transcriptase, the enzyme essential for retrovirus replication in host cells. Rooting the lentivirus tree was difficult due to high sequence divergence between the major retroviral families. I also investigated the evolution of dUTPase, which is found in some lentiviruses, as well as in distantly related RNA and DNA viruses. The pattern of dUTPase expression in these viral groups implies that it was present in a lentiviral ancestor, but has been lost in primate lentiviruses, including HIV.

Table of Contents

List of Tables

List of Figures

CHAPTER 1: Insect Diet Evolution: The relationships between host preference and life history traits

INTRODUCTION

This chapter discusses the relationship between life history traits and host preference in phytophagous insects. A newly-hatched larva's diet during its initial development is entirely dependent on its mother's oviposition decision. The studies presented here examine the mechanistic basis of oviposition decisions; i.e., whether the physiological condition of the female influences the strength of her preference for a particular resource (host plant) and the number of eggs she can lay in a single oviposition bout. The first major section is the result of a collaborative effort between myself and Michael Singer. The second section presents data I collected independently.

The study organism

The study organism is *Euphydryas editha* (Nymphalidae), Edith's checkerspot butterfly. *E. editha* ranges throughout western North America and is found in varied habitats, including chaparral, pine forest, and alpine tundra. Larvae of this species feed on a wide variety of host plants species, primarily members of the Scrophulariaceae. Host preferences (the plant species on which the searching female prefers to oviposit) can vary among populations and even among individuals within a population. Most populations of *E. editha* are univoltine; they produce one generation per year. Adults emerge in the late spring and early summer, they mate and then females lay clusters of eggs on hosts they

find acceptable. An adult female typically lives 10-14 days, and the population flight season generally lasts about 4-6 weeks. Eggs hatch about 7-10 days after oviposition, and the larvae begin feeding immediately on the host plant on which they were laid. In large clusters, larvae spin a communal web on the plant, presumably as a defense from predators. As the larvae grow and molt through instars, they may move off their natal plant. At the fourth instar stage the larvae enter overwintering diapause in the leaf litter. Immediately after snowmelt, the larvae resume feeding and then pupate 7-14 days before eclosion.

The study site

All of the following experiments were conducted at Rabbit Meadow (elev. 2,360m), Tulare County, California, where host preferences were already known. The principal host at this site, *Pedicularis semibarbata* (Scrophulariaceae), is preferred by almost all ovipositing butterflies over *Collinsia torreyi* (Scrophulariaceae). Butterflies vary, however, in their strength of preference for *P. semibarbata*, and eggs are laid on both plant species (Singer, 1983). Female *E. editha* normally lay one egg cluster per day in this population and cluster size can typically vary from 20 to 150 eggs (Moore 1987). The flight season lasts from mid-May to late June. Overnight temperatures typically hover around 0°C. At this altitude, butterflies normally restrict oviposition searches to the period between 11:00 and 16:00.

I. THE RELATIONSHIP BETWEEN FECUNDITY AND HOST PREFERENCE

Evolution is often constrained by mechanism, and the evolution of insect diet may be constrained by the processes involved in oviposition site choice. In the past decade, researchers have focused increasingly on the mechanisms that underlie oviposition behavior. A common theme emerging from this work is that more fecund individuals are more likely to accept low-ranked hosts (Jaenike 1990, Barton Browne 1993). If this inverse relationship between fecundity and host discrimination were generally true and driven by genetic covariance between these traits, then selection on fecundity would constitute an important constraint on the evolution of host use, and vice versa. Evolutionary increases in fecundity would tend to broaden the range of acceptable hosts, even if this host range expansion resulted in the inclusion of less suitable resource types. This would establish an important link between the evolution of resource use and that of life history.

Several optimality models predict that current fecundity drives oviposition decisions (Jaenike 1978, Mangel 1987, 1989a, 1989b). These models predict that when egg load is high, an insect should maximize its oviposition rate because it is not at risk of being egg-limited, while an insect that carries a small egg load should be more discriminating between potential hosts because it runs a greater risk of egg-limitation. A particularly useful physiological model is the *hierarchy-threshold model* developed by Courtney and his coworkers (Courtney et al. 1989). This model identifies critical factors in oviposition decisions and makes clear predictions about both individual and population level responses to those factors.

The model assumes that host ranking is hierarchical and transitive--an insect that accepts a host low in the hierarchy will also accept plants higher in the ranking scheme. This model also assumes that the decision to oviposit is driven by threshold levels of current fecundity--an insect that is willing to lay on a host of a particular rank has reached a particular threshold egg load. The threshold egg load is progressively higher for plants further down the hierarchy of potential hosts (Fig. 1.1). The model is applied to both individuals and populations: changes in host acceptance within an individual are abrupt--a new host is added to the diet when a particular threshold number of eggs is reached. Due to variation in thresholds across individuals, the population response curves will be much gentler, but still show the same trend. The hierarchy-threshold model also predicts that fecundity and oviposition on low-ranked hosts should be genetically correlated (Prediction 15 of Courtney et al., 1989).

A number of empirical studies supports these predictions (reviewed by Minkenberg et al. 1992). Rosenheim and Rosen (1991) found that egg load was inversely related to time taken for host discovery and handling in *Aphytis*. Fitt (1986) showed that the acceptance of low-ranked hosts increases with time if the tephritid fly *Dacus tysoni* is denied oviposition opportunities. This species is known to mature eggs continuously. However, in closely-related species that do not mature eggs continuously, individuals are not more likely to accept low-ranked hosts with increased time. Odendaal's (1989) egg manipulation study indicated that female *Battus philenor* butterflies with many eggs are likely to spend a larger proportion of their time searching for hosts than females with fewer

eggs. Using the same species, Pilson and Rausher (1988) showed that both the probability that females would oviposit on low-ranked hosts and the size of the resulting egg cluster increased with the time since the last oviposition. The authors suggested increasing egg load as a potential causal factor. In laboratory studies of both *Drosophila suboccidentalis* and *Callosobruchus maculatus*, more fecund individuals were more willing to accept low-ranking hosts (Courtney and Chen 1988, Wasserman and Futuyma 1981). Perhaps most importantly, Courtney and Hard (1990) showed positive genetic covariance between fecundity and acceptance of low-ranked hosts in *Drosophila busckii*, establishing a genetic link between these traits. There are few detailed studies (Prokopy et al. 1994, Fitt 1986) in which ovipositing females did not become less discriminating among hosts as egg load increased; most others have drawn the conclusion that egg load and discrimination are negatively associated.

The prediction that fecundity drives the acceptance of low-ranked hosts appears to be well-supported. However, most of the evidence for this model involves observed correlations between egg load and host acceptance: insects with high egg load tend to accept low-ranked hosts. How likely is it that these correlations represent cause and effect? If eggs accumulate during time periods when oviposition is not occurring, any trait that changes unidirectionally with time since the most recent oviposition will automatically be correlated with current egg load, independent of causality. Here, we use field experiments with the checkerspot butterfly, *Euphydryas editha*, to address questions about the relationship between fecundity and diet breadth. We ask whether differences

among individuals in host specificity are correlated with (1) the current fecundity at the first acceptance of a low-ranked host and (2) the rate of egg accumulation.

Models, Predictions and Methods

Free-flying insects are likely to encounter potential host plants individually during an oviposition search. In our study insect the range of acceptable plants broadens during each search, so that the likelihood that an insect encounters an acceptable plant increases over time until such an encounter occurs and oviposition is triggered. During the "discrimination phase" the searching insect accepts only the first-ranked host, rejecting the second-ranked host at each encounter. The discrimination phase ends when the second-ranked host is also accepted if it is encountered. Insects that are more willing to accept lower-ranked hosts have very short discrimination phases, while insects that strongly prefer the first-ranked host have longer discrimination phases. In a population for which the host ranking was already known and effectively invariant, we measured each butterfly's strength of oviposition preference by measuring the length of the discrimination phase (Singer et al., 1992). Using techniques described below, we also estimated current fecundity in order to test the predictions of the following models relating variation in egg load to variation in host acceptance.

Model 1: Short discrimination phases are caused by high rates of egg accumulation.

Model 1 is patterned after the hierarchy-threshold model of Courtney et al. (1989). It assumes that a critical threshold egg number induces oviposition. This

threshold is different for each host but invariant among insects in the same population. For example, all individuals accept the second-ranked host at an egg load of about 100 eggs. Intrapopulation variation in discrimination phase length is therefore determined by variation in egg accumulation rate.

Model 2: Short discrimination phases are caused by low threshold egg numbers.

In Model 2 threshold egg loads for host acceptance vary among individuals in a population but egg accumulation rates remain constant. Consequently, individuals with a threshold of 30 eggs will have shorter discrimination phases than individuals with thresholds of 100 eggs because they will reach the lower threshold number sooner. In this case, variation in discrimination phase length is determined by variation in threshold egg number.

Experimental Designs:

Experiment 1: What physiological mechanisms determine discrimination phase length? If relative fecundity is driving oviposition decisions, then current fecundity should also play an important role in determining the length of the discrimination phase. In the simplest of scenarios, we predict discrimination phase is determined by internal egg number and the rate of egg accumulation. Our first experiment investigates the relationship between discrimination phase and fecundity. We measure each butterfly's strength of oviposition preference and then estimate current fecundity at the end of each discrimination phase. Under Model 1, we expect no relationship between fecundity and discrimination phase length, because the egg load required to stimulate acceptance of the second-

ranked host is constant across the population. Individuals with short discrimination phases are simply maturing eggs faster and reaching this threshold earlier. Under Model 2, we expect a positive relationship between fecundity and discrimination phase length (Fig. 1.2a, b).

Experiment 2: Our second experiment estimates egg load when a standard length of time has elapsed since the beginning of the discrimination phase. Both models presented above have corollary predictions when the length of the oviposition search is experimentally controlled in this manner (Fig. 1.2c, d). In Model 1, short discrimination phases are caused by rapid egg accumulation rates. In this scenario, an insect with a short discrimination phase that is denied oviposition opportunities will continue to accumulate eggs at a rapid rate until some point when eggs can no longer be maintained, resulting in larger egg clusters at the end of a denial period of standard length. Insects that have long discrimination phases (and slow accumulation rates) would have a lower egg number at the end of the same denial period (Fig. 1.2c). In Model 2, we would expect no variation of egg number at all, nor any relationship between egg number and discrimination phase length because insects with short discrimination phases would continue to increase their egg number during the denial period, while insects with long discrimination phases would only just be reaching the end of their discrimination phases at the end of the denial period (Fig. 1.2d). In other words, if accumulation rates are constant and all individuals are held for the same amount of time, all individuals should have the same number of eggs at the end of the denial period.

Measuring host discrimination:

We measured discrimination phases of freshly-captured butterflies on naturally-growing plants. We captured individuals of all ages and each individual's age and condition (wing wear) were noted each day of the experiment. We allowed the insects used in these experiments access to natural nectar plants several times a day, and provided moistened mud plates in their cages as a water and salt source. Discrimination phases were only measured between 11:00 and 16:00. Female insects were caught while basking between 9:00 and 10:30 each morning. Starting at 11:00, each female was repeatedly placed on a *P. semibarbata* plant at 15 min intervals until the plant was accepted (indicated by abdominal curling and ovipositor extrusion). Upon acceptance, we removed the butterfly from the plant before any eggs were laid and then repeatedly placed it on a *C. torreyi* plant, also at 15 min intervals, until the butterfly accepted this host as well. Insects that accepted *P. semibarbata* on their very first encounter were excluded, as their discrimination phase had already begun. The time interval between the first acceptance of *P. semibarbata* and the last recorded rejection of *C. torreyi* is the minimum length of the discrimination phase and is the value used in our analyses indicating strength of preference.

Measuring fecundity:

We used egg cluster size as a measure of current fecundity. To ensure that females were not withholding eggs at oviposition, we performed a series of dissections on wild-caught butterflies not included in the two experimental designs. We dissected 17 females after they had laid on the first-ranked host and

15 females after they had laid on the second-ranked host. In both sets of females, the oviducts were empty of mature (chorionated) eggs, except in 5 individuals, all of which had four or fewer mature eggs remaining. This indicates that all available eggs were usually laid at each oviposition. Twelve additional females that were allowed to oviposit at the end of the day, then were fed, held overnight, and dissected the following morning also had no eggs in the oviducts, indicating that the insects do not mature eggs overnight.

Experiments:

Experiment 1: We allowed each insect to oviposit on the first-ranked host, *P. semibarbata*, immediately at the end of the discrimination phase. Egg cluster size was counted for each butterfly, and then plotted against length of its discrimination phase. This experiment was conducted in the flight seasons of both 1993 and 1994.

Experiment 2: For a second group of freshly-caught females we measured the discrimination phase of each insect, but then did not allow the butterfly to oviposit on *P. semibarbata* until exactly 48 hours after its discrimination phase began. Egg cluster size was counted for each butterfly, and then plotted against the length of its discrimination phase. This experiment was conducted in the flight seasons of 1993 and 1995. Due to time constraints in the 1995 season, discrimination phases were estimated every third hour, instead of every hour.

Results

For both replicates of Experiment 1, in which insects were allowed to oviposit at the end of the discrimination phase, the length of discrimination phase was positively correlated with cluster size (1993: $N = 33$, $r = 0.51$, $p < 0.01$, mean cluster size = 97.55; 1994: $N = 43$, $r = 0.56$, $p < 0.01$, mean cluster size = 70.35; Fig. 1.3).

For Experiment 2, in which all insects were held a standard 48 hours after the acceptance of the first-ranked host, there was no significant relationship between the length of discrimination phase and cluster size (1993: $N = 32$, $r = -0.27$, $p = 0.14$, NS, mean cluster size = 76.66; 1995: $N = 41$, $r = -0.18$, $p = 0.27$, NS, mean cluster size = 58.85; Fig. 1.4). Because discrimination phases are only measured during five hours of the day, the estimates of discrimination phase length in this design were truncated to 10 hours (two days of active insect time).

Females varied at the outset of our experiments, i.e., there was a range of ages, nutritional conditions, and experience with hosts in the free-flying population. There does not appear to be a bias in discrimination phase according to age or condition. In 1993, Experiment 2 was performed immediately after Experiment 1 during the flight season, but the mean age of the butterflies (according to wing wear) in Experiment 2 was greater. The difference in mean cluster size between 1993 experiments may be due to insect age. The results of the dissections confirm that cluster size is normally an accurate estimate of current fecundity, and that females are not accumulating eggs overnight.

Discussion

In Model 1 the low-ranked host is accepted at the same egg load by all individuals, and variation in strength of preference (discrimination phase length) is driven by variation in time taken to reach that egg load. This model predicts no relationship between discrimination phase and egg load in Experiment 1, but a negative correlation between these same parameters in Experiment 2. We obtained a positive relationship in Experiment 1 and none in Experiment 2, in both replicates of both experiments. These results are inconsistent with the predictions of Model 1.

On the basis of the experiments described here, we cannot reject Model 2, in which egg maturation rate is constant, and in which variation in strength of preference is driven entirely by variation in threshold egg load for acceptance of the lower-ranked host. However, the principal assumption of this model was not met. Egg maturation rate was highly variable. Three sets of observations combine to generate this conclusion: 1) cluster size was highly variable even in Experiment 2, when all butterflies were allowed the same length of time to mature eggs; 2) oviposition occurs strictly once per day in the study population (Moore 1987); 3) at each oviposition, all currently mature eggs were laid. Because the rate of egg accumulation was highly variable, we would expect any such variation would tend to bias the predictions towards those of Model 1. No such tendency appears in the data. We should therefore consider the possibility that neither model is correct, and that discrimination phases are not driven directly by egg load at all.

Suppose that discrimination phases result mechanistically from the passage of time, or some factor that changes with time, instead of egg accumulation itself. Since both egg load and the acceptance of low-ranked hosts increase over time, we would still expect them to be positively correlated in the field where different individuals will have been searching for different lengths of time. Those individuals that have already undertaken long searches will tend to have high egg load and accept low-ranked hosts, whereas those that are just beginning their searches will contain fewer eggs and be more discriminating. We would expect exactly the same phenotypic association between egg load and host choice that other authors have found with their study insects (Fitt 1986, Odendaal 1989, Odendaal and Rausher 1990), but our experimental results are not consistent with a *cause-effect* relationship between these traits in *E. editha*.

The widespread result indicating that more fecund insects are more likely to accept low-ranked hosts has fostered an assumption that observed phenotypic associations between egg load and host choice reflect direct cause-effect relationships: egg load drives host acceptance. Such effects do indeed occur, in at least one case due to genetic covariance between fecundity and host choice (Courtney and Chen 1988, Courtney and Hard 1990). However, these effects do not seem to be universal, the evidence presented here argues against acceptance of a low-ranked host being driven by egg load. We suggest that future work on this topic should ask how best to categorize insects in terms of their relationships between life history and host acceptance, rather than seeking complete generality in such relationships.

II. THE RELATIONSHIPS BETWEEN FECUNDITY, HOST PREFERENCE, AND AGE

In the previous section we showed that internal egg load itself is not the causal factor that determines the length of the discrimination phase in *E. editha*. In this section, I present data addressing the relationship between an individual butterfly's age and her fecundity and strength of host preference.

The insect life history theory literature provides conflicting predictions of how strength of host preference should vary with age. Under dynamic state variable models of optimal oviposition behavior, as insects age the likelihood that they will not survive to their next oviposition increases. Therefore, older insects should be less discriminating among potential hosts. The key assumption in this scenario is that older females are at risk of being time-limited and should maximize the number of eggs laid before they die (Jaenike 1990, Mangel 1989a). Consistent with these optimality models, Stanek et al. (1987) found that *Rhagoletis* females become more likely to accept less-preferred hosts with increasing age.

In contrast, under physiological state models, the female is at risk of being egg-limited. As a female ages, her ability to acquire nutrients and convert them into eggs decreases (Zalucki 1981). It takes more time and resources to develop a cluster of eggs for oviposition. These models suggest that because older females carry fewer eggs, (or require longer periods of time between ovipositions), they should be more discriminating among hosts.

In natural systems, however, the relationship between age and strength of preference is almost certainly more complex and probably varies among species, and perhaps even within species. It would be difficult to design appropriate experiments addressing the effects of individual age in natural populations that could only give limited sample sizes. Ideally, strength of preference, fecundity and other factors would be measured throughout an individual's life, and this would be done for many individuals in different flight seasons and populations. However, captive *E. editha* butterflies usually do not survive as long as their free-flying counterparts, and repeated handling biases age estimates due to increased wing wear. Mark-recapture studies suffer from inadequate sample sizes in any given year. During the field season of 1994 I collected data on female age, strength of preference and fecundity at the Rabbit Meadow study population. Although these data are not ideal in that they do not follow particular individuals through their lives, they represent snapshots of free-flying individual life histories.

Methods

I captured 63 female butterflies and estimated each individual's age, measured its discrimination phase, and counted the number of eggs laid. Thirty-eight of the females in this study were concurrently used in Experiment 1 described above, but because that experiment was measuring egg cluster size and strength of preference through one oviposition bout without manipulating the

amount of time between the end of the discrimination phase and oviposition, use of these individuals in both studies should not affect the results.

Estimating individual age:

The age of individual insects was estimated from the degree of wing wear at capture. This method has been consistently used in other studies (May 1992, Watt, et al. 1977, Ehrlich and Gilbert 1973) as a reliable estimator of age. I kept individual dead specimens from previous collections representing each wing wear category in transparent glassine envelopes to use as wing wear standards in the field. The age categories, in order from youngest to oldest, are: Fresh; Fresh-Intermediate; Intermediate; Intermediate-Worn; and Worn. Butterflies in the Fresh category were often captured while still teneral, and rarely showed motivation to lay, either because they had not yet mated, or perhaps because they had not yet developed enough mature eggs for oviposition. Butterflies in the Worn category often had tattered wings or were missing wings entirely.

Measuring discrimination phase:

Discrimination phases were measured according to the method described above. Rather than use absolute time to quantify length of discrimination phase, each butterfly was assigned to one of four categories based on its strength of preference as follows: 1 = neutral (accepted both host plants); 2 = accepted second-ranked host on the same day as accepted first-ranked host; 3 = accepted second-ranked host on the day after accepted first-ranked host; 4 = did not accept second-ranked host until after two days after accepted first-ranked host. I used

the same individual host plants of both *Pedicularis semibarbata* and *Collinsia torreyi* used in both experiments above.

Counting cluster size:

Butterflies were allowed to oviposit immediately at the end of their discrimination phase on sprigs of their most preferred host placed in plastic cups in order to isolate each egg cluster from any others that may be on a plant and to ease egg counting. Although not as natural as having the insects lay directly on the host plant, the number of eggs deposited on sprigs seems to reflect the number the insect would have laid on a natural plant. Additionally, this method controls for any clutch size adjustment the female may make according to preexisting clusters on the plant.

Results

Table 1.1 presents mean preference scores for each age category. Table 1.2 presents egg cluster sizes for each age category.

Discussion

The data presented in Tables 1.1 and 1.2 must be interpreted with caution due to the small sample sizes in the youngest and oldest age categories. In this sample of 63 Rabbit Meadow butterflies, both strength of preference and mean cluster size decrease with individual age. These data support Boggs' 1997 finding that daily reproductive effort declines with age in Colorado populations of *E.*

editha. These data are also consistent with the dynamic state variable models of optimal oviposition behavior, i.e., older butterflies are less discriminating among potential host plants. The data do not support the physiological state models that predict individual females would become more discriminating with age.

The data I collected from Rabbit Meadow conflict with similar data on the relationship between age and strength of preference collected by M. Singer at another *E. editha* population. At Frenchman Lake (Plumas County) in northeastern California, Singer found that the preference strength did not deteriorate significantly with age (M. Singer, pers. comm, unpublished data).

Clearly, more data are needed from both populations to bolster small sample sizes in the young and old age categories. If this population difference in age-host preference is supported, then further studies of underlying mechanisms may be warranted. Perhaps Rabbit Meadow butterflies really are more likely to be time-limited and Frenchman Lake insects are more likely to be egg-limited. Or perhaps this difference in preference strength with age is the result of a complex interaction between host preferences and the physiological state of the individual. For example, perhaps females adjust the number of eggs laid on a particular host according to host quality or some optimal cluster size that maximizes larval survival. Further investigations of the relationship between life history traits and oviposition behavior are needed to determine whether population differences, if real, reflect underlying differences in the mechanisms of oviposition behavior, and how these differences might contribute to the evolution of insect diet.

Table 1.1. Mean preference scores for each age category. N = number of individuals.

Wing wear	Mean preference score	N
Fresh	3.33	3
Fresh-Intermediate	2.95	20
Intermediate	3.00	30
Intermediate-Worn	2.43	7
Worn	1.43	3

Table 1.2. Mean egg cluster size for each age category. N = number of individuals, *s* = standard deviation.

Wing wear	Mean cluster size	N	s	standard error
Fresh	116.00	3	31.19	18.01
Fresh-Intermediate	101.90	20	42.43	9.49
Intermediate	77.63	30	41.94	7.66
Intermediate-Worn	95.00	7	42.12	15.92
Worn	57.67	3	41.41	23.91

Figure 1.1. The hierarchy-threshold model (after Courtney et al., 1989). During an oviposition search, an insect will accept a potential host only when a particular threshold number of eggs has been reached. Hosts are ranked in a preference hierarchy such that oviposition on a low-ranked host occurs only at high egg numbers and after more time has been spent searching for suitable hosts.

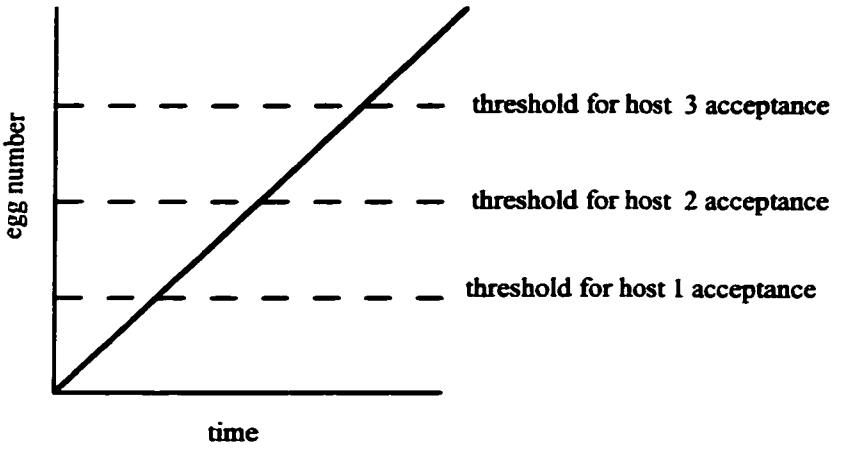

egg number
time
threshold for host 3 acceptance
threshold for host 2 acceptance
threshold for host 1 acceptance

Figure 1.2. Representation of models, predictions and experimental designs. The graphs in the first row represent predictions made by Model 1 under different experimental conditions, the graphs in the second row represent the predictions made by Model 2. The first column of graphs represents model predictions under Experiment 1 design, the second column of graphs represents model predictions under Experiment 2 design. Individuals A and B represent two hypothetical females in the same population with varying egg thresholds and/or egg accumulation rates.

Column 1: Expt. 1

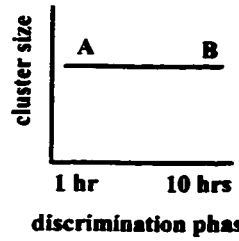
Unmanipulated

Column 2: Expt. 2

Search time standardized to 10 hrs

Model 1: Egg threshold numbers constant, Accumulation rates vary

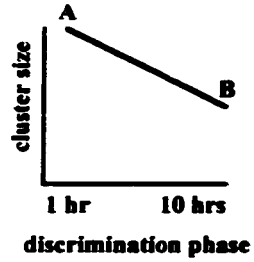
a.



individual A : threshold = 100 eggs
rate = 100 eggs/hr

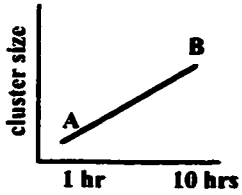
individual B : threshold = 100 eggs
rate = 10 eggs/hr

c.



Model 2: Egg threshold number varies, Accumulation rate constant

b.



individual A : threshold = 10 eggs
rate = 10 eggs/hr

individual B : threshold = 100 eggs
rate = 10 eggs/hr

d.

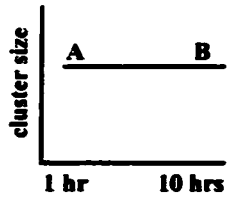


Figure 1.3. Correlations between discrimination phase length and egg cluster size for both replicates of Experiment 1. The lines of best fit for each year are shown to better distinguish each year's data; they are not meant to imply a regression.

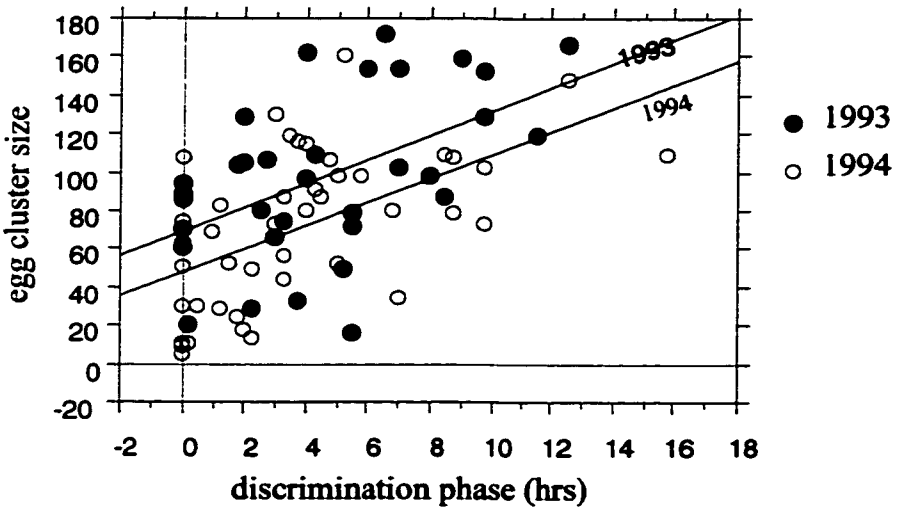
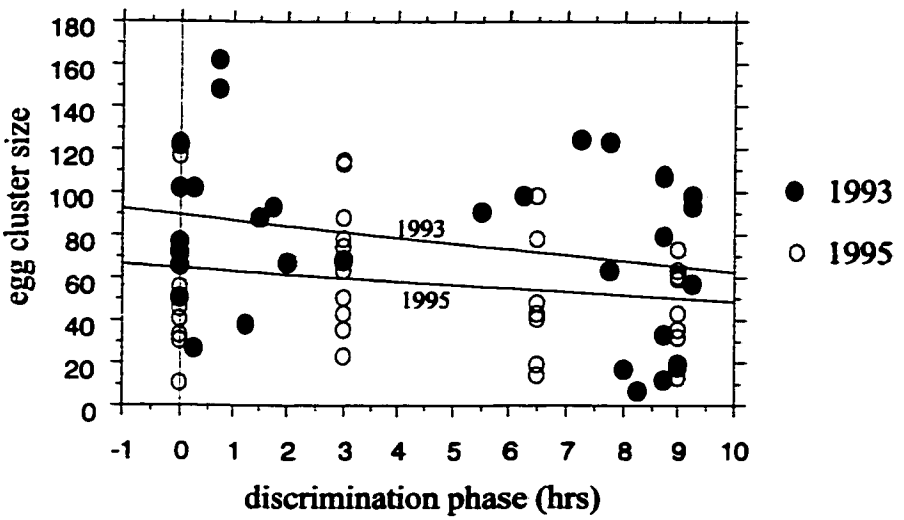

egg cluster size
180
160
140
120
100
80
60
40
20
0
-20
-2
0
2
4
6
8
10
12
14
16
18
discrimination phase (hrs)
1993
1994
1993
1994

Figure 1.4. Correlations between discrimination phase length and egg cluster size for both replicates of Experiment 2. The lines of best fit for each year are shown to better distinguish each year's data. In 1995 discrimination phases were estimated every third hour instead of every hour.


egg cluster size
180
160
140
120
100
80
60
40
20
0
-1
0
1
2
3
4
5
6
7
8
9
10
discrimination phase (hrs)
1993
1995
1993
1995

Chapter 2: Molecular phylogeography of American pikas, *Ochotona princeps (Ochotonidae)*

INTRODUCTION

Overview

Recent interest in the effects of climate change has been heightened by dire predictions of global warming and a rapidly expanding human population. However, environmental change, albeit at a slower rate, is not new. Western North America, in particular, has undergone a remarkable landscape transformation due to climate change in the (geologically) brief time period between the Pleistocene and the present. During the most extensive glaciation (150,000 years ago), continental ice sheets more than 1 km thick extended from the Arctic to northern New Mexico (Kottlowski et al. 1965, Pielou 1992). Since then, the climate and landscape have obviously changed dramatically.

The plants and animals currently found in the mountains of western North America provide an opportunity to examine the effects of such widespread climatic change on species distribution and the process of large-scale population differentiation over heterogeneous landscapes. To address these issues at the molecular level, the concordance between sequence divergence and environmental change must be established. This can be very difficult in groups of species that have ancient divergences influenced by environmental shifts--more recent genetic or environmental events may mask the effects of previous influences. However, for relatively recent divergences, relationships between

environmental change and divergence are easier to detect (daSilva and Patton 1993, Riddle 1996, Zamudio et al. 1997). One species characteristic that is amenable to such analysis is disjunct distribution correlated with habitat specificity.

The "alpine" mammals are members of a group of small mammal species typically restricted to temperate mountain tops and boreal tundra (Fitzgerald et al. 1994, Table 2.1). The alpine habitats on North American mountain tops form a series of habitat islands, similar to oceanic islands in that they are relatively isolated from other such habitats and they support ecosystems that differ from the surrounding lower elevations. The question of how alpine mammals arrived in montane habitat islands has been debated (Brown 1971, Lawlor 1998). Three models for their current distribution are:

1) Alpine mammals were once widespread in ice free periglacial areas and followed the tundra habitat up altitudinal and latitudinal gradients as the ice receded. They have since died out over most of their earlier range;
2) Mammals found on mountaintops today have only recently (since the last glacial maximum, ~18,000 years ago) dispersed there from boreal habitats along mountain corridors;
3) Alpine mammals have survived since the Pliocene in isolated populations on nunataks, high mountaintops that were never completely covered by ice.

I chose the American pika (*Ochotona princeps*) as a model organism for a study of the historical biogeography of North American alpine mammals for four reasons. First, pikas have an extensive geographic distribution, currently ranging from western Canada to northern New Mexico. Second, pikas have a good North American fossil record, indicating when and where they were present at different time points during the Pleistocene. Third, extant pikas are ecologically restricted to remote habitats that are relatively undisturbed by humans. Fourth, pikas' limited dispersal ability and regional variation within the species (particularly in behavioral characters) suggest that populations are well-differentiated and can be readily identified with common molecular markers.

The purpose of this study is threefold: 1) to describe the phylogeographic pattern of a widespread species, *O. princeps*; 2) to interpret the intraspecific phylogeny in light of geologic and climatic changes in western North America since the Pliocene; and 3) to examine the taxonomic validity of the 36 *O. princeps* subspecies currently described. Overall, this study is intended to illustrate how monophyletic assemblages can be used to recognize evolutionary lineages among populations of widespread species.

Brief geologic history of western North America during the Pleistocene

The Pleistocene glacial cycles were the most recent of a long procession of dramatic geologic forces that shaped the highly varied topography of western North America. There were four major Pleistocene glaciations (Table 2.2). Each was characterized by continental ice sheets that covered all but the highest

mountain peaks (nunataks). During the Illinoian, the ice sheet extended into the southern Rocky Mountains, then retreated to well north of the present U.S.-Canada border during the brief Sangamon Interglacial. During the following (and final) Wisconsin glaciation, the margin of the Cordilleran ice sheet extended to only just south of the international border (Fig. 2.1). High mountain ranges farther south, however, were carved by high valley hanging glaciers during this period. Average snow line altitude during the Wisconsin is estimated to have been >1660 m lower than present in the Cascades of Washington state, and about 1250 m lower in the Colorado Rockies (Flint 1971).

The land not covered by ice was dominated by periglacial tundra and coniferous forest. During interglacial periods and after the Wisconsin, the tundra habitat followed the receding ice up altitudinal and latitudinal gradients. The enormous quantities of glacial melt water formed extensive networks of shallow lakes, which changed shapes and positions as the ice front retreated (Pielou 1992). Finally, during the warmest part of the present interglacial (the hypsithermal, about 8,000 years ago) the tree and snow lines were higher than they are today. It is likely that many nunataks that had supported tundra habitat during the glaciations (and whose peaks are tundra today), would have been forested during the hypsithermal.

Biogeographic history

The biogeography of alpine mammals has been the topic of long-standing debate. In a landmark paper, Brown (1971) suggested that the diversity and

distribution of small mammals on montane islands in the Great Basin was not in equilibrium between colonization and extinction (sensu MacArthur and Wilson 1963, 1967). Rather, alpine mammals had reached all the mountain tops during the Pleistocene, but since then there has been extinction without re-colonization. The species found on mountain tops today are therefore relicts of a widespread Pleistocene tundra-adapted mammal fauna (Brown 1971).

Recently, however, Lawlor (1998) re-examined Brown's data and analyzed new data on alpine mammal species ranges and population structure. Contrary to Brown's "remnant fauna" hypothesis, Lawlor suggested that the characteristics of modern alpine mammal faunas are the result of both extensive intermountain dispersal and "extinction resistance" of remaining relict species (Lawlor 1998).

Molecular phylogeograhic studies using mitochondrial DNA may help resolve the degree to which alpine mammal populations are truly isolated from gene flow. mtDNA is maternally inherited, can be amplified with a well-developed battery of primers, and contains regions evolving at different rates, making it useful for both species and population-level studies. Different phylogenetic patterns are predicted for varying amounts of gene flow. If populations are isolated, they are expected to have strongly differentiated mtDNA sequences (associated with stable biogeographic barriers to gene flow). If, however, populations are genetically connected through ongoing or recent dispersal, they are expected to share haplotypes with other populations (Hadly et al. 1998, Riddle 1996, Avise et al. 1987).

Molecular studies of alpine mammal biogeography have been limited to regional examinations and have not addressed the phylogeography of species with continental distributions. The regional studies provide support for both the evolution in isolation and the large-scale gene flow models. For example, in the highlands of Central America, Sullivan et al. (1997) found that some members of the *Peromyscus aztecus* species group are extremely genetically differentiated from their nearest geographical neighbors, while other populations that were widely geographically separated shared haplotypes. However, these latter populations were all located on the spine of the highest cordillera, and dispersal along these habitat corridors is not unlikely. In another study, Dembowski (1999) found substantial population differentiation in Nearctic shrews (*Sorex* spp.) despite the lack of obvious dispersal barriers in tundra habitats.

In western North America there are many geographic features that are potential biogeographic barriers for terrestrial alpine mammals. Restricted to cool, mesic mountain tops, these species are not expected to disperse frequently across deep river gorges or wide, dry basins and deserts. The major basins (the Great Basin, Harney Basin, and the Wyoming Basin) and river systems (Colorado, Snake and Columbia) have been identified as dispersal barriers for other organisms (see Fig. 2.2) (Zamudio et al. 1997, Riddle 1996, Kenagy *in press*). However, the crests of mountain chains may provide habitat corridors for gene flow if the elevational displacement between smaller ranges within chains is not too great (daSilva and Patton 1993).

Pikas: the study organism

The American pika is a possible indicator species for reconstructing the biogeographic history of western North America alpine biota. Pikas are small lagomorphs (mean adult weight ≈ 150 g) found in patchily-distributed habitats ranging from southwestern Canada to northern New Mexico (see range map in Fig. 2.3).

Pikas have a good fossil record in North America. Illinoian-age fossils from the Great Basin, Appalachia, and the eastern Great Lakes region indicate *O. princeps* was found more broadly across North America. However, Wisconsin-age pika fossils have only been found in the western half of the continent, primarily in the Great, Wyoming, and Harney Basins (Fig. 2.3). Although these low-lying areas are dry and warm today, they were periglacial tundra and cool steppe during glacial maxima (Morrison 1965, Scott 1965). The southernmost Wisconsin-age record is from San Bernadino County, California, more than 100 km south of pikas' present distribution (Mead 1987).

Pikas are currently restricted to alpine talus slopes (fields of large granite boulders) at or above timberline. Talus habitat is usually formed through glacial activity (felsenmeer) and is maintained through cycles of alternating freeze and thaw conditions. Pikas live in dens under the jumbled rocks and spend the short alpine summer gathering vegetation in nearby meadows and storing it in hay piles in the talus. Pikas are active all year and feed on the hay caches through the winter under the deep snowpack.

Although each individual has its own territory, considerable territory overlap can occur between males and females (Smith and Weston 1990). Pikas are highly philopatric, tending to establish their territories near the site of their own birth, but females are more likely to disperse as a result of unsuccessful competition with male siblings (Whitworth and Southwick 1984). Mating behavior studies indicate that pikas mate with second-order relatives in greater proportion than would be expected by chance (Peacock and Smith 1997). Pikas also have a highly developed repertoire of vocalizations used in territory maintenance, male-male and male-female interactions, and predator alerts. There are geographic differences in call dialects throughout their range (Somers 1973, Conner 1982, B. Foley, pers. comm.). The major factors limiting population growth are overwinter mortality and predation by golden eagles (*Aquila chrysaetos*) and pine martens (*Martes americana)* (Johnson 1967).

The American pika is currently split into 36 subspecies, most described on the basis of pelage color variation, collection locality, and a few cranial morphology characters (Hall 1981). An early allozyme study of pikas found levels of heterozygosity that were among the lowest for any broadly-distributed mammal species ever reported (Tolliver et al. 1985). This study suggested that the low heterozygosity levels were due to population bottlenecks, particularly in the southern part of the range, which had been isolated from other sources of gene flow longer than northern populations.

A later, more extensive pika allozyme study (Hafner and Sullivan 1995) identified four major genetic units corresponding to the major mountain ranges of

the west: the northern Rocky Mountains, the southern Rocky Mountains, the Cascade Range, and the Sierra Nevada. They suggested that pikas initially spread south along cordilleran dispersal corridors in a glacial stage before the Wisconsin, and then were fragmented into isolated montane refugia during an interglacial. The Wisconsin Glaciation allowed range re-expansion and secondary contact in northern populations, but southern populations remained isolated. The resulting distance trees group the Cascade Range and the northern Rocky Mountains as each other's closest genetic relative, and do the same for the Sierra Nevada and the southern Rocky Mountains.

In this study, I use mitochondrial DNA sequence variation from 57 individual *O. princeps* to infer a phylogeny using parsimony and likelihood methods. The resulting trees are used to test hypotheses about patterns of divergence and degree of isolation of North American pika populations.

MATERIALS AND METHODS

Population sampling

Fifty-seven American pikas from locations throughout western North America were sampled in this study (Fig. 2.4). Pikas collected by me were either live-trapped using fresh vegetation bait, or were shot with a pneumatic rifle. A patch of ear tissue (< 1 cm^2) was collected from each live-trapped specimen and was stored in tissue preservation buffer (Dutton 1995) in the field for transport to

a -80°C laboratory freezer. Fresh carcasses were packed in ice for transport to the laboratory for tissue preparation, where liver, spleen, and muscle tissues were collected. Additionally, DNA was extracted from frozen tissues in the collection at the New Mexico Museum of Natural History in Albuquerque, New Mexico.

The samples include at least one representative of each of the 36 described subspecies given by Hall (1981). Throughout this study, samples will be identified by their state or province and collection location number, as shown in both Fig. 2.4 and in Appendix A, which lists specimen identifications and collection location descriptions. More than one pika was collected from each of five locations. A NMMNH sample from the sister species, the collared pika (*Ochotona collaris*), was used as an outgroup. Collared pikas are the only other New World pika species and are limited to Alaska and northwestern Canada. The two species' ranges are separated by more than 800 km.

Laboratory protocols

Tissue samples were coded so that sequencing and phylogenetic analysis could be conducted blind. DNA was extracted from slivers of frozen liver tissue using either a standard phenol/chloroform extraction protocol (Palumbi 1996), or by using 500µl of 5% Chelex resin (Bio-Rad Laboratories). Chelexed samples were incubated at 56°C for one hour, vortexed for 15 sec, heated to 100°C in a constant temperature block for 15 min, vortexed again, then stored at -20°C.

Just prior to PCR, Chelxed samples were thawed and centrifuged for 5 min to pellet the resin. Two µL of supernatant from each extraction were used in

each PCR amplification. Primers were located in the cytochrome b gene (L15774t: 5'-CAT GAA TTG GAG GAC AAC CAG T-3') and in a conserved sequence block of the control region (H16498: 5'-CCT GAA CTA GGA ACC AGA TG-3') (Shields and Kocher 1991) and consistently amplified about 720 base pairs (Fig. 2.5). The control region is a non-coding stretch of mtDNA that contains the origin of replication (displacement, or D-loop). It is the fastest-evolving region of the mitochondrial genome and is often used for studies of population-level divergence.

PCR consisted of an initial denaturation at 95°C for 5 min, followed by 35 cycles of: denaturation at 95°C for 1 min, annealing at 50°C for 30 sec, and extension at 72°C for 1 min; and then a final extension at 72°C for 5 min. PCR products were purified using Wizard PCRPreps (Promega) and sequenced in both directions on an ABI automated sequencer. For most specimens, 668 bp could be reliably read for both strands.

Lagomorphs have notoriously high levels of heteroplasmy (more than one mitochondrial haplotype within an individual) (Casane et al. 1994). PCR could potentially amplify different mtDNA sequences from the same individual. Additionally, sequence electropherogram peaks usually attributed to background could actually be a minority population of another haplotype sequence present in the sample. I chose ten individuals at random from all the samples and re-extracted DNA, then amplified and sequenced the light strand to check for multiple haplotypes within the same individual.

Data analyses

I aligned 57 *O. princeps* and one *O. collaris* light-strand sequences using the Clustal algorithm in the MEGALIGN program (part of DNASTAR, v. 1.02), (Appendix B). Gaps and gap lengths were both assigned the default penalty of 10. All sequences and the alignment have been deposited in GENBANK under accession numbers XXXXXX-XXXXXX.

There was one region of variable alignment at the beginning of the control region fragment (base pairs 246-278 in Appendix B). This 32bp indel is in the Termination Associated Sequence, which can fold into a cloverleaf structure that signals DNA synthesis termination (Brown et al. 1986). The indel was excluded from the alignment used in most phylogenetic analyses, but because it does contain some information about inter-population relationships, it was coded as an unordered character for use in the parsimony analyses (see below).

In order to estimate the distribution and amount of sequence variation, pairwise sequence comparisons were made using MEGALIGN. The g_1 statistic was calculated from a distribution of 10,000 trees to determine whether significant phylogenetic signal were present in the data. This statistic indicates the skewness of the tree length distribution: a significant g_1 indicates phylogenetic signal present in the data is distinguishable from the noise generated if rates of change are high enough to effectively randomize character states with respect to phylogenetic history (Hillis 1991). Significance levels for the g_1 statistic were assessed using the critical values given in Hillis and Huelsenbeck (1992).

Phylogenetic analyses were conducted under both parsimony and maximum likelihood frameworks using PAUP* (Phylogenetic Analysis Using Parsimony, Swofford 1998). All PAUP* analyses employed the heuristic search option with random addition of taxa, tree bisection-reconnection branch swapping, zero-length branches collapsed to yield polytomies, and the steepest descent option not in effect. Additionally, several analyses were conducted with GAML (Genetic Algorithm for Maximum Likelihood phylogeny inference; Lewis 1998) and the results compared to those from PAUP*.

Parsimony analyses

The optimality criterion in parsimony is minimal tree length. I searched for most-parsimonious trees using two weighting schemes: 1) equal weighting for all characters; 2) transversions weighted four times as heavily as transitions (ti:tv = 1:4), based on the observed transversion bias in the data. The unordered character representing the 32bp indel at the 5' end of the control region was weighted as a transversion in the weighted analysis. Tree support in both schemes was assessed using the nonparametric bootstrap (1000 replicates each).

Maximum Likelihood analyses

The maximum likelihood method chooses the hypothesis that maximizes the likelihood function for the observed data (Edwards 1972). The likelihood score provides an objective criterion for comparing specific models of sequence evolution, as well as testing phylogenetic hypotheses. I used maximum likelihood

to infer the pika tree topolgy and branch lengths, to choose which model of character evolution best fit the data, and to test biogeographic hypotheses.

In order to determine which model of evolution was appropriate for this data set, I examined a hierarchy of increasingly complex models using the likelihood ratio test statistic following the method outlined by Huelsenbeck and Crandall (1997). The likelihood ratio is

$$\Lambda = \frac{\max[L_0(\text{Null Model} \mid \text{Data})]}{\max[L_1(\text{Alternative Model} \mid \text{Data})]}$$

Starting with the simplest model (i.e., equal base frequencies, equal rates of change among all nucleotides), parameters were added and the more parameter-rich model compared to its simpler predecessor at every step until the likelihood ratio test could not detect a significant improvement in the likelihood score of the phylogenetic estimate. For each model tested, I inferred the relevant parameters on the best tree estimate from the previous model, then conducted a heuristic search constrained to the inferred parameters.

I considered the following substitution models: Jukes-Cantor (JC69; Jukes and Cantor 1969), Kimura two-parameter (K2P; Kimura 1980), Hasegawa-Kishino-Yano (HKY85; Hasegawa et al. 1985), and general time reversible (GTR; Lanave et al. 1984). I also considered three models of among-site rate variation: equal rates at all sites; rates at all sites assumed to follow a discrete approximation of the gamma distribution (Γ; Yang 1994); and a proportion of sites estimated to be invariable with gamma-distributed rates at variable sites (I +

Γ; Hasegawa et al. 1985, Gu et al. 1995). Finally, I examined whether rates among lineages are constant through time (i.e., whether there is a molecular clock).

I assumed the likelihood ratio statistic was χ^2-distributed with q degrees of freedom, where q = the difference in the number of free parameters between the models being compared. The Mathematica program was used to approximate the χ^2distribution and calculate the significance values (Wolfram Research, v. 4.0, 1998). Although the likelihood ratio test statistic may not be χ^2-distributed in some instances (Goldman 1993), Yang et al. (1995) showed that χ^2 is a reliable approximation when comparing likelihood differences between two similar, nested models.

Genetic Algorithm using Maximum Likelihood (GAML) analyses

Because maximum likelihood analyses are computationally intensive, particularly with the large number of taxa in this study, I also used GAML to search for maximum likelihood solutions with this data set. GAML is a heuristic search method recently developed by Lewis (1998) that uses a genetic algorithm to reduce the time required for maximum likelihood phylogenetic inference. Genetic algorithms are simulations of evolution by natural selection. In phylogenetic problems, each possible tree is analogous to an individual within a population, and each tree's "fitness" is proportional to its likelihood score. "Natural selection" increases the average likelihood in the tree population by only allowing the best trees to contribute to the next generation's population. "Recombination" and "mutation" are simulated by swapping branches between

trees, and by allowing random changes in tree topology and branch lengths, respectively.

Preliminary analyses on large data sets indicate that GAML reduces the amount of time necessary to conduct a maximum likelihood search by 94% compared to standard TBR maximum likelihood heuristic searches conducted in PAUP* (Lewis 1998). However, the consistency of results between different GAML searches and between GAML and PAUP* search methods has yet to be determined. I conducted four GAML analyses on the pika data set with the tree population size, mutation rates and recombination rates set to the default values used by Lewis (1998). GAML uses the HKY85 model of nucleotide substitution (Hasegawa et al. 1985). In order to test what effect number of generations the algorithm searched had on final population tree score, I used 5,000 generations for the first run, 8,000 generations for the second and third runs, and 10,000 generations for the fourth run.

Testing biogeographic hypotheses

All three methods (parsimony, likelihood, and GAML) consistently resulted in the surprising placement of some pika populations (see Results). These populations did not group with their nearest geographic neighbors, but instead were included in entirely different mountain groups traditionally thought to be isolated by substantial biogeographic barriers.

Because maximum likelihood provides a framework for the objective comparison of phylogenetic hypotheses, I constrained trees to reflect relationships predicted by biogeography (the null hypothesis) and asked whether the data I

collected could reject this null. Using the parameters from the best model of sequence evolution found, I constrained tree topologies so that relationships among populations reflected those predicted by traditional biogeographic hypotheses. In each case, the constraint tree was searched and the parameters re-estimated, resulting in a maximum likelihood tree consistent with relationships predicted by biogeography (the $H_{\varnothing}$).

The constrained tree score ($H_{\varnothing}$) was compared to the best unconstrained tree (H_1), and the likelihood ratio calculated. However, unlike the hierarchical comparison of models used earlier to find the best model of sequence evolution, the χ^2 distribution is not appropriate for this comparison of biogeographic hypotheses because the models in question are not nested.

In order to construct an appropriate null distribution, Monte Carlo simulations were used to generate 100 random data sets. Likelihoods were estimated from each simulated data set for both constrained ($H_{\varnothing}$) and unconstrained (H_1) topologies. The resulting distribution of the difference between null and alternative hypotheses was used to estimate critical values (α = 0.05) for the comparison of the constrained and unconstrained trees from the real (observed) data.

Calculating divergence times

The calculation of divergence times from DNA sequence data is controversial (Hillis et al. 1996). However, in this study, it is a useful exercise to calculate divergence times and their confidence limits for the major mountain clades, if only to illustrate how widely variable these estimates can be. These

divergence dates based on sequence data can then be compared to the dates derived from geological evidence for mountain chain isolation. The pika data set conforms to the assumptions of a "local molecular clock" (Li 1993). All the ingroup taxa are currently recognized as the same species, and have similar life histories, metabolic rates, generation times, and the rate of evolution in the control region is likely to be relatively stable across the taxa compared.

The maximum likelihood branch lengths were estimated (HKY85 + Γ + I tree with the clock enforced) and the 95% confidence limits for each point were calculated using the Poisson expectation (Sokal and Rohlf 1981). Three different divergence rates were considered: 1) 2% per million years, the predicted average mitochondrial rate (Brown et al. 1986); 2) 5% per million years, based on the inverse relationship between divergence rate and homeotherm body mass (pikas average ~ 0.150 kg, see Fig. 2 in Martin and Palumbi 1993); and 3) 11.3% per million years, the highest rate reported for mitochondrial sequence evolution (from *Mus* species, body mass ~ 0.020 kg, Martin and Palumbi 1993).

RESULTS

Sequence variation

Remarkably, of the 58 individuals sampled, only two shared a common haplotype. The remaining individuals had unique haplotypes. I found no evidence of heteroplasmy in the pikas sampled in this study: all sequences were unambiguous and ten random re-extractions had sequences identical to the

original extractions. The ranges of sequence divergence are presented in Tables 2.3 and 2.4. The tree length distribution for 10,000 randomly generated trees was strongly left-skewed ($g_1 = -0.28$, $p < 0.01$), suggesting significant phylogenetic signal in the data (Hillis and Huelsenbeck 1992).

Phylogenetic Analyses

All phylogenetic methods used identified five well-resolved clades, each corresponding to a major mountain region of western North America: the Cascades, the Northern Rocky Mountains, the Southern Rocky Mountains, the Wasatch Range, and the Sierra Nevada Range. The Sierra Nevada clade includes the Shoshone and Toquima Ranges of Nevada and the Pavant Range of Utah (Fig. 2.6). Each population sampled consistently fell into the same mountain group, and relationships within each mountain group were relatively stable across all methods used. Deeper-level relationships between mountain groups, however, were poorly resolved, and varied between different methods.

Parsimony analyses

The unweighted parsimony search recovered 7,650 equally most parsimonious trees with a length of 486 steps (Fig. 2.7). The weighted parsimony search (ti:tv = 1:4) recovered 389 equally most parsimonious trees with a length of 774 steps (Fig. 2.8). Of the 658 characters in the data set, 124 were parsimony-informative.

Maximum likelihood analyses

The HKY85 + Γ + I model was the most complex model that provided a significant improvement over the simpler models (Table 2.5). As in the parsimony analyses, pika populations clustered into five major clades corresponding to mountain ranges (Fig. 2.9).

The likelihood search incorporating the molecular clock parameter was too computationally intensive to include in this analysis. However, the initial parameter inference on the HKY85 + Γ + I tree with the clock enforced resulted in an estimated likelihood score of –3544.44 (Table 2.5). Based on the other comparisons made in this study, the estimated tree score is a reliable indicator of the score resulting from a full search. The best tree yet found (-Ln likelihood = 3471.75) was significantly better than the estimated tree with the clock enforced (a model with fewer parameters), so the hypothesis that rates among lineages are constant through time could be rejected ($H_\varnothing$ = HKY85 + Γ + I + clock vs. H_1 = HKY85 + Γ + I; -2 log Λ = 145.38, $p < 0.001$, df = 56).

GAML analyses

Each of the four GAML searches resulted in different tree topologies (Figs. 2.10-2.13). Although each population was consistently placed in the same mountain clade, the deeper relationships varied between analyses. The likelihood scores improved with the number of generations the algorithm spent searching, from –Ln l = 3835.59 for 5,000 generations, to –Ln l = 3798.37 for 10,000 generations.

Comparing trees from different search methods

Deep branch resolution

Both parsimony and maximum likelihood analyses resulted in trees with poorly resolved deep branches separating major mountain groups. In order to test whether significant phylogenetic signal were present to resolve these basal relationships, I constrained the best maximum likelihood tree to have the same within-mountain group topologies, but collapsed the deeper branches separating the five mountain groups into a polytomy. I set parameters to the maximum likelihood estimates, and performed an exhaustive search, retaining only those trees compatible with the constraint tree. I then calculated the g_1 statistic (Sokal and Rohlf 1981) from the resulting distribution of tree scores. I considered the data set to have 658 characters (the number of nucleotides), but only 6 taxa, since the branching relationships in question were the five mountain clades and the outgroup. The g_1 statistic was not significant ($g_1 = -0.4054$, $P > 0.05$), indicating no distinguishable phylogenetic signal sufficient to resolve relationships between mountain groups.

Comparing tree topologies

The positions of the major mountain clades relative to one another changed dramatically depending on the search method used, while the relationships within mountain groups was relatively stable across methods. Table 2.6 presents the major topological conflicts for ease of comparison between search methods.

Although GAML dramatically reduced computation time to complete each analysis, the reliability of the search method is questionable. The GAML results were inconsistent between runs with different generation times. There were also inconsistencies between GAML results and those obtained with heuristic maximum likelihood searches in PAUP*.

Testing biogeographic hypotheses

All phylogenetic analyses resulted in the surprising placement of two clusters of populations. In both instances, these populations were not most closely associated with their nearest geographic neighbors, but instead clustered with groups isolated by considerable distances across what are thought to be significant zoogeographic barriers. The two most unexpected results are:

1) Pavant Range populations (Utah) group with the Sierra Nevada

The Pavant Range in central Utah lies just west of the spine of the Wasatch Range (Fig. 2.8). However, populations sampled from the Pavant Range (UT 9, 10, 11) do not group with the Wasatch clade, but instead consistently fall out with the Sierra Nevada clade (Fig. 2.9). This is a surprising result: the geographically nearest Wasatch population sampled is less than 40 km to the east, while the nearest Sierra Nevada clade member is the population from the Toquima Range in central Nevada, over 360 km to the west across a valley with greater than 1700 m minimum elevational displacement (1:500,000 map, USGS 1991; Fig. 2.2). Furthermore, both the Pavant and Wasatch Range are thought to

belong to the same zoogeographic unit and are both located east of the prehistoric limits of Lake Bonneville (Lawlor 1998, USGS 1:800,000 Map 1984, Morrison 1965).

2) Medicine Bow/White River populations group with the Northern Rockies

Pika populations in the Medicine Bow Range of southern Wyoming (WY 5, WY 6), together with a population from the headwaters of the White River in northwestern Colorado (CO 1), consistently grouped with the Northern Rockies clade (Fig. 2.8, 2.9). These populations are separated from the nearest Northern Rocky clade population by more than 370 km, while it is only 98 km to the nearest Southern Rocky population. Furthermore, the Wyoming Basin (including the Washakie and Great Divide Basins) lies between WY5, WY 6, and CO1 and the southernmost Northern Rocky population, with an elevational displacement of about 1300 m (1:500,000 map, USGS 1991; Fig. 2.2). This series of deep, dry basins is thought to form a significant zoogeographic barrier (Lundelius et al. 1983, Hibbard et al. 1965).

In order to test whether these unexpected results were spurious placements, I constrained tree topologies so that relationships among populations reflected those predicted by traditional biogeographic hypotheses (i.e., the Pavant Range was forced with the Wasatch Range, and the Medicine Bow group was constrained to the Southern Rockies clade). I conducted searches on the constrained trees and calculated the likelihood ratio. In order to determine the significance of the difference between constrained and unconstrained trees, 100

random data sets were simulated to generate a distribution of likelihood ratios for each biogeographic hypothesis tested (Figs. 2.14 and 2.15).

In both the Pavant Range and Medicine Bow cases, the null hypothesis was rejected ($p < 0.05$ for both cases, Table 2.7). For both biogeographic questions, the observed value was much greater than *all* of the simulated values. These results indicate that the surprising clustering of these groups with distant populations in different mountain ranges is well-supported by the data.

Divergence times

Divergence times with 95% confidence limits were calculated under three different rates of sequence evolution between each of the major mountain clades and between *O. collaris* (outgroup) and *O. princeps* (see Table 2.8). Estimates of divergence time ranged from a minimum of 0.56 mybp for the Cascades-Southern Rockies split (assuming 11.3% divergence per million years), to a maximum of 11.97 mybp for the divergence between the two New World pika species (assuming 2.0% divergence per million years).

DISCUSSION

Current pika distribution as a result of range retraction

The models for the current distribution of pikas in western North America include: range retraction since glacial maxima; recent (since end of glaciations) dispersal to current locations; and survival on nunataks during glacial maxima.

Of these three explanations, the distribution of Pleistocene pika fossils throughout the basins of the West indicate that a range retraction following glacial retreat is most likely. If American pikas had colonized western North America before the Pleistocene and then survived the major glaciations on nunataks, or, alternatively, if pikas had migrated south from a northern source since the end of the Pleistocene, we would not expect to find pika fossils dating from the major glaciations at lower elevations. (However, the collared pika [*O. collaris*] *is* currently found on isolated nunataks in the ice fields of Alaska and northwestern Canada. How they colonized these habitat patches, some isolated from the nearest suitable habitat by more than 150 km, remains a mystery [D. Hik, pers. comm., Krajick 1998].)

The molecular phylogeny is also consistent with a rapid range retraction into isolated fragments of alpine habitat after climatic warming. In all phylogenetic analyses conducted, pika populations formed five clades that correspond to major mountain chains of the west: the Northern Rocky Mountains, the Southern Rocky Mountains, the Cascades, the Sierra Nevada, and the Wasatch Range (Fig. 2.8). Each mountain group clade is well supported, as indicated by bootstrap values for mountain clades of 85% or greater (Fig. 2.16). However, the relationships between mountain group clades are poorly resolved and provide no information about the order of colonization of different mountain regions. The lack of resolution between the major mountain groups indicates a nearly simultaneous divergence (or isolation) of these groups. The first four of these mountain regions had previously been suggested as major genetic groups of pikas

(Hafner and Sullivan 1995). The pika populations in the Wasatch Range of Utah form a discrete genetic group not identified by earlier allozyme studies.

Parsimony, maximum likelihood and GAML were used to estimate pika population phylogeny. Although the arrangement of the major clades differed under the simplest models for each method, the most complex model employed for each method (i.e., most parameter-rich or highest generation time) resulted in a common tree topology (Table 2.6, Figs. 2.7, 2.9, and 2.13). The HKY + Γ + I model of nucleotide substitution was found to be the best fit to the data under maximum likelihood. The weighted parsimony analysis resulted in the same tree topology as the HKY + Γ + I model. The GAML method found the same topology as the other two methods only in the simulation that employed 10,000 generations of tree "evolution." This 10,000 generation GAML estimate also had the best likelihood score of the four simulations performed.

Biogeographic implications

In biogeographic studies it is reasonable to expect that nearest geographic neighbors will be most closely related. However, this was not the case for the major mountain groups of pikas in this study. For example, the Southern Rocky Mountains were not sister to the Wasatch Range, but instead clustered with the Cascades. The Northern Rocky Mountains would be predicted to be sister to the Cascades, based on geographic proximity, and because secondary contact along periglacial corridors would have been possible during the Wisconsin (Waitt et al. 1983). The tree topology from this study conflicts with that of Hafner and

Sullivan (1985), which used allozyme data and distance methods to group the northern populations (Northern Rocky Mountains and Cascades) together, and do the same for the southern populations (Southern Rockies and Sierra Nevada).

Furthermore, in a fine-scale analysis of population placement, several populations did not cluster with their nearest geographic neighbors in the same mountain group, but instead with other mountain groups separated by recognized biogeographic barriers. The Utah populations in the Pavant Range consistently grouped with the Sierra Nevada clade, despite the fact that the Pavant Range populations lie less than 40 km from the nearest Wasatch population, separated only by a deep, narrow valley. Both the Wasatch Range and the Pavant Range would have been on the eastern side of prehistoric Lake Bonneville, which covered most of western Utah and eastern Nevada after the Wisconsin glaciation (1:800,000 map, USGS 1984). Forcing the Pavant Range into the Wasatch group resulted in a tree with a significantly worse maximum likelihood score than the best tree found.

Similarly, the Medicine Bow Range population (WY 5, 6) of south central Wyoming, as well as a population from the headwaters of the White River in northwestern Colorado (CO 1) consistently grouped with the Northern Rocky Mountains clade. The Wyoming Basin separates the Northern Rocky Mountains from the Southern Rockies. Since the end of the Illinoian Glaciation 132,000 years ago, the highest dispersal corridor between the Northern and Southern Rockies would have required an elevational displacement of more than 1000 meters (Scott 1965). However, a constraint tree forcing the Medicine Bow/White

River populations with the Southern Rockies was also statistically rejected in favor of the topology that places these populations in the Northern Rocky Mountains clade.

In addition to the populations bordering the Wyoming Basin just described, the Great Basin and the Harney Basin also contain pika populations with unexpected clade affiliations. In the Great Basin, Nevada populations 1 and 2, from the Shoshone Range and Toquima Range, respectively, cluster with the Sierra Nevada group, but Nevada population 3, from the Ruby Mountains in the northeastern quadrant of the state, groups with the Northern Rocky Mountain clade. Although the Ruby Mountains of northeastern Nevada are geographically closer to other Sierra Nevada populations, Lake Bonneville and a series of pluvial lakes in the north central Nevada (present 25,000 to 10,000 ybp) might have blocked all dispersal corridors except those to the northeast (USGS 1:800,000 Map 1984). Similarly, Oregon populations 4 and 5 in the Harney Basin cluster with the Sierra Nevada clade. These basin mountain ranges are surrounded by vast expanses of xeric scrubland, and modern pika dispersal between ranges is highly unlikely. However, these same basins contain the richest pika fossil deposits (Mead 1987). The best explanation is that the widespread and isolated populations that belong to the Sierra Nevada clade are relicts of a once widespread group that has since been restricted to habitat islands on mountain tops.

The Snake and Columbia Rivers do not split clades. The pika populations south of the Columbia form a monophyletic group within the Cascade Range

clade. The Colorado River does separate UT 12 from the other Utah populations, but more extensive sampling in the region would be required to determine whether the Colorado was an effective barrier to gene flow between the Southern Rockies and the Wasatch Range.

Because few studies have examined the molecular phylogeography of other alpine mammals in western North America, it is difficult to assess how well the pattern seen in American pikas represents other species. Kenagy (in press) found that the Colmbia River gorge was not a barrier to dispersal in three species of ground squirrels (*Spermophilus)*, two of which are alpine and sub-alpine species. Dembowski (1999) found levels of sequence divergence in two Alaskan tundra shrew species (*Sorex monticolus* and *S. palustris)* similar to those found between the mountain groups of pikas studied here. Although not a North American species, rock hyrax (*Procavia capensis)* populations also show strong geographic patterning with deep divergences between major South African mountain ranges. Based on these results, Prinsloo and Robinson (1992) suggested there may be more than one hyrax species in South Africa.

Sequence divergence and the molecular clock

Because mtDNA is maternally inherited, mitochondrial markers can not detect gene flow due to male dispersal. In pikas, however, females are more likely to disperse from their native territory than their brothers (Whitworth and Southwick 1984). The mitochondrial markers used in this study are therefore not likely to underestimate gene flow due to undetected male dispersal.

With the exception of a pair of individuals from the same population, every individual sampled in this study had a unique haplotype. This is a surprising result given early results from allozyme studies that indicated extremely low variation between pika populations (Tolliver et al. 1985, Hafner and Sullivan 1995). The mtDNA sequence data used in this study suggest pika populations are strongly differentiated and isolated from significant gene flow.

Divergence dates and their 95% confidence limits were calculated for each major mountain group clade under three published mitochondrial divergence rates (Table 2.8). Even if the expectations of a perfect molecular clock were assumed to be true (i.e., change is linear with time, with substitutions following a Poisson distribution, the rate of change is equal across all sites and all lineages, the phylogeny is correct, there is no error in data collection, etc.), the large error around the estimates for any rate complicates efforts to pinpoint an informative divergence time for each group.

For example, the divergence times estimated for the split between *O. collaris* and *O. princeps* from 11.97 MYA (upper limit at 2.0%) to 1.55 MYA (lower limit at 11.3%). A solid ice sheet that spread from the Arctic Ocean to the British Columbia coast during the late Pliocene (2.4 MYA) would have separated the current ranges of the two pika species (Shackleton et al. 1984). However, there is no direct evidence that this geological event caused the speciation.

The divergence time estimates are further confounded by other factors. The likelihood ratio test indicates there is undoubtedly rate heterogeneity in the pika lineages studied here. Published divergence rates of mitochondrial evolution

(Brown et al.1986, Martin and Palumbi 1993) are based on RFLP estimates averaged across protein coding genes, but unconstrained control region sequences are likely to evolve at much faster rates.

Taxonomic issues

There are currently two pika species recognized in the New World: *Ochotona collaris* (the collared pika of Alaska and northwestern Canada); and *Ochotona princeps* (the American pika). *O. princeps* is further split into 36 subspecies. Many subspecies are known only from the type locality, and few characters are given to identify individual specimens. The molecular data collected in this study indicate that the clades identified are distinct evolutionary lineages and correspond to major mountain groups of western North America. Each of the five mountain group clades identified here corresponds to subspecies that were originally described as full species (Hall 1981): Cascades (*O. p. fenisex*); Northern Rocky Mountains (*O. p. princeps);* Southern Rocky Mountains (*O. p. saxatilis*); Wasatch Range (*O. p. uinta*); and Sierra Nevada (*O. p. schistoceps*). Furthermore, recent studies have recorded diagnosable vocalization differences between populations studied in the Sierra Nevada, Northern Rocky Mountains and Southern Rockies that correspond to the clades identified in this study (Conner 1982, Foley 1998, T. Lawlor, pers. comm., pers. obser.). Vocalizations in the other mountain groups await further study.

Whether these groups should be elevated to species status, recognized as five subspecies, or remain as they are is the subject of spirited debate (Hall 1981).

However, these five taxa names provide a useful way to recognize and refer to groups with distinct evolutionary histories and which, in at least some cases, are defined by diagnosable behavioral characters. The ability to identify species is of particular concern when considering conservation issues and the policies designed to protect species and habitats at risk.

Because pika populations are isolated on habitat patches in metapopulations within mountain ranges, they are subject to both accelerated differentiation and heightened probability of extinction. With low vagility and high philopatry, pikas are particularly vulnerable to extinction due to climatic warming or stochastic natural processes. Ironically, the habitat loss predicted by rapid human-induced climatic warming may be offset by other, more direct human impacts, including mining, logging and the construction of ski runs, all of which create eroded rock slides and, in the case of maintained ski runs, foster stable patches of alpine summer meadow.

CONCLUSIONS

The phylogeographic patterns resulting from all analyses of this data set are consistent with those predicted by long-term extrinsic barriers to gene flow and/or extinctions of intermediate genotypes in species with limited gene flow (Avise et al. 1987). Although phylogenetic gaps within *Ochotona princeps* are generally geographically concordant with boundaries between traditionally

recognized zoogeographic provinces, there are some unexpected associations between some pika populations.

There are four major conclusions from this study of *Ochotona princeps* intraspecific phylogeography:

1) All analyses identified five major groups of pikas corresponding to major mountain groups in western North America, including the Wasatch mountain group, not recognized in previous allozyme studies;
2) Relationships within mountain groups are well resolved, but deeper levels in the tree are poorly resolved, resulting from a lack of significant phylogenetic signal that distinguishes relationships between mountain groups. This pattern is consistent with rapid fragmentation of a widespread species into relict mountain groups that have evolved in isolation since then.
3) There were some unexpected associations between pika populations. In particular, a) the affinity of the Pavant Range populations of western Utah with the Sierra Nevada clade, and b) the grouping of the Medicine Bow Range of southern Wyoming and northern Colorado with the Northern Rocky Mountain clade. Alternative topologies forcing these aberrant groups into clades predicted by traditional biogeography were strongly rejected.
4) Each of the 36 subspecies of *O. princeps* currently described can be assigned to one of five identifiable phylogeographic groups. These groupings are consistent with proposed classifications based on

behavioral characters. These mountain groups are most likely strongly isolated from gene flow from other groups and may help identify which groups are distinct evolutionary lineages.

Table 2.1. Alpine mammals.

Scientific name	Common name	Scientific name	Common name
Insectivores		Rodents	
Sorex monticolus	Montane shrew	*Aplodontia rufa*	Mountain beaver
Lagomorphs		*Tamais umbrinus*	Uinta chipmunk
Ochotona princeps	American pika	*Marmota flaviventris*	Yellow-bellied marmot
Sylvilagus nuttallii	Mountain cottontail	*Marmota caligata*	Hoary marmot
Lepus americana	Snowshoe hare	*Spermophilus beldingi*	Belding's ground squirrel
Carnivores		*Neotoma cinerea*	Bushy-tailed woodrat
Martes americana	American marten	*Tamiasciurus hudsonicus*	Pine squirrel
Mustela ermina	Ermine	*Microtus montanus*	Montane vole
Gulo gulo	Wolverine	*Zapus princeps*	Western jumping mouse
Felis lynx	Lynx		

Table 2.2. Pleistocene timeline. Dates are after Richmond (1965), Porter et al. (1983).

Period	Epoch	Event	approximate dates
Quaternary	Holocene		began 5-8,000 ya
	Pleistocene	Wisconsin Glaciation	11,000 - 70,000 ya
		(In western North America: Pinedale Glaciation	6,500-25,000 ya
		Bull Lake Glaciation)	32,000-~70,000 ya
		Sangamon Interglacial	
		Illinoian Glaciation	132,000 - 302,000 ya
		(Sacagawea Ridge Glaciation in w. N. A.)	
		Yarmouth Interglacial	
		Kansan Glaciation	began 610,000 ya
		(Cedar Ridge Glaciation in w. N. A.)	
		Aftonian Interglacial	
		Nebraskan Glaciation	began 788,000 ya
		(Washakie Point Glaciation in w. N. A.)	
Tertiary	Pliocene		ended 1,650,000 ya

Table 2.3. Ranges of sequence divergence within and between pika species and populations.

Comparison		Sequence divergence	Samples
Between species			
	Low	8.9%	*O. collaris* and BC 1
	High	11.0%	*O. collaris* and WY 6
Between *O. princeps* populations			
	Low	0.2%	NV 1 and NV 2
	High	9.7%	CA 5 and MT 3
Within *O. princeps* populations			
	Low	0.0%	OR 2 and OR 3
	High	1.8%	UT 1 and UT 2

Table 2.4. Sequence divergence between representative pika populations from each of the major mountain groups.

	Cascades	Northern Rockies	Southern Rockies	Sierra Nevada	Wasatch Range	*O. collaris*
Cascades	---					
Northern Rockies	7.8%	---				
Southern Rockies	5.4%	6.8%	---			
Sierra Nevada	6.8%	7.2%	7.1%	---		
Wasatch Range	6.5%	7.2%	6.5%	6.4%	---	
O. collaris	10.0%	9.7%	10.1%	10.0%	10.0%	---

Table 2.5 Legend: Results of the likelihood ratio tests. L_0 and L_1 represent the likelihoods under the null (H_0) and alternative (H_1) hypotheses, respectively. P represents the probability of obtaining the observed value of the likelihood ratio test statistic if the null hypothesis were true. Degrees of freedom are equal to the difference in the number of free parameters between the hypotheses being compared. Significance values are adjusted using a Bonferroni correction because multiple comparisons are made.

Table 2.5. **Results of the likelihood ratio tests.**

Null Hypothesis	Models compared	-Ln L	-2 log Λ	df	P
Transition rate equals transversion rate	H_0: JC69	4070.80	436.68	1	$<1x10^{-15}$
	H_1: K2P	3852.46			
Equal rates among sites	H_0: K2P	3852.46	603.62	1	$<1x10^{-15}$
	H_1: K2P + Γ	3550.65			
Equal base frequencies	H_0: K2P + Γ	3550.65	133.10	3	$<1x10^{-15}$
	H_1: HKY85 + Γ	3484.10			
A proportion of sites are invariant	H_0: HKY85 + Γ	3484.10	22.34	1	$2.3x10^{-6}$
	H_1: HKY + Γ + I	3472.93			
Two rate s among substitution types	H_0: HKY85 + Γ + I	3472.93	2.36	4	0.67
	H_1: GTR + Γ + I	3471.75			
Molecular clock	H_0: HKY85 + Γ + I + clock	est. 3544.44	est. 145.38	56	est. <0.001
	H_1: HKY85 + Γ + I	3472.93			

Table 2.6. Comparison of tree topologies among different search methods. Cas = Cascades; NR = Northern Rockies; SR = Southern Rockies; Was = Wasatch; SiNv = Sierra Nevada.

Method	Relationships among mountain groups	Tree score
Unweighted parsimony	(((SR, NR),(SiNv, Was)) Cas)	486 steps
Weighted parsimony	((((SR, Cas) NR) Was) SiNv)	774 steps
Maximum likelihood	((((SR, Cas) NR) Was) SiNv)	-Ln l = 3471.75
GAML 1 (5,000 gens)	(((SiNv, Was), (SR, Cas))NR)	-Ln l = 3835.59
GAML 2 (8,000 gens)	((((SiNv, Was) SR) NR) Was)	-Ln l = 3801.28
GAML 3 (8,000 gens)	((((SR, Cas) NR) Was) SiNv)	-Ln l = 3801.35
GAML 4 (10,000 gens)	((((SR, Cas) NR) Was) SiNv)	-Ln l = 3798.37

Table 2.7. Comparison of null and alternative models for both biogeographic hypotheses tested.

	Null Model: Tree expected from biogeography predictions	Alternative model: Tree observed from data	Δ	p
Are the Pavant Range populations (UT 9, 10, 11) members of the Wasatch Range clade?	Pavant Range constrained to Wasatch Range -Ln l = 3501.60	Pavant Range constrained to Sierra Nevada -Ln l = 3473.14	28.46	<.05
Are the Medicine Bow and White River populations (WY 5, 6, CO 1) members of the Southern Rockies clade?	Medicine Bow/ White River constrained to Southern Rockies -Ln l = 3534.49	Medicine Bow/ White River constrained to Northern Rockies -Ln l = 3473.33	61.16	<.05

Table 2.8 Divergence times between major mountain clades based on three rates of sequence evolution. Divergences are given in millions of years before present; the first and third estimate in each rate column represent the lower and higher 95% confidence limits, respectively, while the second estimate (in bold) represents the best divergence time. SR = Southern Rockies; Cas = Cascades; NR = Northern Rocky Mountains; Was = Wasatch Range; SiNv = Sierra Nevada.

Clade comparison	2.0%	5.0%	11.3%
SR , Cas	3.15	1.66	0.56
	4.16	**2.14**	**0.74**
	5.34	0.90	0.94
NR (SR, Cas)	3.76	1.50	0.66
	4.83	**1.93**	**0.86**
	6.12	2.45	1.08
Was (NR(SR, Cas))	4.01	1.61	0.71
	5.14	**2.06**	**0.91**
	6.46	2.58	1.14
SiNv(Was (NR(SR, Cas))	4.35	1.74	0.77
	5.53	**2.21**	**0.98**
	6.88	2.75	1.22
O. collaris vs. *O. princeps*	8.78	3.51	1.55
	10.16	**4.06**	**1.80**
	11.97	4.79	2.12

Figure 2.1. Map of western United States and Canada indicating extent of glaciation and lake coverage during the Wisconsin Glaciation. (After Porter et al. 1983, Smith and Street-Perrott 1983).

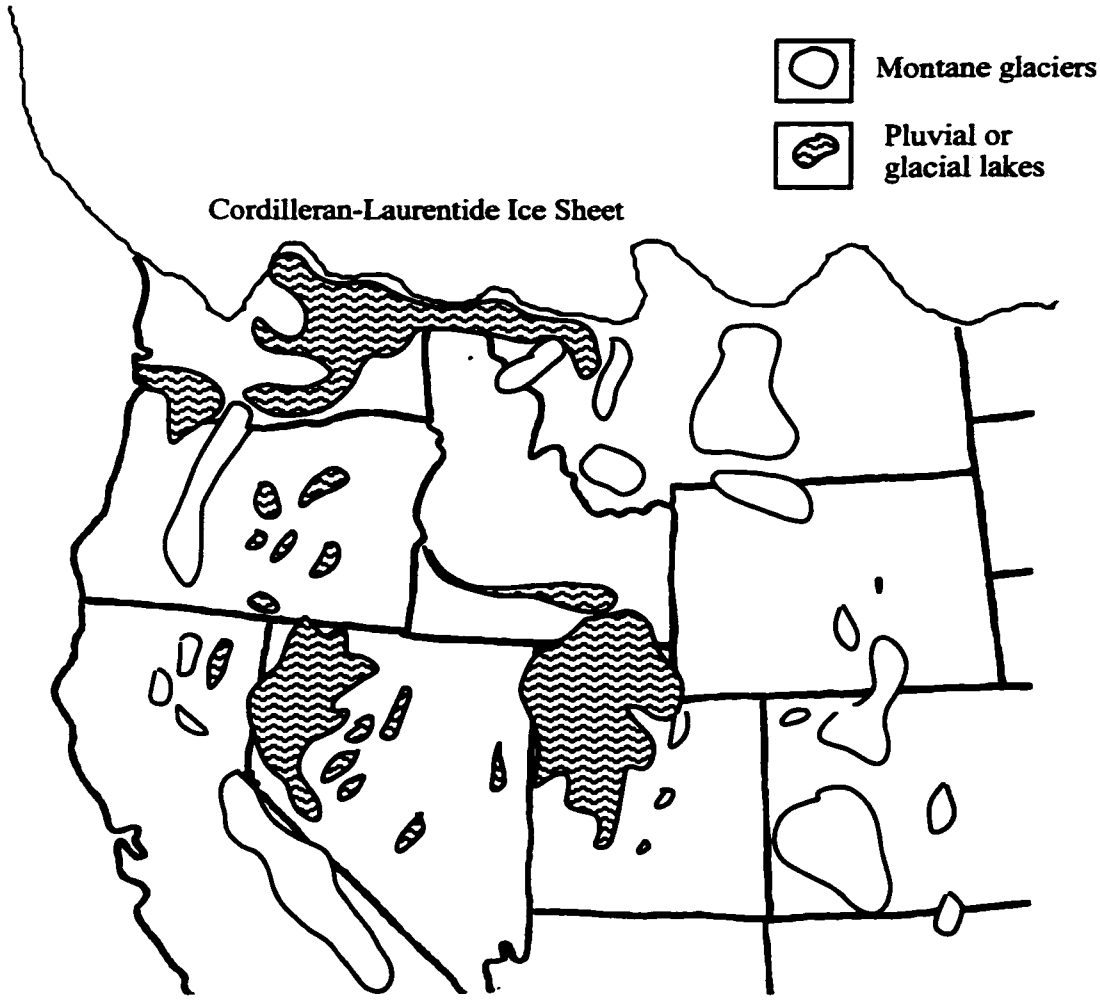

Montane glaciers
Pluvial or
glacial lakes
Cordilleran-Laurentide Ice Sheet

Figure 2.2. Major geographic features of western United States and Canada. Mountain ranges within the Great Basin are indicated by the following abbreviations: RM = Ruby Mountains, SH = Shoshone Range, TQ = Toquima Range. The Great Basin, Harney Basin, and Wyoming Basin, as well as the Snake River, Columbia River and Colorado River are all recognized zoogeographic barriers (Kenagy *in press*, Hafner and Sullivan 1995).


NORTHERN
CASCADE RANGE
Columbia R.
ROCKY
MOUNTAINS
Harney
Basin
Snake R.
Wyoming
Basin
Great
Basin
RM
SH
TQ
SIERRA NEVADA
WASATCH
SOUTHERN
ROCKY
MOUNTAINS
Colorado R.

Figure 2.3. Current distribution of the American pika, *Ochotona princeps*(after Hall 1981), shaded. Wisconsin Glaciation-age fossils of *O. princeps* (Mead 1987), black circles.

Figure 2.4. Map of western United States and Canada showing pika collection locations. Samples are identified throughout this study by state/province and number. Appendix A contains collection location descriptions, subspecies names and NMMNH specimen numbers.

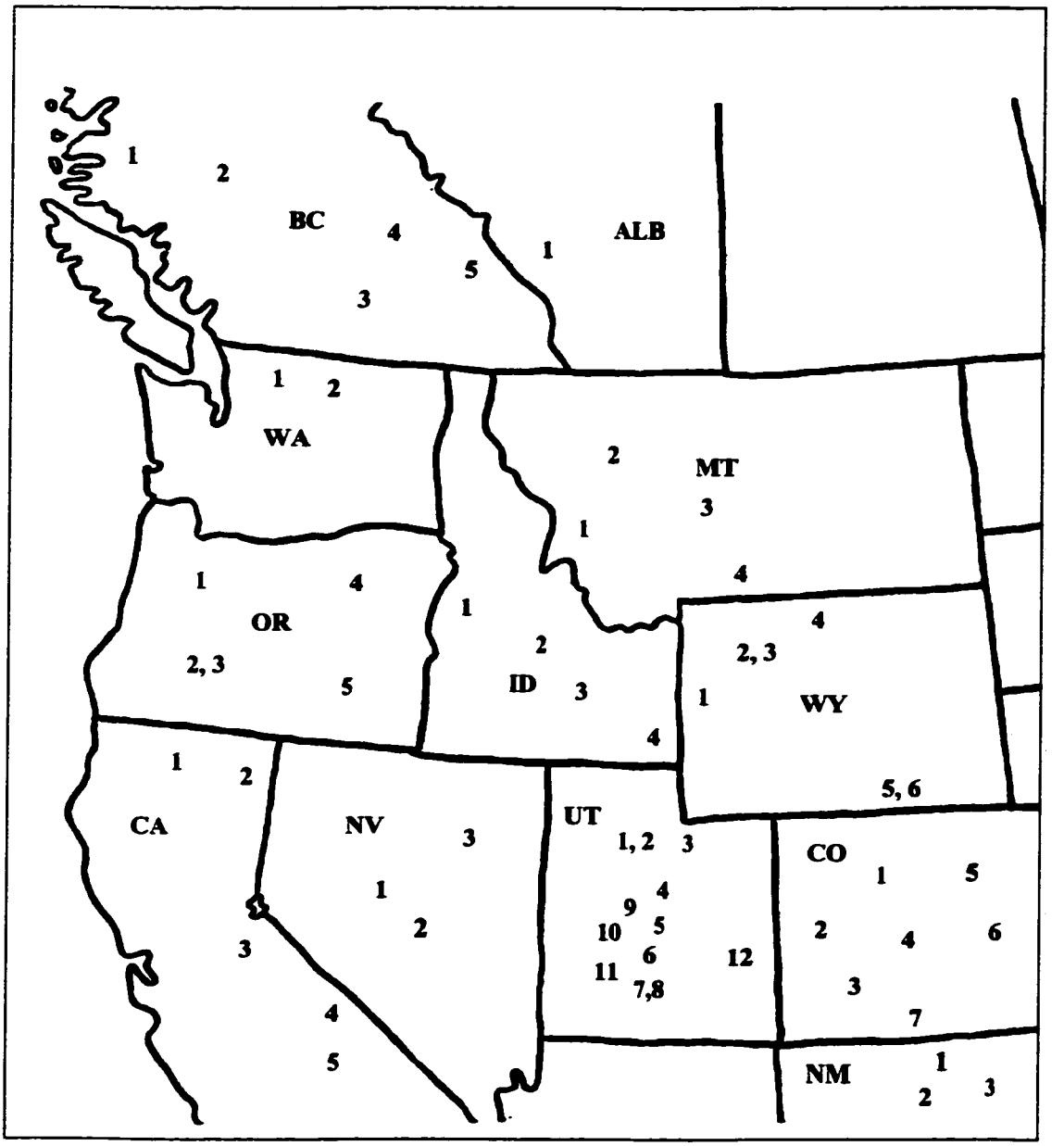

1
2
BC
4
5
3
1
ALB
1
2
WA
2
MT
3
1
4
1
4
OR
2, 3
5
1
2
ID
3
4
4
2, 3
1
WY
5, 6
1
2
CA
NV
3
1
2
3
4
5
UT
1, 2
3
4
9
10
5
6
11
7,8
12
CO
1
5
2
4
6
3
7
NM
1
2
3

Figure 2.5. Region of mitochondrial DNA amplified, showing location of primers. TAS = Termination Associated Sequence, CSB = Conserved Sequence Block. Thr and Pro indicate the tRNAs for threonine and proline, respectively.

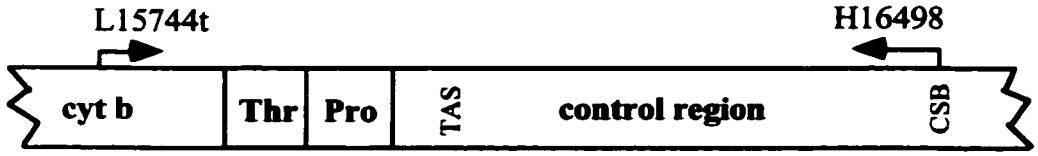

L15744t
H16498
cyt b
Thr
Pro
TAS
control region
CSB

Figure 2.6. The five major clades of *O. princeps* populations, corresponding to the mountain ranges (clockwise from upper left): Cascades, Northern Rocky Mountains, Southern Rockies, Wasatch, Sierra Nevada. These clades were found by all phylogenetic methods used. Populations UT 9, 10, 11 represent the Pavant Range of central Utah. Populations WY 5, 6 are in the Medicine Bow Range of Wyoming, and CO 1 is found at the headwaters of the White River on the Colorado Plateau.

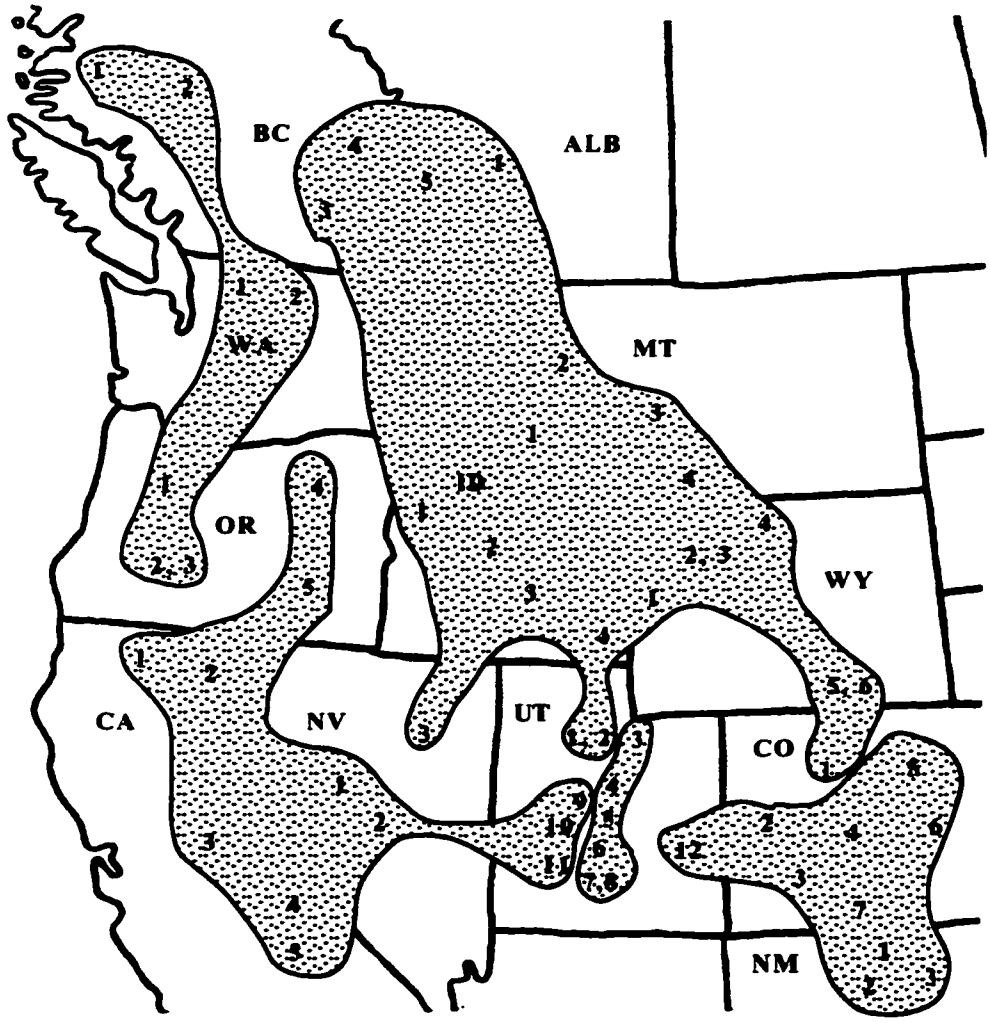

BC
ALB
WA
MT
OR
ID
WY
CA
NV
UT
CO
NM

Figure 2.7. Strict consensus of 7,650 equally most parsimonious trees (length = 486 steps) resulting from unweighted analysis. Numbers above branches are nonparametric bootstrap proportions based on 1000 replicates.

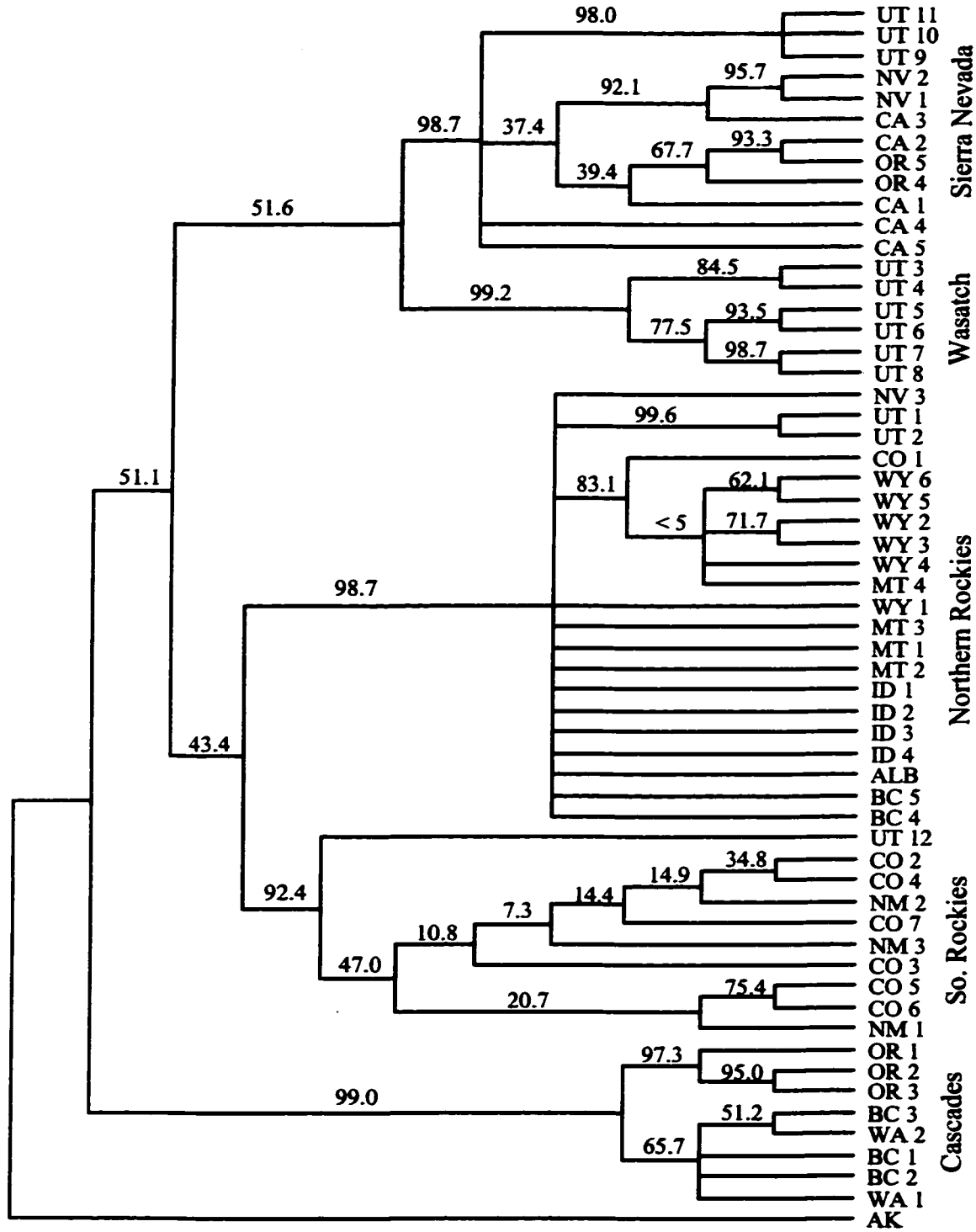

98.0
95.7
92.1
98.7
37.4
93.3
67.7
39.4
51.6
84.5
99.2
93.5
77.5
98.7
99.6
51.1
83.1
62.1
< 5
71.7
98.7
43.4
34.8
14.9
92.4
14.4
7.3
10.8
47.0
75.4
20.7
97.3
95.0
99.0
51.2
65.7
UT 11
UT 10
UT 9
NV 2
NV 1
CA 3
CA 2
OR 5
OR 4
CA 1
CA 4
CA 5
UT 3
UT 4
UT 5
UT 6
UT 7
UT 8
NV 3
UT 1
UT 2
CO 1
WY 6
WY 5
WY 2
WY 3
WY 4
MT 4
WY 1
MT 3
MT 1
MT 2
ID 1
ID 2
ID 3
ID 4
ALB
BC 5
BC 4
UT 12
CO 2
CO 4
NM 2
CO 7
NM 3
CO 3
CO 5
CO 6
NM 1
OR 1
OR 2
OR 3
BC 3
WA 2
BC 1
BC 2
WA 1
AK
Sierra Nevada
Wasatch
Northern Rockies
So. Rockies
Cascades

Figure 2.8. Strict consensus of 389 equally most parsimonious trees (length = 774 steps) resulting from weighted analysis (ti:tv = 1:4). Numbers above branches are nonparametric bootstrap proportions based on 1000 replicates.

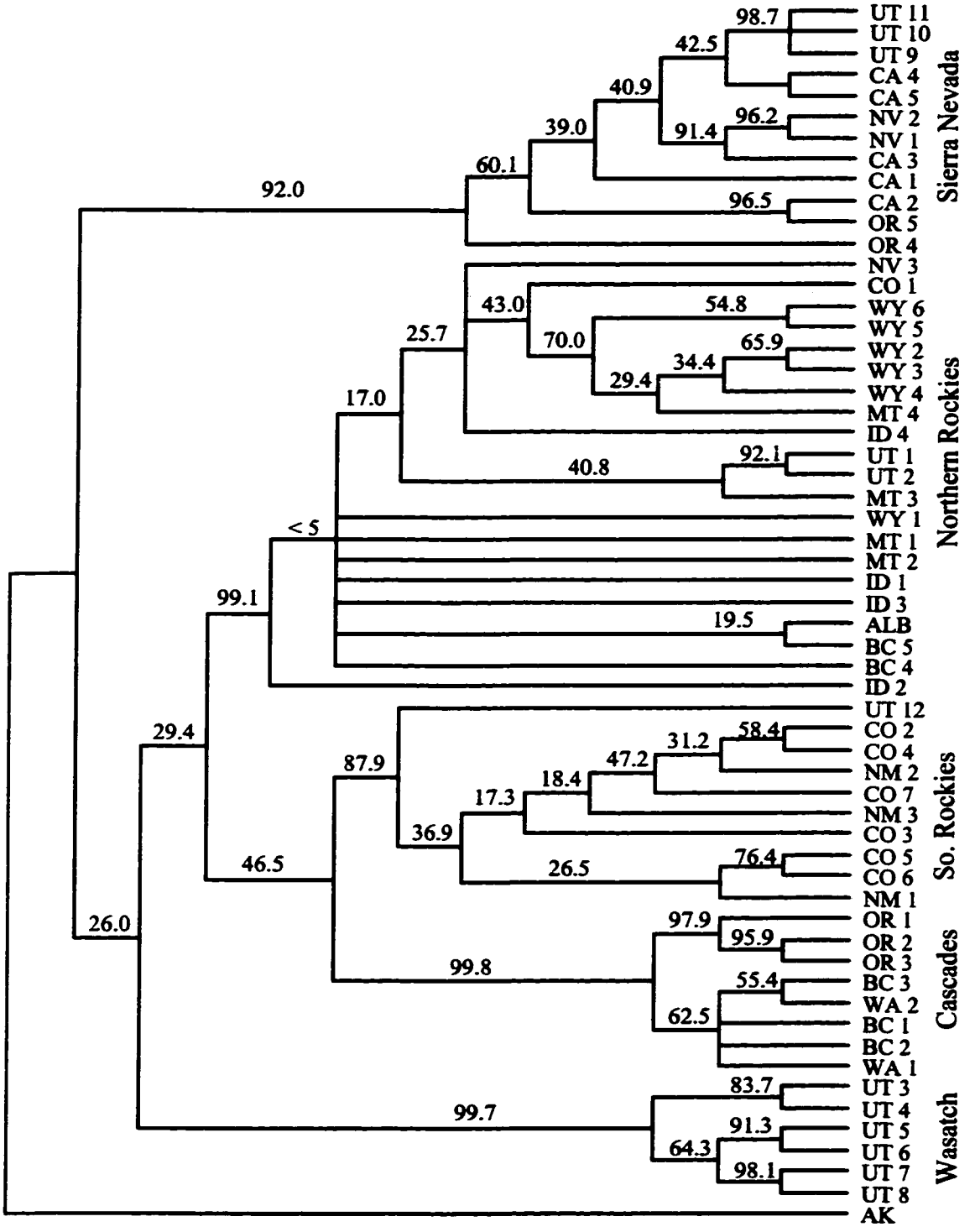

UT 11
UT 10
UT 9
CA 4
CA 5
NV 2
NV 1
CA 3
CA 1
CA 2
OR 5
OR 4
NV 3
CO 1
WY 6
WY 5
WY 2
WY 3
WY 4
MT 4
ID 4
UT 1
UT 2
MT 3
WY 1
MT 1
MT 2
ID 1
ID 3
ALB
BC 5
BC 4
ID 2
UT 12
CO 2
CO 4
NM 2
CO 7
NM 3
CO 3
CO 5
CO 6
NM 1
OR 1
OR 2
OR 3
BC 3
WA 2
BC 1
BC 2
WA 1
UT 3
UT 4
UT 5
UT 6
UT 7
UT 8
AK
98.7
42.5
40.9
96.2
91.4
39.0
60.1
92.0
96.5
43.0
54.8
25.7
70.0
65.9
34.4
29.4
17.0
92.1
40.8
< 5
99.1
19.5
29.4
87.9
31.2
58.4
47.2
18.4
17.3
36.9
46.5
26.5
76.4
26.0
97.9
95.9
99.8
55.4
62.5
83.7
99.7
91.3
64.3
98.1
Sierra Nevada
Northern Rockies
So. Rockies
Cascades
Wasatch

Figure 2.9. **Maximum likelihood tree, -Ln likelihood = 3472.93. The HKY85 + Γ + I model was used. Ti:tv = 4.80, α = 0.459, I = 0.462, f(A) = 0.337, f(C) = 0.286, f(G) = 0.112, f(T) = 0.266.**

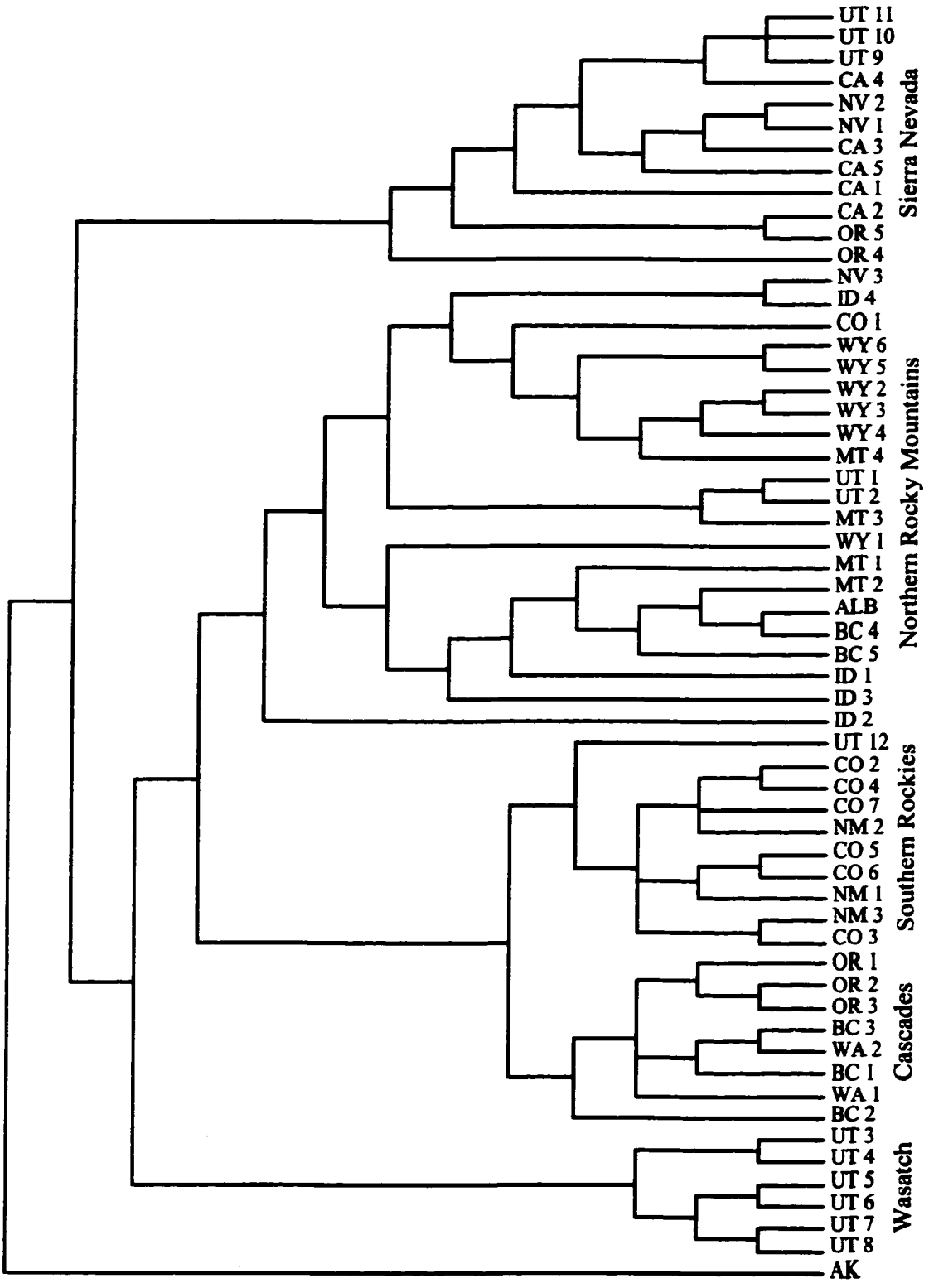

UT 11
UT 10
UT 9
CA 4
NV 2
NV 1
CA 3
CA 5
CA 1
CA 2
OR 5
OR 4
NV 3
ID 4
CO 1
WY 6
WY 5
WY 2
WY 3
WY 4
MT 4
UT 1
UT 2
MT 3
WY 1
MT 1
MT 2
ALB
BC 4
BC 5
ID 1
ID 3
ID 2
UT 12
CO 2
CO 4
CO 7
NM 2
CO 5
CO 6
NM 1
NM 3
CO 3
OR 1
OR 2
OR 3
BC 3
WA 2
BC 1
WA 1
BC 2
UT 3
UT 4
UT 5
UT 6
UT 7
UT 8
AK
Sierra Nevada
Northern Rocky Mountains
Southern Rockies
Cascades
Wasatch

Figure 2.10. GAML tree based on 5,000 generations search. -Ln likelihood = 3835.59.

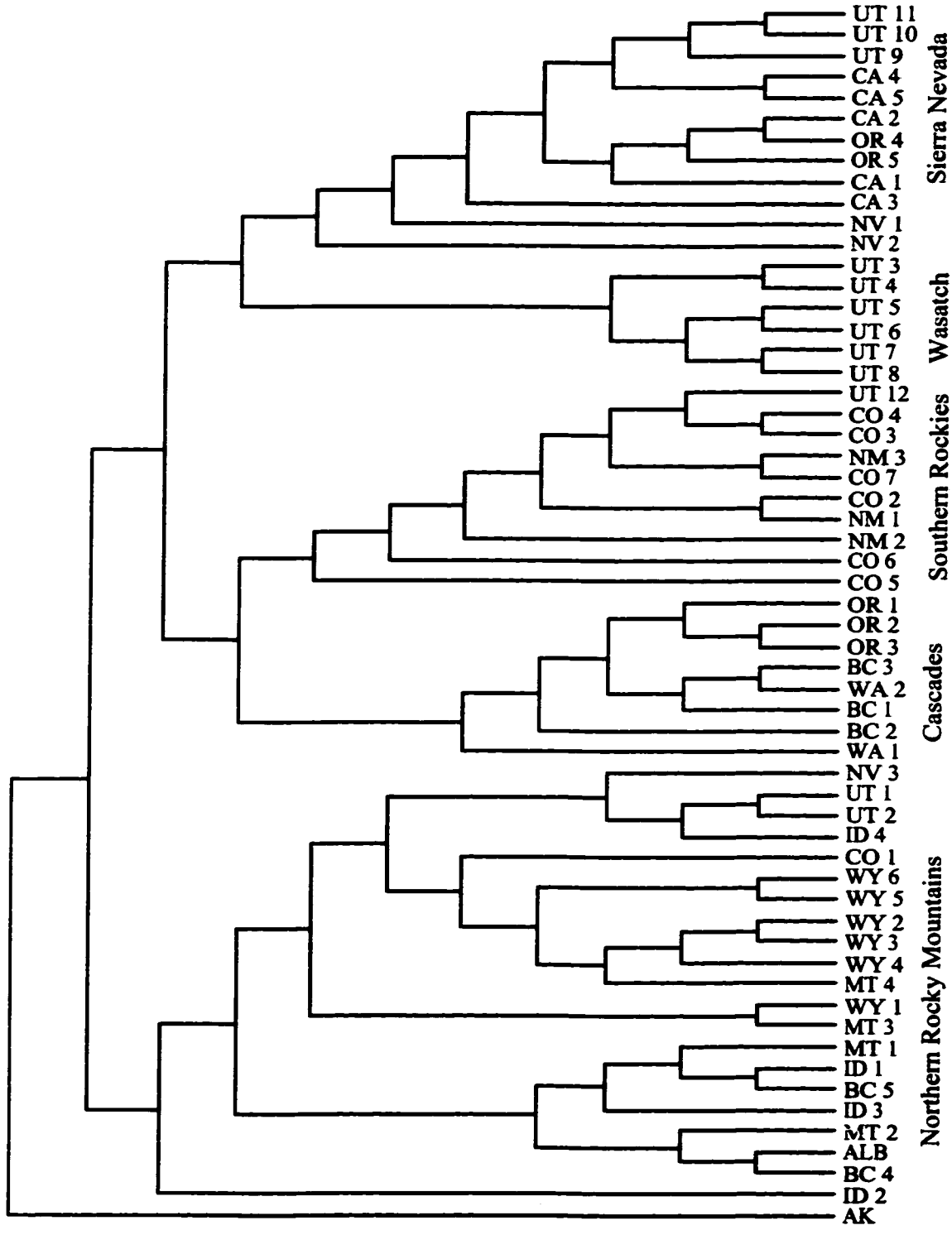

UT 11
UT 10
UT 9
CA 4
CA 5
CA 2
OR 4
OR 5
CA 1
CA 3
NV 1
NV 2
UT 3
UT 4
UT 5
UT 6
UT 7
UT 8
UT 12
CO 4
CO 3
NM 3
CO 7
CO 2
NM 1
NM 2
CO 6
CO 5
OR 1
OR 2
OR 3
BC 3
WA 2
BC 1
BC 2
WA 1
NV 3
UT 1
UT 2
ID 4
CO 1
WY 6
WY 5
WY 2
WY 3
WY 4
MT 4
WY 1
MT 3
MT 1
ID 1
BC 5
ID 3
MT 2
ALB
BC 4
ID 2
AK
Sierra Nevada
Wasatch
Southern Rockies
Cascades
Northern Rocky Mountains

Figure 2.11. GAML tree based on 8,000 generations search, first replicate. -Ln likelihood = 3801.28.

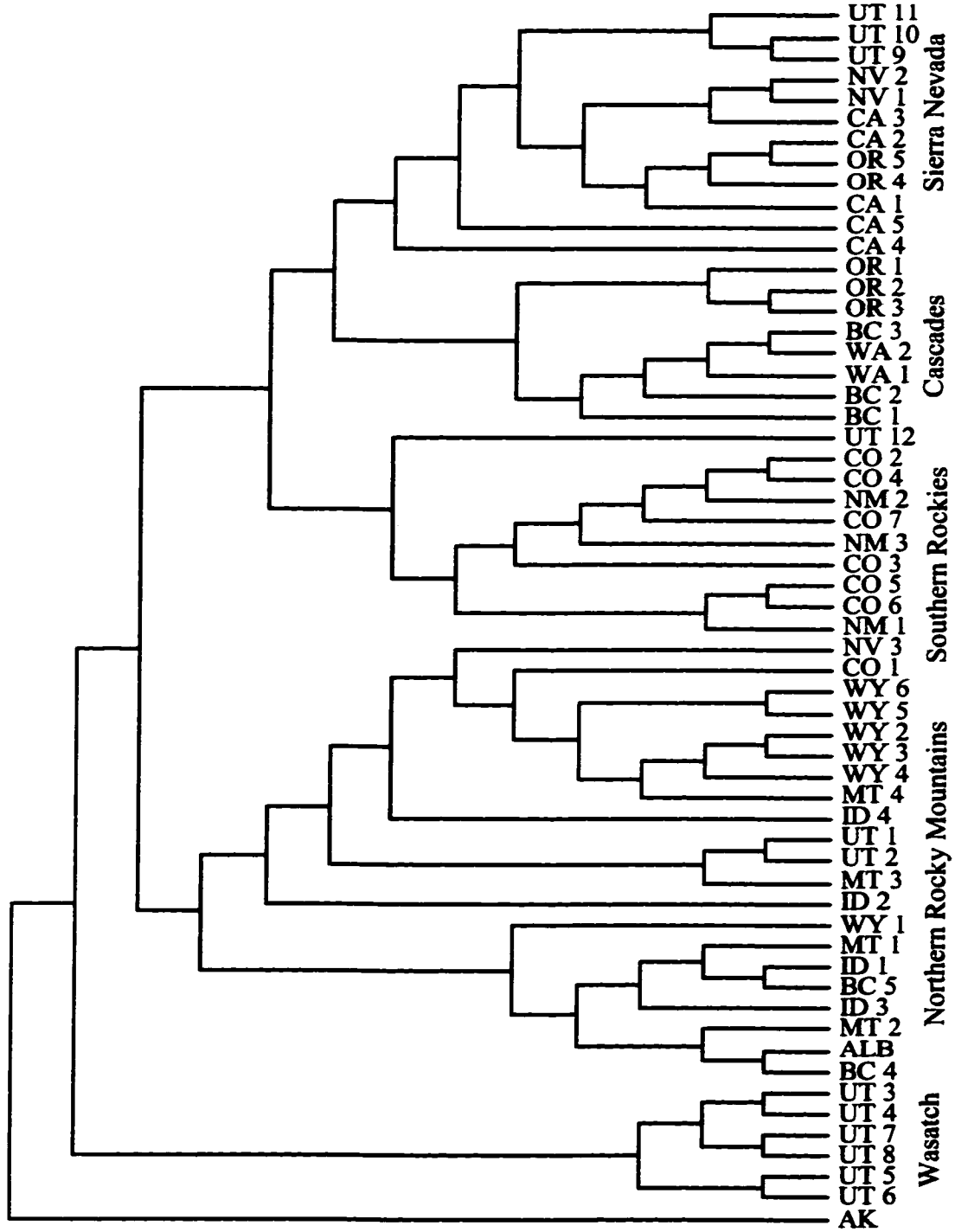

UT 11
UT 10
UT 9
NV 2
NV 1
CA 3
CA 2
OR 5
OR 4
CA 1
CA 5
CA 4
OR 1
OR 2
OR 3
BC 3
WA 2
WA 1
BC 2
BC 1
UT 12
CO 2
CO 4
NM 2
CO 7
NM 3
CO 3
CO 5
CO 6
NM 1
NV 3
CO 1
WY 6
WY 5
WY 2
WY 3
WY 4
MT 4
ID 4
UT 1
UT 2
MT 3
ID 2
WY 1
MT 1
ID 1
BC 5
ID 3
MT 2
ALB
BC 4
UT 3
UT 4
UT 7
UT 8
UT 5
UT 6
AK
Sierra Nevada
Cascades
Southern Rockies
Northern Rocky Mountains
Wasatch

Figure 2.12. GAML tree based on 8,000 generations search. -Ln likelihood = 3801.35.

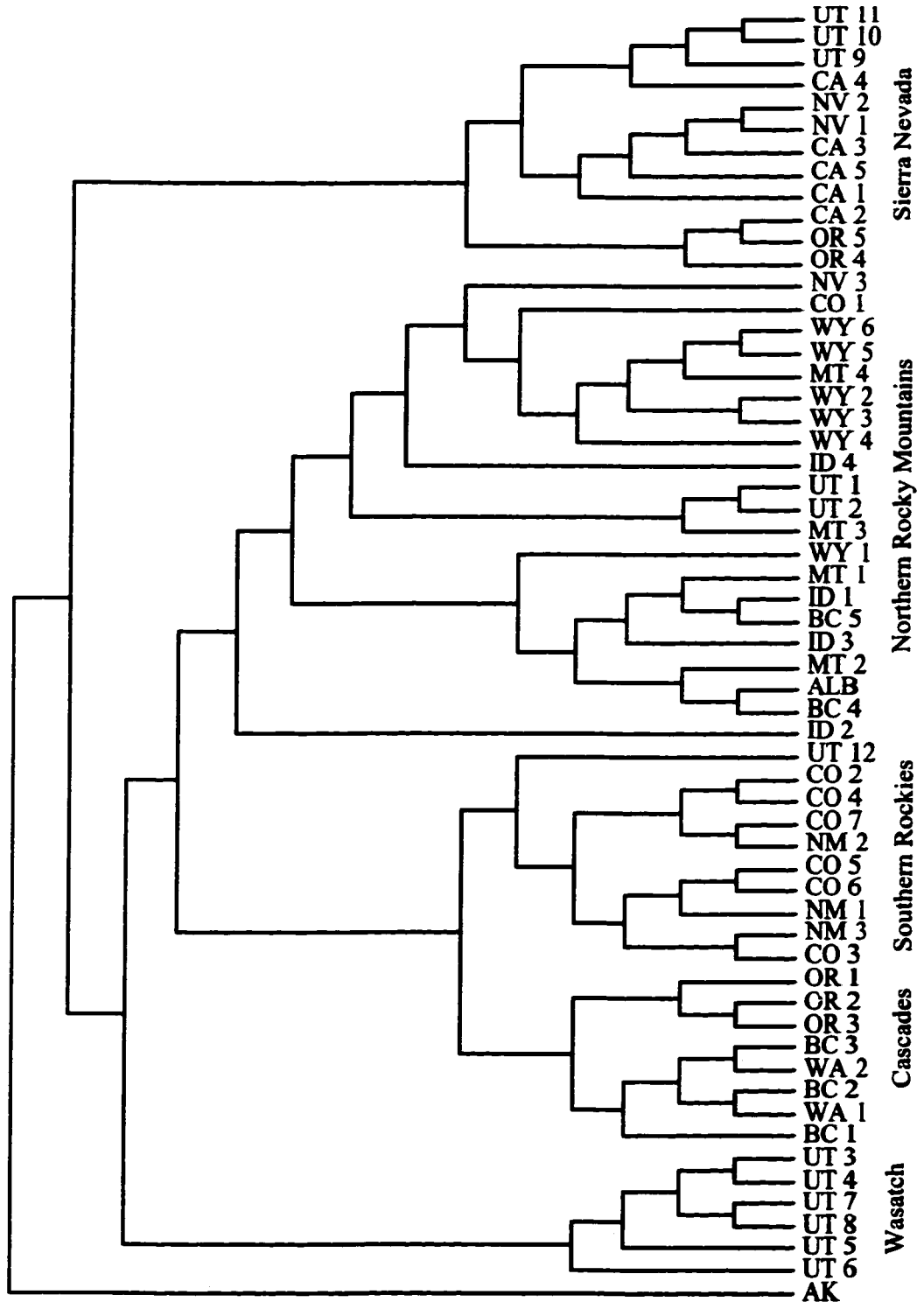

UT 11
UT 10
UT 9
CA 4
NV 2
NV 1
CA 3
CA 5
CA 1
CA 2
OR 5
OR 4
Sierra Nevada
NV 3
CO 1
WY 6
WY 5
MT 4
WY 2
WY 3
WY 4
ID 4
UT 1
UT 2
MT 3
WY 1
MT 1
ID 1
BC 5
ID 3
MT 2
ALB
BC 4
ID 2
Northern Rocky Mountains
UT 12
CO 2
CO 4
CO 7
NM 2
CO 5
CO 6
NM 1
NM 3
CO 3
Southern Rockies
OR 1
OR 2
OR 3
BC 3
WA 2
BC 2
WA 1
BC 1
Cascades
UT 3
UT 4
UT 7
UT 8
UT 5
UT 6
Wasatch
AK

Figure 2.13. GAML tree based on 10,000 generations search. -Ln likelihood = 3798.37.

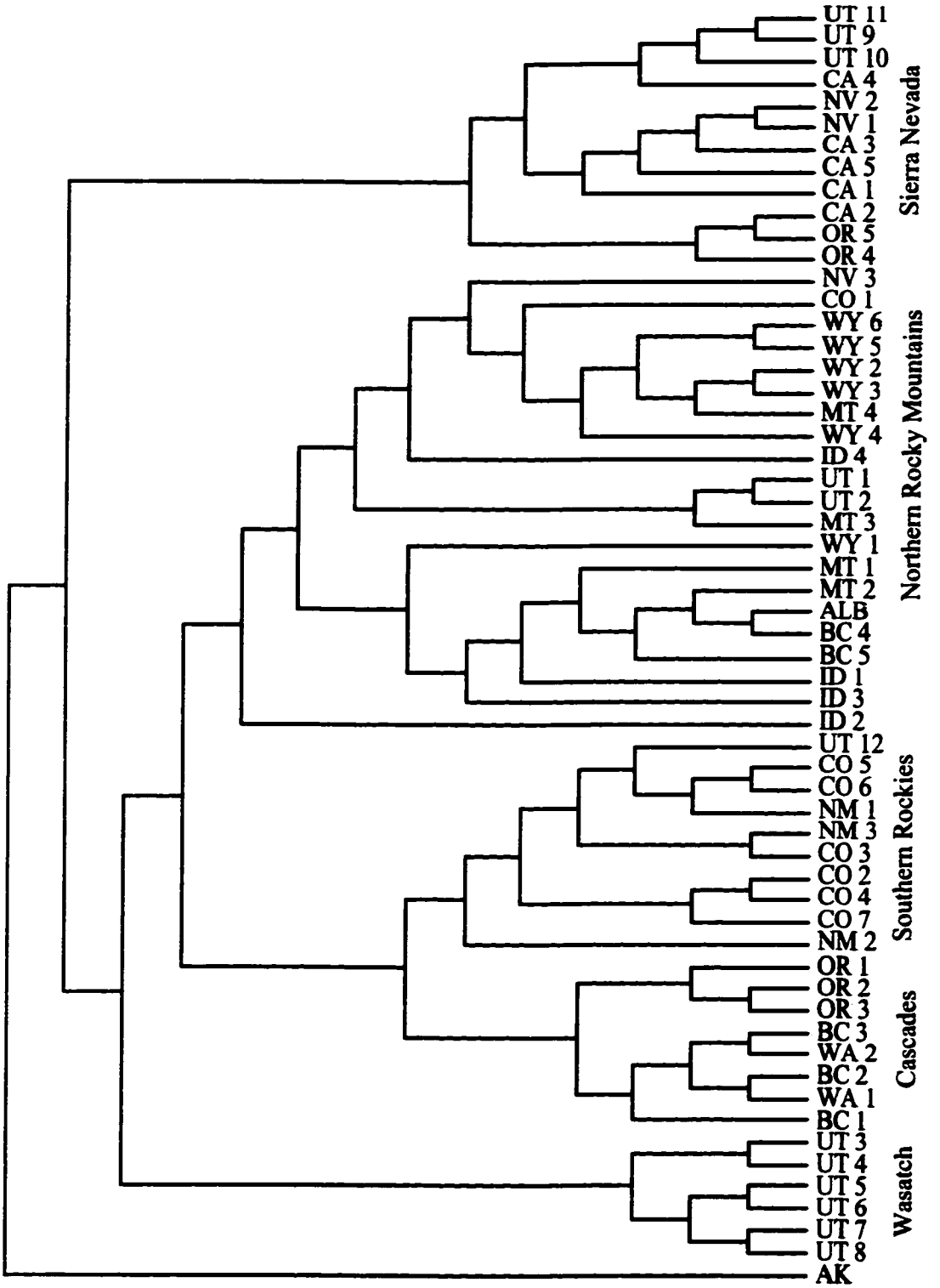

UT 11
UT 9
UT 10
CA 4
NV 2
NV 1
CA 3
CA 5
CA 1
CA 2
OR 5
OR 4
Sierra Nevada
NV 3
CO 1
WY 6
WY 5
WY 2
WY 3
MT 4
WY 4
ID 4
UT 1
UT 2
MT 3
WY 1
MT 1
MT 2
ALB
BC 4
BC 5
ID 1
ID 3
ID 2
Northern Rocky Mountains
UT 12
CO 5
CO 6
NM 1
NM 3
CO 3
CO 2
CO 4
CO 7
NM 2
Southern Rockies
OR 1
OR 2
OR 3
BC 3
WA 2
BC 2
WA 1
BC 1
Cascades
UT 3
UT 4
UT 5
UT 6
UT 7
UT 8
Wasatch
AK

Figure 2.14. Distribution of tree score differences between the null and alternative models from 500 simulated data sets. Null model: Pavant Range popualtions (UT 9, 10,11) constrained to Wasatch Range clade. Alternative model: best tree found (global optimum). The distribution is used to estimate the critical value for comparison between the constraint tree and the best tree found from the observed data.

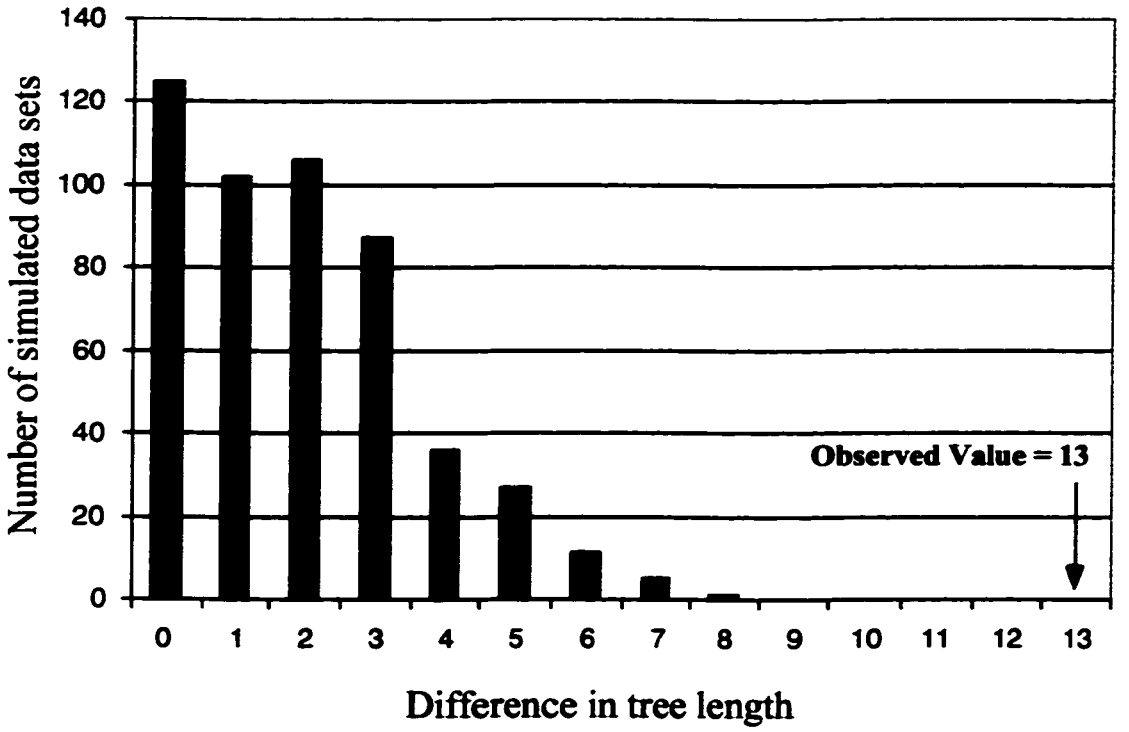

140
120
100
80
60
40
20
0
Number of simulated data sets
Observed Value = 13
0 1 2 3 4 5 6 7 8 9 10 11 12 13
Difference in tree length

Figure 2.15. Distribution of tree score differences between the null and alternative models from 500 simulated data sets. Null model: Medicine Bow Range (WY 5, 6) and White River (CO 1) populations constrained to Southern Rockies clade. Alternative model: best tree found (global optimum). The distribution is used to estimate the critical value for comparison between the constraint tree and the best tree found from the observed data.

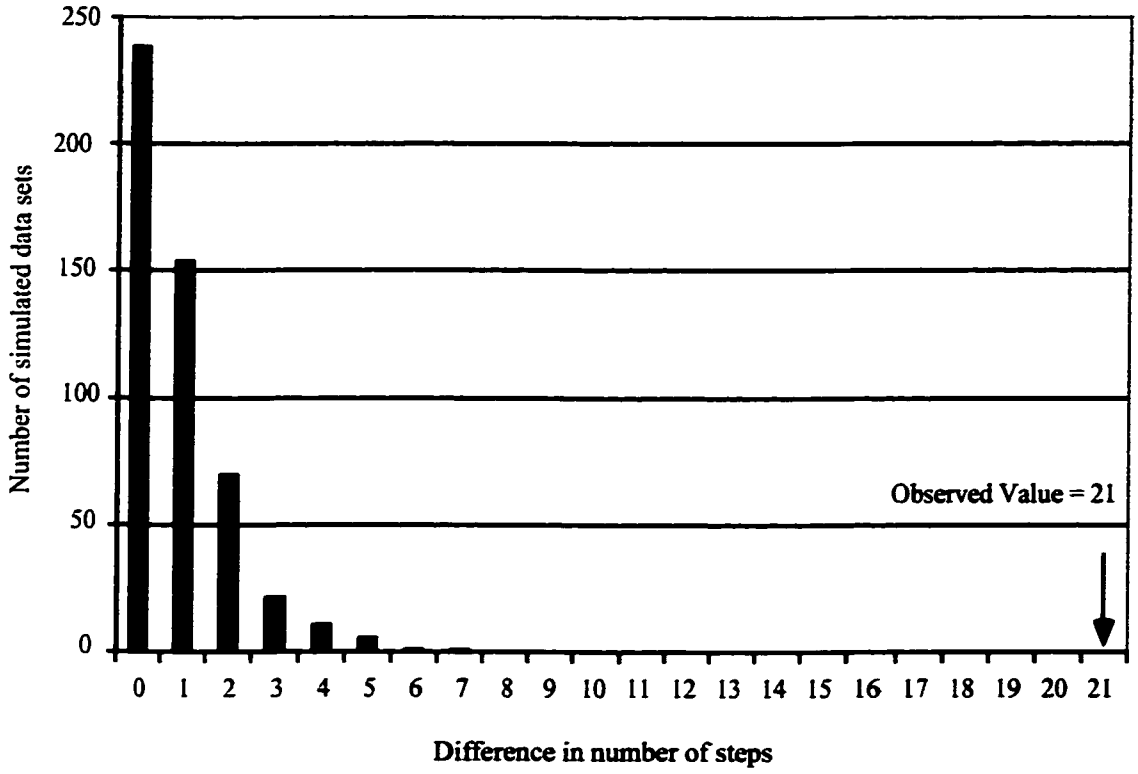
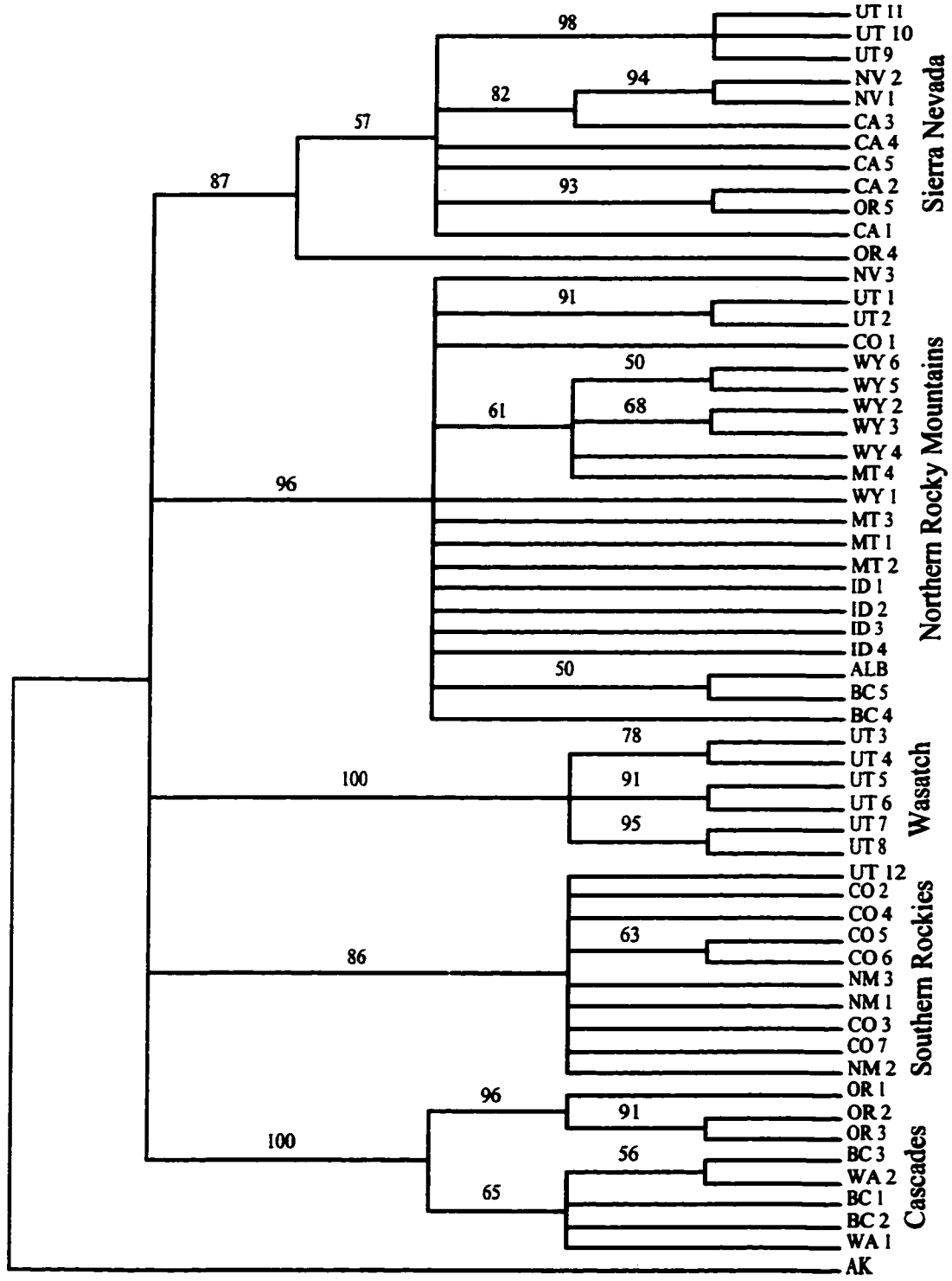

250
200
150
100
50
0
Number of simulated data sets
Observed Value = 21
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21
Difference in number of steps

Figure 2.16. **Strict consensus maximum parsimony tree with bootstrap values indicated for nodes with ≥ 50% support.**


UT 11
UT 10
UT 9
NV 2
NV 1
CA 3
CA 4
CA 5
CA 2
OR 5
CA 1
OR 4
NV 3
UT 1
UT 2
CO 1
WY 6
WY 5
WY 2
WY 3
WY 4
MT 4
WY 1
MT 3
MT 1
MT 2
ID 1
ID 2
ID 3
ID 4
ALB
BC 5
BC 4
UT 3
UT 4
UT 5
UT 6
UT 7
UT 8
UT 12
CO 2
CO 4
CO 5
CO 6
NM 3
NM 1
CO 3
CO 7
NM 2
OR 1
OR 2
OR 3
BC 3
WA 2
BC 1
BC 2
WA 1
AK
98
94
82
57
87
93
91
50
61
68
96
50
78
100
91
95
63
86
96
91
100
56
65
Sierra Nevada
Northern Rocky Mountains
Wasatch
Southern Rockies
Cascades

Chapter 3: Lentivirus phylogeny and sequence evolution

INTRODUCTION

Lentiviruses are exceptionally complex retroviruses that infect mammals. Compared to other known viruses, they are genetically unstable in many regions of their genome, undergoing substantial change in a very short time. Presumably, much of this change is driven by host immune systems. Lentiviruses also have conserved gene regions essential for viral replication. The purpose of this study is twofold: 1) to estimate the phylogeny of known lentiviruses using conserved gene regions; and 2) to examine the origin and evolution of a gene fragment in the lentiviral polymerase gene previously suggested to have been acquired by horizontal transfer.

System background

The Lentiviridae is a group of double-stranded RNA viruses that includes Human Immunodeficiency Virus (HIV), Simian Immunodeficiency Virus (SIV), and Equine Infectious Anemia Virus (EIAV). Lentiviruses are part of the Retroviridae, the family of retroviruses that also includes the "foamy viruses" (Spumaviridae) and the "tumor viruses" (Oncoviridae). Lentiviruses are only known to infect mammals, and to date have been described in primates, cattle, sheep, goats, horses, and cats (Table 3.1). "Lenti-" means "slow" and unlike the oncoviruses, lentiviruses do not usually cause rapidly-growing cancers, but

instead result in slower courses of disease, including immune dysfunction, anemia, and central nervous system disorders.

Lentiviruses are extremely complex and have been historically important in the study of viral evolution. The first virus ever described (in 1904) was a lentivirus (EIAV) (Temin 1992). Continued study led to the discovery of reverse transcriptase, the polymerase used by retroviruses, retroelements, and retrotransposons to generate DNA from RNA. Of course today, considerable resources are devoted to the study of HIV and the resulting disease, AIDS.

Viral life cycle

Of the viruses making up the Retroviridae, most of the oncoviruses are considered "simple," while the lentiviruses and the spumaviruses are "complex." Simple viruses enter the host cell, reverse transcribe their three major genes in the cytoplasm and then integrate into the host genome. The viral genome then uses the host cell machinery to transcribe, splice and translate. The resulting viral proteins are assembled in the cytoplasm, then the virion particle leaves the host, usually by budding out of the plasma membrane. Complex viruses follow a pathway similar to that of simple viruses, except that they also have regulatory proteins, which are transcribed and translated first, before *gag*, *pol* and *env*. Regulatory proteins then control the production of other viral structural and non-structural proteins (Fig. 3.1).

Genome organization and base composition

Both simple and complex retroviruses have three major genes: *gag*, *pol*, and *env*. *gag* and *env* are both structural proteins: *gag* is responsible for capsid proteins, while *env* codes for the glycoproteins that are presented on the virion surface. *pol* encodes a large polyprotein that includes protease, reverse transcriptase, RNaseH, and integrase. The protease is thought to cleave the *pol* polyprotein from *gag* after both have been translated. The reverse transcriptase is the polymerase that generates DNA from the RNA viral template. RNaseH digests the RNA template moiety of the DNA-RNA hybrid so that it can be replaced by another strand of DNA. Integrase cuts the host DNA and ligates the viral DNA genome (made by reverse transcription) into the host genome.

In lentiviruses, the *gag*, *pol*, and *env* genes have remarkably high adenine contents, on the order of 35% A, while the regulatory proteins (*tat*, *nef*, *rev*) have nearly normal A contents (i.e., about 25%). Almost all other RNA and DNA-based organisms, including other retroviruses, also have about 25% A (Myers and Pavlakis 1992).

***pol* gene insertion/deletion**

McClure, et al. (1987) described an indel of approximately 130 amino acids in the *pol* gene (which they interpreted as an insertion) of Visna and EIAV that was absent in primate lentiviruses. Based on amino acid sequence similarity, they suggested that the inserted region was a duplication of the protease gene

from oncoviruses that had been horizontally transferred to the lentiviruses sometime since the primate and non-primate lentiviruses diverged (Fig. 3.2).

Since McClure and her colleagues published their finding, many more lentiviruses have been sequenced, and the anomalous region has been identified as deoxyuridine triphosphatase (dUTPase) in FIV (Elder et al. 1992). dUTPase is involved in nucleotide metabolism, degrading uridine residues during nucleotide incorporation in DNA polymerization. To date, no primate virus has been found to have the dUTPase region.

In this study, I used *pol* sequence from 17 lentiviruses to estimate a phylogeny of the lentiviruses. The *pol* protein is the most evolutionarily conserved across all retroviruses. Previous studies of relationships among lentiviruses were hampered by the small number of sequences available and limited computational power (Sharp et al. 1995, Myers and Pavlakis 1992).

In order to reexamine the origin of the insertion/deletion in lentivirus *pol* genes, I conducted database searches for homologous sequences. I then compared the base composition of lentiviral *pol* sequences with that of similar sequences recovered in the homology searches. If the dUTPase region present in some lentiviruses was recently horizontally transferred from oncoviruses, the inserted region is expected to have a lower A content than the rest of the lentivirus *pol* gene.

MATERIALS AND METHODS

Lentiviral Phylogeny

All sequences were from GenBank. Table 3.2 lists the seventeen lentiviral sequences used. Protein sequences from the *pol* gene of each virus were aligned using the Clustal W alignment program, and then were back-translated into their corresponding GenBank nucleotide sequences using a program written by M. Holder. Regions at the beginning and end of the sequences that had missing data, and the large insertion/deletion region that was present in some viruses but not in others (see below) were temporarily excluded from the data set for phylogenetic analysis. The complete nucleotide alignment (including the excluded regions, set off by brackets) for 17 lentiviral *pol* genes is included in Appendix C.

All phylogenetic analyses were conducted with maximum likelihood using PAUP* (Phylogenetic Analysis Using Parsimony, Swofford 1998). All PAUP* analyses employed the heuristic search option with random addition of taxa, tree bisection-reconnection branch swapping, zero-length branches collapsed to yield polytomies, and the steepest descent option not in effect.

Early studies of lentiviral phylogeny used several different viruses as outgroups, including the Moloney Murine Leukemia Virus (MoMLV), Human T-cell lymphotropic viruses type 1 and 2 (HTLV-1, HTLV-2), and the human spumaretrovirus (HSRV) based on their assumed membership in the other retroviral groups that were not lentiviruses (Sharp et al. 1995, Myers and Pavlakis 1992). Attempts to align *pol* protein sequences from these viruses with lentiviral *pol* sequences were frustrated by the extensive sequence divergence that has taken

place between lentiviruses and other retroviruses. Using such a distant outgroup can be problematic because the number of changes along the outgroup branch may be so great that the location of the root is effectively randomized (Swofford et al. 1996). Although not ideal, an unrooted estimate is preferred to a tree that is rooted with an outgroup as distant as these other viruses.

Because maximum likelihood provides a statistical framework for comparing alternative models, the nucleotide sequences from the 17 lentiviruses were first analyzed under the simplest model of sequence evolution (equal base frequencies, equal rates of change), and then compared to increasingly more complex models. I considered the following substitution models: Jukes-Cantor (JC69, base frequencies and rates of nucleotide change are equal; Jukes and Cantor 1969), Kimura two-parameter (K2P, transitions not equal to transversions; Kimura 1980), Hasegawa-Kishino-Yano (HKY85, base frequencies vary; Hasegawa et al. 1985), and general time reversible (GTR, rates vary among substitution types; Lanave et al. 1984). I also considered three models of among-site rate variation: equal rates at all sites; rates at all sites assumed to follow a discrete approximation of the gamma distribution (Γ; Yang 1994); and a proportion of sites estimated to be invariable with gamma-distributed rates at variable sites ($I + \Gamma$; Hasegawa et al. 1985, Gu et al. 1995). Finally, I examined whether the data fit the model of constant rates through time (i.e., whether there is a molecular clock).

The likelihood ratio test was used at each step in the hierarchy of model complexity to determine whether the additional parameters significantly improved

the tree score (Huelsenbeck and Crandall, 1997). The test statistic was assumed to follow a χ^2 distribution, with the degrees of freedom equal to the difference in the number of free parameters between the competing hypotheses. Critical values were Bonferroni corrected to adjust for multiple comparisons. Finally, the reliability of individual tree branches was assessed with the non-parametric bootstrap (1000 replicates).

pol gene insertion/deletion

In order to reexamine the origin and evolution of the 130 amino acid region found in these lentiviruses, I compared fragment length variation and sequence similarity within the non-primate lentiviruses. Then, using the National Center for Biotechnology Information (NCBI) BLAST search, I searched the GenBank database for similar sequences.

When deciding whether a given alignment constitutes evidence for homology, it is useful to know how strong an alignment can be expected from chance alone. Traditionally, the significance of sequence similarity has been calculated by generating many random sequence pairs of appropriate length, calculating an alignment score for each pair, and then expressing the score of interest in terms of standard deviations from the mean (McClure et al. 1987). However, these scores can not be assumed to be normally distributed, and therefore the standardized *z* scores can not be converted to p values (Gonnet et al. 1992). The recently updated BLAST local sequence comparison instead calculates E values as the expected number of high-scoring segment pairs in a

large database. When E values are < 0.01, p values and E values are nearly identical (Altschul et al. 1994).

After searching the database for protein sequences most similar to the anomalous fragment, alignments proved too difficult to generate a phylogeny of the fragments due to high sequence divergence and limited nucleotide sequence availability. Because lentiviruses characteristically have remarkably high A contents, I compared the base composition of the dUTPase region in lentiviruses to other lentiviral sequence regions, as well as to similar sequence fragments recovered from the BLAST search.

RESULTS

Lentiviral phylogeny

Likelihood ratio tests indicated that the most complex model considered, GTR + Γ + I was the best model of sequence evolution and rate heterogeneity (Table 3.3, Fig. 3.3). Outgroup designation (which was necessary for the moleuclar clock tests) was done by assigning primate viruses to the outgroup and the remaining viruses to the ingroup, since all previous models tested resulted in a monophyletic primate virus group. The molecular clock assumption was rejected (see Table 3.3).

The primate lentiviruses formed a monophyletic group, but the two strains of HIV were not monophyletic with respect to non-human primate viruses. The

equine lentivirus strains sampled were basal to the clade comprised of sheep, goat and bovine viruses. The feline virus strains formed a separate clade entirely (Fig. 3.3).

***pol* gene insertion/deletion**

The dUTPase sequence fragment is much shorter in BIV (73 amino acids) than it is in all other non-primate lentiviruses (128 aa each for Visna, OVL, and CAEV; 129 aa for FIV, and 131 aa in EIAV, see Appendix C). In all non-primate lentiviruses except BIV, the fragment has a continuous open reading frame, in frame with the RNaseH sequence upstream.

Not surprisingly, an advanced BLAST search of the NCBI database using the anomalous fragment from the Visna virus as the query sequence found the most similar matches to be other non-primate lentivirus anomalous regions, with the exception of BIV, which did not appear in the 102 most similar hits (Table 3.4). However, the most similar non-lentiviral sequences were all dUTPase regions from DNA-based organisms, including poxviruses, adenoviruses, and dUTPase from rats and humans. Less similar, but still scoring significant E-values, were other RNA viruses, including Type B and D oncoviruses.

Base composition

Within a lentivirus genome, the A content of the dUTPase region was similar to that of the *gag* gene (EIAV dUTPase = 36.9% A , EIAV *gag* = 36.2% A). When lentivirus dUTPase sequence was compared to other similar regions recovered from the BLAST search, lentiviral sequences had the highest adenine

content (36.9% for EIAV), while the A content of other organisms varied from 15.7% (Avian Adenovirus, Adenoviridae) to a surprisingly high 35.0% (vaccinia virus, Poxviridae). The dUTPase sequence from *Rattus norvegicus* was 24.6% A (Table 3.5).

DISCUSSION

The phylogeny of lentiviruses presented here is the first to include viruses from all hosts in which they have been isolated and sequenced to date. The viruses that infect primates form a distinct clade, separate from viruses that infect non-primates. The positions of the HIV-2 and SIVb viruses are consistent with the results of Gao et al. (1992), which identified the sooty mangabey (*Cercocebus atys*) as the primate reservoir of HIV-2. While the lentiviral phylogeny presented here is consistent with earlier partial phylogenies (Sharp et al. 1995), it conflicts with the most complete previous lentivirus phylogeny (Myers and Pavlakis 1992), where EIAV and BIV formed a monophyletic group and EIAV was identified as the most rapidly evolving lentiviral lineage. All previous studies that included felid viruses have placed them in a monophyletic group with the primate viruses (see Myers and Pavlakis 1992, McClure et al. 1987). This result may be an artifact of long branch attraction problems forcing the felids and primates together when the tree is rooted with a highly divergent oncovirus or spumavirus sequence.

McClure et al. (1987) investigated the origin of an anomalous 130 amino acid fragment located between reverse transcriptase and integrase in both EIAV

and Visna virus *pol* polyproteins. Based on slight similarities to aspartate proteases found in other retroviruses, they concluded that a protease-like fragment from an oncovirus was inserted into the non-primate lentiviral lineage after it diverged from the primate viral lineage, but before the non-primate lineages diverged from one another (Fig. 3.2). Since their study was published, many more lentiviruses have been sequenced, and Elder et al. (1992) showed that this region encodes dUTPase, an enzyme involved in nucleotide catabolism, in FIV.

Although the protein sequences of dUTPase regions have been determined in many taxa, relatively few nucleotide sequences are available that can be unambiguously aligned to construct a gene tree. I instead chose 1) to compare the base compositions of lentiviral dUTPase regions with that of other lentiviral genes, and 2) to compare lentiviral dUTPase regions with other sequence fragments recovered in the BLAST search that were similar to lentiviral dUTPase sequences. Because lentivirus nucleotide sequences of the three major genes have unusually high adenine contents, the dUTPase region in non-primate lentiviruses would be expected to have a base composition similar to other lentiviral genes if the dUTPase has been evolving in lentiviral genomes for some time. Alternatively, if the lentiviral dUTPase region is the result of a recent horizontal transfer from another retorvirus since primate and non-primate viruses diverged (as proposed by McClure et al [1987]), then the dUTPase region would be expected to have base compositions characteristic of other retroviral groups.

The lentiviral dUTPase region had a base composition similar to other regions (*gag* genes) in its own genome (Table 3.5). When comparing between

viruses, the lentiviral A content was the highest surveyed, but other dUTPases from widely divergent viruses, like vaccinia virus, also had extremely high A content. The A contents of other organisms varied widely (Table 3.5), providing no evidence for the origin of the lentiviral fragments, or how long they have been evolving in lentiviral genomes.

dUTPase is encoded in a different region in type D and B retroviruses (oncoviruses), located at the beginning of the *pol* polyprotein gene rather than between RnaseH and integrase, as it is in lentiviruses. Although clearly similar to one another (Table 3.4), the proteins in each retrovirus subfamily have diverged substantially. Their distinct locations imply horizontal acquisition, rather than vertical inheritance from a common retrovirus ancestor. The presence of similar proteins encoded by some poxviruses, adenoviruses, and herpesviruses, is consistent with this interpretation. The dUTPases encoded by some of these DNA viruses are actually more similar to lentiviral dUTPases than are retroviral enzymes.

The exact origin of the dUTPase region in lentiviruses remains unknown. Within lentiviruses, dUTPase is found in the non-primate lentiviruses (FIV, EIAV, CAEV, OVL, and Visna), but is lacking in HIV and SIV, and is only a non-functional remnant in BIV. All eukaryotic cells express dUTPase (Wagaman et al. 1993), so perhaps the primate lentiviruses commandeer this function from host cells. The fact that non-primate lentiviruses have retained the protein in their relatively small (8-9 kb) genomes suggests dUTPase has a functional role in these viruses.

Although the dUTPase region present in non-primate lentiviruses may indeed be the result of a horizontal transfer event, there is no compelling evidence that the fragment came directly from oncoviruses, as suggested by McClure et al. (1987). Rather, the pattern of dUTPase expression in lentiviruses and in other retroviruses suggests that it was present in a lentiviral ancestor, but has since been lost in primate lentiviruses and is currently a nonfunctional remnant in BIV.

Table 3.1. Lentiviruses, their hosts, and clinical symptoms.

Virus		Host	Disease
HIV	Human Immunodeficiency Virus	Humans	immune dysfunction; several strains, probably of simian origin; usually fatal
SIV	Simian Immunodeficiency Virus	non-human primates	immune dysfunction; severity of disease varies among hosts
EIAV	Equine Infectious Anemia Virus	horses, ponies, and asses	anemia, returning febrile periods; usually fatal; transmitted by biting flies
BIV	Bovine Immunodeficiency Virus	cattle	immune dysfunction
CAEV	Caprine Arthritic Encephalitis Virus	goats	immune dysfunction
Visna	Visna Virus	sheep	central nervous system disorders
OVL	Ovine Lentivirus	sheep	immune dysfunction
FIV	Feline Immunodeficiency Virus	domestic cats, also mountain lions, (*Felis concolor*)	immune dysfunction; transmitted by body fluids, usually fatal

Table 3.2. Lentivirus sequences used in this study for phylogeny inference. The letters 'a' and 'b' identify different samples of the same virus.

Virus	Accession Number	Source
HIV 2 a	AF082339	Guinea-Bissau patient
HIV 2 b	M31113	Senegal patient
SIV b	AF077017	sooty mangabey (*Cercocebus atys*)
SIV cl	AF075269	L'Hoest monkey (*Cercopithecus l'hoesti*)
SIV a	M29975	African green monkey (*Cercopithecus aethiops*)
HIV 1 a	AF005496	Central African Republic patient
HIV 1 b	AF005495	Brazil patient
BIV a	L04972	unknown
BIV b	M32690	unknown
CAEV	M33677	Clements isolate
OVL	M31646	South African sheep
Visna a	L06906	Icelandic sheep
Visna b	M60610	Icelandic sheep
FIV a	M25381	Petaluma, CA feral cat
FIV b	U11820	unknown
EIAV a	AF028232	unknown
EIAV b	M16575	Wyoming blood sample

Table 3.3 Legend: Results of the likelihood ratio tests. L_0 and L_1 represent the likelihoods under the null (H_0) and alternative (H_1) hypotheses, respectively. P represents the probability of obtaining the observed value of the likelihood ratio test statistic if the null hypothesis were true. Degrees of freedom are equal to the difference in the number of free parameters between the hypotheses being compared. Significance values were calculated with Mathematica (Wolfram 1998): P values $< 1 \times 10^{-16}$ indicate the probability is less than the smallest value that can be calculated by Mathematica. Critical values are adjusted using a Bonferroni correction because multiple comparisons were made.

Table 3.3. Results of likelihood ratio tests.

Null Hypothesis	Models compared	-Ln L	-2 log Λ	df	P
Transition rate equals transversion rate	H_0: JC69	23,712.92	633.04	1	$< 1 \times 10^{-16}$
	H_1: K2P	23,396.40			
Equal rates among sites	H_0: K2P	23,396.40	1035.92	1	$< 1 \times 10^{-16}$
	H_1: K2P + Γ	22,878.44			
Equal base frequencies	H_0:K2P + Γ	22,878.44	936.02	3	$< 1 \times 10^{-16}$
	H_1: HKY85 + Γ	22,410.43			
A proportion of sites are invariant	H_0: HKY85 + Γ	22,410.43	50.98	1	9.33×10^{-13}
	H_1: HKY85 + Γ + I	22,384.94			
Two rates among substitution types	H_0: HKY85 + Γ + I	22,384.94	109.90	4	$< 1 \times 10^{-16}$
	H_1: GTR + Γ + I	22,329.99			
Molecular clock	H_0: GTR + Γ + I + clock	22350.73	41.48	15	2.70×10^{-4}
	H_1: GTR + Γ + I	22,329.99			

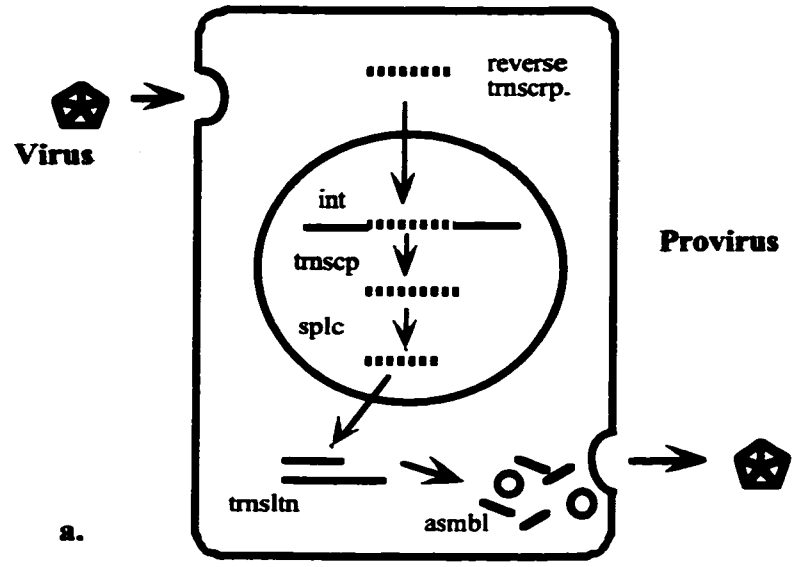
Table 3.4. List of top 15 unique sequence similarity hits in GenBank after query with Visna virus fragment. E-values (from the BLAST searches) are the probability of a match this strong expected from chance alone: the smaller the value, the greater the similarity between sequences.

	E-value
Other Visna virus fragment	4×10^{-82}
OVL fragment	6×10^{-71}
EIAV fragment	2×10^{-30}
FIV dUTPase	5×10^{-26}
Vaccina virus dUTPase	2×10^{-15}
Cowpox dUTPase	3×10^{-15}
Avian Adenovirus dUTPase	3×10^{-12}
Rattus norvegicus dUTPase	5×10^{-12}
Homo sapiens dUTPase	1×10^{-11}
Simian Retrovirus (Type D oncovirus)	5×10^{-9}
Mason-Pfizer Monkey virus (Type D oncovirus)	5×10^{-9}
Mouse Mammary Tumor Virus (Type B oncovirus)	1×10^{-8}
Simian Sarcoma virus (Type C oncovirus	8×10^{-7}
Ovine Pulmonary Adenocarcinoma virus	1×10^{-6}
Salmonid herpes virus	2×10^{-5}
BIV	No match in top 102 hits

Table 3.5. Base composition of different dUTPase sequences. EIAV *gag* is included for comparison to a non-dUTPase fragment from the same virus. AAV = Avian Adenovirus, OPAV = Ovine Pulmonary Adenocarcinoma Virus.

Sequence	Taxonomic affiliation	GenBank accession	Length	% A	% G	% T/U	%C
EIAV dUTPase	Lentivirus ds RNA	M87583	1490	36.9	21.8	25.4	15.8
EIAV *gag*	Lentivirus ds RNA	M87583	1460	36.2	23.5	22.9	17.5
SRV dUTPase	Oncovirus ds RNA	AF126467	944	33.6	19.5	23.8	23.1
AAV dUTPase	Adenovirus ds DNA	AF021253	491	15.7	28.5	26.7	29.1
Rattus norvegicus dUTPase	mammal	U64030	952	24.6	26.2	27.1	22.0
Poxvirus D1701 dUTPase	Poxvirus ds DNA	AF056304	482	23.0	27.2	23.7	26.1
Vaccinia virus dUTPase	Poxvirus ds DNA	M34368	443	35.0	20.5	28.0	16.5
OPAV	Type D retrovirus	M80216	869	29.0	21.4	27.6	22.0
salmonid herpesvirus	Herpesvirus ds DNA	AF023673	545	24.8	26.6	24.8	23.9

Figure. 3.1. Retroviral life cycles. a) Simple viruses. b) Complex viruses transcribe regulatory proteins first, which then aid in the manufacture of structural and other non-structural viral proteins. int = integration, trnscp = transcription, splc = splicing, trnsltn = translation, asmbl = assembly.


Virus
reverse
trnscrp.
int
trnscp
splc
Provirus
trnsltn
asmbl
a.

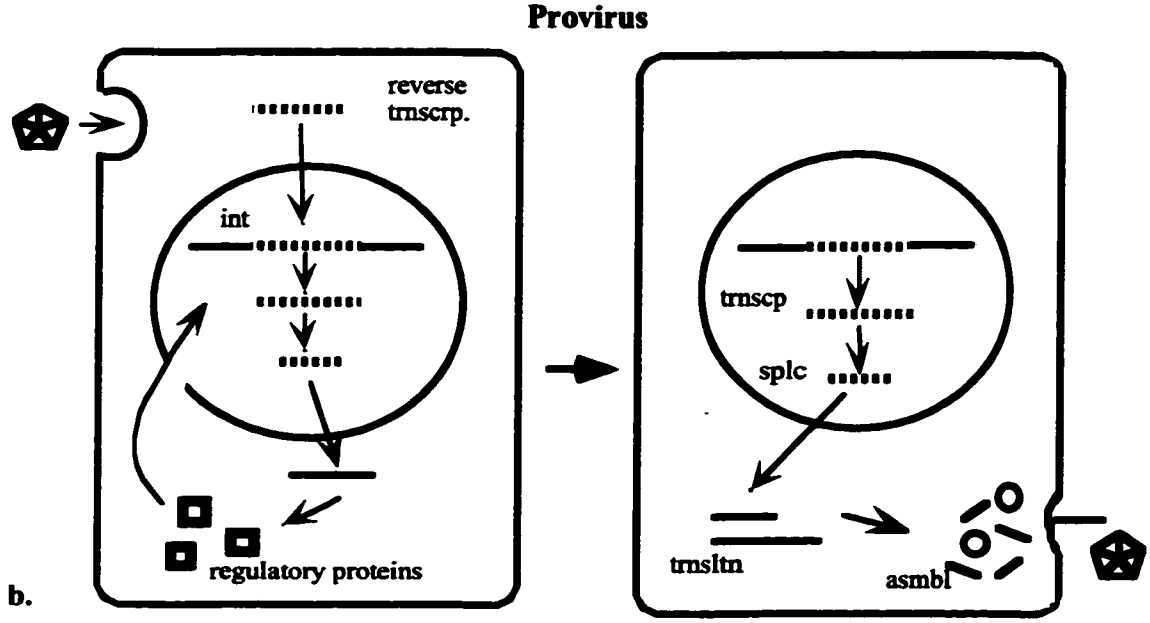

Provirus
reverse
trnscrp.
int
regulatory proteins
trnscp
splc
trnsltn
asmbl
b.

Figure. 3.2. Schematic representing the transfer of a protease-like segment from an oncovirus to non-primate lentiviruses, as proposed by McClure et al. (1987). X1 = protease-like segment (a duplication of protease), P = protease, RT = reverse transcriptase, RH = RNAseH, INT = integrase, X2 = protease-like gene fragment inserted into non-primate lentiviruses.

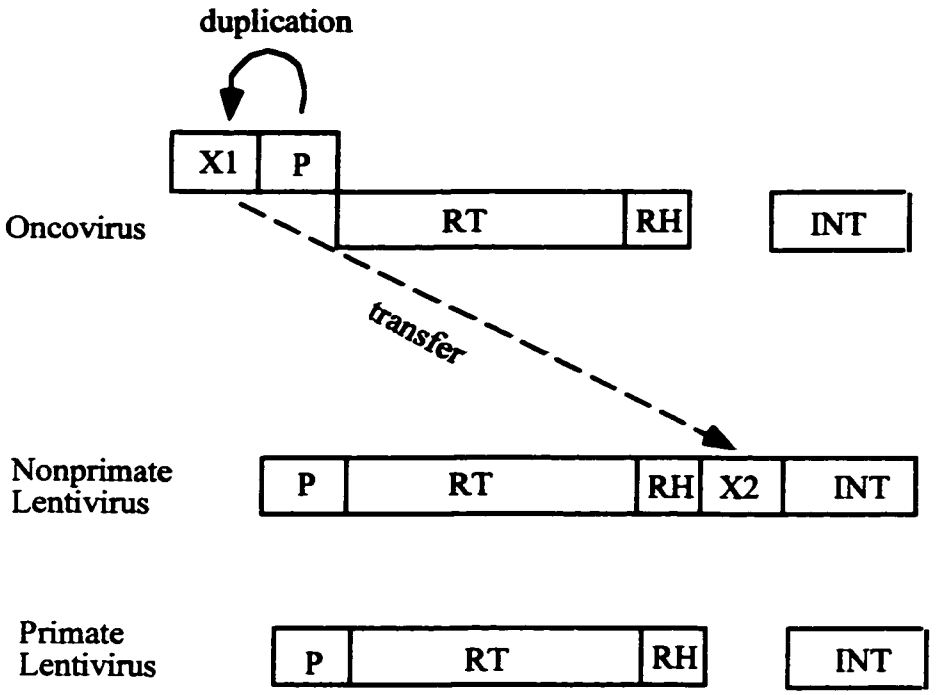
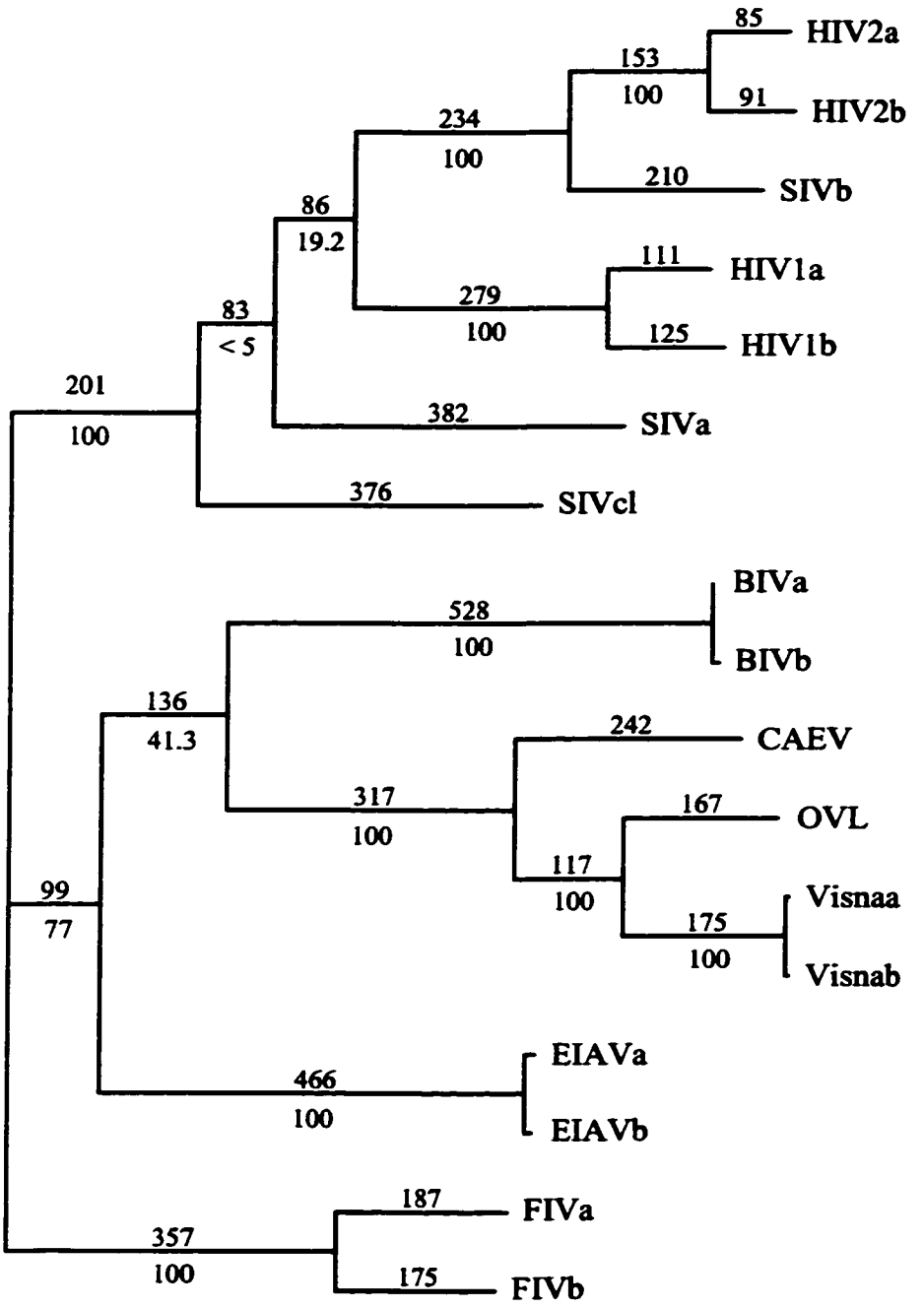

duplication
X1
P
Oncovirus
RT
RH
INT
transfer
Nonprimate
Lentivirus
P
RT
RH
X2
INT
Primate
Lentivirus
P
RT
RH
INT

Figure. 3.3. Maximum likelihood tree of 17 lentiviruses. –Ln likelihood = 22350.73. The GTR + Γ + I model best explained the data. Numbers above the branches indicate number of changes, numbers below the branches are bootstrap support based on 1000 non-parametric replicates.


85
HIV2a
153
100
91
HIV2b
234
100
210
SIVb
86
19.2
111
HIV1a
279
100
83
< 5
125
HIV1b
201
100
382
SIVa
376
SIVcl
BIVa
528
100
BIVb
136
41.3
242
CAEV
317
100
167
OVL
99
77
117
100
Visnaa
175
100
Visnab
EIAVa
466
100
EIAVb
187
FIVa
357
100
175
FIVb

Appendix A

List of pika collection locations, subspecies identification, and NMMNH specimen number. Map abbreviations correspond to locations in Figure 2.4.

Map abbreviation	Collection location	Subspecies	NMMNH #
BC 1	Hagensborg, BC	O. princeps littoralis	978
BC 2	Little Itcha Mtns., BC	O. p. septentrionalis	1000
BC 3	Westwold, BC	O. p. brooksi	934
BC 4	Clearwater, BC	O. p. saturata	953
BC 5	Golden, BC	O. p. cuppes	917
ALB	Buller Mtn., Alberta	O. p. lutescens	894
WA 1	Whatcom Co., WA	O. p. brunnescens	1011
WA 2	Okanogan Co., WA	O. p. fenisex	1028
OR 1	Lane Co., OR	O. p. fumosa	862
OR 2	Jackson Co., OR	O. p. brunnescens	1211
OR 3	Jackson Co., OR	O. p. brunnescens	1212
OR 4	Baker Co., OR	O. p. jewetti	1060
OR 5	Harney Co., OR	O. p. taylori	871
ID 1	Adams Co., ID	O. p. howelli	624
ID 2	Custer Co., ID	O. p. lemhi	642
ID 3	Butte Co., ID	O. p. goldmani	663
ID 4	Bear Lake Co., ID	O. p. clamosa	670
MT 1	Ravalli Co., MT	O. p. princeps	594
MT 2	Powell Co., MT	O. p. princeps	602
MT 3	Cascade Co., MT	O. p. princeps	570
MT 4	Carbon Co., MT	O. p. ventorum	561
WY 1	Teton Co., WY	O. p. ventorum	517
WY 2	Fremont Co., WY	O. p. ventorum	500
WY 3	Fremont Co., WY	O. p. ventorum	503
WY 4	Big Horn Co., WY	O. p. obscura	537
WY 5	Carbon Co., WY	O. p. figginsi	474

Apndx.A,cont.

Map abbreviation	Collection location	Subspecies	NMMNH #
WY 6	Albany Co., WY	O. p. saxatilis	454
CA 1	Siskiyou Co., CA	O. p. schisticeps	854
CA 2	Modoc Co., CA	O. p. schisticeps	847
CA 3	Alpine Co., CA	O. p. muiri	833
CA 4	Mono Co., CA	O. p. sheltoni	819
CA 5	Inyo Co., CA	O. p. albata	826
NV 1	Lander Co., NV	O. p. tutelata	tel2044
NV 2	Nye Co., NV	O. p. tutelata	250
NV 3	Elko Co., NV	O. p. nevadensis	257
UT 1	Salt Lake Co., UT	O. p. wasatchensis	299
UT 2	Salt Lake Co., UT	O. p. wasatchensis	301
UT 3	Summit Co., UT	O. p. uinta	281
UT 4	San Pete Co., UT	O. p. moorei	310
UT 5	Sevier Co., UT	O. p. barnesi	331
UT 6	Wayne Co., UT	O. p. barnesi	1217
UT 7	Wayne Co., UT	O. p. utahensis	881
UT 8	Wayne Co., UT	O. p. utahensis	882
UT 9	Sevier Co., UT	O. p. cinnamomea	1857
UT 10	Beaver Co., UT	O. p. cinnamomea	235
UT 11	Iron Co., UT	O. p. fuscipes	222
UT 12	San Juan Co., UT	O. p. lasalensis	348
CO 1	Garfield Co., CO	O. p. figginsi	405
CO 2	Mesa Co., CO	O. p. figginsi	368
CO 3	Ouray Co., CO	O. p. saxatilis	1062
CO 4	Chaffee Co., CO	O. p. saxatilis	380
CO 5	Grand Co., CO	O. p. saxatilis	436
CO 6	El Paso Co., CO	O. p. saxatilis	683
CO 7	Conejos Co, CO	O. p. saxatilis	1201
NM 1	Taos Co, NM	O. p. incana	729

Apndx.A,cont.

Map abbreviation	Collection location	Subspecies	NMMNH #
NM 2	Rio Arriba Co., NM	O. p. nigrescens	1891
NM 3	San Miguel Co., NM	O. p. incana	711
AK	Alaska Range, AK	O. collaris	1307

Appendix B

Aligned mitochondrial DNA sequences from the cytochrome b, tRNA Threonine, tRNA Proline, and the control region from 58 pika samples. Sample names correspond to those listed in Appendix A. Each row is 60bp long. Total sequence length is 689bp. Due to formatting constraints, the alignment is split across two pages so that the first 30 sequences appear on the first page and the remaining 28 sequences appear on the next page.

Locations of the genes are as follows:

Cytochrome b	1-103
Threonine tRNA	104-171
Proline tRNA	172-238
Control region	239-689

```
                1         2         3         4         5         6
                0         0         0         0         0         0
UT 11CCATTCATCATCATCGGCCAACTAGCATCATTCCTCTATTTCCTCCTAATCCTAGTCCTC
UT 10------------------------------------------------------------
NV 2 CCATTCA-CATCATCGGCCAACTAGCATCATTCCTCTATTTCCTCCTAATCCTAGTCCTC
NV 3 CCATTCATTATCATCGGTCAACTAGCATCATTCCTCTACTTTCTTTTAATCCTTGTCCTC
UT 3 CCATTCATCATCATCGGCCAACTAGCATCATTCCTCTATTTCCTTTTAATCCTAGTCCTC
UT 1 ------------------------------------------------------------
UT 2 CCATTCATTATCATCGGTCAACTAGCATCATTCCTCTACTTCCTTTTAATCCTTGTCCTC
UT 4 CCATTCA-CATCATCGGCCAACTAGCATCATTCCTCTATTTCCTTTTAATCCTAGTCCTC
UT 5 -------------------------CATCATTCCTCTANNTCCTTTTAATCCTAGTCCTC
UT 12CCATTCATCATCATCGGTCAACTAGCATCATTCCTCTATTTCCTTTTAATCCTAGTCCTC
CO 2 -------------------------CATCATTCCTCTATGNCCTTTTAATCCTAGTCCTC
CO 4 -----------------TCAACTAGCATCATTCCTCTATTTCCTTTTAATCCTAGTCCTC
CO 1 CCATTCATTATCATCGGTCAACTAGCATCATTCCTCTACTTTCTTTTANTCCTTGTCCTC
CO 5 CCATTCATCATCATCGGTCAACTAGCATCATTCCTCTATTTCCTTTTAATCCTAGTCCTC
WY 6 CCATTCATTATCATCGGTCAACTAGCATCATTCCTCTACTTTCTTTTAATCCTTGTCCTC
WY 5 CCATTCATTATCATCGGTCAACTAGCATCATTCCTCTACTTTCTTTTAATCCTTGTCCTC
WY 2 --------TATCATCGGTCAACTAGCATCATTCCTCTACTTTCTTTTAATCCTTGTCCTC
WY 3 CCATTCATTATCATCGGTCAACTAGCATCATTCCTCTACTTTCTTTTAATCCTTGTCCTC
WY 1 CCATTCATTATCATCGGTCAACTAGCATCATTCCTGTACTTCCTTTTAATCCTTGTCCTC
WY 4 CCATTCAT-ATCATCGGTCAACTAGCATCATTCCTCTACTTTCTTTTAATCCTTGTCCTC
MT 4 CCATTCATTATCATCGGTCAACTAGCATCATTCCTCTACTTTCTTTTAATCCTTGTCCTC
MT 3 CCATTCATTATCATCGGTCAACTAGCATCATTCCTCTACTTCCTTTTAATCCTTGTCCTC
MT 1 CCATTCATTATCATCGGTCAACTAGCATCATTCCTCTACTTCCTTTTAATCCTTGTCCTC
MT 2 --------------CGGTCAACTAGCATCATTCCTCTACTTCCTTTTAATCCTTGTCCTC
ID 1 CCATTCAT-ATCATCGGCCAACTAGCATCATTCCTCTACTTCCTTTTAATCCTTGTCCTC
ID 2 CCATTCATCGTCATCGGTCAACTAGCATCATTCCTCTACTTCCTTTTAATCCTTGTCCTC
ID 3 CCATTCATTATCATCGGTCAACTAGCATCATTCCTCTACTTCCTTTTAATCCTTGTCCTC
ID 4 CCATTCATTATCATCGGTCAACTAGCATCATTCCTCTACTTTCTTTTAATCCTTGTCCTC
CO 6 CCATTCATCATCATCGGTCAACTAGCATCATTCCTCTATTTCCTTTTAATCCTAGTCCTC
NM 3 CCATTCATCATCATCGGTCAACTAGCATCATTCCTCTATTTCCTTTTAATCCTAGTCCTC
```

```
                1         2         3         4         5         6
                0         0         0         0         0         0
NM  1 CCATTCATCATCATCGGTCAACTAGCATCATTCCTCTATTTCCTTTTAATCCTAGTCCTC
CA  4 CCATTCATCATCATCGGCCAACTAGCATCATTCCTCTATTTCCTCCTAATCCTAGTCCTC
CA  5 ----TCATCATCATCGGCCAACTAGCATCATTCCTCTATTTCCTCCTAATCCTAGTCCTC
CA  3 -------------------------CATCATTCCTCTANNNCCNCCTAATCCTAGTCCTC
CA  2 CCATTCATCATCATCGGCCAACTAGCATCATTCCTCTATTTCCTCCTAATCCTAGTCCTC
CA  1 CCATTCATCATCATCGGCCAACTAGCATCATTCCTCTATTTCCTCCTAATCCTAGTCCTC
OR  1 CCATTCATCATCAT-GGCCAACTAGCATCATTCCTCTATTTCCTTCTAATCCTAGTCCTC
OR  5 ---------------GGCCAACTACCATCATTCCTCTATTTCCTCCTAATCCTAGTCCTC
UT  7 CCATTCATCATCATCG-CCAACTAGCATCATTCCTCTATTTCCTTTTAATCCTAGTCCTC
UT  8 CCATTCATCATCATCGGCCAACTAGCATCATTCCTCTATTTCCTTTTAATCCTAGTCCTC
ALB  ----------------GTCAACTAGCATCATTCCTCTACTTCCTTTTAATCCTTGTCCTC
BC  5 CCATTCATTATCATCGGCCAACTAGCATCATTCCTNNANTTCCTTTTAATCCTTGTCCTC
BC  3 CCATTCATCATCATCGGCCAACTAGCATCATTCCTCTATTTCCTTCTAATCCTAGTCCTC
BC  4 CCATTCATTATCATCGGTCAACTAGCATCATTCCTCTACTTCCTTTTAATCCTTGTCCTC
BC  1 -------------------------CATCATTCCTCTATTTCCTTCTAATCCTAGTCCTC
BC  2 CCATTCATCATCATCGGCCAACTAGCATCATTCCTCTATTTCCTTCTAATCCTAGTCCTC
WA  1 CCATTCATCATCATCGGCCAACTAGCATCATTCCTCTATTTCCTTCTAATCCTAGTCCTC
WA  2 -------------TCGGCCAANTAGCATCATTCCTCTATTTCCTTCTAATCCTAGTCCTC
OR  4 CCATTCATCATCATCGGCCAACTAGCATCATTCCTCTACTTCCTCCTAATCCTAGTCCTC
CO  3 CCATTCATCATCATCGGTCAACTAGCATCATTCCTCTATTTCCTTTTAATCCTAGTCCTC
AK    CCCTTCATCATTATCGGCCAACTAGCATCATTCCTTTACTTCTTCCTAATCCTAGTCCTC
CO  7 -------------------------CATCATTCCTCTNNNNNCTTGTAATCCTAGTCCTC
OR  2 CCATTCATCATCATCGGCCAACTAGCATCATTCCTCTATTTCCTTCTAATCCTAGTCCTC
OR  3 CCATTCATCATCATCGGCCAACTAGCATCATTCCTCTATTTCCTTCTAATCCTAGTCCTC
UT  6 CCATTCATCATCATCGGCCAACTAGCATCATTCCTCTATTTCCTTTTAATCCTAGTCCTC
UT  9 CCATTCATCATCATCGGCCAACTAGCATCATTCCTCTATTTCCTCCTAATCCTAGTCCTC
NM  2 CCATTCATCATCATCGGTCAACTAGCATCATTCCTCTATTTCCTTTTAATCCTAGTCCTC
NV  1 CCATTCATCATCATCGGCCAACTAGCATCATTCCTCTATTTCCTCCTAATCCTAGTCCTC
```

```
                                                  1         1         1
                    7         8         9         0         1         2
                    0         0         0         0         0         0
UT 11ATGCCAGCCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
UT 10--------CTGCAGCCTAATCGAGAACAAATTACTCAAATGAAGGTCCTCGTAGTATAAC
NV 2 ATGCCAGCCTGCAGCCTAATCGAGAACAAATTACTCAAATGAAGGTCCTCGTAGTATAAC
NV 3 ATACCAGTCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
UT 3 ATGCCAGCCTGCAGCCTAATCGAGAACAAATTACTCAAATGAAGGTCCTCGTAGTATAAC
UT 1 ----------------------------------------GAAGGTCCTCGTAGTATAAC
UT 2 ATGCCAGTCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
UT 4 ATGCCAGCCTGCAGCCTAATCGAGAACAAATTACTCAAATGAAGGTCCTCGTAGTATAAC
UT 5 ATGCCAGCCTGCAGCCTAATCGAGAACAAATTACTCAAATGAAGGTCCTCGTAGTATAAC
UT 12ATACCAGCCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
CO 2 ATACCAGCCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
CO 4 ATACCAGCCTGCAGCCTAATCGAGAACAAACTACTCGAATGAAGGTCCTCGTAGTATAAC
CO 1 ATGCCAGTCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
CO 5 ATACCAGCCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
WY 6 ATGCCAGTCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
WY 5 ATGCCAGTCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
WY 2 ATGCCAGTCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTANTATAAC
WY 3 ATGCCAGTCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
WY 1 ATGCCAGTCTGCAGCTTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
WY 4 ATACCAGTCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
MT 4 ATGCCAGTCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
MT 3 ATGCCAGTCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
MT 1 ATGCCAGTCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
MT 2 ATGCCAGTCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
ID 1 ATGCCAGTCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
ID 2 ATGCCAGTCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
ID 3 ATGCCAGTCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
ID 4 ATGCCAGTCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
CO 6 ATACCAGCCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
NM 3 ATACCAGCCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
```

```
                                                    1         1         1
                    7         8         9         0         1         2
                    0         0         0         0         0         0
NM  1 ATACCAGCCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
CA  4 ATGCCAGCCTGCAGCCTAATCGAGAACAAATTACTCAAATGAAGGTCCTCGTAGTATAAC
CA  5 ATGCCAGCCTGCAGCCTAATCGAGAACAAATTACTCAAATGAAGGTCCTCGTAGTATAAC
CA  3 ATGCCAGCCTGCAGCCTAATCGAGAACAAATTACTCAAATGAAGGTCCTCGTAGTATAAC
CA  2 ATGCCAGCCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
CA  1 ATGCCAGCCTGCAGCCTAATCGAGAACAAATTACTCAAATGAAGGTCCTCGTAGTATAAC
OR  1 ATACCAGCCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
OR  5 ATGCCAGCCTGCAGCCTAATCGAGAACAAATTACTCAAATGAAGGTCCTCGTAGTATAAC
UT  7 ATGCCAGCTTGCAGCTTAATCGAGAACAAATTACTCAAATGAAGGTCCTCGTAGTATAAC
UT  8 ATGCCAGCCTGCAGCCTAATCGAGAACAAATTACTCAAATGAAGGTCCTCGTAGTATAAC
ALB   ATGCCAGTCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
BC  5 ATGCCAGTCTGCAGCTCAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
BC  3 ATACCAGCCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
BC  4 ATGCCAGTCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
BC  1 ATACCAGCCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTANTATAAC
BC  2 ATACCAGCCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
WA  1 ATACCAGCCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
WA  2 ATACCAGCCTGCAACCTAATCGAGAACAAACTACTCNAATGAAGGTCCTCGTAATATAAC
OR  4 ATGCCAGCCTGCAGCCTGATCGAGAACAAATTACTCAAATGAAGGTCCTCGTAGTATAAC
CO  3 ATACCAGCCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
AK    ATGCCAGCCTGCAGCTTAATTGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAG
CO  7 ATACCAGCCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
OR  2 ATACCAGCCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
OR  3 ATACCAGCCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
UT  6 ATGCCAGCCTGCAGCCTAATCGAGAACAAATTACTCAAATGAAGGTCCTCGTAGTATAAC
UT  9 ATGCCAGCCTGCAGCCTAATCGAGAACAAATTACTCAAATGAAGGTCCTCGTAGTATAAC
NM  2 ATACCAGCCTGCAGCCTAATCGAGAACAAACTACTCAAATGAAGGTCCTCGTAGTATAAC
NV  1 ATGCCAGCCTGCAGCCTAATCGAGAACAAATTACTCAAATGAAGGTCCTCGTAGTATAAC
```

```
                 1         1         1         1         1         1
                 3         4         5         6         7         8
                 0         0         0         0         0         0
UT 11AATTACCCTGGTCTTGTAAACCAAAAATGAAGTACAACCCTTCCTAGGACATCAAAGAGG
UT 10AATTACCCTGGTCTTGTAAACCAAAAATGAAGTACAACCCTTCCTAGGACATCAAAGAGG
NV 2 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTACAACCCTTCCTAGGACATCAAAGAGG
NV 3 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAATTCTTCCTAGGACATCAAAGAGG
UT 3 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTAGGACATCAAAGAGG
UT 1 AATTACCCTGGTCTTGTAACCCAAAAATGAAGTGCAATTCTTC-TAGGACATCAAGGAGG
UT 2 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAATTCTTCCTAGGACATCAAAGAGG
UT 4 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTAGGACATCAAAGAGG
UT 5 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTAGGACATCAAAGAGG
UT 12AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTAGGACATCAAAGAGG
CO 2 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTAGGACATCAAAGAGG
CO 4 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTAGGACATCAAAGAGG
CO 1 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAATTCTTCCTAGGACATCAAAGAGG
CO 5 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTAGGACATCAAAGAGG
WY 6 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAANNCNNCCTAGGACATCAAAGAGG
WY 5 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAATTCTTCCTAGGACATCAAAGAGG
WY 2 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAATTCTTCCTAGGACATCAAAGAGG
WY 3 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAATTCTTCCTAGGACATCAAAGAGG
WY 1 AATTACTCTGGTCTTGTAAACCAAAAATGAAGTACAATTCTTCATAGGACATCAAAGAGG
WY 4 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAATTCTTCCNNNGACATCAAAGAGG
MT 4 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAATTCTTCCTAGGACATCAAAGAGG
MT 3 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAATTCTTCCTAGGACATCAAAGAGG
MT 1 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAATTCTTCCTAGGACATCAAAGAGG
MT 2 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAATTCTTCCTAGGACATCAAAGAGG
ID 1 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAATTCTTCCTAGGACATCAGAGAGG
ID 2 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAATTCTTCCTAGGACATCAAAGAGG
ID 3 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAATTCTTCCTAGGACATCAAAGAGG
ID 4 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAATTCTTCCTAGGACATCAAAGAGG
CO 6 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTAGGACATCAAAGAGG
NM 3 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTAGGACATCAAAGAGG
```

```
                1         1         1         1         1         1
                3         4         5         6         7         8
                0         0         0         0         0         0
NM  1 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTAGGACATCAAAGAGG
CA  4 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTACAACCCTTCCTAGGACATCAAAGAGG
CA  5 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTACAACCCTTCCTAGGACATCAGAGAGG
CA  3 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTACAACCCTTCCTAGGACATCAAAGAGG
CA  2 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTAGGACATCAGAGAGG
CA  1 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTACAACCCTTCCTAGGACATCAAAGAGG
OR  1 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTAGGACATCAAAGAGG
OR  5 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTAGGACATCAAAGAGG
UT  7 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCNTAGGACATCAAAGAGG
UT  8 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTAGGACATCAAAGAGG
ALB   AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAATTCTTCCTAGGACATCAAAGAGG
BC  5 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAATTCTTCCTAGGACATCAAAGAGG
BC  3 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTANGACATCAAAGAGG
BC  4 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAATTCTTCCTAGGACATCAAAGAGG
BC  1 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTAGGACATCAAAGAGG
BC  2 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTAGGACATCAAAGAGG
WA  1 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTAGGACATCAAAGAGG
WA  2 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTAGGACATCAAAGAGG
OR  4 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTAGGACATCAAAGAGG
CO  3 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTAGGACATCAAAGAGG
AK    AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAAACCTTCCTAGGACATCAAAGAGG
CO  7 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTAGGACATCAAAGAGG
OR  2 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTAGGACATCAAAGAGG
OR  3 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTAGGACATCAAAGAGG
UT  6 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTAGGACATCAAAGAGG
UT  9 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTACAACCCTTCCTAGGACATCAAAGAGG
NM  2 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTGCAACCCTTCCTAGGACATCAAAGAGG
NV  1 AATTACCCTGGTCTTGTAAACCAAAAATGAAGTACAACCCTTCCTAGGACATCAAAGAGG
```

```
                  1         2         2         2         2         2
                  9         0         1         2         3         4
                  0         0         0         0         0         0
UT 11GGACCTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTCATTAAACTACTCTCTGCA
UT 10GGACCTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTCATTAAACTACTCTCTGCA
NV 2 AGACCTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTCATTAAACTACTCTCTGCA
NV 3 AGACTTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
UT 3 AGACCTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
UT 1 AGACTTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
UT 2 AGACTTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
UT 4 AGACCTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
UT 5 AGACCTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
UT 12AGACCTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
CO 2 AGACCTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
CO 4 AGACCTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
CO 1 AGACTTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
CO 5 AGACCTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
WY 6 AGACTTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
WY 5 AGACTTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
WY 2 AGACTTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
WY 3 AGACTTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
WY 1 AGACTTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
WY 4 AGACTTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
MT 4 AGACTTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
MT 3 AGACTTAA-GTCTCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
MT 1 AGACTTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
MT 2 AGACTTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
ID 1 AGACTTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
ID 2 AGACTTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
ID 3 AGACTTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
ID 4 AGACTTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
CO 6 AGACCTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
NM 3 AGACCTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
```

```
                 1         2         2         2         2         2
                 9         0         1         2         3         4
                 0         0         0         0         0         0
NM  1 AGACCTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
CA  4 AGACCTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTCATTAAACTACTCTCTGCA
CA  5 AGACCTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTCATTAAACTACTCTCTGCA
CA  3 AGACCTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTCATTAAACTACTCTCTGCA
CA  2 AGACCTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTCATTAAACTACTCTCTGCA
CA  1 AGACCTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTCATTAAACTACTCTCTGCA
OR  1 AGACCTAAAGTCCCACTATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
OR  5 AGACCTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTCATTAAACTACTCTCTGCA
UT  7 AGACCTAA-GTCTCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
UT  8 AGACCTAA-GTCTCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
ALB   AGACTTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
BC  5 AGACTTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
BC  3 AGACCTAAAGTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
BC  4 AGACTTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
BC  1 AGACCTAAAGTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
BC  2 AGACCTAAAGTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
WA  1 AGACCTAAAGTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
WA  2 AGACCTAAAGTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
OR  4 AGACCTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTCATTAAACTACTCTCTGCA
CO  3 AGACCTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
AK    GGACCCAAGGTCCCGCCATCAGCACCCAAAGCTGAAATTCTTCTTAAACTACTCTCTGTA
CO  7 AGACCTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
OR  2 AGACCTAAAGTCCCACTATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
OR  3 AGACCTAAAGTCCCACTATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
UT  6 AGACCTGA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
UT  9 GGACCTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTCATTAAACTACTCTCTGCA
NM  2 AGACCTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTTATTAAACTACTCTCTGCA
NV  1 AGACCTAA-GTCCCACCATCAGCACCCAAAGCTGAAATTCTCATTAAACTACTCTCTGCA
```

```
                  2         2         2         2         2         3
                  5         6         7         8         9         0
                  0         0         0         0         0         0
UT 11ACTCACTCTCCCAC-----------GCGCA-AGACTCGCTATGTACCTCGTGCATTAATG
UT 10ACTCANTCTCACAC-----------GCGCA-CGACTCGCTATGTACCTCGTGCATTAATG
NV 2 ACTCACTCTCCCAC-----------GCGCA-AGACTCGCTATGTACCTCGTGCATTAATG
NV 3 ACTCATTCCTACACTTCCTC-TCCCACGTACAA-CCCGCTATGTACCTCGTGCATTAATG
UT 3 ACTCACTTCTACGCTC-TTC-TCCCACACACAA-CCCGCTATGTACTTCGTGCATTAATG
UT 1 ACTCATTCCTACACTTCCTC-TCCCACGTACAA-CCCGCTATGTACCTCGTGCATTAATG
UT 2 ACTCATTCCTACACTTCCTC-TCCCACGTACAA-CCCGCTATGTACCTCGTGCATTAATG
UT 4 ACTCACTCCTACACTC-TTC-TCCCACACACAA-CCCGCTATGTACCTCGTGCATTAATG
UT 5 ACTCACTCCTACACTC-TTC-TCCCACACGCAA-CCCGCTATGTACCTCGTGCATTAATG
UT 12ACTCATTCCTACACTCCTTC-TCCCACACATTA-TCCGCTATGTACTTCGTGCATTAATG
CO 2 ACTCATTCCTACACTCCCTC-TCCCACACATTA-TCCGCTATGTACCTCGTGCATTAATG
CO 4 ACTCATTCCTACACTCCCTC-TCCCACACATTA-TCCGCTATGTACCTCGTGCATTAATG
CO 1 ACCCATTCCTACACTTCCCC-TCCCACGTACAA-CCCGCTATGTACCTCGTGCATTAATG
CO 5 ACTCATTCC---------------CACACATTA-TCCGCTATGTACCTCGTGCATTAATG
WY 6 ACCCATTCCTACACTTCCTC-TCCCACGTACAA-CCCGCTATGTACCTCGTGCATTAATG
WY 5 ACCCATTCCTACACTTCCTC-TCCCACGTACAA-CCCGCTATGTACCTCGTGCATTAATG
WY 2 ACCCATTCCTACACTTCCTC-TCCCACGTACAA-CCCGCTATGTACTTCGTGCATTAATG
WY 3 ACCCATTCCTACACTTCCTC-TCCCACGTACAA-CCCGCTATGTACTTCGTGCATTAATG
WY 1 ACTCATTCCTACACTCCCTCTACCCACGCACAA-CCCGCTATGTACCTCGTGCATTAATG
WY 4 ACCCATTCCCACACTTCCTC-TCCCACGTACAA-CCCGCTATGTACCTCGTGCATTAATG
MT 4 ACCCATTCCTACACTTCCTC-TCCCACGTACAA-CCCGCTATGTACCTCGTGCATTAATG
MT 3 ACTCATTCCTACACTCCCTC-TCCCACATACAA-CCCGCTATGTACCTCGTGCATTAATG
MT 1 ACTCATTCCTACACTCCCTCATCCCACGTACAA-CCCGCTATGTACCTCGTGCATTAATG
MT 2 ACTCATTCCTACACTCCCTCACCCCACGTACAA-CCCGCTATGTACCTCGTGCATTAATG
ID 1 ACTCATTCCTATACTCCCTC-TCCCACGTACAA-CCCGCTATGTACCTCGTGCATTAATG
ID 2 ACTCATTCCTACACTCCCTC-TCCCACGTACAA-CCCGCTATGTACCTCGTGCATTAATG
ID 3 ACTCATTCCTACACTCCCTCATCCCACGTACAA-CCCGCTATGTACCTCGTGCATTAATG
ID 4 ACTCATTCCTACACTTCCTC-TCCCACGTACAA-CCCGCTATGTACCTCGTGCATTAATG
CO 6 ACTCATTCC---------------CACACATTA-TCCGCTATGTACCTCGTGCATTAATG
NM 3 ACTCATTCCTACACTCCCTC-TCCCACACATTA-TCCGCTATGTACCTCGTGCATTAATG
```

```
                  2         2         2         2         2         3
                  5         6         7         8         9         0
                  0         0         0         0         0         0
NM  1 ACTCATTCCTACACTCCCTC-TCCCGCACATTA-TCCGCTATGTACCTCGTGCATTAATG
CA  4 ACTCACTCTCCCAC-----------GCGCA-AGACTCGCTATGTACCTCGTGCATTAATG
CA  5 ACTCACTCTCCCAC-----------GCGGG-AGACTCGCTATGTACCTCGTGCATTAATG
CA  3 ACTCACTCTCCCAC-----------GCGCA-AGACTCGCTATGTACCTCGTGCATTAATG
CA  2 ACTCACTCTCCCAC-----------GCGCA-AGACTCGCTATGTACCTCGTGCATTAATG
CA  1 ACTCACTCTCCCAC-----------GCGCA-AGACTCGCTATGTACCTCGTGCATTAATG
OR  1 ACCCAC-CCCACGCCCCC-------------------GCTATGTACCTCGTGCATTAATG
OR  5 ACTCACTCTCCCAC-----------GCGCA-AGACTCGCTATGTACCTCGTGCATTAATG
UT  7 ACTCACTCCTACACTC-TTC-TCCCACACGCAA-CCCGNTATGTACTTCGTGCATTAATG
UT  8 ACTCACTCCTACACTC-TTC-TCCCACACGCAA-CCCGCTATGTACTTCGTGCATTAATG
ALB   ACTCATTCCTACACTCCCTCATCCCACGTACAA-CCCGCTATGTACCTCGTGCATTAATG
BC  5 ACTCATTCATACACTCCCTCATCCCACGTACAAACCCGCTATGTACCTCGTGCATTAATG
BC  3 ACCCAC-CTCACACCCCC-------------------GCTATGTACCTCGTGCATTAATG
BC  4 ACTCATTCCTACACTCCCTCATCCCACGTACAA-CCCGCTATGTACCTCGTGCATTAATG
BC  1 ACCCAC-CCCACACCCCC-------------------GCTATGTACCTCGTGCATTAATG
BC  2 ACCCAC-CCCACACCCCC-------------------GCTATGTACCTCGTGCATTAATG
WA  1 ACCCAC-CCCACACCCCC-------------------GCTATGTACCTCGTGCATTAATG
WA  2 ACCCAC-CCCACACCCCC-------------------GCTATGTACCTCGTGCATTAATG
OR  4 ACTCACTCTCCCAC-----------GCGCA-AGACTCGCTATGTACCTCGTGCATTAATG
CO  3 ACTCATTCCTACACTCCCTC-TCCCACACATTA-TCCGCTATGTACCTCGTGCATTAATG
AK    ATTCAT----ACAGCACATC--TCCGCGCACAG--CCGCTATGTACCTCGTGCATTAATG
CO  7 ACTCATTCCTACACTCCCTC--CCCACACATTA-TCCGCTATGTACCTCGTGCATTAATG
OR  2 ACCCAC-CCCACACCCCC-------------------GCTATGTACCTCGTGCATTAATG
OR  3 ACCCAC-CCCACACCCCC-------------------GCTATGTACCTCGTGCATTAATG
UT  6 ACTCACTCCTACACTC-TTC-TCCCACACGCAA-CCCGCTATGTACCTCGTGCATTAATG
UT  9 ACTCACTCTCCCAC-----------GCGCA-AGACTCGCTATGTACCTCGTGCATTAATG
NM  2 ACTCATTCCTACACTCCCTC-TCCCACACACTA-TCCGCTATGTACCTCGTGCATTAATG
NV  1 ACTCACTCTCCCAC-----------GCGCA-AGACTCGCTATGTACCTCGTGCATTAATG
```

```
                  3         3         3         3         3         3
                  1         2         3         4         5         6
                  0         0         0         0         0         0
UT 11CACCTCCCCATTAATAATGGTACAGTACATATATTTCATAATAGTACATAGCACATCTAT
UT 10CACNTCCCCATTAATAATGGTACAGTACATATATTTCATAATAGTACATAGCACATTTAT
NV 2 CACCTCCCCATTAATAATGGTACAGTACATATATTTCATAATAGTACATAGTACATTTAT
NV 3 CACCTCCCCATTAATAATGGTACAGTACATATATTCCATAACAGTACATAGTACATCTAT
UT 3 CACCTCCCCATTAATAATGGTACAGTACATATATTTCATAATAGTACATAGTACATTTAT
UT 1 CACCTCCCCATTAATAATGGTACAGTACATACATTTCATAACAGTACATAGTACATCTAT
UT 2 CACCTCCCCATTAATAATGGTACAGTACATACATTTCATAACAGTACATAGTACATCTAT
UT 4 CACCTCCCCATTAATAATGGTACAGTACATACATTTCATAATAGTACATAGTACATTTAT
UT 5 CACCTCCCCATTAATAATGGTACAGTACATATATTTCATAATAGTACATAGTACATTTAT
UT 12CACCTCCCCATTAATAATGGTACAGTACATATATTCCATAACAGTACATAGTACATTCAT
CO 2 CACCTCCCCATTAATAATGGTACAGTACATACATTCCATAACAGTACATAGTACATTTAT
CO 4 CACCTCCCCATTAATAATGGTACAGTACATATATTCCATAACAGTACATAGTACATTCAT
CO 1 CACCTCCCCATTAATAATGGTACAGTACATACATTCCATAACAGTACATAGTACATCTAT
CO 5 CACCTCCCCATTAATAATGGTACAGTACATATATTCCATAACAGTACATAGTACATTTAT
WY 6 CACCTCCCCATTAATAATGGTACAGTACATATATTCCATAACAGTACATAGTACATCTAT
WY 5 CACCTCCCCATTAATAATGGTACAGTACATATATTTCATAACAGTACATAGTACATTTAT
WY 2 CACCTCCCCATTAATAATGGTACAGTACATATATTCCATAACAGTACATAGTACATTTAT
WY 3 CACCTCCCCATTAATAATGGTACAGTACATATATTCCATAACAGTACATAGTACATTTAT
WY 1 CACCTCCCCATTAATAATGGTACAGTACATACATTCCATAACAGTACATAGTACATTCAT
WY 4 CACCTCCCCATTAATAATGGTACAGTACATATATTCCATAACAGTACATAGTACATCTAT
MT 4 CACCTCCCCATTAATAATGGTACAGTACATATATTCCATAACAGTACATAGTACATTTAT
MT 3 CACCTCCCCATTAATAATGGTACAGTACATATATTCCATAACAGTACATAGTACATCTAT
MT 1 CACCTCCCCATTAATAATGGTACAGTACATATATTTCATAACAGTACATAGTACATTTAT
MT 2 CACCTCCCCATTAATAATGGTACAGTACATATATTCCATAACAGTACATAGTACATTTAT
ID 1 CACCTCCCCATTAATAATGGTACAGTACATATATTTCATAACAGTACATAGTACATTTAT
ID 2 CACCTCCCCATTAATAATGGTACAGTACATATATTCCATAACAGTACATAGTACATTCAT
ID 3 CACCTCCCCATTAATAATGGTACAGTACATATATTTCATAACAGTACATAGTACATTTAT
ID 4 CACCTCCCCATTAATAATGGTACAGTACATATATTTCATAACAGTACATAGTACATCCAT
CO 6 CACCTCCCCATTAATAATGGTACAGTACATATATTCCATAACAGTACATAGTACATTTAT
NM 3 CACCTCCCCATTAATAATGGTACAGTACATATATTCCATAACAGTACATAGTACATTTAT
```

```
                  3         3         3         3         3         3
                  1         2         3         4         5         6
                  0         0         0         0         0         0
NM 1 CACCTCCCCATTAATAATGGTACAGTACATACATTCCATAACAGTACATAGTACATTTAT
CA 4 CACCTCCCCATTAATAATGGTACAGTACATATATTTCATAATAGTACATAGTACATCTAT
CA 5 CACCTCCCCATTAATAATGGTACAGTACATATATTTCATAATAGTACATAGTACATTTAT
CA 3 CACCTCCCCATTAATAATGGTACAGTACATATATTTCATAATAGTACATAGTACATTTAT
CA 2 CACCTCCCCATTAATAATGGTACAGTACATATATTTCATAATAGTACATAGTACATTTAT
CA 1 CACCTCCCCATTAATAATGGTACAGTACATATATTTCATAATAGTACATAGTACATTTAT
OR 1 CACCTCCCCATTAATAATGGTACAGTACATAAAATTCATAATAGTACATAGTACATCTAT
OR 5 CACCTCCCCATTAATAATGGTACAGTACATATATTTCATAATAGTACATAGTACATTTAT
UT 7 CACCTCCCCATTAATAATGGTACAGTACATATATTTCATAATAGTACATAGTACATTTAT
UT 8 CACCTCCCCATTAATAATGGTACAGTACATATATTTCATAATAGTACATAGTACATTTAT
ALB  CACCTCCCCATTAATAATGGTACAGTACATATATTCCATAACAGTACATAGTACATTTAT
BC 5 CACCTCCCCATTAATAATGGTACAGTACATATATTTCATAACAGTACATAGTACATTTAT
BC 3 CACCTCCCCATTAATAATGGTACAGTACATAAGATTCATAATAGTACATAGTACATCCAT
BC 4 CACCTCCCCATTAATAATGGTACAGTACATATATTCCATAACAGTACATAGTACATTTAT
BC 1 CACCTCCCCATTAATAATGGTACAGTACATAANATTCATAATAGTACATAGTACATCTAT
BC 2 CACCTCCCCATTAATAATGGTACAGTACATAAGATTCATAATAGTACATAGTACATCTAT
WA 1 CACCTCCCCATTAATAATGGTACAGTACATAAGATCCATAATAGTACATAGTACATCTAT
WA 2 CACCTCCCCATTAATAATGGTACAGTACATAAGATTCATAATAGTACATAGTACATCCAT
OR 4 CACCTCCCCATTAATAATGGTACAGTACATATATTTCATAATAGTACATAGTACATTTAT
CO 3 CACCTCCCCATTAATAATGGTACAGTACATATATTCCATAACAGTACATAGTACATTCAT
AK   CACCTCCCCATTAATAATGGTACAGTACATATAATCCATAACAGTACATAGTACATTCAT
CO 7 CACCTCCCCATTAATAATGGTACAGTACATATATTCCATAACAGTACATAGCACATTTAT
OR 2 CACCTCCCCATTAATAATGGTACAGTACATAAGATTCATAATAGTACATAGTACATCTAT
OR 3 CACCTCCCCATTAATAATGGTACAGTACATAAGATTCATAATAGTACATAGTACATCTAT
UT 6 CACCTCCCCATTAATAATGGTACAGTACATATATTTCATAATAGTACATAGTACATTTAT
UT 9 CACCTCCCCATTAATAATGGTACAGTACATATATTTCATAATAGTACATAGCACATTTAT
NM 2 CACCTCCCCATTAATAATGGTACAGTACATATATTCCATAACAGTACATAGTACATTTAT
NV 1 CACCTCCCCATTAATAATGGTACAGTACATATATTTCATAATAGTACATAGTACATTTAT
```

```
                3         3         3         4         4         4
                7         8         9         0         1         2
                0         0         0         0         0         0
UT 11GTTTAATCGTACATTAACTTACTGTCCTCATGCATATCTAGTAAGACATAGAATGTTAAC
UT 10GTTTAATCGTACATTAACTTACTGTCCTCATGCATATCTAGTAAGACATAGAATGTTAAC
NV 2 GTTTAATCGTACATTAACTTACTGTCCTCATGCATATCTAGTAAGACATAAAATGCTAAC
NV 3 GTTTAATCGTACATTAACTTACTGTCCCCATGCATATCTAGTAAGACATAAAACGCTATA
UT 3 GTTTAATCGTACATTAACTTATTATCCCCATGCATATCTAGTAAGACATAAAACTCTATC
UT 1 GTTTAATCGTACATTAACCTACTGTCCTCATGCATATCTAGTAGGACATAAAACGTTATA
UT 2 GTTTAATCGTACATTAACCTACTGTCCTCATGCATATCTAGTAAGACATAAAACGTTATA
UT 4 GTTTAATCGTACATTAACTTATTATCCCCATGCATATCTAGTAAGACATAAAATTCTATC
UT 5 GCTTAATCGTACATTAATTTATTATCCCCATGCATATCTAGTAAGACATAAAACTCTATC
UT 12GTTTAATCGTACATTAACCTATTGTCCCCATGCATATCTAGTAAGACATAAAACGTTAAT
CO 2 GCTTAATCGTACATTAATCTATTACCCCCATGCATATCTAGTAAGACATAAAACGTTAAT
CO 4 GTTTAATCGTACATTAACCTATTGTCCTCATGCATATCTTGTAAGACATAAAACGTTAAT
CO 1 GTTTAATCGTACATTAACCTACTGTCCCCATGCATATCTAGTAAGACATAAAACGCTATA
CO 5 GTTTAATCGTACATTAATTTATCGTCCCCATGCATATCTAGTAAGACATAAAACGTTAAT
WY 6 GTTTAATCGTACATAAACTTACCGTCCCCATGCATATCTAGTAAGACATAGAACGCTATA
WY 5 GTTTAATCGTACATAAACTTACTGTCCCCATGCATATCTAGTAAGACATAAAACGCTATA
WY 2 GTTTAATCGTACATTAACTTACTGTCCCCATGCATATCTAGTAAGACATAAAACGCTATA
WY 3 GTTTAATCGTACATTAACTTACTGTCCCCATGCATATCTAGTAAGACATAAAACGCCATA
WY 1 GCTTAATCGTACATTAACCTACCGTCCCCATGCATATCTAGTAAGACATAAAACGTTATA
WY 4 GTTTAATCGTACATTAACTTACTGTCCCCATGCATATCTAGTAAGACATAAAACGCTATA
MT 4 GTTTAATCGTACATTAACTTACTGTCCCCATGCATATCTAGTAAGACATAAAACGCTATA
MT 3 GTTTAATCGTACATTAATCTACTGTCCCCATGCATATCTAGTAAGACATAAGACGTTATA
MT 1 GTTTAATCGTACATTAACTTACTGTCCCCATGCATATCTAGTAAGACATAGAACGTTATA
MT 2 GTTTAATCGTACATTAACCTACTGTCCCCATGCATATCTAGTAAGACATAGAACGTTATA
ID 1 GTTTAATCGTACATTAACTTACTGTCCCCATGCATATCTAGTAAGACATAGAACGTTATA
ID 2 GTTTAATCGTGCATTAACTTACTGTCCCCATGCATATCTAGTAAGACATAAAACGTTATA
ID 3 GTTTAATCGTACATTAACTTACTGTCCCCATGCATATCTAGTAAGACATAGAACGTTATA
ID 4 GTTTAATCGTACATTAACTTACTGTCCCCATGCATATCTAGTAAGACATAAAACGTTATA
CO 6 GTTTAATCGTACATTAATCTATTGTCCCCATGCATATCTAGTAAGACATAAAACGTTAAT
NM 3 GTTTAATCGTACATTAATCTATTGTCCCCATGCATATCTAGTAAGACATAAAACGCTAAT
```

```
                  3         3         3         4         4         4
                  7         8         9         0         1         2
                  0         0         0         0         0         0
NM  1 GTTTAATCGTACATTAACCTATTATCCCCATGCATATCTAGTAAGACATAAAACGTTAAT
CA  4 GTTTAATCGTACATTAACTTATTGTCCCCATGCATATCTAGTAAGACATAGAATGTTAAT
CA  5 GTTTAATCGTGCATTAACTTATAGTCCCCATGCATATCTAGTAAGACATAGAATGCTAAC
CA  3 GTTTAATCGTACATTAACTTACTGTCCCCATGCATATCTAGTAAGACATAAAATGCTAAC
CA  2 GTTTAATCGTACATTAACTTACTGTCCCCATGCATATCTAGTAAGACATGAGATGTTAAC
CA  1 GTTTAATCGTACATTAAATTACCGTCCCCATGCATATCTAGTAAGACATAAGATGCTAAC
OR  1 GTTTAATCGTACATTAAATTACTGTCCTCATGCATATCTAGTAATACATAGAACGTTAAT
OR  5 GTTTAATCGTACATTAACTTACTGTCCCCATGCATATCTAGTAAGACATGAGATGTTAAC
UT  7 GTTTAATCGTACATTAACTTATTATCCCCATGCATATCTAGTAAGACATAAAACTCTATC
UT  8 GTTTAATCGTACATTAACTTATTATCCCCATGCATATCTAGTAAGACATAAAACTCTATC
ALB   GTTTAATCGTACATTAACTTACTGTCCCCATGCATATCTAGTAAGACATAGAACGTTATA
BC  5 GTTTAATCGTACATTAACTTACTGTCCCCATGCATATCTAGTAAGACATAGAACGTTAAA
BC  3 GCTTAATCGTGCATTAAATTATTGTCCCCATGCATATCTAGTAAGACATAAAACGTTAAT
BC  4 GTTTAATCGTACATTAACTTACTGTCCCCATGCATATCTAGTAAGACATAAAACGTTATA
BC  1 GTTTAATCGTGCATTAAATTATTGTCCCCATGCATATCTAGTAAGACATAAAACGTTAAT
BC  2 GCTTAATCGTGCATTAAATTATTGTCCCCATGCATATCTAGTAAGACATAAAATGTTAAT
WA  1 GCTTAATCGTGCATTAAATTATTATCCCCATGCATATCTAGTAAGACATAAAACGTTAAT
WA  2 GCTTAATCGTGCATTAAATTATTGTCCCCATGCATATCTAGTAAGACATAAAGCGTTAAT
OR  4 GTTTAATCGTACATTAACTTACTGTCCCCATGCATATCTAGTAAGACATAAGATGTTAAC
CO  3 GTTTAATCGTACATTAACCTATTATCCCCATGCATATCTAGTAAGACATAAAACGCTAAT
AK    GTTTAATCGTGCATTAAACTATTGTCCCCATGCATATCTAGTAATACATAAAACGTTATT
CO  7 GTTTAATCGTACATTAATCTATTGTCCCCATGCATATCTAGTAAGACATAAAACGCTAAT
OR  2 GTTTAATCGTACATTAAATTACTGTCCTCATGCATATCTAGTAAGACATAGAACGTTAAT
OR  3 GTTTAATCGTACATTAAATTACTGTCCTCATGCATATCTAGTAAGACATAGAACGTTAAT
UT  6 GCTTAATCGTACATTAACCTATTATCCCCATGCATATCTAGTAAGACATAAAACTTTATC
UT  9 GTTTAATCGTACATTAACTTACTGTCCTCATGCATATCTAGTAAGACATAGAATGTTAAC
NM  2 GTTTAATCGTACATTAATCTATTGTCCCCATGCATATCTAGTAAGACATAAAACGTTAAT
NV  1 GTTTAATCGTACATTAACTTACTGTCCTCATGCATATCTAGTAAGACATAAAATGCTAAC
```

```
                  4         4         4         4         4         4
                  3         4         5         6         7         8
                  0         0         0         0         0         0
UT 11CGTACATAATACATTT-CCTACTTGAATCAACTAAACCA--TATATAACAATACGAATAT
UT 10CGTACATAATACATTT-CCTACTTGAATCAACTAAACTAT-TATATAACAATACGAATAT
NV 2 CGTACATAATACATTT-CCTACTTGAATCAACTAAATTAT-TATATAACAACACGAATAT
NV 3 CGTACATAACACATTC-CTCACTT-AATCAACTAAACAA-ATCCTTAACAATACGAATAT
UT 3 CGTACATAACACATTT-CCTACTT-AATCAACTAAA--A-ATCCTTAACAATACGAATAT
UT 1 CGTACATAACACATTT-CTCACTT-AATCAACTAAA-TAAATCCTCAACAATACGAATAT
UT 2 CGTACATAACACATTT-CTCACTT-AATCAACTAAA-TAAATCCTCAACAATACGAATAT
UT 4 CGTACATAACACATTT-CCTACTT-AATCAACTAAA-TA-ATCCTTAACAATACGAATAT
UT 5 CGTACATAACACATTC-CTCACTT-AATCAACTAAA-TA-ATCCTTAACAATACGAATAT
UT 12CGTACATAAGACATTC-CCTACTT-AATCAACTAA-CTA--TTCTCAACAATACGAATAT
CO 2 CGTACATAAGACATTC-CCTACTT-AATCAACTAA-CTA--TTCCCAACAATACGAATAT
CO 4 CGTACATAAGACATTC-CCTACTT-AATCAACTAA-CTA--TTCCTAACAATACGAATAT
CO 1 CGTACATAACACATTC-CTCACTC-AATCAACTAAA-TAAACCCTTAACAATACGAATAT
CO 5 CGTACATAAGACATTC-CCTACTT-AATCAACTAA-CTA--TTCCTAACAACACGAATAT
WY 6 CGTACATAACACATTC-TTCACTT-AATCAACTAAA-TATATTCTTAACAATACGAATAT
WY 5 CGTACATAACACATTC-TCCACTT-AATCAACTAAA-TATATTCTTAACAATACGAATAT
WY 2 CGTACATAACACATTC-TTCACTT-AATCAACTAAA-TATATCCTTAACAATACGAATAT
WY 3 CGTACATAACACATTC-TTCACTT-AATCAACTAAA-TATATCCTTAACAATACGAATAT
WY 1 CGTACATAACACATTC-TTTACTT-AATCAACTAGACTA-ATCCTTAACAATACGAATAT
WY 4 CGTACATAACACATTC-CTCACTT-AATCAACTAAA-TATATCCTTAACAATACGAATAT
MT 4 CGTACATAACACATTC-TTCACTT-AATCAACTAAA-TATATCCTTAACAATACGAATAT
MT 3 CGTGCATAACACATTC-CTCACTT-AACCAACTGAA-TAAATCCTTAACAATACGAATAT
MT 1 CGTACATAACACATTC-CTTACTT-AATCAACTAGA-TAAATCCTTGACAATACGAATAT
MT 2 CGTGCATAACACATTC-CTTACTT-AATCAACTAGA-TAAATCCTTGACAATACGAATAT
ID 1 CGTACATAACACATTC-CTTACTT-AATCAACTAAA-TAAATCCTTGACAATACGAATAT
ID 2 CGTACATAACACATTC-CTTACTT-AATCAACTAAA-TAAATCCTTAACAATACGAATAT
ID 3 CGTACATAACACATTC-CTTACTT-AATCAACTAGA-TAAATCCTTGACAATACGAATAT
ID 4 CGTACATAACACATTC-CTCACTT-AATCAACTAAA-CAAATCCTTAACAATACGAATAT
CO 6 CGTACATAAGACATTA-CCTACTT-AATCAACTAA-CTA--TTCCTAACAATACGAATAT
NM 3 CGTACATAAGACATTC-CCTACTT-AATCAACTAA-CTA--TTCCTAACAATACGAATAT
```

```
                  4         4         4         4         4         4
                  3         4         5         6         7         8
                  0         0         0         0         0         0
NM  1 CGTACATAAGACATTC-CCTACTT-AATCAACTAA-CTA--TTCCCAACAATACGAATAT
CA  4 CGTACATAATACATTT-CCTACTTGAATCAACTAAACTAT-TATATAACAATACGAATAT
CA  5 CGTACATAATACATTT-CCTACTTGAATCAACTAAACTAT-TATATAACAACACGAATAT
CA  3 CGTACATAATACATTT-TCTACTTGAATCAACTAAATTAT-TATATAACAACACGAATAT
CA  2 CGTACATAATACATTTTCCTACTTGAGTCAACTAAACTAT-TATATAACAATACGAATAT
CA  1 CGTACATAGTACATTTTCCTACTTGAATCAACTAAACTAT-TATATAACAATACGAATAT
OR  1 CGTACATAAGACATTA-CCTACTT-AACCAACTAAA-TA--TTCTTAACAATACGAATAT
OR  5 CGTACATAATACATTTTCCTACTTGAATCAACTAAACTAT-TATATAACAATACGAATAT
UT  7 CGTACATAACACATTC-TTCACTT-AATCAACTAA-CTA-ATCCTTAACAATACGAATAT
UT  8 CGTACATAACACATTC-TTCACTT-AATCAACTAA-CTA-ATCCTTAACAATACGAATAT
ALB   CGTGCATAACACATTC-CTTACTT-AATCAACTAGATAA-ATCCTTGACAATACGAGTAT
BC  5 CGTGCATAACACATTC-CTTACTT-AATCAACTAGATAA-ATCCTTGACAATACGAATAT
BC  3 CGTACATAAGACATTA-CTCACTT-AATCAACTAAA-TA--TTCTTAACAATACGAATAT
BC  4 CGTGCATAACACATTC-CTTACTT-AATCAACTAGA-TAAATCCTTAACAATACGAATAT
BC  1 CGTACATAAGACATTA-CTCACTT-AATCAACTAAA-TA--TTCTTAACAATACGAATAT
BC  2 CGTACATAAGACATTA-CTTACTT-AATCAACTAAA-TA--TTCTTAACAATACGAATAT
WA  1 CGTACATAAGACATTA-CTTACTT-AATCAACTAAA-TA--TTCTTAACAATACGAATAT
WA  2 CGTACATAAGACATTA-CTCACTT-AATCAACTAAA-TA--TTCTCAACAATACGAATAT
OR  4 CGTACATAATACATTTTCCTACTTGAACCAACTAAATTA--TTATATACAATACGAATAT
CO  3 CGTACATAAGACATTC-CCTACTT-AATCAACTAA-CTA--TTCCTAACAATACGAATAT
AK    CGTACATAACACATTT-AACACTT-AACCAACTAAACTA--TCCCTTACAGTACGGATAT
CO  7 CGTACATAAGACATTC-CTTACTT-AATCAACTAA-CTA--TTCCTAACAATACGAATAT
OR  2 CGTACATAAGACATTA-TCTACTT-AACCAACTAAA-TA--TTCTTAACAATACGAATAT
OR  3 CGTACATAAGACATTA-TCTACTT-AACCAACTAAA-TA--TTCTTAACAATACGAATAT
UT  6 CGTACATAACACATTC-CTCACTT-AATCAACTAAA-TA-ATCCTTAACAATACGAATAT
UT  9 CGTACATAATACATTT-CCTACTTGAATCAACTAAACTAT-TATATAACAATACGAATAT
NM  2 CGTACATAAGACATTC-CCTACTT-AACCAACTAA-CTA--TTCCTAACAATACGAATAT
NV  1 CGTACATAATACATTT-CCTACTTGAATCAACTAAATTAT-TATATAACAACACGAATAT
```

```
                 4         5         5         5         5         5
                 9         0         1         2         3         4
                 0         0         0         0         0         0
UT 11TCACGAATACTTGATC-TTCATAATCGTACATAGGACATGAACTCCATAATCGTACATAG
UT 10TCACGAATACTTGATC-TTCATAATCGTACATAGGACATGAACTCCATAATCGTACATAG
NV 2 CCATAAATACTTGACC-TTCATAATCGTACATAGGACATGAACTCCATAATCGTACATAG
NV 3 TCATAAACACTTGATC-TTCATAATCTTACATAGGACATGAACTCCATAATCGTACATAG
UT 3 TCATTAATACTTGACC-TTCATAATCGTACATAGTACATGAACTCCGTAATCGTACATAG
UT 1 TCATAAACACTTGATC-TTCATAATCTTACATAGGACATGAACTCCATAATCGTACATAG
UT 2 TCATAAACACTTGATC-TTCATAATCTTACATAGGACATGAACTCCATAATCGTACATAG
UT 4 TCATTAATACTTGACC-TTCATAATCGTACATAGGACATGAACTCCGTAATCGTGCATAG
UT 5 TCATTAATACTTGACC-TTCATAATCGTACATAGGACATGAACTCCATAATCGTACATAG
UT 12TCACGATTACTTGATC-TTCATAATCGTACATAGGACATGAACTCCATAATCGTACATAG
CO 2 TCATGATTACTTGATC-TTCATAATCGTACATAGGACATGAACTCCATAATCGTACATAG
CO 4 TCATGATTACTTGATC-TTCATAATCGTACATAGGACATGAACTCCGTAATCGTACATAG
CO 1 TCATAAATACTTGATC-TTCATAATTTTACATAGGACATGGACTCCATAATCGTACATAG
CO 5 TCATAATTACTTGACC-TTCATAATCGTACATAGGACATGAATTCCATAATCGTACATAG
WY 6 TCATAAATACTTGATC-TTCATAATTTTACATAGGACATGGACTCCATAATCGTACATAG
WY 5 TCACGAATACTTGATC-TTCATAATTTTACATAGGACATGGACTCCATAATCGTACATAG
WY 2 TCATAAATACTTGATC-TTCATAATTTTACATAGGACATGGACTCCATAATCGTACATAG
WY 3 TCATAAATACTTGATC-TTCATAATTTTACATAGGACATGGACTCCATAATCGTACATAG
WY 1 CCATAAATACTTGATC-TTCATAATTTTACATAGGACATGAACTCCATAATCGTACATAG
WY 4 TCATAAATACTTGATC-TTCATAATTTTACATAGGACATGGACTCCATAATCGTACATAG
MT 4 TCATAAATACTTGATC-TTCATAATTTTACATAGGACATGGACTCCATAATCGTACATAG
MT 3 CCATAAATACTTGATC-TTCATAATTTTACATAGGACATGAACTCCATAATCGTACATAG
MT 1 TCATAAATACTTGATC-TTCATAATCTTACATAGGACATGAACTCCATAATCGTACATAG
MT 2 TCATAAATACTTGATC-TTCATAATCTTACATAGGACATGAACTCCATAATCGTACATAG
ID 1 TCATAAATACTTGATC-TTCATAATCTTGCATAGGACATGAACTCCATAATCGTACATAG
ID 2 TCATAGATACTTGATC-TTCATAATCTTACATAGGACATGAACTCCATAATCGTACATAG
ID 3 TCATAAATACTTGATC-TTCATAATCTTACATAGGACATGAACTCCATAATCGTACATAG
ID 4 TCATAAACACTTGACC-TTCATAATTTTACATAGGACATGAACTCCATAATCGTACATAG
CO 6 TCATAATTACTTGACC-TTCATAATCGTACATAGGACATAAATTCCATAATCGTACATAG
NM 3 TCATGATTACTTGATC-TTCATAATCGTACATAGGACATGAACTCCATAATCGTACATAG
```

```
                  4         5         5         5         5         5
                  9         0         1         2         3         4
                  0         0         0         0         0         0
NM  1 TCATAATTACTTGACC-TTCATAATCGTACATAGGACATGAACTCCATAATCGTACATAG
CA  4 TCATGAATACTTGATC-TTCATAATCGTACATAGGACATGAACTCCATAATCGTACATAG
CA  5 TCACTAATACTTGATC-TTCATAATCGTACATAGGACATGAACTCCATAATCGTACATAG
CA  3 TCATAAATACTTGACC-TTCATAATCGTACATAGGACATGAACTCCATAATCGTACATAA
CA  2 TCATAAATACTTGACC-TTCATAATCGTACATAGGACATGAACTCCATAATCGTACATAG
CA  1 TCATAATTACTTGACC-TTCATAATCGTACATAGGACATGAACTCCATAATCGTACATAG
OR  1 TCACCTATACTTGACC-TTCATAATCGTACATAGGACATGAATTCCTTAATCGTGCATAG
OR  5 TCATAGATACTTGACC-TTCATAATCGTACATAGGACATGAACTCCATAATCGTACATAG
UT  7 TCATTAATACTTGACC-TTCATAATCGTACATAGGACATGAACTCCATAATCGTACATAG
UT  8 TCATTAATACTTGACC-TTCATAATCGTACATAGGACATGAACTCCATAATCGTACATAG
ALB   TCATAAATACTTGATC-TTCATAATCTTACATAGGACATGAACTCCATAACCGTACATAG
BC  5 TCATAAATACTTGACC-TTCATAATCTTACATAGGACATGAACTCCATAATCGTACATAG
BC  3 TCACCTATACTTGATC-TTCATAATCGTACATAGGACATGAATTCCTTAATCGTACATAG
BC  4 CCATAAATACTTGATC-TTCATAATCTTACATAGGACATGAACTCCATAACCGTACATAG
BC  1 TCACCTATACTTGAAC-TTCATAATCGTACATGGGACATGAATTCCTTAATCGTACATAG
BC  2 TCACCTATACTTGATC-TTCATAATCGTACATAGGACATGAACTCCTTAATCGTACATAG
WA  1 TCACCTATACTTGATC-TTCATAATCGTACATAGGACATGAATTCCTTAATCGTACATAG
WA  2 TCATCTATACTTGACC-TTCATAATCGTACATAGGACATGAATTCCTTAATCGTACATAG
OR  4 TCATAATTACTTGACC-TTCATAATCGTACATAGGACATGAACTCCATAATCGTACATAG
CO  3 TCATGATTACTTGATC-TTCATAATCGTACATAGGACATGAACTCCGTAATCGTACATAG
AK    CCATCAACACTTGAAAATTCATAATCGTACATGGGACATAAATTCCATA-TCGTACATAG
CO  7 TCATGATTACTTGATC-TTCATAATCGTACATAGGACATGAACTCCATAATCGTGCATAG
OR  2 TCATCTATACTTGATC-TTCATAATCGTACATAGGACATGAATTCCTTAATCGTGCATAG
OR  3 TCATCTATACTTGATC-TTCATAATCGTACATAGGACATGAATTCCTTAATCGTGCATAG
UT  6 TCATTAATACTTGACC-TTCATAATCGTACATAGGACATGAACTCCATAATCGTACATAG
UT  9 TCACGAATACTTGATC-TTCATAATCGTACATAGGACATGAACTCCATAATCGTACATAG
NM  2 TCATGATTACTTGATC-TTCATAATCGTACATAGGACATAAATTCCATAATCGTACATAG
NV  1 CCATAAATACTTGACC-TTCATAATCGTACATAGGACATGAACTCCATAATCGTACATAG
```

```
                  5         5         5         5         5         6
                  5         6         7         8         9         0
                  0         0         0         0         0         0
UT 11ACCATATCATCCAAACTGACATTAAAACCCTC-ACAACATGACTATCCCCTTCCCCAGGA
UT 10ACCATATCATCCAAACTGACATTAAAACCCTT-ACAACATGACTATCCCCTTCCCCAGGA
NV 2 ACCATAACATCCAAACTGACATTACAATCCTT-ACAACATGACTATCCCCTTCCCCAGGA
NV 3 ACCATACCATCCAAACTGACATTACAACCCTC-TCAACATGACTATCCCCCTCCCCCAGG
UT 3 ACCATAATACCCAAACTGACATTAAAACCTCC-TCAACATGCCTATCCCCTACCCCCAGG
UT 1 ACCATATCATCCAAACTGACATCATAACCTCC-CCAACATGACTATCCCCCTCCCCCAGG
UT 2 ACCATATCATCCAAACTGACATCATAACCTCC-CCAACATGACTATCCCCCTCCCCCANG
UT 4 ACCATAACACCCAAACTGACATTACAACCTCC-TCAACATGCCTATCCCCCACCCCCAGG
UT 5 ACCATAACATCCAAACTGACATTACAACCTTC-TCAACATGTCTATCCTCTACCCTCAGG
UT 12ACCATAACATCCAAACTGACATCACAACCCCC-TCAACATGACTATCCCCTTCCCCCAGG
CO 2 ACCATAACACCCAAACTGACATCACAACCCCC-CCAACATGACTATCCCCTTCCCCCAGG
CO 4 ACCATAACACCCAAACTGACATCACAACCCCC-CCAACATGACTATCCCCTTCCCCCAGG
CO 1 ACCATACCATCCAAACTGACATTACAACCCTG-TCAACATGACTATCCCCCTCCCCCAGG
CO 5 ACCATAACACCCAAACTGACATCACAACCTCT-TCAACATGACTATCCCCTTCCCCCAGG
WY 6 ACCATACCATCCAAACTGACATTATAACCCTC-TCAACATGACTATCCCCCTCCCCCANG
WY 5 ACCATACCATCCAAACTGACATTATAACCCTC-TCAACATGACTATCCCCCTCCCCCAGG
WY 2 ACCATACTATCCAAACTGACATTATAACCCTC-TCAACATGACTATCCCCCTCCCCCAGG
WY 3 ACCATACTATCCAAACTGACATTATAACCCTC-TCAACATGACTA-CCCCTTCCCCCAGG
WY 1 ACCATACCATCCAAACTGACATTACAACCCCT-TCAACATGACTATCCCCCTCCCCCAGG
WY 4 ACCATACTATCCAAACTGACATTATAACCCTC-TCAACATGACTATCCCCCTCCCCCAGG
MT 4 ACCATACTATCCAAACTGACATTATAACCCTC-TCAACATGACTATCCCCCTCCCCCANG
MT 3 ACCATAAAATCCAAACTGACATCACAACCCTT-CCAACATGACTATCCCCCTCCCCCAGG
MT 1 ACCATATCATCCAAACTGACATTACAACCCTT-TCAACATGACTATCCCTCTCCCCCAGG
MT 2 ACCATATCATCCAAACTGACATTACAACCCTC-TCAACATGACTATCCCTCTCCCCCAGG
ID 1 ACCATACCATCCAAACTGACATTACAACCCTT-TCAACATGACTATCCCTCTCCCCTAGG
ID 2 ACCATACCATCCAAACTGACATTA-AACCCTCACCAACATGACTATCCCCCTCCCCCAGG
ID 3 ACCATACCATCCAAACTGACATTACAACCCTC-TCAACATGACTATCCCTCTCCCCAGGA
ID 4 ACCATACCATCCAAACTGACATTACAACCCTC-TCAACATGACTATCCCCCTCCCCCAGG
CO 6 ACCATAACACCCAAACTGACATCACAACCCTC-TCAACATGACTATCCCCTTCCCCCAGG
NM 3 ACCATAACACCCAAACTGACATCACAACCCCC-TCAACATGACTATCCCCTTCCCCCAGG
```

```
                 5         5         5         5         5         6
                 5         6         7         8         9         0
                 0         0         0         0         0         0
NM 1 ACCATAACACCCAAACTGACATCACAACCCCC-TCAACATGACTATCCCCTTCCCCCAGG
CA 4 ACCATAACATCCAAACTGACATTACAACCCCC-ACAACATGACTATCCCCTTCCCCAGGA
CA 5 ACCATAACGTCCAAACTGACATTACAATCCCC-ACAACATGACTATCCCCTTCCCCAGAG
CA 3 ACCATAACATCCAAACTGACATTACAATCCTC-ACAACATGACTATCCCCTTCCCCAGGA
CA 2 ACCATAACGTCCAAACTGACATTACGACCTTC-ACAACATGACTATCCCCTTCCCCAGGG
CA 1 ACCATAACATCCAAACTGACATTACAACCCTCAACAACATGACTATCCCCTTCCCCAGGG
OR 1 ACCATAACATCTAAACTGACATTACAACCTCC-TAACCATGACTATCCCCTTCCCCTAGG
OR 5 ACCATAACGTCCAAACTGACATTACGACCTTC-ACAACATGACTATCCCCTTCCCCAGGG
UT 7 ACCATAATATCCAAACTGACATTACAACCTCC-TCAACATGTCTATCCCCCACCCCCAGG
UT 8 ACCATAATATCCAAACTGACATTACAACCTCC-TCAACATGTCTATCCCCCACCCCCAGG
ALB  ACCATATCATCCAAACTGACATTACAACCCTCGTCAACATGACTATCCCTCTCCCCCAGG
BC 5 ACCATATTATCCAAACTGACATTACGACCCTT-TCAACATGACTATCCCTCTCCCCCAGG
BC 3 ACCATAACATCTAAACTGACATTACAACCTCC-TAACCATGACTATCCCCTTCCCCCAGG
BC 4 ACCATATCATCCAAACTGGCATTACAACCCCC-TCAACATGACTATCCCTCTCCCCCAGG
BC 1 ACCATAACATCTAAACTGACATTACAACCTCC-TAACCATGACTATCCCCTTCCCCCAGG
BC 2 ACCATAACACCTAAACTGACATTACAACCTCC-TAACCATGACTATCCCCTTCCCCCAGG
WA 1 ACCATAACATCTAAACTGACATTACAACCTCC-TAACCATGACTATCCCCCTCCCCCAAG
WA 2 ACCATAACATCTAAACTGACATTACAACCTCC-TAACCATGACTATCCCCTTCCCCCAGG
OR 4 ACCATAACGTCCAAACTGACATTACGACCCTC-ACAACATGACTATCCCCTTCCCCAGGG
CO 3 ACCATAACACCCAAACTGACATCACAACCCCT-TCAACATGACTATCCCCTTCCCCCAGG
AK   ACCATACTATCCAAACTGACATTACACCCTCC-ACAACATGACTATCCCCTTCCCCCCGG
CO 7 ACCATAACACCCAAACTGACATCACAACCCCC-TCAACATGACTATCCCCTTCCCCCAGG
OR 2 ACCATAACATCTAAACTGACATTACAACCTCC-TAACCATGACTATCCCCTTCCCCTAAG
OR 3 ACCATAACATCTAAACTGACATTACAACCTCC-TAACCATGACTATCCCCTTCCCCTAAG
UT 6 ACCATAACATCCAAACTGACATTACAACCTTC-CCAACATGTCTATCCTCTACCCTCAGG
UT 9 ACCATATCATCCAAACTGACATTAAAACCCTC-ACAACATGACTATCCCCTTCCCCAGGA
NM 2 ACCATAACACCCAAACTGACATTACAACCCCT-TCAACATGACTATCCCCTTCCCCCAGG
NV 1 ACCATAACATCCAAACTGACATTACAATCCTT-ACAACATGACTATCCCCTTCCCCAGGA
```

```
                  6         6         6         6         6         6
                  1         2         3         4         5         6
                  0         0         0         0         0         0
UT 11GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
UT 10GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
NV 2 GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
NV 3 GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
UT 3 AATCCCTTGATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
UT 1 GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATTCGT-ATCCCTCTT
UT 2 GGTCCCTTAATCTAACATCCTCCGTGAAACCAAGAAACCGCCCAATTCGT-ATCCCTCCT
UT 4 AATCCCTTGATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
UT 5 AATCCCTTGATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
UT 12AGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGG-ATCCCTCTT
CO 2 AGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGG-ATCCCCCTT
CO 4 AGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGG-ANNCCCCTT
CO 1 AGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
CO 5 AGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATGCGGGATCCCTCTT
WY 6 GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
WY 5 GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
WY 2 GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
WY 3 GGTCC-TTAATCTACCATCCTCCGTGAAACCAGCAACCCGCC-AATAAGT-ATCCCTCTT
WY 1 GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
WY 4 GGTCNCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
MT 4 GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
MT 3 GATCTCTTAATCTAACATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
MT 1 GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
MT 2 GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
ID 1 GGTCNCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
ID 2 GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
ID 3 GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
ID 4 GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
CO 6 AGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATGCGGGATCCCTCTT
NM 3 AGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGGGATCCCTCTT
```

```
                 6         6         6         6         6         6
                 1         2         3         4         5         6
                 0         0         0         0         0         0
NM  1 AGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGG-ATCCCTCTT
CA  4 AGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
CA  5 GGTCTCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
CA  3 GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
CA  2 GGTCCCTTAATCTACCATCCTCCGTGAAATCAGCAACCCGCCCAATACGT-ATCCCTCTT
CA  1 GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
OR  1 GGTCTCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
OR  5 AGTCCCTTAATCTACCATCCTCCGTGAAATCAGCAACCCGCCCAATACGT-ATCCCTCTT
UT  7 AATCCCTTGATCTACCATCCTCCGTGANACCAGCAACCCGCCCAATNCGT-ATCCCTCTT
UT  8 AATCCCTTGATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
ALB   GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
BC  5 GGTCCCTTAATCTACCATCCTCCGTGAGACCAGCAACCCGCCCAATGCGT-ATCCCTCTT
BC  3 GGTCNCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
BC  4 GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
BC  1 GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
BC  2 GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
WA  1 GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCC-TCTT
WA  2 GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
OR  4 GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
CO  3 AGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGG-ATCCCTCTT
AK    GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCTTCTT
CO  7 AGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGG-ATCCCTCTT
OR  2 GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
OR  3 GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
UT  6 AATCCCTTGATCTACCATCCTCCGTGAAANNAGTAACCCGCCCAATACGT-ATCCCTCTT
UT  9 GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
NM  2 AGTCCTT--ATCTACAATCCTCCGTGAAACAAGCAACCCGCCCAATACGG-ATCCC-CTT
NV  1 GGTCCCTTAATCTACCATCCTCCGTGAAACCAGCAACCCGCCCAATACGT-ATCCCTCTT
```

```
                    6         6         6
                    7         8         8
                    0         0         9
UT 11      CTTGCTCT--GAGCCCA-TGAC-CCTTGG
UT 10      CTTGCTCT--GAGCCCA-TGAC-CCTTGG
NV 2       CTTGCTCT--GAGCCCA-TGAC-CCTTGG
NV 3       CTTGCTCT--GCGCCCA-TGAC-CTCTGG
UT 3       CTTGCTCT--GAGCCCA-TAAC-CCTTGG
UT 1       CTTGCTCT--GCGCCCA-TGAC-CTTTGG
UT 2       CCTGCTCT--GCGCCCAATGAC-CTCTGG
UT 4       CTTGCTCT--GAGCCCA-TAAC-TCTTGG
UT 5       CCTGCTCC--GAGCCCA-TAAC-CCTTGG
UT 12      CTTGCTCT--GCGCCCA-TAAC-CCTTGG
CO 2       CTCGCTCT--GCGCCCA-TGAA-CCTTGG
CO 4       CTTGCTCT--GCGCCCA-TGAA-CCTTGG
CO 1       CTTGCTCT--GCGCCCA-TGAC-CTCTGG
CO 5       CTTGCTCT--GCGCCCA-TGAC-CCTTGG
WY 6       CTTGCTCT--GCGCCCA-TGAA-CTCTGG
WY 5       CTTGCTCT--GTGCCCA-TGAC-GTCTGG
WY 2       CTTGCTCT--GCGCCCA-TGAA-CTCTGG
WY 3       CT-GCTCT--GCGCCCA-TGAA-CTCTGG
WY 1       CTTGCTCT--GCGCCCA-TGAA-CTCTGG
WY 4       CTTGCTCT--GCGCCCA-TGAA-CTCTGG
MT 4       CCTGCTCT--GCGCCAA-TGAC-CTCTGG
MT 3       CTTGCTCT--GCGCCCA-TGAC-CTCTGG
MT 1       CTTGCTCT--GCGCCCA-TGAC-CTCTGG
MT 2       CTTGCTCT--GCGCCCA-TGAC--TCTGG
ID 1       CTTGCTCT--GCGCCCA-TGAA-CTCTGG
ID 2       CTTGCTCT--GCGCCCA-TGAC-CTCTGG
ID 3       CTTGCTCT--GCGCCCA-TGAC-CTTTGG
ID 4       CTTGCTCT--GCGCCCA-TGAC-CTCTGG
CO 6       CTTGCTCT--GCGCCCA-TGAC-CCTTGG
NM 3       CTTGCTCT--GCGCCCA-TGAC-C-TTGG
```

```
                    6         6         6
                    7         8         8
                    0         0         9
NM 1        CTTGCTCT--GCGCCCA-TGAC-CCTTGG
CA 4        CTTGCTCT--GAGCCCN-TGAC-CCTTGG
CA 5        CTTGCTCT--GAGCCCA-TGAC-CCTTGG
CA 3        CTTGCTCT--GAGCCCA-TGAC-C-TTGG
CA 2        CTTGCTCT--GAGCCCA-TGAC-CCTTGG
CA 1        CTTGCTCT--GAGCCCA-TGAC-CCTTGG
OR 1        CTCGCTCT--GCGCCCA-TGAA-CCTTGG
OR 5        CTTGCTCT--GAGCCCA-TGAC-CCTTGG
UT 7        CTTGCTCT--GAGCCCA-TAAC-CCTTGG
UT 8        CTTGCTCT--GAGCCCA-TAAC-CCTTGG
ALB         CTTGCTCT--GCGCCCA-TGAC-CTCTGG
BC 5        CTTGCT----------------------
BC 3        CTCGCTCT--GCGCCCA-TGAA-CCTTGG
BC 4        CTTGCTCT--GCGCCCA-TGACCTCCTGG
BC 1        CCCGCTCT--GCGCCCA-TGAA-CCTTGG
BC 2        CTCGCTCT--GCGCCCA-TGAA-CCTTGG
WA 1        CTCGCTCT--GCGCC-A-TGAA-CCTTGG
WA 2        CTCGCTCT--GCGCCCA-TGAA-CCTTGG
OR 4        CTTGCTCT--GAACCCA-TGAA-CCTTGG
CO 3        CTTGCTCT--GCGCCCA-TGACCCTTTGG
AK          CTCGCTC-GTGAGCCCA-TGAC-ACTTGG
CO 7        CTTGCTCT--GCGCCCA-TGAA-CCTTGG
OR 2        CTCGCTCT--GCGCCCA-TGAA-CCTTGG
OR 3        CTCGCTCT--GCGCCCA-TGAA-CCNTGG
UT 6        CT-GGTCT--GAG-CCA-TA---CCTTGG
UT 9        CTTGCTCT--GAG-CCA-TGA-CCCTTGG
NM 2        CT-GCTCT--GCGCCCAATGAACCCTTGG
NV 1        CCTGCTCT--GAGCCCAATGA-CCCTTGG
```

Appendix C

Aligned DNA sequences from the polymerase polygene (*pol*) from 17 lentiviruses. Sample names and GenBank accession numbers correspond to those listed in Table 3.2. Each row is 60bp long. Total sequence length is 3035bp. Nucleotide sequences were back translated from protein sequences using GenBank nucleotide data. Protein alignments were carried out with Clustal W and employed the BLOSUM protein matrix. Gap opening penalty was 10, gap extension penalty was 0.05.

```
                    10        20        30        40        50        60
                    .         .         .         .         .         .
HIV2a     ATTTTTGGCAGAAACATTTTGACAGCCTTAGGCATGTCATTAAACCTACCA---------
HIV2b     ATTTTTGGCAGAAACATTCTGACAGCCTTAGGCATGTCATTAAATCTACCA---------
SIVb      ATTTTTGGCAGAAATCTGTTAACAGCTATGGGCATGTCCTTAAATCTCCCC---------
SIVcl     ATTGTAGGGAGGAATATTTTAACTGCAATGGGGGCCAAATTGATTTTAGCTCAGTTA---
SIVa      ATCATAGGAAGAAACTTTTTAGCCCAGGCAGGAGCCAAATTAGTGATGGGGCAATTG---
HIV1a     ATAATTGGAAGGAATATATTGACTCAAATTGGTTGCACCTTAAATTTTCCA---------
HIV1b     ATAATTGGAAGAAATCTGTTGACTCAGATTGGCTGCACTTTAAATTTTCCC---------
BIVa      CTTTTTGGGAGATCTCTTCTACGTAGCATAGTGACTTGCTTCACCCTACTTATTCACACA
BIVb      CTTTTTGGGAGATCTCTTCTACGTAGCATAGTGACTTGCTTCACCCTACTTGTTCACACA
CAEV      GTATTAGGACGAGATAACATGGCCCGATTTGGAATAAAGATAATAATGGCAAATTTAGAG
OVL       GTATTAGGAAGGGATAATATGGCAAAATTAGACATAGGAATAATTATGGCAAATTTAGAA
Visnaa    GTATTAGGAAGAGATAATATGAGAGAATTGGGAATAGGATTAATTATGGCAAATTTAGAA
Visnab    GTATTAGGAAGAGATAATATGAGAGAATTAGGAATAGGATTAATTATGGCAAATTTAGAA
FIVa      TTATTAGGGAGAGATAATATGATTAAATTCAATATTAGGTTAGTAATGGCTCAAATTTCT
FIVb      TTATTAGGAAGAGATAACATGATTAAGTTTAATATAAGATTGGTAATGGCTCAGATTTCA
EIAVa     ATTTTGGGACGAGATATTCTTCAGGACTTAGGTGCAAAATTGGTTTTGGCACAGCTCTCC
EIAVb     ATTTTGGGACGAGATATTCTTCAGGACTTAGGTGCAAAATTGGTTTTGGCACAGCTCTCC

                    70        80        90        100       110       120
                    .         .         .         .         .         .
HIV2a     GTTGCCAAGATAGAGCCAATAGAGGTAAGATTAAAGCCAGGAAAAGACGGGCCAAAATTA
HIV2b     GTCGCCAAGATAGAACCAATAAAAATAATGCTGAAGCCAGGAAAGGATGGACCAAAACTG
SIVb      GTAGCTAAGGTGGAGCCTATAAAAGTAACACTAAAACCAGGAAAAGACGGACCAAAATTG
SIVcl     AGTGATAAAATTCCTATTACAAAGGTATCTTTAAAACCTGGGTGTGATGGACCAAGAGTA
SIVa      TCGCAGACAATACCAATCACCCCGGTACGCTTAAAGGAAGGGGCCAGAGGACCACGATTG
HIV1a     ATTAGTCCTATTGAAACTGTACCAGTAAAATTAAAGCCAGGAATGGATGGCCCAAAGGTT
HIV1b     ATTAGTCCTATTGAAACTGTACCAGTAAAATTGAAGCCAGGAATGGATGGCCCAAGGGTT
BIVa      GAAAAA---ATCGAACCCCTACCCGTCAAGGTAAGG---GGA---CCAGGGCCTAAGGTA
BIVb      GAAAAA---ATCGAACCCCTACCCGTCAAGGTAAGG---GGA---CCAGGGCCTAAGGTA
CAEV      GAAAAAAGAATCCCAATTACAAAAGTAAAATTGAAAGAGGGATGTACGGGTCCACATGTC
OVL       GAAAAGAAAATTCCCATTACACAGGTAAAATTAAAAGAAGGCTGTAAGGGTCCTCATATA
Visnaa    GAAAAGAAAATTCCCAGTACAAGAGTAAGATTAAAAGAGGGATGTAAGGGACCCCACATA
Visnab    GAAAAGAAAATTCCCAGTACAAGAGTAAGATTAAAAGAGGGATGTAAGGGACCCCACATA
FIVa      GATAAG---ATTCCAGTAGTAAAAGTAAAAATGAAGGATCCTAATAAAGGACCTCAAATA
FIVb      GAAAAA---ATTCCAATAGTAAAGGTAAGGATGAGGGACCCTATTCAAGGGCCTCAGGTA
EIAVa     AAGGAA---ATAAAATTTAGAAAAATAGAGTTAAAAGAGGGCACAATGGGGCCAAAAATT
EIAVb     AAGGAA---ATAAAATTTAGAAAAATAGAGTTAAAAGAGGGCACAATGGGGCCAAAAATT
```

```
                  130        140        150        160        170        180
                   .          .          .          .          .          .
HIV2a       AGACAATGGCCCTTAACAAAAGAAAAAATAGAGGCACTAAAAGAAATCTGTGAAAAAACG
HIV2b       AGACAATGGCCCTTAACAAAAGAAAAAATAGAGGCACTAAAAGAGATCTGTGAGAAAATG
SIVb        AGACAGTGGCCGCTATCAAAAGAAAAAATAGTTGCATTAAGAGAAATCTGTGAAAAAATG
SIVcl       AAGCAGTGGCCTCTATCAAAAGAGAAAATAGAAGGCCTACAAGCTATTTGTGATAGGTTA
SIVa        AAGCAATGGCCACTCTCTAAAGAAAAAATAATAGCCCTGCAAGAAATTTGCAAAACATTA
HIV1a       AAACAATGGCCATTGACAGAAGAAAAAATAAAAGCATTAACGGAAATTTGTACAGAGATG
HIV1b       AAACAATGGCCATTGACAGAAGAAAAAATAAAAGCATTAACAGAAATATGTACAGAAATG
BIVa        CCCCAGTGGCCCTTGACAAAAGAAAAGTATCAGGCTCTTAAGGAAATTGTGAAAGATCTT
BIVb        CCCCAGTGGCCCTTGACAAAAGAAAAGTATCAGGCTCTTAAGGAAATTGTGAAAGATCTT
CAEV        CCACAATGGCCATTAACAGAAGAGAAATTAAAAGGTCTAACAGAAATCATAGATAAATTA
OVL         GCACAATGGCCTTTAACTCAAGAAAAATTAGAAGGCTTAAAAGAAATAGTGGACAAGTTA
Visnaa      GCGCAATGGCCTTTGACGCAAGAAAAATTAGAGGGATTAAAAGAAATAGTAGACAGATTA
Visnab      GCGCAATGGCCTTTGACGCAAGAAAAATTAGAGGGATTAAAAGAAATAGTAGACAGATTA
FIVa        AAACAATGGCCATTAACAAATGAAAAAATTGAAGCCTTAACAGAAATAGTAGAAAGACTA
FIVb        AAACAGTGGCCATTATCAAATGAAAAGATTGAAGCTTTAACAGACATAGTAGAAAGATTA
EIAVa       CCTCAATGGCCACTCACTAAGGAGAAACTAGAAGGGGCCAAAGAGATAGTCCAAAGACTA
EIAVb       CCTCAATGGCCACTCACTAAGGAGAAACTAGAAGGGGCTAAAGAGATAGTCCAAAGACTA

                  190        200        210        220        230        240
                   .          .          .          .          .          .
HIV2a       GAAAGAGAAGGCCAATTAGAGGAGGCACCTCCAACTAACCCTTATAATACCCCCACATTT
HIV2b       GAAAGAGAGGGCCAGCTAGAGGAGGCACCTCCAACTAATCCTTATAATACCCCCACATTT
SIVb        GAAAAAGATGGCCAGTTAGAGGAAGCCCCTCCAACCAATCCATATAACACCCCCACATTT
SIVcl       GAAAAAGAAGGAAAAATATCTCCAGTGGATCCAGGGAACCCATACAATACTCCAATATTT
SIVa        GAGGAAGAAGGAAAATTAAGCAGGGTAGGGGGAGACAATGCATACAATACACCAGTATTC
HIV1a       GAAAAAGAAGGAAAAATCTCAAGAATAGGGCCTGAGAATCCATACAGCACTCCAATATTT
HIV1b       GAAAAAGAAGGAAAAATTTCAAAAATTGGGCCCGAAAATCCATACAATACTCCAGTATTT
BIVa        TTAGCAGAAGGAAAAATTTCCGAAGCTGCTTGGGATAACCCATATAATACCCCAGTTTTT
BIVb        TTAGCAGAAGGAAAAATTTCCGAAGCTGCTTGGGATAACCCATATAATACCCCAGTTTTT
CAEV        GTGGAAGAAGGAAAACTAGGAAAGGCACCCCCACATTGGACATGTAATACTCCAATCTTT
OVL         GAAAAGGAAGGAAAGGTAGGTAGAGCGCCGCCACATTGGACATGTAATACTCCTATATTT
Visnaa      GAGAAGGAAGGGAAAGTAGGAAGAGCGCCCCCACACTGGACTTGTAATACCCCTATATTT
Visnab      GAGAAGGAAGGGAAAGTAGGAAGAGCGCCCCCACACTGGACTTGTAATACCCCTATATTT
FIVa        GAAAGAGAAGGGAAAGTAAAAAGAGCAGATCCAAATAATCCATGGAATACACCAGTATTT
FIVb        GAATCAGAAGGAAAAGTGAAAAGGGCTGACCCGAATAATCCTTGGAATACCCCAGTATTC
EIAVa       TTGTCAGAGGGAAAAATATCAGAAGCTAGTGACAATAATCCTTATAATTCACCCATATTT
EIAVb       TTGTCAGAGGGAAAAATATCAGAAGCTAGTGACAATAATCCTTATAATTCACCCATATTT
```

```
                  250        260        270        280        290        300
                   .          .          .          .          .          .
HIV2a      GCAATAAAGAAGAAGGACAAAAACAAATGGAGAATGCTAATAGATTTTAGAGAATTAAAC
HIV2b      GCAATCAAGAAAAAGGACAAAAACAAATGGAGAATGCTAATAGATTTTAGAGAACTAAAC
SIVb       GCTATAAAAAAGAAAGACAAAAATAAATGGAGGATGCTAATAGATTTTAGAGAATTGAAT
SIVcl      GCTATCAAGAAGAAAGACAAAAATGAATGGAGAAAGTTAATTGATTTCAGAAAGCTCAAT
SIVa       TGTATAAGGAAAAAAGACAAATCACAGTGGAGAATGCTGGTAGATTTCAGGGAACTCAAC
HIV1a      GCCATAAAAAAGAAGGATAGTACTAAATGGAGAAAATTAGTGGATTTCAGAGAACTCAAT
HIV1b      GCCATAAAGAAAAAAGATAGTACTAAATGGAGAAAATTAGTAGATTTCAGAGAACTTAAT
BIVa       GTTATAAAGAAAAAGGGAACGGGAAAATGGAGGATGCTAATGGATTTTAGGGAATTAAAT
BIVb       GTTATAAAGAAAAAGGGAACGGGAAGATGGAGGATGCTAATGGATTTTAGGGAATTAAAT
CAEV       TGCATAAAAAAGAAA---TCAGGGAAGTGGAGAATGTTAATAGATTTCAGAGAATTGAAC
OVL        TGCATCAAGAAAAAA---TCAGGGAAATGGAGAATGTTAATAGATTTTAGGGAATTAAAT
Visnaa     TGTATTAAGAAGAAA---TCAGGAAAATGGAGGATGTTAATAGATTTTAGAGAATTAAAT
Visnab     TGTATTAAGAAGAAA---TCAGGAAAATGGAGGATGTTAATAGATTTTAGAGAATTAAAT
FIVa       GCTATAAAAAAGAAA---AGTGGAAAATGGAGAATGCTCATAGATTTTAGAGAATTAAAC
FIVb       GCTATTAAGAAGAAG---AGTGGTAAATGGCGGATGCTCATAGATTTTAGGGTCCTAAAT
EIAVa      GTAATAAAAAAGAGG---TCTGGCAAATGGAGGTTATTACAAGATCTGAGAGAATTAAAC
EIAVb      GTAATAAAAAAGAGG---TCTGGCAAATGGAGGTTATTACAAGATCTGAGAGAATTAAAC

                  310        320        330        340        350        360
                   .          .          .          .          .          .
HIV2a      AAGGTAACTCAAGATTTCACAGAGATTCAGTTAGGGATTCCACATCCAGCAGGATTAGCC
HIV2b      AAGGTAACTCAAGACTTCACAGAAATCCAGTTAGGAATTCCACACCCAGCAGGACTAGCC
SIVb       AGGGTTACTCAAGATTTTACAGAAGTACAGTTAGGGATACCACACCCTGCAGGACTAGCA
SIVcl      GAATTAACTCAAGACTTTCATGAGCTACAATTAGGTATTCCACACCCAGCAGGTATCAAG
SIVa       AAAGCTACACAAGACTTCTTTGAAGTCCAATTAGGTATACCCCATCCAGCAGGGTTAAAG
HIV1a      AAAAGAACTCAAGACTTCTGGGAAGTTCAGTTAGGAATACCACACCCAGCAGGGTTAAAA
HIV1b      AAGAGAACTCAAGACTTCTGGGAAGTTCAGTTAGGGATACCACATCCCGCAGGGTTAAAG
BIVa       AAGATAACAGTTAAAGGACAAGAATTCTCTACAGGCTTACCTTACCCTCCAGGAATTAAG
BIVb       AAGATAACAGTTAAAGGACAAGAATTCTCTACAGGCTTACCTTACCCTCCAGGAATTAAG
CAEV       AAACAGACAGAAGATTTAACAGAAGCGCAGTTAGGACTCCCGCATCCGGGAGGACTACAA
OVL        AAGCAAACAGAAGATTTGGCGGAGGCACAATTAGGTTTGCCGCATCCAGGGGGATTACAG
Visnaa     AAGCAAACAGAAGATTTAGCAGAAGCACAGTTAGGGTTACCGCATCCAGGAGGATTACAG
Visnab     AAGCAAACAGAAGATTTAGCAGAAGCACAGTTAGGGTTACCGCATCCAGGAGGATTACAG
FIVa       AAACTAACTGAGAAAGGAGCAGAGGTCCAGTTGGGACTACCTCATCCTGCTGGTTTACAA
FIVb       AAATTAACAGACAAAGGGGCAGAAGTCCAGTTAGGACTTCCTCATCCTGCTGGATTACAA
EIAVa      AAAACAGTACAAGTAGGAACGGAAATATCCAGAGGATTGCCTCACCCGGGAGGATTAATT
EIAVb      AAAACAGTACAAGTAGGAACGGAAATATCCAGAGGATTGCCTCACCCGGGAGGATTAATT
```

```
                370        380        390        400        410        420
                .          .          .          .          .          .
HIV2a      AAGAAAAGAAGAATCACTGTGCTGGATGTAGGGGATGCTTACTTTTCCATACCACTGCAT
HIV2b      AAGAAGAAACGAATTACTGTCCTAGATGTAGGGGATGCTTACTTTTCCATACCACTACAT
SIVb       AAGAGGAGAAGAATCACAGTATTGGATGTAGGTGATGCATATTTCTCCATACCTCTAGAT
SIVcl      AAATGCAAACAGATTACAGTAGTAGACATAGGAGATGCCTATTTCAGTATCCCTTTAGAT
SIVa       AAAATGAAGCAAATAACCATTATAGATGTGGGGGATGCATATTATAGCATACCACTGGAT
HIV1a      AAGAAAAAATCAGTATCAGTACTGGATGTGGGGGATGCATATTTTTCAGTCCCTTTAGAT
HIV1b      AAGAAAAAATCAGTAACAGTACTGGATGTGGGTGATGCATATTTTTCAGTTCCATTAGAT
BIVa       GAATGTGAACACTTAACTGCAATAGATATAAAAGATGCCTACTTTACTATCCCTTTACAT
BIVb       GAATGTGAACACTTAACTGCAATAGATATAAAAGATGCCTACTTTACTATCCCTTTACAT
CAEV       AAGAAAAAACATGTTACAATATTGGACATAGGAGATGCATATTTTACTATACCCCTATAT
OVL        AAAAAGAAGCATGTAACAATACTTGACATAGGGGATGCATATTTTACAATACCATTGTAT
Visnaa     AGAAAGAAACATGTAACAATATTAGATATAGGAGATGCATATTTTACAATACCATTATAT
Visnab     AGAAAGAAACATGTAACAATATTAGATATAGGAGATGCATATTTTACAATACCATTATAT
FIVa       ATAAAAAAACAAGTAACAGTATTAGATATAGGGGATGCATATTTCACCATTCCTCTTGAT
FIVb       ATGAAAAAACAAGTAACTGTGTTGGATATAGGGGATGCATATTTTACCATTCCTCTAGAT
EIAVa      AAATGTAAACACATGACTGTATTAGATATTGGAGATGCATATTTCACTATACCCTTAGAT
EIAVb      AAATGTAAACACATGACTGTATTAGATATTGGAGATGCATATTTCACTATACCCTTAGAT

                430        440        450        460        470        480
                .          .          .          .          .          .
HIV2a      GAGAGCTTTAGACAGTATACTGCATTTACTCTACCATCAGTAAACAATGCAGAACCAGGA
HIV2b      GAGGATTTTAGACAGTATACTGCATTTACTCTACCATCAATAAACAATGCTGAACCAGGA
SIVb       GAAGAATTCAGGCAATACACTGCTTTTACTTTACCATCAGTAAACAATGCAGAACCAGGA
SIVcl      CCAAATTACAGAAAGTATACAGCATTTACTATACCTTCTCTCAATAATCAAGAGCCAGGA
SIVa       CCTGAGTTTAGAAAATACACAGCTTTCACCATCCCTACGGTAAACAATGAGGGACCAGGC
HIV1a      AAAGAATTCAGAAAGTATACTGCATTCACCATACCTAGTATAAACAATGAGACACCAGGG
HIV1b      AAAGACTTCAGGAAGTATACTGCATTTACCATACCTAGTACAAACAATGAAACACCAGGG
BIVa       GAGGACTTTAGACCCTTTACAGCCTTCTCTGTAGTCCCTGTAAATCGAGAAGGACCTATA
BIVb       GAGGACTTTAGACCCTTTACAGCCTTCTCTGTAGTCCCTGTAAATCGAGAAGGACCTATA
CAEV       GAACCATATCGAGAGTACACATGTTTTACTCTATTAAGTCCTAATAATCTAGGACCATGT
OVL        GAGCCATATAGACCATATACATGTTTTACCATGTTAAGTCCAAATAATTTGGGACCATGT
Visnaa     GAGCCATATAGACAATATACATGCTTTACCATGTTAAGTCCAAATAATTTAGGACCATGT
Visnab     GAGCCATATAGACAATATACATGCTTTACCATGTTAAGTCCAAATAATTTAGGACCATGT
FIVa       CCAGATTATGCTCCTTATACAGCATTTACTTTACCTAGAAAAAATAATGCGGGACCAGGA
FIVb       CCAGATTATGCTCCTTATACTGCATTTACATTACCCAGAAAGAATAATGCAGGGCCAGGG
EIAVa      CCAGAGTTTAGACCATATACAGCTTTCACTATTCCCTCCATTAATCATCAAGAACCAGAT
EIAVb      CCAGAGTTTAGACCATATACAGCTTTCACTATTCCCTCCATTAATCATCAAGAACCAGAT
```

```
                490       500       510       520       530       540
                .         .         .         .         .         .
HIV2a      AAAAGATATATATATAAAGTCTTACCGCAGGGATGGAAGGGATCACCAGCAATTTTTCAA
HIV2b      AAAAGATACATATATAAAGTCTCACCACAGGGATGGAAGGGATCACCAGCAATTTTTCAG
SIVb       AAAAGATACATCTATAAGGTATTACCTCAAGGGTGGAAAGGGTCACCAGCTATTTTTCAG
SIVcl      AAAAGATATCAGTACAATGTGCTGCCACAAGGGTGGAAGGGGAGCCCTTGTATATTTCAG
SIVa       ATAAGATATCAATTTAATTGCCTACCGCAGGGCTGGAAGGGATCCCCGACAATTTTCCAA
HIV1a      ATTAGATATCAGTATAATGTGCTTCCACAGGGATGGAAAGGATCACCAGCAATATTCCAG
HIV1b      CTTAGATATCAGTACAATGTGCTTCCACAGGGGTGGAAAGGATCACCAGCAATATTCCAA
BIVa       GAAAGGTTCCAGTGGAATGTTCTACCACAAGGATGGATATGTAGCCCTGCCATTTATCAG
BIVb       GAGAGGTTCCAGTGGAATGTTCTACCACAAGGATGGGTATGTAGCCCTGCCATTTATCAG
CAEV       AAAAGATACTATTGGAAAGTGCTGCCACAAGGTTGGAAATTGAGTCCATCTGTATATCAA
OVL        ACACGGTATTATTGGAAAGTACTACCACAAGGATGGAAGTTGAGTCCCTCAGTGTATCAA
Visnaa     GTAAGATATTATTGGAAAGTGTTACCACAAGGATGGAAATTAAGTCCTGCAGTGTATCAA
Visnab     GTAAGATATTATTGGAAAGTGTTACCACAAGGATGGAAATTAAGTCCTGCAGTGTATCAA
FIVa       AGGAGATTTGTGTGGTGTAGTCTACCACAAGGCTGGATTTTAAGTCCATTGATATATCAA
FIVb       AGGAGATATGTATGGTGCAGTTTACCACAGGGGTGGGTTCTAAGTCCATTAATATATCAG
EIAVa      AAAAGATATGTGTGGAAATGTTTACCACAAGGATTCGTGTTGAGCCCATATATATATCAG
EIAVb      AAAAGATATGTGTGGAATTGTTTACCACAAGGATTCGTGTTGAGCCCATATATATATCAG

                550       560       570       580       590       600
                .         .         .         .         .         .
HIV2a      CACACAATGAGACAGATCTTAGAGCCATTCAGAAAGGCAAACCAGGATGTCATTCTCATT
HIV2b      TACACAATGAGGCAGGTCTTAGAACCATTCAGAAAAGCAAACCCGGATATCATTCTCATT
SIVb       CATACTATGAGAAATGTCTTAGGACCTTTCAGAAAAGCAAATCCAGATGTGACCCTGATC
SIVcl      GGAACAGTAGCAGGACTTCTCTCAGAGTTTAGGAAATTAAATCCAGACATGATCATTTAC
SIVa       AACACAGCATCAAAAATTCTAGAAGAAATAAAGAAAGAATTAAAACAGCTGACGATTGTC
HIV1a      AGTAGCATGACAAAAATCTTAGCGCCCTTTAGAGAACAAAATCCTGAAATGGTTATTTAC
HIV1b      AGTAGCATGACAAAAATCTTAGAGCCTTTTAGAAAACAAAATCCAGACATAGTTATCTAT
BIVa       ACTACCACCCAGAAGATTATAGAAAACATTAAAAAGAGTCACCCAGATGTCATGTTGTAT
BIVb       ACTACCACCCAGAAGATTATAGAAAACATTAAAAAGAGTCACCCAGATGTCATGTTGTAT
CAEV       TTTACTATGCAGGAGATCTTAGAGGATTGGATACAGCAGCATCCAGAAATTCAATTTGGC
OVL        TTTACAATGCAAGAAATATTAAGGGATTGGATAGCGAAACATCCTATGATACAATTTGGA
Visnaa     TTTACAATGCAAAAAATATTAAGAGGATGGATAGAAGAACACCCTATGATACAATTTGGA
Visnab     TTTACAATGCAAAAAATATTAAGAGGATGGATAGAAGAACACCCTATGATACAATTTGGA
FIVa       AGTACATTAGATAATATAATACAACCTTTTATTAGACAAAATCCTCAATTAGATATTTAC
FIVb       AGCACCTTAGATAATATATTACAGCCATTTATTAAACAAAATTCTGAGTTAGATATTTAT
EIAVa      AAAACATTACAGGAAATTTTACAACCTTTTAGGGAAAGATATCCTGAAGTACAATTGTAT
EIAVb      AAAACATTACAGGAAATTTTACAACCTTTTAGGGAAAGATATCCTGAAGTACAATTGTAT
```

```
                    610         620         630         640         650         660
                    .           .           .           .           .           .
HIV2a       CAATACATGGATGATATCTTAATAGCTAGTGACAGGACAGATTTAGAACATGACAAGGTG
HIV2b       CAGTACATGGATGATATCTTGATAGCCAGCGACAGGACAGATTTAGAACATGACAGAGTG
SIVb        CAATACATGGATGACATCCTAATAGCTAGTGACAGAACAGATTTAGAGCATGACAGGGTA
SIVcl       CAATATATGGATGATTTATTCATAGGATCAGATAGAGAGAGAAAAGGACATGATCAGGCA
SIVa        CAGTACATGGATGACCTCTGGGTAGGATCACAAGAAGAGGGTCCAAAGCATGATCAGCTA
HIV1a       CAATACATGGATGATTTGTATGTAGGATCTGACTTAGAAATAGGGCAACATAGAGCAAAA
HIV1b       CAATACATGGATGATTTGTATGTAGGATCTGACTTAGAAATAGGGCAGCATAGAACTAAG
BIVa        CAATATATGGATGATTTGTTGATTGGGTCTAAT------AGGGATGATCATAAGCAAATA
BIVb        CAATATATGGATGATTTGTTGATTGGGTCTAAT------AGGGATGATCATAAGCAAATA
CAEV        ATATATATGGATGATATTTACATAGGAAGTGATTTAGAAATTAAAAAGCATAGAGAAATA
OVL         ATATACATGGATGATATTTATATAGGGAGTGATTTGGATATAATGAAACACAGAGAGATA
Visnaa      ATATACATGGATGATATCTATATAGGGAGTGATTTAGGACTAGAAGAGCACAGGGGTATC
Visnab      ATATACATGGATGATATCTATATAGGGAGTGATTTAGGACTAGAAGAGCACAGGGGTATC
FIVa        CAATATATGGATGACATTTATATAGGATCAAATTTAAGTAAAAAGGAGCATAAAGAAAAG
FIVb        CAATATATGGATGATATATATATAGGATCAAATTTAAGTAAAAAAGAACATAAACAAAAA
EIAVa       CAATATATGGATGATTTGTTCATGGGAAGTAATGGTTCTAAAAAACAACACAAAGAGTTA
EIAVb       CAATATATGGATGATTTGTTCGTGGGAAGTAATGGTTCTAAAAAACAACACAAAGAGTTA

                    670         680         690         700         710         720
                    .           .           .           .           .           .
HIV2a       GTCCTGCAGTTAAAGGAACTCCTAAATGGCCTAGGATTTTCCACCCCAGATGAGAAGTTC
HIV2b       GTTCTGCAGCTAAAGGAACTTCTAAATGGCCTGGGATTTTCCACCCCAGATGAGAAGTTC
SIVb        GTTTTACAGTTAAAGGAACTTCTGAACAGCATAGGATTTTCCACCCCAGAAGAGAAGTTC
SIVcl       GTAAAAGAACTCAGAGAACTTCTTATGACATGGAACTTAGAGACACCAGAAAAGAAGTTC
SIVa        GTACAAACACTTAGGAATAGATTGCAAGAATGGGGATTAGAAACACCAGAGAAAAAGGTG
HIV1a       ATAGAGGAGTTAAGAGCTCATTTGTTGAAATGGGGATTTACCACACCAGACAAAAAACAT
HIV1b       ATAGAGGAATTGAGACAGCATTTGTTGAGGTGGGGATTTACCACACCAGACAAAAAACAT
BIVa        GTGCAGGAAATCAGGGATAAGTTAGGATCATATGGTTTCAAGACTCCAGATGAAAAGGTC
BIVb        GTGCAGGAAATCAGGGATAAGTTAGGATCATATGGTTTCAAGACTCCAGATGAAAAGGTC
CAEV        GTGAAAGATTTAGCCAATTATATTGCCCAATATGGATTCACTCTGCCAGAAGAGAAGAGA
OVL         GTAGAAGAACTAGCTAGCTATATTGCCCAGTATGGATTTATGTTACCAGAAGAAAAGAGA
Visnaa      GTGAACGAACTAGCATCATATATAGCGCAATATGGATTCATGCTGCCTGAAGATAAGAGG
Visnab      GTGAACGAACTAGCATCATATATAGCGCAATATGGATTCATGCTGCCTGAAGATAAGCGG
FIVa        GTAGAAGAATTAAGAAAATTACTATTATGGTGGGGATTTGAAACTCCAGAAGATAAATTA
FIVb        GTAGAAGAATTAAGAAAATTGTTATTATGGTGGGGATTTGAAACCCCGGAAGATAAATTA
EIAVa       ATCATAGAATTAAGGGCGATCTTACTGGAAAAGGGTTTTGAGACACCAGATGATAAATTA
EIAVb       ATCATAGAATTAAGGGCAATCTTACTGGAAGAGGGTTTTGAGACACCAGATGATAAATTA
```

```
                730       740       750       760       770       780
                .         .         .         .         .         .
HIV2a      CAAAAAGACCCTCCATACAAATGGATGGGCTATGGACTGTGGCCAACTAAATGGAAGCTG
HIV2b      CAAAAAGACCCTCCATACCAATGGATGGGCTATGAACTGTGGCCAACTAAATGGAAGCTG
SIVb       CAGAAAGATCCCCCATTCCAGTGGATGGGATATGAATTGTGGCCAACCAAATGGAAACTG
SIVcl      CAAGCAGAGCCACCCTATCATTGGATGGGTTATGTACTGCATCCTGATAGGTGGGAAATA
SIVa       CAAAGAGAACCTCCCTTTGAGTGGATGGGATATAAATTATGGCCTCATAAATGGAAGTTA
HIV1a      CAGAAAGAACCCCCATTTCTTTGGATGGGATATGAACTCCATCCTGACAAATGGACAGTA
HIV1b      CAGAAAGAACCTCCATTCCTTTGGATGGGTTATGAACTCCATCCTGATAAATGGACAGTA
BIVa       CAGGAAGAGAGA---GTGAAATGGATCGGTTTTGAGCTCACACCCAAGAAATGGCGTTTT
BIVb       CAGGAAGAGAGA---GTGAAATGGATCGGTTTTGAGCTCACACCCAAGAAATGGCGTTTT
CAEV       CAAAAGGGATATCCAGCAAAATGGCTAGGATTTGAACTACACCCGCAGACCTGGAAATTT
OVL        CAAGAAGGGTATCCAGCAAAGTGGCTTGGATTTGAATTGCACCCAGAGAAATGGAGATTT
Visnaa     CAAGAAGGATACCCGGCTAAATGGCTTGGATTTGAATTGCATCCGGAGAAATGGAAATTT
Visnab     CAAGAAGGATACCCGGCTAAATGGCTTGGATTTGAATTGCATCCGGAGAAATGGAAATTT
FIVa       CAGGAAGAACCCCCATATACATGGATGGGTTATGAATTACATCCATTAACATGGACAATA
FIVb       CAAGAAGAACCTCCATATAAGTGGATGGGCTATGAATTACATCCATTAACATGGTCAATA
EIAVa      CAAGAAGTGCCACCTTATAGCTGGCTAGGTTATCAACTTTGTCCTGAAAATTGGAAAGTA
EIAVb      CAAGAAGTGCCACCTTATAGCTGGCTAGGTTATCAACTTTGTCCTGAAAATTGGAAAGTA

                790       800       810       820       830
                .         .         .         .         .
HIV2a      CAAAAAATACAATTG[CCCCAGAAAGAA------GTATGG]ACAGTCAATGACATCCAAA
HIV2b      CAAAGAATACAATTG[CCCCAAAAGGAA------GTATGG]ACAGTCAATGACATCCAAA
SIVb       CAGAAAATAGAGTTG[CCACAAAGAGAA------ACCTGG]ACAGTAAATGACATACAAA
SIVcl      GAAAAGATTAAATTA[CCAGAGATGGATCTAACAAAAACT]ACAGTAAATCAAATACAGA
SIVa       CAAAGTATAGAATTA[GAGAAGAAAGAA------CAATGG]ACAGTGAATGATCTTCAGA
HIV1a      CAGACTGTAAAACTG[CCAGAAAAAGAC------AGCTGG]ACTGTCAATGATATACAGA
HIV1b      CAGCCTATAGTGCTG[CCAGAAAAAGAC------AGCTGG]ACTGTCAATGACATACAGA
BIVa       CAGCCCAGGCAACTA[------AAGATAAAAAACCCACTC]ACAGTAAATGAATTACAGC
BIVb       CAGCCCAGGCAACTA[------AAGATAAAAAACCCACTC]ACAGTAAATGAATTACAGC
CAEV       CAGAAGCATACATTA[CCTGAATTAACAAAGGGAACAATA]ACATTAAATAAATTACAGA
OVL        CAGAAACATACACTT[CCAGAAATAAAGGAAGGGACCATA]ACATTAAATAAATTACAAA
Visnaa     CAAAAACATACGCTC[CCAGAGATTACAGAAGGACCCATA]ACCTTGAATAAACTACAGA
Visnab     CAAAAACATACGCTC[CCAGAGATTACAGAAGGACCCATA]ACCTTGAATAAACTACAGA
FIVa       CAACAGAAACAGTTA[---GACATTCCAGAACAGCCC---]ACTCTAAATGAGTTGCAAA
FIVb       CAACAAAAACAATTA[---GAAATTCCAGAAAGACCC---]ACATTGAATGAACTACAGA
EIAVa      CAAAAAATGCAATTA[---GACATGGTAAAGAATCCA---]ACCCTTAATGATGTGCAAA
EIAVb      CAAAAAATGCAATTA[---GACATGGTAAAGAATCCA---]ACCCTTAATGATGTGCAAA
```

```
                840         850         860         870         880         890
                .           .           .           .           .           .
HIV2a           AACATGTGGGTGTCCTAAATTGGGCAGCACAAATCTACCCAGGAATAAAGACCAAACACT
HIV2b           AACTGGTGGGTGTCCTAAATTGGGCAGCACAAATCTACCCAGGGATAAAGACCAGAAACT
SIVb            AATTAGTAGGAGTACTAAATTGGGCAGCACAAATTTATCCAGGAATAAAGACTAAACATC
SIVcl           AACTGGTGGGAGTACTTAATTGGGCAGCTCAATTGTATGATGGTATTAGGACAAAAGAAC
SIVa            AATTGGTAGGGAAATTAAATTGGGCAGCACAATTATATCCAGGATTGAGAACAAAAAATA
HIV1a           AGTTAGTGGGAAAACTAAATTGGGCAAGTCAGATTTATCCAAATATTAAAGTAAAGCAAC
HIV1b           AGTTAGTGGGAAAATTGAATTGGGCAAGTCAGATTTATGCAGGGATTAAAGTAAGGCAAT
BIVa            AATTAGTAGGTAATTGTGTTTGGGTA---CAGCCAGAAGTAAAAATCCCTCTATACCCCT
BIVb            AATTAGTAGGTAATTGTGTTTGGGTA---CAGCCAGAAGTAAAAATCCCTCTATACCCCT
CAEV            AATTAGTAGGAGAATTAGTATGGAGA---CAATCCATAATTGGGAAAAGCATTCCTAACA
OVL             AATTAGTAGGAGATTTAGTCTGGAGA---CAATCATTAATAGGAAAAAGCATACCTAATA
Visnaa          AATTAGTAGGAGATTTAGTTTGGAGA---CAATCCCTAATAGGAAAAAGCATCCCAAATA
Visnab          AATTAGTAGGAGATCTAGTTTGGAGA---CAATCCCTAATAGGAAAAAGCATCCCAAATA
FIVa            AATTAGCAGGAAAAATTAATTGGGCTAGCCAAGCTATTCCAGACTTGAGTATAAAAGCAT
FIVb            AATTAGCAGGGAAGATAAACTGGGCCAGTCAAACTATCCCAGACTTAAGTATAAAAGAAC
EIAVa           AATTAATGGGGAATATAACATGGATGAGCTCAGGGATCCCAGGGTTGACAGTAAAACACA
EIAVb           AATTAATGGGGAATATAACATGGATGAGCTCAGGGGTCCCAGGGTTGACAGTAAAACACA

                900         910         920         930         940         950
                .           .           .           .           .           .
HIV2a           TATGTAGACTAATTAGAGGAAAAATGACACTCACAGAAGGAGTGCAGTGGACAGAACTAG
HIV2b           TATGTAGGTTAATCAGAGGAAAAATGACACTCACAGAAGAGGTACAGTGGACAGAATTAG
SIVb            TTTGCAAATTAATCAGAGGAAAAATGACTTTAACAGAAGAGGTTCAGTGGACTGAGATGG
SIVcl           TCTGCAAATTAATAAGGGGAGTAAAACCCTTGGAAGAAATCATAAACTGGACAGAGGAAG
SIVa            TCTGTAAGCTACTTAGAGGAAAGAAAAATTTATTAGACGTGGTAGAATGGACCCCAGAGG
HIV1a           TATGTAAACTCCTTAGGGGGGCCAAAGCATTAACAGACATAATACCACTGACAAAAGAGG
HIV1b           TATGTAAACTCCTTAGGGGAACCAAAGCACTAACAGAAGTAGTACCACTAACAGCAGAGG
BIVa            TAACCGATCTACTGAGGGATAAGACCAATCTCCAAGAAAAGATACAACTAACACCAGAAG
BIVb            TAACCGATCTACTGAGGGATAAGACCAATCTCCAAGAAAAGATACAACTAACACCAGAAG
CAEV            TTCTGAAATTAATGGAAGGAGATAGAGAATTACAAAGTGAAAGAAAAATTGAAGAAGTAC
OVL             TACTAAAGTTAATGGAAGGGGATAGGGCGCTCCAAAGTGAAAGAAGGATAGAGCTCAGAC
Visnaa          TCTTAAAATTAATGGAAGGAGATAGGGCTTTACAAAGTGAAAGATACATAGAGAGTATAC
Visnab          TCTTAAAATTAATGGAAGGAGATAGGGCTTTACAAAGTGAAAGATACATAGAGAGTATAC
FIVa            TAACTAACATGATGAGAGGAAATCAAAACCTAAATTCAACAAGACAATGGACTAAAGAAG
FIVb            TAACTAACATGATGAGAGGAGATCAGAAGTTAGACTCAATAAGAGAATGGACTGTAGAGG
EIAVa           TAGCAGCTACTACTAAGGGATGTTTAGAGTTGAATCAAAAAGTAATTTGGACGGAAGAGG
EIAVb           TAGCAGCTACTACTAAGGGATGTTTAGAGTTGAATCAAAAAGTAATTTGGACGGAAGAGG
```

```
                960        970        980        990        1000       1010
                .          .          .          .          .          .
HIV2a      CAGAAGCAGAACTAGAGGAGAACAGAATTATCTTAAGTCAG[GAACAAGAGGGG]CACTA
HIV2b      CAGAAGCGGAACTAGAAGAAAACAAAATCATCTTAAGCCAG[GAACAAGAAGGA]TGCTA
SIVb       CAGAAGCAGAATATGAAGAAAACAAGATAATTCTCAGTCAA[GAACAGGAGGGA]TGTTA
SIVcl      CCTTGGAAGAATATGGGCAGAACAAAGAGGTACTTAAAGAA[AAGATGCAGGGA]GCCTA
SIVa       CAGAAGCAGAGTACGAAGAAAACAAGGAGATCCTAAAAACA[GAGCAAGAAGGT]ACTTA
HIV1a      CAGAATTGGAATTGGCAGAAAACAGGGAGATTCTGAGAGAA[CCAATACATGGA]GTATA
HIV1b      CAGAGCTAGAACTGGCAGAAAACAGGGAGATTCTAAAAGAA[CCAGTACATGGA]GTGTA
BIVa       CCATCAAGTGTGTAGAAGAATTCAATCTAAAACTAAAAGAT[CCAGAATGGAAA]GATAG
BIVb       CCATCAAGTGTGTAGAAGAATTCAATCTAAAACTAAAAGAT[CCAGAATGGAAA]GATAG
CAEV       ATGTGAAAGAATGGGAAGCATGTAGGAAAAAATTAGAAGAA[ATGGAAGGA---]AATTA
OVL        ATGTAAAGGAATGGGAGGAATGTAGAAGAAAATTAGCAGAA[ATGGAAGGA---]AATTA
Visnaa     ATGTAAGAGAATGGGAAGCCTGTAGACAAAAGCTGAAGGAA[ATGGAAGGA---]AATTA
Visnab     ATGTAAGAGAATGGGAAGCCTGTAGACAAAAGCTGAAGGAA[ATGGAAGGA---]AATTA
FIVa       CTCGACTGGAAGTACAAAAGGCAAAAAAGGCTATAGAAGAA[CAAGTACAACTA]GGATA
FIVb       CCAAGAGAGAAGTACAAAAAGCCAAGGAAGCTATTGAAAAG[CAAGCACAGCTA]AATTA
EIAVa      CACAAAAAGAGTTAGAAGAAAATAATGAGAAGATTAAAAAT[GCTCAAGGGTTA]CAATA
EIAVb      CACAAAAAGAGTTAGAAGAAAATAATGAGAAGATTAAAAAT[GCTCAAGGGTTA]CAATA

                1020       1030       1050       1060       1070       1080
                .          .          .          .          .          .
HIV2a      TTACCAAGAAGAAAAGGAGTTAGAAGCAACAGTCCAA[AAAGAT]CAAGACAATCAATGG
HIV2b      TTACCAAGAGGAAAAGGAGCTAGAAGCAACAGTCCAA[AAAGAT]CAAGACAATCAGTGG
SIVb       CTACCAAGAGGGAAAGCCATTAGAGGCAACAGTAATA[AAGAGT]CAGGATAATCAATGG
SIVcl      TTATGACCCAGAAAAGGAACTTATTGTCAGGGTACAG[CAAAAC]AAAAAGGGGATAATT
SIVa       TTATGCACCAGAAAAACCCCTTAGGGCAGCAGTACAG[AAATTA]GGAGATGGGCAATGG
HIV1a      TTATGATCCATCAAAAGACTTAATAGCAGAAATACGG[AAGCAA]GGGCAAGGCCAATGG
HIV1b      TTATGACCCCTCAAAAGACTTAATAGCAGAAATACAG[AAACAG]GGGCAAGGCCAATGG
BIVa       AATAAGAGAAGGAGCAGAATTAGTCATAAAAATACAG[---ATG]GTTCCTCGGGGCATA
BIVb       AATAAGAGAAGGAGCAGAATTAGTCATAAAAATACAG[---ATG]GTTCCTCGGGGCATA
CAEV       TTATAATAAAGACAAAGATGTCTATGGACAATTGGCT[---TGG]GGAGACAAAGCTATA
OVL        CTATGATGAAGAGAAGGATGTATATGGACAAATAGAT[---TGG]GGAGATAAGGCAATA
Visnaa     TTATGATGAAGAGAAGGATATCTATGGGCAACTAGAT[---TGG]GGAAATAAAGCAATA
Visnab     TTATGATGAAGAGAAGGATATCTATGGGCAACTAGAT[---TGG]GGAAATAAAGCAATA
FIVa       CTATGACCCCAGTAAGGAGTTATATGCTAAATTAAGT[TTGGTG]GGACCACATCAAATA
FIVb       TTATGATCCCAACCGAGGATTATATGCAAAATTGAGT[TTAGTG]GGACCACATCAAATA
EIAVa      TTATAATCCAGAAGAAGAAATGTTATGTGAGGTTGAA[ATTACA]AAAAATTATGAGGCA
EIAVb      TTATAATCCAGAAGAAGAAATGTTATGTGAGGTTGAA[ATTACA]AAAAATTATGAGGCA
```

```
                1090        1100        1110        1120        1130        1140
                .           .           .           .           .           .
HIV2a      ACATATAAAATACACCAG[GGA------GAAAAA]ATTCTAAAAGTGGAAAAGTATGCAA
HIV2b      ACATATAAGATACACCAG[GGA------GGAAAA]ATTCTAAAAGTAGGAAAATATGCAA
SIVb       TCATATAAAATCCACCAA[GAA------GACAAG]ATACTGAAAGTAGGCAAATTTGCAA
SIVcl      ACTTTCCAGTGGAGACAA[GGA------AATAAC]ATCTTAAGAGCTGGGAGGTATCAAA
SIVa       TCATACCAATTCAAGCAG[GAA------GGAAAA]ATCTTAAAGGTAGGGAAGTTCGCCA
HIV1a      ACATATCAAATTTATCAG[GAGCCA---TTTAAA]AATCTGAAGACAGGAAAATATGCAA
HIV1b      ACATATCAAATTTATCAA[GAGCCA---TATAAA]AATTTGAAAACAGGAAAGTATGCAA
BIVa       GTATTTGATCTGTTGCAA[GATGGAAAT------]CCCATATGGGGAGGAGTAAAAGGAC
BIVb       GTATTTGATCTGTTGCAA[GATGGAAAT------]CCCATATGGGGAGGAGTAAAAGGAC
CAEV       GAATATATAGTGTATCAG[GAGAAAGGG---AAA]CCATTATGGGTAAATGTGGTTCACA
OVL        GAGTACATAGTGTTTCAA[GAGAGAGGA---AAA]CCTTTATGGGTAAATGTAGTACATA
Visnaa     GAATACATAGTATTTCAA[GAAAAAGGA---AAA]CCTTTATGGGTAAATGTAGTACATA
Visnab     GAATACATAGTATTTCAA[GAAAAAGGA---AAA]CCTTTATGGGTAAATGTAGTACATA
FIVa       AGTTATCAAGTATATCAG[AAGGATCCAGAAAAG]ATACTATGGTATGGAAAAATGAGTA
FIVb       TGTTATCAAGTGTATCAA[AAGAACCCAGAACAC]ATTTTATGGTATGGTAAGATGAATA
EIAVa      ACTTATGTTATAAAACAA[TCACAAGGA------]ATCCTATGGGCAGGTAAAAAGATTA
EIAVb      ACTTATGTTATAAAACAA[TCACAAGGA------]ATCCTATGGGCAGGTAAAAAGATTA

                1150        1160        1170        1180        1190        1200
                .           .           .           .           .           .
HIV2a      AAATGAAAAATACCCATACC[AACGGGGTC]AGATTGTTAGCACAGGTAGTTCAAAAAAT
HIV2b      AGGTAAAAAATACCCACACC[AACGGAGTC]AGACTCCTAGCACAAGTAGTTCAAAAAAT
SIVb       AGATTAAAAATACACATACA[AATGGAGTC]AGATTATTAGCACATGTAGTCGAGAAAAT
SIVcl      GACAGAAGGCAGCACACACA[AATCCCCTA]CAGAAATTAGTAGAAGCTATTCAGAAGAT
SIVa       AACAGAAAGCTACTCACACC[AATGAGTTG]CGTGTACTAGCAGGAGTAGTACAGAAAAT
HIV1a      AAATGAGAACTGCCCACACT[AATGATATA]AAACAATTAACAGAAGCAGTGCAAAAGAT
HIV1b      GGATGAGGGGTGCCCACACT[AATGATGTA]AAACAACTAACAGAGGCAGTGCAAAAAAT
BIVa       TAAATTATGATCATTCAAAC[---AAAATA]AAAAAGATACTTAGAACTATGAATGAGCT
BIVb       TAAATTATGATCATTCAAAC[---AAAATA]AAAAAGATACTTAGAACTATGAATGAGCT
CAEV       ATATAAAGAACCTAAGCATC[------CCG]CAACAGGTTATTAAAGCAGCGCAAAAATT
OVL        ATATTAAAAAACCTCAGTCAA[------TCA]CAGCAAATTATTAAAGCAGCACAAAAACT
Visnaa     GTATTAAGAATTTGAGTCAA[------GCC]CAACAAATTATCAAAGCAGCACAAAAACT
Visnab     GTATTAAGAATTTGAGTCAA[------GCC]CAACAAATTATCAAAGCAGCACAAAAACT
FIVa       GACAAAAGAAAAAGGCAGAA[AATACATGT]GATATAGCCTTAAGAGCATGCTATAAGAT
FIVb       GACAAAAGAAAAAAGCAGAA[AATACCTGT]GATATAGCTCTAAGGGCATGTTATAAAAT
EIAVa      TGAAGGCTAATAAGGGATGG[TCAACAGTA]AAAAATTTAATGTTATTGTTGCAACATGT
EIAVb      TGAAGGCTAATAAGGGATGG[TCAACAGTA]AAAAATTTAATGTTACTGTTGCAACATGT
```

```
                1210        1220        1230        1240        1250        1260
                 .           .           .           .           .           .
HIV2a      AGGAAAAGAAGCACTGGTCATTTGGGGACGAATACCAAGATTTCACCTACCAGTAGAAAG
HIV2b      AGGAAAAGAAGCACTAGTCATTTGGGGACGAATACCAAAATTTCACCTACCAGTAGAAAG
SIVb       AGGGAAAGAAGCAATAGTAATTTGGGGACAGGTGCCAAAATTTCATTTGCCAGTAGAGAG
SIVcl      AGGAAAAGAAAGCATAGTCATCTGGGGCTTTGTGCCAAAAATTCAAGTCCCAGTGACTAG
SIVa       AGGGAAAGAGGCCCTAGTAATTTGGGGACAATTACCCACTTTTGAACTCCCAGTGGAGAG
HIV1a      ATCTACAGAAAGCATAGTAATATGGGGAAAAATTCCTAAATTTAGACTACCTATACAAAA
HIV1b      AACCACAGAAAGCATAGTAATATGGGGAAAGATTCCTAAATTTAAACTACCCATACAAAA
BIVa       GAACAGAACAGTGGTAATTATGACAGGAAGAGAAGCTAGTTTCCTGCTTCCTGGGTCTTC
BIVb       GAACAGAACAGTGGTAATTATGACAGGAAGAGAAGCTAGTTTCCTGCTTCCTGGGTCTTC
CAEV       AACCCAAGAAGTCATCATTAGGACAGGAAAAATACCATGGATATTGTTGCCAGGGAAAGA
OVL        TACGCAAGAGGTAATAATAAGGATAGGAAAAATACCATGGATACTATTACCAGGAAAGGA
Visnaa     GACACAAGAAGTAATAATAAGAACAGGCAAGATACCCTGGATTTTGTTGCCGGGAAGGGA
Visnab     GACACAAGAAGTAATAATAAGAACAGGCAAGATACCCTGGATTTTGTTGCCGGGAAGGGA
FIVa       AAGAGAAGAGTCTATTATAAGAATAGGAAAAGAACCAAGATATGAAATACCTACTTCTAG
FIVb       AAGAGAAGAGTCTATTATAAGAATAGGAAAAGAACCAATGTATGAAATACCTGCATCCAG
EIAVa      GGCAACAGAAAGTATTACTAGAGTAGGAAAATGTCCAACGTTTAAGGTACCATTTACCAA
EIAVb      GGCAACAGAAAGTATTACTAGAGTAGGAAAATGTCCAACGTTTAAGGTACCATTTACCAA

                1270        1280        1290        1300        1310
                 .           .           .           .           .
HIV2a      AGAAACCTGGGAACAGTGGTGGGATGAC[TACTGGCAAGTGACATGGATCCCA------G
HIV2b      AGATACCTGGGAACAGTGGTGGGATAAC[TACTGGCAAGTGACATGGATCCCA------G
SIVb       GGAAATTTGGGAACAATGGTGGACAGAT[TATTGGCAAGTAACCTGGATACCA------G
SIVcl      AGAAGTCTGGGAGCACTGGTGGAGCGAC[CACTGGCAGGTTACATGGATTCCA------G
SIVa       GGACACATGGGAACAATGGTGGGCAGAC[TATTGGCAAGTCAGTTGGATACCC------G
HIV1a      AGAAACATGGGAGACCTGGTGGACAGAG[TATTGGCAAGCCACATGGATTCCT------G
HIV1b      AGAAACGTGGGAAGCATGGTGGATAGAG[TATTGGCAAGCCACCTGGATTCCT------G
BIVa       TGAAGATTGGGAAGCGGCACTCCAGAAG[GAGGAAAGTCTAACACAAATATTCCCA---G
BIVb       TGAAGATTGGGAAGCGGCACTCCAGAAG[GAAGAAAGTCTAACACAAATATTCCCA---G
CAEV       AGAAGATTGGAGACTAGAATTGCAATTA[GGG---AACATCACATGGATGCCAAAATTTT
OVL        GGAAGATTGGATCTTAGAGTTGCAAATA[GGA---AATATAACATGGATGCCCTCATTTT
Visnaa     AGAAGATTGGATATTAGAGTTACAAATG[GGA---AACATAAATTGGATGCCATCATTTT
Visnab     AGAAGATTGGATATTAGAGTTACAAATG[GGA---AACATAAATTGGATGCCATCATTTT
FIVa       AGAAGCCTGGGAATCAAATTTAATTAAT[TCA---CCATATCTTAAGGCCCCACCTCCTG
FIVb       AGAGGCCTGGGAATCAAATCTAATTAGA[TCT---CCATACCTTAAGGCCCCACCACCTG
EIAVa      AGAGCAAGTAATGTGGGAAATGCAAAAA[GGATGGTATTATTCTTGGCTCCCA------G
EIAVb      AGAGCAAGTAATGTGGGAAATGCAAAAA[GGATGGTATTATTCTTGGCTCCCA------G
```

```
          1320       1330       1340       1350       1360       1370       1380
          .          .          .          .          .          .          .
HIV2a     ACTGGGACTTTGTATCTACCCCACCACTGGTCAGGCTAGCATTTAACCTAGTAAAAGATC
HIV2b     ACTGGGACTTCATATCTACCCCGCCACTGGTCAGATTAGTATTTAACCTGGTGAAAGATC
SIVb      AATGGGACTTTGTGTCAACACCTCCCTTAGTCAGATTAGTCTTCAACCTAGTAAAAGAAC
SIVcl     ACTTAGAATTCATTTCTACCCCGCAATTAGAACAAGAGTGGTACATTTGGGAGGCAGAAC
SIVa      AATGGGACTTTGTCAGTGTTCCGCCCTTAGTAACTTTGTGGTATACACTGACTAAGGAAC
HIV1a     AATGGGAGTTTGTTAACACCCCTCATCTAGTAAAATTATGGTATCAGTTAGAAACAGAGC
HIV1b     AGTGGGAGTTTGTCAATACCCCTCCCTTAGTGAAATTATGGTACCAGTTAGAGAAAGAAC
BIVa      TAAAGTTTTATAGGCACTCCTGCAGATGGACCTCCATATGTGGGCCTGTAAGAGAAAATC
BIVb      TAAAGTTTTATAGGCACTCCTGCAGATGGACCTCCATATGTGGGCCAGTAAGAGAAAATC
CAEV      GGTCCTGTTATCGAGGACATACAAGATGGAGAAAAAGA------AATATAATAGAAGAAG
OVL       GGTCGTGTTATAGGGGGTCAATAAGGTGGAAAAAGAGA------AATGTAATAACAGAAG
Visnaa    GGTCATGTTATAAAGGCTCGGTGAGGTGGAAAAAGAGG------AATGTAATAGCGGAAG
Visnab    GGTCATGTTATAAAGGCTCGGTGAGGTGGAAAAAGAGG------AATGTAATAGCGGAAG
FIVa      AGGTAGAATATATCCATGCTGCTTTGAATATAAAGAGAGCGTTAAGTATGATAAAAGATG
FIVb      AGGTGGAATTTATACATGCTGCCTTAAGTATAAAAAGGGCTCTAAGCATGATACAAGATG
EIAVa     AAATAGTATATACACATCAAGTAGTTCATGATGATTGGAGAATGAAATTGGTAGAAGAAC
EIAVb     AAATAGTATATACACATCAAGTAGTTCATGATGATTGGAGAATGAAATTGGTAGAAGAAC

                    1390       1400       1410       1420       1430
                    .          .          .          .          .
HIV2a     CT---ATACTAGGCGCA]GAGACCTTCTACACAGACGGG[CCCTGTAATAGGCAATCAAA
HIV2b     CC---ATACTAGGCGCA]GAAACCTTCTACACAGATGGA[TCCTGCAATAAGCAATCAAG
SIVb      CT---ATACAGGGAGCA]GAGACATTTTATGTAGATGGA[TCCTGTAATAGGCAATCAAG
SIVcl     CC---ATAATAGGGGTA]GACACCTACTATGTAGATGGA[GCAGCAGAAAAGGTAGGAAA
SIVa      CC---ATCCCGGGAGAG]GATGTCTACTATGTAGATGGA[GCCTGTAATAGACAGTCGAA
HIV1a     CC---ATAGCAGGAGCA]GAAACTTACTATATAGATGGG[GCAGCTAATAGGGAAACTAA
HIV1b     CC---ATAGTAGGAGCA]GAAACTTTCTATGTAGATGGG[GCAGCTAATAGGGAAACTAA
BIVa      TA---------------]ACCACCTACTATACTGACGGA[GGGAAGAAA------GGGAA
BIVb      TA---------------]ACCACCTACTATACTGACGGA[GGGAAGAAA------GGGAA
CAEV      TA---GTAGAAGGG---]CCTACATATTATACAGATGGA[GGAAAAAAGAAT---AAAGT
OVL       TA---GTAGAAGGC---]CCGACATATTATACAGATGGA[GGTAAGAAGAAT---GGAAA
Visnaa    TA---GTCCCAGGA---]CCAACATATTATACCGATGGA[GGAAAGAAAAAT---GGGCG
Visnab    TA---GTCCCAGGA---]CCAACATATTATACCGATGGA[GGAAAGAAAAAT---GGGCG
FIVa      CTCCAATACCAGGAGCA]GAAACATGGTATATAGATGGA[GGTAGAAAGCTA---GGAAA
FIVb      CCCCTATAACAGGAGCA]GAAACATGGTATATAGATGGG[AGTAGAAAACAA---GGAAA
EIAVa     CT---ACATCAGGA---]ATAACAATATACACTGATGGG[GGAAAACAAAAT---GGAGA
EIAVb     CT---ACATCAGGA---]ATAACAATATACACTGATGGG[GGAAAACAAAAT---GGAGA
```

```
                1440        1450        1460        1470        1480        1490
                .           .           .           .           .           .
HIV2a           AGAAGGAAAAGCAGGATATATAACAGATAGAGGGAGAGACAAGGTGAAGGTACTAGAACA
HIV2b           AGAAGGAAAAGCAGGATACATAACAGATAGAGGAAGAGACAAGGTGAGGCTATTAGAGCA
SIVb            AGAAGGAAAGGCAGGCTATGTAACAGATAGAGGCAGAGACAAGGCAAAACCTTTGGAACA
SIVcl           AACAGGAAAAGCAGGATATATAACACAATCAGGGAAAGAGAAAGTAAAGGAGTTAACTGA
SIVa            AGAGGGAAAAGCAGGCTACATAACCCAACAAGGCAAACAAAGAGTACAACAGCTAGAAAA
HIV1a           ATTAGGAAAAGCAGGATATGTCACTGATAGAGGAAAGCAAAAAGTTGTCTCCCTAACGGA
HIV1b           ATTAGGAAAAGCAGGATATGTGACTGACAGAGGAAGACAAAAAGTTGTCCCCCTAACGGA
BIVa            AACAGCTGCAGCAGTATATTGGTGTGAA---GGAAGGACTAAATCAAAGGTATTTCCA--
BIVb            AACAGCTGCAGCAGTATATTGGTGTGAA---GGAAGGACTAAGTCAAAGGTATTTCCA--
CAEV            AGGAAGT---CTAGGGTTCATAGTATCAACAGGGGAAAAATTTAGAAAGCATGAAGAG--
OVL             AGGAAGT---CTAGGATTCATTGCCTCTACAGGCGTTAAATTTAGAAAACACGAAGAG--
Visnaa          GGGAAGC---CTGGGGTATATTACCTCCACAGGTGAAAAGTTTAGAATACATGAAGAA--
Visnab          GGGAAGC---CTGGGGTATATTGCCTCCACAGGTGAAAAGTTTAGAATACATGAAGAA--
FIVa            AGCAGCAAAAGCAGCCTATTGGACAGATACAGGAAAGTGGCAAGTGATGGAATTAGAA--
FIVb            AGCAGCAAGAGCAGCCTATTGGACAGATACAGGTAAATGGCAGATAATGGAGATAGAA--
EIAVa           AGGAATA---GCAGCTTATGTGACCAGTAATGGGAGAACTAAACAGAAAAGGTTAGGACC
EIAVb           AGGAATA---GCAGCTTATGTGACCAGTAATGGGAGAACTAAACAGAAAAGGTTAGGACC

                1500        1510        1520        1530        1540        1550
                .           .           .           .           .           .
HIV2a           AACT]ACCAATCAGCAAGCAGAATTAGAAGCCTTCGCGCTGGCAGTAACAGACTCAGGTC
HIV2b           AACC]ACCAATCAGCAAGCAGAATTAGAAGCCTTTGCGATGGCAGTAACAGACTCAGGTC
SIVb            GACT]ACCAATCAACAAGCAGAGTTGGAAGCCTTCTATCTAGCCTTAGCAGATTCAGGAC
SIVcl           CACC]ACAAATCAGCAAGCAGAACTAGAGGCAGTTCTAATGGCATTACAAGATAGCAATA
SIVa            CACA]ACAAATCAACAAGCTGAACTGACAGCCATAAAAATGGCCTTGGAGGATAGCGGCC
HIV1a           AACA]ACAAATCAGAAGACTGAATTACAAGCAATTTATCTAGCTTTGCAAGATTCAGGGT
HIV1b           CACA]ACAAATCAGAAAACTGAGTTACAAGCAATTCATCTAGCTTTGCAGGATTCGGGAT
BIVa            -GGA]ACCAATCAACAGGCGGAATTGAAGGCCATATGCATGGCTCTCTTGGATGGACCAC
BIVb            -GGA]ACCAATCAACAGGCGGAATTGAAGGCCATATGCATGGCTCTCTTGGATGGACCAC
CAEV            -GGC]ACAAACCAGCAACTAGAATTAAGAGCCATAGAGGAAGCTCTAAAACAAGGGCCTC
OVL             -GGA]ACAAATCAACAATTAGAATTAAGAGCAATAGAAGAGGCGTGCAAACAGGGACCAG
Visnaa          -GGG]ACAAATCAGCAGTTAGAATTGAGGGCCATAGAAGAGGCATGTAAACAGGGACCAG
Visnab          -GGG]ACAAATCAGCAGTTAGAATTGAGGGCCATAGAAGAGGCATGTAAACAGGGACCAG
FIVa            -GGC]AGTAATCAGAAGGCAGAAATACAAGCATTATTATTGGCATTAAAAGCAGGATCAG
FIVb            -GGA]AGTAATCAAAAAGCAGAGGTGCAGGCTTTATTATTGGCCTTAAAAGCAGGATCAG
EIAVa           TGTC]ACTCATCAAGTTGCTGAAAGAATGGCAATACAAATGGCATTAGAGGATACCAGAG
EIAVb           TGTC]ACTCATCAAGTTGCTGAAAGAATGGCAATACAAATGGCATTAGAGGATACCAGAG
```

```
                1560        1570        1580        1590        1600        1610
                .           .           .           .           .           .
HIV2a      CAAAA---GCCAATATTATAGTAGATTCACAGTATGTAATGGGAATAGTAGCAGGC[CAG
HIV2b      CAAAG---GCCAACATTATAGTAGACTCACAATATGTAATGGGAATAGTAGCAGGC[CAA
SIVb       CAAAG---GCCAATATCATAGTAGATTCCCAATATGTTATGGGCATAATAGCAGGT[CAA
SIVcl      GTAAA---GTAAATATAGTAACAGATTCACAATATGTAATGAAAATATTGTCACAA[AGA
SIVa       CTAAA---GTCAATATAGTAACAGATTCACAATATGCGATGGGCATATTGACAGCA[CAG
HIV1a      TAGAA---GTGAACATAGTGACAGATTCACAGTATGCACTAGGAATCATTCAAGCA[CAA
HIV1b      TAGAA---GTAAACATAGTAACAGACTCACAATATGCATTAGGAATCATTCAAGCA[CAA
BIVa       CAAAA---ATGAATATCATAACAGATAGTAGATACGCCTATGAGGGAATGAGAGAA[GAA
BIVb       CAAAA---ATGAATATCATAACAGATAGTAGATACGCCTATGAGGGAATGAGAGAA[GAA
CAEV       AAACA---ATGAATTTAGTAACAGATAGTAGATATGCATTTGAATTTTTATTAAGA[AAT
OVL        AGAAA---ATGAATATAGTAACAGATAGCAGATATGCATATGAATTTATGAGAAGA[AAT
Visnaa     AAAAA---ATGAATATAGTAACAGATAGCAGGTATGCATATGAATTTATGTTGCGG[AAC
Visnab     AAAAA---ATGAATATAGTAACAGATAGCAGGTATGCATATGAATTTATGTTGCGG[AAC
FIVa       AGGAG---ATGAATATTATAACAGATTCACAATATGTTATAAATATTATTCTTCAA[CAA
FIVb       AAGAA---ATGAATATCATAACAGATTCACAATATATTTTGAATATTATCAATCAA[CAA
EIAVa      ATAAACAAGTAAATATAGTAACTGATAGTTATTATTGTTGGAAAAATATTACAGAA[GGA
EIAVb      ATAAACAAGTAAATATAGTAACTGATAGTTATTATTGTTGGAAAAATATTACAGAA[GGA

                1620        1630        1640        1650        1660        1670
                .           .           .           .           .           .
HIV2a      CCAACAGAA------TCAGAAAATAGAATAGTAAATCAAATCATAGAAGAAATGATAAAA
HIV2b      CCAACAGAG------TCAGAGAGTAAAATAGTAAATCAAATCATAGAAGAAATGATAAAA
SIVb       CCCACTGAA------TCAGAAAGTAGATTAGTAAACCAGATAATAGAGGAAATGATTAAA
SIVcl      CCAACAGAA------ACAGAACACCCCATAGTAAAAGACATCATAGAACAATGCAAGCAG
SIVa       CCCACACAG------AGTGACTCCCCACTAGTAGAACAAATAATAGCACAGATGGTACAG
HIV1a      CCCGATAAG------AGTGAATCAGAGTTAGTTAATCAAATAATAGAGGAATTAATAAAG
HIV1b      CCAGATAAG------AGTGAATTAGAAATAGTCAATCAAATAATAGAGCAGTTAATAAAA
BIVa       CCAGAA---ACGTGGGCCAGGGAAGGAATCTGGCTGGAGATTGCCAAGATATTGCCCTTT
BIVb       CCAGAA---ACGTGGGCCAGGGAAGGAATCTGGCTGGAGATTGCCAAGATATTGCCCTTT
CAEV       TGGGATGAAGAAGTAATAAAGAATCCAATTCAAGCAAGAATTATGGAAATTGCCCACAAG
OVL        TGGGATGAAGAAGTTATAAAGAACCCAATACAGGCTAGAATTATGAAATTAGTGCATGAT
Visnaa     TGGGACGAAGAAGTAATAAGAAACCCTATACAAGCGAGAATCATGGAATTAGTGCATAAT
Visnab     TGGGACGAAGAAGTAATAAGAAACCCTATACAAGCGAGAATTATGGAATTAGTGCATAAT
FIVa       CCAGAT---------ATGATGGAGGGAATCTGGCAAGAAGTTTTAGAAGAATTGGAGAAG
FIVb       CCAGAT---------TTGATGGAAGGACTATGGCAAGAGGTCTTAGAAGAAATGGAAAAG
EIAVa      TTAGGTTTAGAA---GGACCACAAAGTCCTTGGTGGCCTATAATACAAAATATACGAGAA
EIAVb      TTAGGTTTAGAA---GGACCACAAAGTCCTTGGTGGCCTATAATACAAAATATACGAGAA
```

```
                1680        1690        1700        1710        1720        1730
                .           .           .           .           .           .
HIV2a       ]AAGGAAGCCATCTATGTTGCGTGGGTCCCAGCCCACAAAGGCATAGGGGGAAAT[----
HIV2b       ]AAGGAAGCAATCTATGTTGCATGGGTCCCAGCCCATAAAGGCATAGGAGGAAAT[----
SIVb        ]AAGGAAGCAATTTATGTGGCATGGGTACCTGCTCATAAAGGAATAGGAGGAAAT[----
SIVcl       ]AAAGATCAAGTTTATCTAGGATGGGTGCCTGCTCATAAAGGAATAGGAGGTAAT[----
SIVa        ]AAAGAAGCCATCTATCTGCAATGGGTACCTGCTCATAAAGGTATAGGGGGCAAT[----
HIV1a       ]AAGGAAAAGGTCTACCTGTCATGGGTACCAGCACACAAAGGAATTGGAGGAAAT[----
HIV1b       ]AAGGAAAAGATCTACCTGGCATGGGTACCAGCACACAAAGGAATTGGAGGAAAT[----
BIVa        ]AAGCAGTACGTGGGGATCGGGTGGGTGCCTGCACATAAAGGGATAGGAGGAAAT[----
BIVb        ]AAGCAGTACGTGGGGGTCGGGTGGGTGCCTGCACATAAAGGGATAGGAGGAAAT[----
CAEV        ]AAAGATAGGATAGGAGTGCATTGGGTGCCAGGACATAAAGGGATTCCCCAAAAT[----
OVL         ]AAGGAACAGATAGGGGTACATTGGGTACCTGGACACAAAGGAATTCCTCAAAAT[----
Visnaa      ]AAAGAAAAAATAGGGGTACATTGGGTGCCTGGACACAAGGGGATTCCTCAAAAT[----
Visnab      ]AAAGAAAAAATAGGGGTACATTGGGTGCCTGGACACAAGGGGATTCCTCAAAAT[----
FIVa        ]AAAACAGCAATATTTATAGATTGGGTCCCAGGACATAAAGGTATTCCAGGAAAT[----
FIVb        ]AAAATAGCAATTTTTATAGACTGGGTTCCTGGACATAAAGGTATACCAGGAAAT[----
EIAVa       ]AAAGAGATAGTTTATTTTGCTTGGGTACCTGGTCACAAAGGGATATATGGTAAT[CAAT
EIAVb       ]AAAGAGATAGTTTATTTTGCTTGGGTACCTGGTCACAAAGGGATATGTGGTAAT[CAAT

                1740        1750        1760        1770        1780        1790
                .           .           .           .           .           .
HIV2a       ------------------------------------------------------------
HIV2b       ------------------------------------------------------------
SIVb        ------------------------------------------------------------
SIVcl       ------------------------------------------------------------
SIVa        ------------------------------------------------------------
HIV1a       ------------------------------------------------------------
HIV1b       ------------------------------------------------------------
BIVa        ----------------------------------ACAGAGGCAGATGAAGGA-------
BIVb        ----------------------------------ACAGAGGCAGATGAAGGA-------
CAEV        -----GAAGAAATAGACAAATATATTTCGGAAATATTTCTTGCAAAAGAAGGAGAAGGAA
OVL         -----GAAGAAATAGATAAATATATTTCAGAAATATTTTTAGCAAGAGAAGGAAGCGGAA
Visnaa      -----GAAGAAATAGATAGGTATATCTCAGAAATATTTTTAGCAAAAGAGGGAAGAGGGA
Visnab      -----GAAGAAATAGATAGGTATATCTCAGAAATATTTTTAGCAAAAGAGGGAAGAGGGA
FIVa        -----GAGGAAGTAGATAAGCTTTGTCAAACAATGATGATAATAGAA---GGGGATGGGA
FIVb        -----GAAGAAGTAGATAAACTTTGCCAAACAATGATGATTATAGAA---GGTGAGGGGA
EIAVa       TGGCAGATGAAGCCGCAAAAATAAAAGAAGAAATCATGCTAGCATACCAAGGCACACAAA
EIAVb       TGGCAGATGAAGCCGCAAAAATAAAAGAAGAAATCATGCTAGCATACCAAGGCACACAAA
```

```
                1800       1810       1820       1830       1840       1850
                .          .          .          .          .          .
HIV2a     ----------CAGGAAGTAGACCAT----------------------------------
HIV2b     ----------CAGGAGGTAGATCAC----------------------------------
SIVb      ----------CAAGAAGTAGATCAC----------------------------------
SIVcl     ----------CAAGAGGTAGACCAC----------------------------------
SIVa      ----------GAAGAAATAGACAAA----------------------------------
HIV1a     ----------GAACAAGTAGATAAA----------------------------------
HIV1b     ----------GAACAAGTAGACAAA----------------------------------
BIVa      -----GTTAAGAAAGCCTTAGAACAGATGGCCCCGTGTAGCCCTCCTGAGGCCATTCTAT
BIVb      -----GTTAAGAAAGCCTTAGAACAGATGGCCCCGTGTAGCCCTCCTGAGGCCATTCTAT
CAEV      TTCTCCCAAAAAGAGAAGAGGATGCAGGGTATGATTTAATATGCCCAGAAGAGGTTACCA
OVL       TTCTTCCAAAAAGAGCGGAAGATGCAGGGTATGATCTCATATGTCCGCAGGAAGTGTGTA
Visnaa    TTTTACAAAAAAGGGCAGAAGATGCTGGATATGACTTAATATGTCCACAGGAGATAAGCA
Visnab    TTTTACAAAAAAGGGCAGAAGATGCTGGATATGACTTAATATGTCCACAGGAGATAAGCA
FIVa      TATTAGATAAAAGGTCAGAAGATGCAGGATATGATTTATTAGCTGCAAAAGAAATACATT
FIVb      TATTAGATAAAAGATCAGAGGATGCAGGATATGACTTACTAGCTGCACAGGAAACACATT
EIAVa     TTAAAGAGAAAAGAGATGAAGATGCAGGGTTTGACTTATGTGTTCCTTATGACATCATGA
EIAVb     TTAAAGAGAAAAGAGATGAAGATGCAGGGTTTGACTTATGTGTTCCTTATGACATCATGA

                1860       1870       1880       1890       1900       1910
                .          .          .          .          .          .
HIV2a     ----------------------------------------------------------
HIV2b     ----------------------------------------------------------
SIVb      ----------------------------------------------------------
SIVcl     ----------------------------------------------------------
SIVa      ----------------------------------------------------------
HIV1a     ----------------------------------------------------------
HIV1b     ----------------------------------------------------------
BIVa      TAAAACCAGGAGAAAAACAAAATCTGGAGACAGGGATCTACATGCAGGGGCTTAGACCAC
BIVb      TAAAACCAGGAGAAAAACAAAATCTGGAGACAGGGATCTACATGCAGGGGCTTAGACCAC
CAEV      TAGAGCCAGGACAAGTGAAATGCATCCCCATAGAGCTAAGATTAAAT---TTAAAGAAAT
OVL       TTCCAGCGGGGCAAGTAAGAAAAATTCCAATTAATCTAAGAATAAAC---TTAAAGGAGG
Visnaa    TTCCGGCGGGGCAAGTGAAAAGGATAGCAATTGACTTGAAAATAAAT---TTGAAAAAGG
Visnab    TTCCGGCGGGACAAGTGAAAAGGATAGCAATTGACTTGAAAATAAAT---TTGAAAAAGG
FIVa      TATTGCCAGGAGAGGTAAAAGTAATACCAACAGGGGTAAAGCTAATG---TTGCCTAAAG
FIVb      TCTTGCCTGGAGAGGTAAGAATAGTACCAACAAAAACAAGAATAATG---CTACCAAAAG
EIAVa     TACCTGTATCTGACACAAAAATCATACCCACAGATGTAAAAATTCAA---GTTCCTCCTA
EIAVb     TACCTGTATCTGACACAAAAATCATACCCACAGATGTAAAAATTCAA---GTTCCTCCTA
```

```
                1920        1930        1940        1950        1960        1970
                .           .           .           .           .           .
HIV2a     ----------------------------------------------TTAGTAAGT---C
HIV2b     ----------------------------------------------TTAGTAAGT---C
SIVb      ----------------------------------------------CTGGTTAGT---C
SIVcl     ----------------------------------------------CTAGTAAGT---A
SIVa      ----------------------------------------------TTAGTAAGC---A
HIV1a     ----------------------------------------------TTAGTTAGT---T
HIV1b     ----------------------------------------------TTAGTCAGT---T
BIVa      AA------AGCTTCCTCCCA---AGAGCAGACTTACCA---------GTAGCCATCACAG
BIVb      AA------AGCTTCCTCCCA---AGAGCAGACTTACCA---------GTAGCCATCACAG
CAEV      CACAATGGGCTATGATTGCTACAAAAAGCAGCATGGCTGCCAAAGGAGTGTTCACACAAG
OVL       ATCAGTGGGCCATGGTAGGGACGAAAAGTAGTTTTGCAAGCAAGGGAGTATTTGTACAAG
Visnaa    ACCAGTGGGCCATGATAGGGACCAAAAGCAGTTTTGCAAATAAGGGAGTATTCGTACAAG
Visnab    ACCAGTGGGCCATGATAGGGACCAAAAGCAGTTTTGCAAATAAGGGAGTATTCGTACAAG
FIVa      GATATTGGGGATTAATAATAGGAAAAAGCTCGATAGGGAGTAAAGGATTGGATGTATTAG
FIVb      GACACTGGGGACTAATAATGGGAAAAAGCTCAATAGGAAGTAAAGGGATGGATGTATTAG
EIAVa     ATAGCTTTGGATGGGTCACTGGGAAATCATCAATGGCAAAACAGGGGTTATTAATTAATG
EIAVb     ATAGCTTTGGATGGGTCACTGGGAAATCATCAATGGCAAAACAGGGGTTATTAATTAATG

                1980        1990        2000        2010        2020        2030
                .           .           .           .           .           .
HIV2a     AGGGCATCAGA-----------------------CAAGTATTGTTTCTAGAAAAA----
HIV2b     AGGGCATCAGA-----------------------CAAGTATTATTCCTAGAGAAA----
SIVb      AAGGAATTAGA-----------------------CAAGTCCTATTCTTAGAAAAA----
SIVcl     AAGGTATAAGACAGAAA-----------------CAGGTCATGTTTCTAGAAAAG----
SIVa      AGGGAGTTAGA-----------------------AGAATATTGTTCATTGGCAGG----
HIV1a     CTGGAGTCAGA-----------------------AAAGTGCTATTTCTAGATGGG----
HIV1b     CTGGAATCAGG-----------------------AAAGTACTATTTTTAGATGGA----
BIVa      GAACCATGGTAGATTCA-----------GAGCTACAGCTACAGCTACTTAACATAGGAA
BIVb      GAACCATGGTAGATTCA-----------GAGCTACAGCTACAGCTACTTAACATAGGAA
CAEV      GAGGAATCATAGACTCAGGATATCAGGGACAAATACAGGTAATAATGTATAATAGCAATA
OVL       GGGGAATAATAGATTCAGGATATCAAGGCATTATACAGGTAGTAGTATATAACAGCAATG
Visnaa    GAGGTATCATAGATTCGGGATATCAAGGAACAATACAAGTAGTGATATATAATAGTAATA
Visnab    GAGGTATCATAGATTCGGGATATCAAGGAACAATACAAGTAGTAATATATAATAGTAATA
FIVa      GAGGGGTAATAGACGAAGGATATCGAGGTGAAATTGGAGTAATAATGATTAATGTATCAA
FIVb      GAGGAGTTATAGATGAAGGATATAGAGGAGAATTAGGAGTGATTATGATCAATTTAACAA
EIAVa     GAGGAATAATTGATGAAGGATATACAGGAGAAATACAAGTGATATGTACTAATATTGGAA
EIAVb     GAGGAATAATTGATGAAGGATATACAGGAGAAATACAAGTGATATGTACTAATATTGGAA
```

```
                2040        2050        2060        2070        2080        2090
                .           .           .           .           .           .
HIV2a           ------------------------------------------------------------
HIV2b           ------------------------------------------------------------
SIVb            ------------------------------------------------------------
SIVcl           ------------------------------------------------------------
SIVa            ------------------------------------------------------------
HIV1a           ------------------------------------------------------------
HIV1b           ------------------------------------------------------------
BIVa            CTGAGCATATAAGAATCCAAAAAGATGAG--------------GTCTTCATGACCTGTT
BIVb            CTGAGCATATAAGAATCCAAAAAGATGAG--------------GTCTTCATGACCTGTT
CAEV            AAATAGCAGTAGTCATACCCCAAGGGAGAAAATTTGCACAATTAATATTAATGGATAAAA
OVL             ACAAGGAAGTCATTATCCCACAAGGGAGAAAATTTGCACAATTAATTCTCATGCCTTTAA
Visnaa          ATAAAGAAGTAGTAATACCACAGGGAAGAAAATTTGCACAGTTGATCCTCATGCCTCTAA
Visnab          ATAAAGAAGTAGTAATACCACAGGGAAGAAAATTTGCACAGTTGATCCTCATGCCTCTAA
FIVa            GAAAATCAATCACCTTAATGGAACGACAAAAGATAGCACAATTAATAATATTGCCTTGTA
FIVb            AGAAATCAATAACTATATTAGAAAAGCAAAAAGTAGCACAATTGATAATATTGCCTTGTA
EIAVa           AAAGTAATATTAAATTAATAGAGGGACAAAAATTTGCACAATTAATTATACTACAGCATC
EIAVb           AAAGTAATATTAAATTAATAGAGGGACAAAAATTTGCACAATTAATTATACTACAGCATC

                2100        2110        2120        2130        2140        2150
                .           .           .           .           .           .
HIV2a           ------------------------------------------------------------
HIV2b           ------------------------------------------------------------
SIVb            ------------------------------------------------------------
SIVcl           ------------------------------------------------------------
SIVa            ------------------------------------------------------------
HIV1a           ------------------------------------------------------------
HIV1b           ------------------------------------------------------------
BIVa            TC----------------------------------------------------------
BIVb            TC----------------------------------------------------------
CAEV            AGCATGGAAAATTGGAACCCTGGGGGGAAAGCAGAAAA---ACAGAAAGGGGAGAAAAAG
OVL             TACATGAAGACCTAGAAGCTTGGGGGGAAACTAGAAGG---ACAGAAAGAGGAAACCAAG
Visnaa          TACATGAAGAGTTGGAGCCATGGGGAGAAACAAGAAAA---ACAGAAAGAGGGGAACAAG
Visnab          TACATGAAGAGTTGAAGCCATGGGGAGAAACAAGAAAA---ACAGAAAGAGGGGAACAAG
FIVa            AACATGAAGTATTAGAA---CAAGGAAAAGTAGTAATGGATTCAGAGAGAGGAGACAATG
FIVb            GACACGAAAGCCTACAA---CAAGGAGAAATACAAATGGATTCAGAAAGAGGAGAAAAGG
EIAVa           ACTCAAATTCCAGACAGCCTTGGGATGAAAATAAAATA---TCTCAGAGAGGGGATAAAG
EIAVb           ACTCAAATTCCAGACAGCCTTGGGATGAAAATAAAATA---TCTCAGAGAGGGGATAAAG
```

```
                2160        2170        2180        2190        2200        2210
                 .           .           .           .           .           .
HIV2a     ----------------------------------------]ATAGAGCCAGCTCAGGAA
HIV2b     ----------------------------------------]ATAGAACCCGCTCAGGAG
SIVb      ----------------------------------------]ATAGAACCTGCACAAGAA
SIVcl     ----------------------------------------]ATAGAACCTGCAGTAGAA
SIVa      ----------------------------------------]ATAGAAGAAGCACAAGAA
HIV1a     ----------------------------------------]ATAGATAAAGCTCAAGAA
HIV1b     ----------------------------------------]ATAGATAAGGCCCAAGAA
BIVa      --CTA-----------------------------GAAAAT]ATCCCCTCAGCCACTGAA
BIVb      --CTA-----------------------------GAAAAT]ATCCCCTCAGCCACTGAA
CAEV      GATTTGGGTCTACAGGAATGTAT------TGGATAGAAAAT]ATTCCTCTGGCAGAGGAA
OVL       GATTTGGATCAACGGGAGCATAT------TGGATTGAAAAT]ATTCCCCTAGCAGAGGAA
Visnaa    GATTTGGATCAACAGGGATGTAT------TGGATAGAAAAT]ATTCCCCTAGCAGAAGAA
Visnab    GATTTGGATCAACAGGGATGTAT------TGGATAGAAAAT]ATTCCCCTAGCAGAAGAA
FIVa      GTTATGGGTCAACAGGAGTATTCTCCTCTTGGGTTGACAGA]ATTGAGGAAGCAGAAATA
FIVb      GATTCGGATCAACAGGAGTTTTTTCATCATGGGTAGATAGA]ATTGAAGAAGCAGAGTTG
EIAVa     GATTTGGAAGTACAGGAGTATTC------TGGGTAGAAAAT]ATTCAGGAAGCACAAGAT
EIAVb     GATTTGGAAGTACAGGAGTATTC------TGGGTAGAAAAT]ATTCAGGAAGCACAAGAT

                2220        2230        2240        2250        2260        2270
                 .           .           .           .           .           .
HIV2a     GAACATGAAAAGTATCATAGCAATGTGAAAGAACTATCCCATAAATTTGGATTACCCAAT
HIV2b     GAACATGAAAAATATCATAGCAATGTAAAAGAACTATCCCATAAATTTGGACTGCCCAAA
SIVb      GAGCATGAAAAGTACCATAGCAATGTAAAAGAATTGGTATTCAAATTTGGTATACCTAGG
SIVcl     GAACATAGCAAATTCCATAACAATGCAAAGGATCTAGAAGAAAAATTTAACCTACCCCCC
SIVa      GAACATGATAGGTATCACAGTAACTGGAGAAATCTAGCAGACACATTTGGATTGCCACAA
HIV1a     GAACATGAAAGGTATCATAACAATTGGAGAGCAGTGGCTAGTGATTTTAATCTACCACCT
HIV1b     GAACATGAGAAATATCACAATAATTGGAGAGCAATGGCTAGTGACTTTAACATACCACCT
BIVa      GATCATGAGAGATGGCATACCTCACCAGACATTTTGGTTAGGCAGTTCCATCTCCCGAGG
BIVb      GATCATGAGAGATGGCATACCTCACCAGACATTTTGGTTAGGCAGTTCCATCTCCCTAAG
CAEV      GACCACACAAAATGGCATCAAGATGCCCGATCATTGCATCTAGAATTTGAAATTCCAAGA
OVL       GATCACAGTAAATGGCATCAAGATGCTGGGTCATTACACTTAGACTTTGGGATACCCCGA
Visnaa    GAGCATAACAAATGGCATCAAGATGCTGTGTCCTTGCATTTAGAATTTGGGATTCCCAGG
Visnab    GAGCATAACAAATGGCATCAAGATGCTGTGTCCTTGCATTTAGAATTTGGGATTCCCAGG
FIVa      AATCATGAAAAATTTCACTCAGATCCACAGTACTTAAGGACTGAATTTAATTTACCTAAA
FIVb      AATCATGAAAAATTTCATTCAGACCCACAATATTTAAGGACAGAGTTCAATCTACCTAGA
EIAVa     GAACATGAGAATTGGCATACATCACCAAAGATATTGGCAAGAAATTATAAGATACCATTG
EIAVb     GAACATGAGAATTGGCATACATCACCAAAGATATTGGCAAGAAATTATAAGATACCATTG
```

```
                2280        2290        2300        2310        2320        2330
                  .           .           .           .           .           .
HIV2a     CTGGTGGCAAGACAGATAGTAAACACATGTGCCCAATGTCAG[CAGAAG------GGAGA
HIV2b     TTAGTGGCAAGACAAATAGTAAACACATGCACCCAATGTCAG[CAGAAA------GGGGA
SIVb      CTAGTAGCAAAACAGATAGTAGACACATGTCATAAATGCCAC[CAGAAA------GGAGA
SIVcl     ATGGTAGCCAAACAAATTGTCAATGACTGTGCAAACTGTCAA[AAGAAA------GGAGA
SIVa      ATAGTAGCTAAAGAAATTGTAGCAATGTGCCCAAAATGTCAA[GTAAAA------GGGGA
HIV1a     ATAGTAGCAAAAGAAATAGTAGCTAGCTGTGATAAATGTCAG[CTAAAA------GGGGA
HIV1b     GTAGTAGCAAAAGAAATAGTAGCCAGCTGTGATAAATGTCAG[CTAAAA------GGAGA
BIVa      AGAATAGCTAAAGAGATAGTAGCCAGATGCCAAGAATGTAAA[AGGACA---ACCGCTAG
BIVb      AGAATAGCTAAAGAGATAGTAGCCAGATGCCAAGAATGTAAA[AGGACA---ACCACTAG
CAEV      ACAGCAGCAGAAGACATAGTAAATCAATGTGAAATATGCAAA[GAAGCGAGGACACCTGC
OVL       ACTGCAGCTGAGGATATTGTACAACAATGTGAAGTATGTCAA[GAAAATAAAATGCCCAG
Visnaa    ACAGCTGCAGAAGATATAGTTCAACAATGTGATGTGTGTCAA[GAAAATAAAATGCCTAG
Visnab    ACAGCTGCAGAAGATATAGTTCAACAATGTGATGTGTGTCAA[GAAAATAAAATGCCTAG
FIVa      ATGGTAGCAGAAGAGATAAGACGAAAATGCCCAGTATGCAGA[ATCAGA------GGAGA
FIVb      ATAGTAGCAGAAGAAATAAAGAGAAAATGTCCCCTATGCAGA[ATTAGA------GGCGA
EIAVa     ACTGTAGCAAAACAGATAACTCAAGAATGTCCTCATTGCACT[AAGCAA------GGATC
EIAVb     ACTGTAGCAAAACAGATAACTCAAGAATGTCCTCATTGCACT[AAGCAA------GGATC

                2340        2350        2360        2370        2380        2390
                  .           .           .           .           .           .
HIV2a     GGCTATACAT]GGGCAAGTGAATGCAGAACTAGGCACTTGGCAAATGGACTGCACACACT
HIV2b     GGCTATACAT]GGGCAAGTAAATGCAGAATTAGGCACTTGGCAAATGGACTGCACACACT
SIVb      AGCCATACAT]GGGCAAGTAAATGCAGAACTAGGGACTTGGCAAATGGACTGTACACACC
SIVcl     AGCCATAACA]GGACAAGTGGATGTCTCAGTGGGTATTTGGCAGCTAGACTGTACTCACT
SIVa      ACCAATACAT]GGACAAGTAGATGCTTCACCAGGAGTGTGGCAGATGGACTGCACACATA
HIV1a     AGCCATGCAT]GGACAAGTAGACTGTAGCCCAGGAATATGGCAATTAGATTGCACACATT
HIV1b     AGCCATGCAT]GGACAAGTAGACTGTAGTCCAGGAATATGGCAGCTAGATTGTACACACT
BIVa      CCCAGTCAGA]GGAACAAACCCCAGAGGTCGATTCTTATGGCAGATGGACAATACTCACT
BIVb      CCCAGTCAGA]GGAACAAACCCCAGAGGTCGATTCTTATGGCAGATGGACAATACTCACT
CAEV      AGTAATTAGA]GGCGGAAACAAAAGGGGGGTAAATCATTGGCAAGTGGATTATACCCATT
OVL       CACAATAAGG]GGAAGCAATAGGAGAGGAATAGATCATTGGCAAGTAGACTATACACATT
Visnaa    TACATTAAGA]GGCAGTAATAAAAGGGGCATAGACCATTGGCAAGTAGACTATACCCACT
Visnab    TACATTAAGA]GGCAGTAATAAAAGGGGCATAGACCATTGGCAAGTAGACTATACCCACT
FIVa      ACAAGTGGGA]GGACAATTGAAAATAGGGCCTGGTATCTGGCAAATGGATTGCACACACT
FIVb      ACAAGTAGGG]GGACAATTAAAGATTGGGCCTGGAATATGGCAAATGGATTGCACACATT
EIAVa     AGGACCTGCA]GGTTGTGTCATGAGATCTCCTAATCATTGGCAGGCAGATTGCACACATT
EIAVb     AGGACCTGCA]GGTTGTGTCATGAGATCTCCTAATCATTGGCAGGCAGATTGCACACATT
```

```
                2400        2410        2420        2430        2440        2450
                 .           .           .           .           .           .
HIV2a      TAGAAGGAAAAGTCATCATAATAGCAGTGCATGTTGCCAGTGGATTCATAGAAGCAGAGG
HIV2b      TAGAAGGAAAAATCATTATAGTAGCAGTACATGTTGCAAGTGGATTTATAGAAGCAGAAG
SIVb       TAGAGGGCAAAATAATCATAGTAGCAGTACATGTGGCTAGTGGATTCATAGAGGCAGAAG
SIVcl      TAGAAGGACAGGTCATTATAAATGCAGTCCATGTGGCCTCAGGCTTCATGGTTGCAGAAG
SIVa       TAGAAGGAAAAATAGTGATAGTAGCGGTCCATGTAGCCAGTGGGTTTATAGAAGCAGAGG
HIV1a      TGGAAGGACAAGTTATTCTGGTAGCAGTCCATGTAGCCAGTGGCTATATAGAAGCAGAAG
HIV1b      TAGAAGGAAAAGTTATCCTGGTAGCAGTGCATGTAGCCGGTGGATATATAGAAGCAGAAG
BIVa       GGAATAAAACAATTATTTGGGTAGCAGTAGAGACAAATTCAGGATTAGTGGAAGCTCAGG
BIVb       GGAATAAAACAATTATTTGGGTAGCAGTAGAGACAAATTCAGGATTAGTGGAAGCTCAGG
CAEV       ATGAAAATATCATACTATTAGTATGGGTAGAAACAAATTCAGGACTAATATATGCAGAAA
OVL        ATGAGGACAAGATAATATTAGTATGGGTAGAAACAAATTCAGGATTAATATATGCAGAAA
Visnaa     ATGAAGACAAGATAATATTGGTCTGGGTAGAAACAAATTCAGGACTAATCTATGCAGAGA
Visnab     ATGAAGACAAGATAATATTGGTCTGGGTAGAAACAAATTCAGGACTAATCTATGCAGAAA
FIVa       TTGATGGCAAAATAATTCTTGTGGGTATACATGTGGAATCAGGATATATATGGGCACAAA
FIVb       TTAATGGTAAAATAATCATTGTAGCAGTGCATGTGGAATCAGGATTTTTATGGGCACAGA
EIAVa      TGGACAATAAGATAATATTGACTTTTGTAGAGTCAAATTCAGGATACATACATGCTACAT
EIAVb      TGGACAATAAGATAATATTGACTTTTGTAGAGTCAAATTCAGGATACATACATGCTACAT

                2460        2470        2480        2490        2500        2510
                 .           .           .           .           .           .
HIV2a      TCATCCCGCAGGAATCAGGAAGACAAACAGCACTCTTCCTATTAAAACTGGCTAGTAGAT
HIV2b      TCATCCCACAGGAATCAGGAAGGCAAACGGCACTCTTCCTACTAAAACTGGCCAGTAGGT
SIVb       TAATTCCGCAGGAAACAGGAAGGCAAACAGCACTGTTTCTGCTAAAATTAGCTAGCAGAT
SIVcl      TTATACCAGATGAGACAGGAAAAACAACGTCACATTTCCTGTTAAAATTATGCAGTAGAT
SIVa       TTATCCCTAGGGAAACAGGAAAAGAGACAGCAAAGTTCTTGTTAAAAATAATAGGAAGAT
HIV1a      TCATCCCAGCAGAAACAGGAAAGGAAACAGCATACTTCCTGTTGAAACTAGCAAGCAGAT
HIV1b      TTATTCCAGCAGAGACAGGGCAAGAAACAGCATACTTTCTCTTAAAATTAGCAGGAAGAT
BIVa       TGATCCCTGAAGAAACAGCACTACAAGTAGCTCTCTGCATTTTACAGCTAATCCAGAGAT
BIVb       TGATCCCTGAAGAAACAGCACTACAAGTAGCTCTCTGCATTTTACAGCTAATCCAGAGAT
CAEV       AAGTAAAAGGAGAATCAGGGCAAGAATTCAGAATAAAAGTGATGCATTGGTATGCATTAT
OVL        GAGTAAAAGGGGAAACAGGACAAGAATTTAGAATCATGACTATAAGGTGGTATGGTCTGT
Visnaa     GGGTAAAAGGAGAAACAGGACAAGAATTTAGGGTGCAAACTATGAAATGGTATGCGATGT
Visnab     GGGTAAAAGGAGAAACAGGACAAGAATTTAGGGTGCAAACTATGAAATGGTATGCGATGT
FIVa       TAATTTCTCAAGAAACTGCTGACTGTACAGTTAAAGCTGTCTTACAATTGTTGAGTGCTC
FIVb       TAATTCCACAGGAGACCGCGGAGTGTACAGTCAAGGCTCTTTTGCAACTTATATGTGCTC
EIAVa      TATTGTCAAAAGAAAATGCATTATGTACTTCATTGGCTATTTTAGAATGGGCAAGATTGT
EIAVb      TATTATCAAAAGAAAATGCATTATGTACTTCATTGGCTATTTTAGAATGGGCAAGATTGT
```

```
                2520        2530        2540        2550        2560        2570
                 .           .           .           .           .           .
HIV2a      GGCCAATAACACACTTGCACACAGATAGTGGTGTCAACTTCACCTCACAGGAAGTAAAGA
HIV2b      GGCCAATAACACATTTGCACACAGACAATGGTGCCAACTTCACTTCACAGGAAGTAAAGA
SIVb       GGCCCATCACACATCTGCATACTGATAATGGTGCCAATTTCACATCACAAGAAGTGAAAA
SIVcl      GGCCTGTAAAACAAATACATACAGACAATGGTCCAAATTTTGTAAGTAAGGAGGTACAGG
SIVa       GGCCCATCACTCACCTCCATACAGATAATGGACCAAATTTCACTTCTCAGGAAGTAGCTG
HIV1a      GGCCAGTAAAAGTAATACATACAGACAATGGCAGCAATTTCACGAGTGCTGCGGTTAAGG
HIV1b      GGCCAGTAAAAACAATACACACAGACAATGGCAGCAATTTCACCAGTACTACAGTCAAGG
BIVa       ATACAGTTCTTCACTTACATAGTGACAACGGGCCGTGCTTTACTGCACACAGGATAGAAA
BIVb       ATACAGTTCTTCACTTACATAGTGACAACGGGCCGTGCTTTACTGCACACAGGATAGAAA
CAEV       TTGGTCCAGAGTCATTGCAGTCAGACAATGGACCTGCATTTGCAGCAGAGCCCACACAGC
OVL        TTGCCCCAAAGTCATTGCAGTCTGACAATGGACCAGCATTTGTAGCAGAGCCAACACAGC
Visnaa     TTGCCCCGAAATCATTGCAGTCTGATAACGGACCAGCATTTGTAGCAGAATCTACTCAGC
Visnab     TTGCCCCGAAATCATTGCAGTCTGATAACGGACCAGCATTTGTAGCAGAATCTACTCAGC
FIVa       ATAATGTTACTGAATTACAAACAGATAATGGACCAAATTTTAAAAATCAAAAGATGGAAG
FIVb       ATAATGTTACAGAATTACAAACAGATAATGGACCAAATTTTAAAAATCAGAAAATGGAAG
EIAVa      TTTCACCAAAGTCCTTACACACAGATAACGGCACTAATTTTGTGGCAGAACCAGTTGTAA
EIAVb      TTTCACCAAAGTCCTTACACACAGATAACGGCACTAATTTTGTGGCAGAACCAGTTGTAA

                2580        2590        2600        2610        2620        2630
                 .           .           .           .           .           .
HIV2a      TGGTAGCATGGTGGGTTGGTATAGAGCAATCCTTTGGAGTACCTTACAATCCACAAAGCC
HIV2b      TGGTGGCATGGTGGATAGGTATAGAACAATCCTTCGGAGTACCTTACAATCCACAAAGCC
SIVb       TGGTTGCCTGGTGGGCAGGAATTGAGCAGACCTTCGGGGTACCTTATAATCCACAGAGCC
SIVcl      CAGTAACCTGGTGGATAGGGATAGAACATACAACTGGGATACCTTATAACCCTCAGAGTC
SIVa       CTATGTGCTGGTGGGGAAAGGTAGAACACACAACGGGGGTACCATATAATCCACAGTCCC
HIV1a      CAGCCTGTTGGTGGGCAGATATCCAACAGGAATTTGGGATTCCCTACAATCCCCAAAGTC
BIVa       ATCTATGTAAGTATCTGGGGATCACAAAAACTACGGGAATACCCTACAACCCACAATCCC
BIVb       ATCTATGTAAGTATCTGGGGATCACAAAAACTACGGGAATACCCTACAACCCACAATCCC
CAEV       TGTTAATGCAATACCTAGGAGTAAAACACACAACAGGCATACCTTGGAATCCACAGTCTC
OVL        TACTAATGAAATATTTGGGGATAACACACACAACAGGGATACCATGGAATCCCCAATCGC
Visnaa     TCTTAATGAAATATTTGGGCATAGAACATACTACAGGGATCCCCTGGAACCCACAATCTC
Visnab     TCTTAATGAAATATTTGGGCATAGAACATACTACAGGGATCCCCTGGAACCCACAATCTC
FIVa       GAGTACTCAATTACATGGGTGTGAAACATAAGTTTGGTATCCCAGGGAACCCACAGTCAC
FIVb       GATTATTAAATTATATGGGAATAAAACATAAATTGGGTATACCAGGTAACCCACAATCAC
EIAVa      ATTTGTTGAAGTTCCTAAAGATAGCACATACCACAGGAATACCATATCATCCAGAAAGTC
EIAVb      ATTTGTTGAAGTTCCTAAAGATAGCACATACCACAGGAATACCATATCATCCAGAAAGTC
```

```
                2640        2650        2660        2670        2680        2690
                 .           .           .           .           .           .
HIV2a      AAGGAGTAGTAGAAGCAATGAATCACCACCTAAAAAATCAGATAAGCAGAATTAGAGAGC
HIV2b      AAGGAGTAGTGGAAGCAATGAATCACCACCTAAAAAATCAGATAAGCAGAATTAGAGAGC
SIVb       AAGGAGTAGTGGAAGCAATGAACCATCATCTAAAAACCCAGATAGATAGAATTAGAGAAC
SIVcl      AAGGGGTTGTAGAAGCAAAAAACAAAGTTTTAAAAAGTATTATAGAAAGAGTAAGAGAAG
SIVa       AGGGATCTATAGAAAGTATGAACAAACAATTGAAAGAGATAATTGGAAAAATAAGAGATG
HIV1a      AGGGAGTAGTAGAATCTATGAATAAAGAATTAAAGAAGATCATAGGGCAGGTAAGAGACC
HIV1b      AAGGAGTAATAGAATCTATGAATAAAGAATTAAAGAAAATTATAGGACAGGTAAGGGATC
BIVa       AGGGAGTTGTAGAAAGAGCCCACAGAGATCTAAAAGACAGATTGGCAGCTTATCAGGGAG
BIVb       AGGGAGTTGTAGAAAGAGCCCACAGAGATCTAAAAGACAGATTGGCAGCTTATCAGGGAG
CAEV       AGGCTATAGTAGAAAGGGCACATCAACTATTGAAAAGCACTTTAAAGAAGTTCCAGCCAC
OVL        AAGCACTAGTAGAAAGAACTCATCAAACTCTAAAAAATACAATAGAAAAATTTGTTTCTA
Visnaa     AAGCATTAGTAGAGAGAACACATCAGACGTTAAAGAATACATTAGAAAAACTTATACCTA
Visnab     AAGCATTAGTAGAGAGAACACATCAGACGTTAAAGAATACATTAGAAAAACTTATACCTA
FIVa       AAGCATTAGTTGAAAATGTAAATCATACATTAAAAGTTTGGATTCGGAAATTTTTGCCTG
FIVb       AGGCATTAGTGGAAAATGCTAATAACACATTGAAAGCTTGGATTCAGAAATTCCTACCAG
EIAVa      AGGGTATTGTAGAAAGGGCAAATAGGACCTTGAAAGAGAAGATTCAAAGTCATAGAGACA
EIAVb      AGGGTATTGTAGAAAGGGCAAATAGGACCTTGAAAGAGAAGATTCAAAGTCATAGAGACA

                2700        2710        2720        2730        2740        2750
                 .           .           .           .           .           .
HIV2a      AGGCAAATACAGTGGAAACAATAGTACTAATGGCAGTTCATTGCATGAAT[TTTAAAAGA
HIV2b      AGGCAAACACAGTAGAAACAATAGTACTAATGGCAGTTCATTGCATGAAT[TTTAAAAGG
SIVb       AGGCAAATTCAATAGAGACTATAGTACTAATGGCAGTTCATTGCATGAAT[TTTAAAAGA
SIVcl      ATGCACAGCAACTAAAAACAGCAGTACTAATGGCAGTCCACATTCATAAT[TTTAAACAA
SIVa       ACTGTCAATATACAGAAACAGCAGTACTTATGGCCTGCCACATTCACAAT[TTTAAAAGA
HIV1a      AAGCAGAACACCTTAAGACAGCAGTACAAATGGCAGTATTCATTCACAAT[TTTAAAAGA
HIV1b      AGGCTGAACATCTTAAGACAGCAGTACAAACGGCAGTATTCATCCACAAT[TTTAAAAGA
BIVa       ATTGTGAAACCGTAGAAGCAGCCCTTAGCCTCGCATTAGTTTCTTTAAAT[---AAAAAA
BIVb       ATTGTGAAACCGTAGAAGCAGCCCTTAGCCTCGCATTAGTTTCTTTAAAT[---AAAAAA
CAEV       AATTTGTCGCTGTAGAATCAGCCATAGCAGCAGCCCTAGTCGCCATAAAT[ATAAAAAGA
OVL        TGTTTGCCTCATTTGATTCAGCAATAGCCGCAGCATTAATCACACTAAAT[ATAAAAAGA
Visnaa     TGTTTAACGCGTTTGAATCAGCCCTCGCAGGGACCCTCATTACTCTAAAT[ATAAAAAGA
Visnab     TGTTTAACGCGTTTGAATCAGCCCTCGCAGGGACCCTCATTACTCTAAAT[ATAAAAAGA
FIVa       AAACAACCTCCTTGGATAATGCCTTATCTCTCGCTGTACATAGTCTCAAT[TTTAAAAGA
FIVb       AGACTACTTCTCTGGATAATGCTTTGGCCCTAGCCCTGCATTGTCTCAAC[TTTAAACAA
EIAVa      ACACTCAAACACTGGAGGCAGCTTTACAACTTGCTCTCATTACTTGTAAC[---AAAGGG
EIAVb      ACACTCAAACACTGGAGGCAGCTTTACAACTTGCTCTCATTACTTGTAAC[---AAAGGG
```

```
                2760        2770        2780        2790        2800
                 .           .           .           .           .
HIV2a      AGGGGAGGAATAGGGGATATG]ACCCCAGCAGAAAGACTCATCAAT[ATGATCTCCACAG
HIV2b      AGGGGAGGAATAGGGGATATG]ACCCCAGCAGAAAGACTAATCAAT[ATGGTCACTGCAG
SIVb       AGGGGAGGAATAGGGGATATG]ACTCCAGCAGAAAGATTAGTCAAT[ATGATCACCACAG
SIVcl      AGGGGAGGACTAGGGGGGCTT]ACACCAGCAGAGAGGTTTATTAAT[ATGATTAATGCAG
SIVa       AAGGGAGGAATAGGGGGGCTA]ACAGCTGCAGAGAGACTAATAAAT[ATGATAACAACAC
HIV1a      AAAGGGGGGATTGGGGGGTAC]AGTGCAGGGGAAAGAATAATAGAC[ATAATAGCAACAG
HIV1b      AAAGGGGGGATTGGGGGGTAC]AGTGCAGGGGAAAGAATAGTAGAC[ATAATAGCAACAG
BIVa       AGAGGGGGAATAGGGGGCCAT]ACACCATATGAAATATACCTAGAA[---TCAGAACATA
BIVb       AGAGGGGGAATAGGGGGCCAT]ACACCATATGAAATATACCTAGAA[---TCAGAACATA
CAEV       AAGGGTGGGCTGGGG---ACA]AGCCCTATGGATATTTTTATATAT[AATAAAGAACAGA
OVL        AAGGGTGGGCTAGGG---ACA]AGCCCTATGGATATATTCATATTT[AATAAGGAACAGC
Visnaa     AAGGGTGGGCTAGGG---ACA]AGCCCTATGGATATATTTATATTT[AATAAGGAACAAC
Visnab     AAGGGTGGGCTAGGG---ACA]AGCCCTATGGATATATTTATATTT[AATAAGGAACAAC
FIVa       AGAGGTAGGATAGGAGGGATG]GCCCCTTATGAATTATTAGCACAA[---CAAGAATCCT
FIVb       AGGGGTAGACTAGGGAAGATG]GCCCCTTATGAGTTATACATACAA[---CAAGAATCAT
EIAVa      AGGGAAAGTATGGGAGGACAG]ACACCATGGGAAGTATTTATCACT[---AATCAAGCAC
EIAVb      AGGGAAAGTATGGGAGGACAG]ACACCATGGGAAGTATTTATCACT[---AATCAAGCAC

                2810        2820        2830        2840        2850        2860
                 .           .           .           .           .           .
HIV2a      AACAAGAAATACAATTCCTCCAAACAAAAAATTTGAAA---TTTAAAAATTTCCCGGTCT
HIV2b      AACAGGAAATACAATTCCTCCAAGCAAAAAATTCAAAA---TTACAAAATTTTCGGGTCT
SIVb       AACAAGAAATACAATTCCAACAATCAAAAAATTCAAAA---TTTAAAAATTTTCGGGTCT
SIVcl      AACTAGAAACACAATATCTACAAAAATTAAATTCAAAA---ATTTTAAAATTTAAGGTTT
SIVa       AATTAGAAATCAACACTCTACAAACCAAAATCCAAAAA---ATTTTGAATTTTAGAGTCT
HIV1a      ACATACAAACTAAAGAATTACAAAAACAAATTTCAAAC---ATTCAAAAATTTCGGGTTT
HIV1b      ACATACAGACTAAAGAATTACAAAAACAAATTACAAAA---ATTCAAAATTTTCGGGTTT
BIVa       CCAAATACCAAGACCAACTAGAACAACAATTTTCAAAACAAAAAATTGAAAAGTGGTGTT
BIVb       CCAAATACCAAGACCAACTAGAACAACAATTTTCAAAACAAAAAATTGAAAAGTGGTGTT
CAEV       AAAGAATAAATAATAAATATAATAAAAATTCTCAAAAA---------ATTCAATTCTGTT
OVL        AAAGAATACAACAGCAATCTACAAGAAATCAATCAAAA---------TTTCGATTTTGTT
Visnaa     AAAGAATACAGCAACAAAGTAAATCAAAACAAGAAAAA---------ATTCGATTTTGTT
Visnab     AAAGAATACAGCAACAAAGTAAATCAAAACAAGAAAAA---------ATTCGATTTTGTT
FIVa       TAAGAATACAAGATTATTTTTCTGCAATACCACAAAAATTG---CAAGCACAGTGGATTT
FIVb       TAAGAATACAAGACTATTTTTCGCAGATTCCACAAAAGTTA---ATGATGCAATGGGTGT
EIAVa      AAGTAATACATGAGAAACTTTTACTACAGCAAGCACAATCC---TCCAAAAAATTTTGTT
EIAVb      AAGTAATACATGAGAAACTTTTACTACAGCAAGCACAATCC---TCCAAAAAATTTTGTT
```

```
            2870        2880        2890        2900        2910        2920
             .           .           .           .           .           .
HIV2a       ATTACAGGGAAGGCAGA---------GATCAGCTGTGGAAAGGACCTGGGGAGCTACTGT
HIV2b       ATTTCAGAGAAGGCAGA---------GATCAGCTGTGGAAAGGACCTGGGGAACTACTGT
SIVb        ATTACAGAGAAGGCAGA---------GACCAGCTGTGGAAAGGACCCGGTGAGCTATTGT
SIVcl       ATTACAGACAAGGAAGA---------GATCCTCAGTGGAAAGGACCAGCGCAACTTCTGT
SIVa        ACTACAGAGAAGGCAGA---------GATCCAGTGTGGAAGGGACCTGCTCGCCTGATCT
HIV1a       ATTACAGGGACAGCAGA---------GACCCAATTTGGAAAGGACCAGCAAAACTCCTCT
HIV1b       ATTACAGAGACAGCAGA---------GATCCACTTTGGAAAGGACCAGCAAAGCTTCTCT
BIVa        ACGTAAGGAACAGAAGA------------AAGGAATGGAAAGGCCCCTACAAAGTGTTGT
BIVb        ACGTAAGGAACAGAAGA------------AAGGAATGGAAAGGACCCTACAAAGTGTTGT
CAEV        ATTACAGAATAAGGAAAAGAGGACATCAGGAGAGTGGAAAGGACCAACCCAGGTACTGTG
OVL         ATTACAGAGTCAGGAAAAGAGGACACCCAGGCGAGTGGCTGGGACCAACACAGGTACTCT
Visnaa      ATTACAGAACAAGAAAAAGAGGGCATCCAGGAGAGTGGCAAGGACCAACACAGGTACTTT
Visnab      ATTACAGAACAAGAAAAAGAGGGCATCCAGGAGAGTGGCAAGGACCAACACAGGTACTTT
FIVa        ATTATAAAGATCAAAAA---------GATAAGAAATGGAAAGGACCAATGAGAGTAGAAT
FIVb        ATTACAAAGATCAAAAA---------GATAAAAAATGGAAGGGACCAATGAGAGTGGAAT
EIAVa       TTTACAAAATCCCTGGT---------GAACATGATTGGAAGGGACCTACTAGGGTGCTGT
EIAVb       TTTACAAAATCCCTGGT---------GAACATGATTGGAAGGGACCTACTAGGGTGCTGT

            2930        2940        2950        2960        2970        2980
             .           .           .           .           .           .
HIV2a       GGAAAGGGGACGGA]GCAGTCATAGTTAAGGTAGGG[ACA------GAC]ATAAAAGTAG
HIV2b       GGAAGGGGGACGGA]GCAGTCATAGTCAAGGTAGGG[GCT------GAC]ATAAAAATAA
SIVb        GGAAAGGGGAAGGA]GCAGTCATCTTAAAGGTAGGG[ACA------GAG]ATCAAGGTAG
SIVcl       GGAAAGGAGAAGGT]GCTGTAGTGGTAAAAGAAGGG[GAG------AAC]ATCTTCTCAG
SIVa        GGAAAGGAGAAGGC]GCGGTAGTTCTCAAGGAAGGT[GAA------GAA]CTGAAGGTAG
HIV1a       GGAAAGGTGAAGGG]GCAGTAGTAATACAAGACAAT[AGT------GAA]ATAAAAGTAG
HIV1b       GGAAAGGTGAAGGG]GCAGTAGTAATACAAGATAAT[AGT------GAC]ATAAAAGTAG
BIVa        GGGACGGAGACGGG]GCAGCAGTAATAGAGGAAGAG[GGA------AAA]ACAGCCTTAT
BIVb        GGGACGGAGACGGG]GCAGCAGTAATAGAGGAAGAG[GGA------AAA]ACAGCCTTAT
CAEV        GAAAGGGGAAGGAG]CCAATTGTGGTAAAGGATATA[GAAAGTGAAAAG]TATTTAGTAA
OVL         GGGAAGGGGAAGGA]GCAATCGTAATAAAAGATAAA[AATCTAGAAAAG]TATTTAGTCA
Visnaa      GGGGCGGGGACGGT]GCGATTGTAGTGAAAGACAGA[GGCACAGATAGA]TATCTGGTGA
Visnab      GGGGCGGGGACGGT]GCGATTGTAGTGAAAGACAGA[GGCACAGATAGA]TATCTGGTGA
FIVa        ACTGGGGACAGGGA]TCAGTATTATTAAAGGATGAA[GAG---AAGGGA]TATTTTCTTA
FIVb        ATTGGGGACAAGGA]TCAGTATTATTAAAGGATGAA[GAG---AAGGGA]TATTTTCTTG
EIAVa       GGAAGGGTGATGGT]GCAGTAGTAGTTAATGATGAA[GGA---AAGGGA]ATAATTGCTG
EIAVb       GGAAGGGTGATGGT]GCAGTAGTAGTTAATGATGAA[GGA---AAGGGA]ATAATTGCTG
```

```
                2990        3000        3010        3020
                .           .           .           .
HIV2a       TACCAAGAAGGAAGGCCAAGATCATCAGAGACTAT
HIV2b       TACCAAGAAGGAAAGCTAAGATCATCAAAGACTAT
SIVb        TACCAAGAAGGAAAGCTAAAATTATCAAAGATTAT
SIVcl       TCCCCAGAAGAAAAGCAAAACTAGTAAAAGATTAT
SIVa        TTCCGAGAAGGAAAGCAAAAATCATAAAAGACTAT
HIV1a       TACCAAGAAGAGAGGCAAAAATCATTAGGGATTAT
HIV1b       TGCCAAGAAGAAAAGTAAAGATCATTAGGGATTAT
BIVa        ATCCACACCGTCATATGCGCTTCATCCCCCCCCCA
BIVb        ATCCACACCGTCATATGCGCTTCATCCCCCCCCCA
CAEV        TACCTTACAAAGATGCAAAATTCATCCCGCCACCA
OVL         TAGCAAAAAAGGATGTTAAGTTCATACCGCAACCA
Visnaa      TAGCTAACAAAGATGTTAAGTTCATACCGCCACCA
Visnab      TAGCTAACAAAGATGTTAAGTTCATACCGCCACCA
FIVa        TACCTAGGAGACACATAAGGAGAGTTCCAGAACCC
FIVb        TACCTAGGAGACACATAAGAAGAGTCCCAGAACCC
EIAVa       TACCATTAACCAGGACTAAGTTACTAATAAAACCA
EIAVb       TACCATTAACCAGGACTAAGTTACTAATAAAGCCA
```

REFERENCES

Altschul, S. F., M. S. Boguskai, W. Gish, and J. C. Wootton. 1994. Issues in searching molecular databases. Nature Genetics 6: 119-129.

Avise, J.C., J. Arnold, R. Martin Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annual Reviews of Ecology and Systematics 18: 489-522.

Barton Browne, L. 1993. Physiologically induced changes in resource-oriented behavior. Annual Review of Entomology 38: 1-25.

Boggs, C. L. 1997. Dynamics of reproductive allocation from juvenile and adult feeding: Radiotracer studies. Ecology 78: 192-202.

Brown, J. H. 1971. Mammals on mountaintops: nonequilibrium insular biogeography. American Naturalist 105: 467-478.

Brown, G. G., G. Gadaleta, G. Pepe, C. Saccone, and E. Sbisa. 1986. Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA. Journal of Molecular Biology 192: 503-511.

Casane, D., N. Dennebouy, H. deRochambeau, J. C. Mounlou, and M. Monnerot. 1994. Genetic analysis of systematic mitochondrial heteroplasmy in rabbits. Genetics 138: 471-480.

Conner, D. A. 1982. Geographic variation in short calls of pikas (*Ochotona princeps*). Journal of Mammalogy 63: 48-52.

Courtney, S. P. and G. K. Chen. 1988. Genetic and environmental variation in oviposition behaviour in the mycophagous *Drosophila suboccidentalis* Spcr. Functional Ecology 2: 521-528.

Courtney, S. P., G. K. Chen, and A. Gardner. 1989. A general model for individual host selection. Oikos 55: 55-65.

Courtney, S. P., and J. Hard. 1990. Host acceptance and life-history traits in *Drosophila busckii*: tests of the hierarchy-threshold model. Heredity 64: 371-375.

daSilva, M. N. F., and J. L. Patton. 1993. Amazonian phylogeography: mtDNA sequence variation in arboreal echimyid rodents (Caviomorpha). Molecular Phylogenetics and Evolution 2: 243-255.

Dembowski, J.R. 1999. Speciation and phylogeography within the dusky shrew (*Sorex monticolus*) species complex. Journal of Mammalogy, *in press.*

Dobson, F. S., A. T. Smith, and W. X. Gao. 1998. Social and ecological influences on dispersal and philopatry in the plateau pika (*Ochotona curzoniae*). Behavioral Ecology 9: 622-635.

Dutton, P. H. 1996. Methods for collection and preservation of samples for sea turtle genetic studies, pp. 17-24. *In:* B.W. Bowen and W. N. Witzell (eds.), *Proceedings of the International Symposium on Sea Turtle Conservation Genetics,* NOAA Technical Memorandum NMFS-SEFSC-396, Miami, Florida.

Edwards, A. W. F. 1972. *Likelihood.* Cambridge: Cambridge University Press.

Ehrlich, P. and L. E. Gilbert. 1973. Population structure and dynamics of the tropical butterfly, *Heliconius ethilla.* Biotropica 5:69-82.

Elder, J. H., D. L. Lerner, C. S. Hasselkus-Light, D. J. Fontenot, E. Hunter, P. A. Luciw, R. C. Montelaro, and T. R. Phillips. 1992. Distinct subsets of retroviruses encode dUTPase. Journal of Virology 66: 1791-1794.

Fitt, G. P. 1986. The influence of a shortage of hosts on the specificity of oviposition behaviour in species of *Dacus* (Diptera, Tephritidae). Physiological Entomology 11: 133-143.

Fitzgerald, J. P., C. A. Meaney, and D. M. Armstrong. 1994. *Mammals of Colorado*. University Press of Colorado, Niwot, Colorado.

Flint, R. F. 1971. *Glacial and Quarternary Geology*. John Wiley and Sons, New York.

Gao, F., L. Yue, A. T. White, P. G. Pappas, J. Barchue, A. P. Hanson, B. M. Greene, P. M. Sharp, G. M. Shaw, and B. H. Hahn. 1992. Human infection by genetically diverse SIVsm-related HIV-2 in West Africa. Nature 358: 495-499.

Goldman, N. 1993. Statistical test of models of DNA substitution. Journal of Molecular Evolution 36:182-198.

Gonnet, G.H., M. A. Cohen, and S. A. Benner. 1992. Exhaustive matching of the entire protein sequence database. Science 256:1443-1445.

Gu, X., Y.-X. Fu, and W.-H. Li. 1995. Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. Molecular Biology and Evolution. 12: 546-557.

Hadly, E. A., M. H. Kohn, J. A. Leonard, and R. K. Wayne. 1998. A genetic record of population isolation in pocket gophers during Holocene climatic change. Proceedings of the National Academy of Sciences, USA 95: 6893-6896.

Hafner, D. J., and R. M. Sullivan. 1995. Historical and ecological biogeography of nearctic pikas (Lagomorpha: Ochotonidae). Journal of Mammalogy 76: 302-321.

Hall, E. R. 1981. *The Mammals of North America*. 2nd ed. John Wiley and Sons, New York. pp. 286-292.

Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating the human-ape split by a moleuclar clock of mitochondrial DNA. Journal of Molecular Evolution 22: 160-174.

Hibbard, C. W., D. E. Ray, D. E. Savage, D. W. Taylor, J. E. Guilday. 1965. Quaternary mammals of North America. pp. 509-526. *In* H. E. Wright and D. G. Frey (eds.). *The Quaternary of the United States*, Princeton University Press, Princeton, NJ.

Hillis, D. M., and J. P. Huelsenbeck. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. Journal of Heredity 83: 189-195.

Hillis, D. M., B. K. Mable, and C. Moritz. 1996. Applications of molecular systematics. pp. 515-543. *In* D. M. Hillis, C. Moritz, and B. K. Mable (eds.). *Molecular Systematics*. Sinauer Associates, Inc., Sunderland, Massachusetts.

Huelsenbeck, J. P., and K. A. Crandall. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. Annual Reviews of Ecology and Systematics 28: 437-466.

Jaenike, J. 1978. On optimal oviposition behavior in phytophagous insects. Theoretical Population Biology 14: 350-56.

Jaenike, J. 1990. Host specialization in phytophagous insects. Annual Reviews of Ecology and Systematics 21: 243-273.

Johnson, D. R. 1967. Diet and reproduction of Colorado pika. Journal of Mammalogy 48: 311-315.

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

Kenagy, G. J. (*in press*). Role of the Columbia River Gorge in shaping ground squirrel phylogeography: vicariance or dispersal barrier? Journal of Mammalogy.

Kimura, M. 1980. A simple model for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16: 111-120.

Kottlowski, F. E., M. E. Cooley, and R. V. Ruhe. 1965. Quaternary geology of the Southwest. pp. 287-298. *In* H. E. Wright and D. G. Frey (eds.). *The Quaternary of the United States*, Princeton University Press, Princeton, NJ.

Krajick, K. 1998. Nunataks. National Geographic 194: 60-71.

Lanave, C., G. Preparata, C. Saccone, and B. Serio. 1984. A new method for calculating evolutionary substitution rates. Journal of Molecular Evolution 20: 86-93.

Lawlor, T. E. 1998. Biogeography of Great Basin mammals: Paradigm lost? Journal of Mammalogy 79: 1111-1130.

Lewis, P. O. 1988. A genetic algorithm for maximum-likelihood phylogney inference using nucleotide sequence data. Molecular Biology and Evolution 15: 277-283.

Li, W. –H. 1997. *Molecular Evolution.* Sinauer Associates, Inc., Sunderland, Massachusetts. 487pp.

Lundelius, E. L., Jr., R .W. Graham, E. Anderson, J. Guilday, J. A. Holman, D. W. Steadman, and S. D. Webb. 1983. Terrestrial Vertebrate Faunas. pp. 311-353. *In* S. C. Porter (ed.), *Late-Quaternary Environments of the United States,* University of Minnesota Press, Minneapolis.

MacArthur, R. H. and E. O. Wilson. 1963. An equilibrium theory of insular biogeography. Evolution 17: 373-387.

MacArthur, R. H. and E. O. Wilson. 1967. *The theory of island biogeography.* Princeton University Press, Princeton, NJ.

Mangel, M. 1987. Oviposition site selection and clutch size in insects. Journal of Mathematical Biology 25: 1-22.

Mangel, M. 1989a. An evolutionary interpretation of the "motivation to oviposit." Journal of Evolutionary Biology 2: 157-72.

Mangel, M. 1989b. Evolution of host selection in parasitoids: does the state of the parasitoid matter? American Naturalist 133: 688-705.

Martin, A. P. and S. P. Palumbi. 1993. Body size, metabolic rate, generation time, and the molecular clock. Proceedings of the National Academy of Science, USA 90: 4087-4091.

May, P. G. 1992. Flower selection and the dynamics of lipid reserves in two nectarivorous butterflies. Ecology 73: 2181-2191.

McClure, M. A., M. S. Johnson, and R. F. Doolittle. 1987. Relocation of a protease-like gene segment between two retroviruses. Proceedings of the National Academy of Sciences, USA 84: 2693-2697.

McClure, M. A., M. S. Johnson, D. -F. Feng, and R. F. Doolittle. 1988. Sequence comparisons of retroviral proteins: Relative rates of change and general phylogeny. Proceedings of the National Academy of Sciences, USA 85: 2469-2473.

Mead, J. I. 1987. Quaternary records of pika, *Ochotona*, in North America. Boreas 16:165-171.

Minkenberg, O. P. J. M., M. Tatar, and J. A. Rosenheim. 1992. Egg load as a major source of variability in insect foraging and oviposition behavior. Oikos 65: 134-142.

Moore, R. A. 1987. Patterns and consequences of within population variation in reproductive strategies. Ph.D. dissertation, University of Texas at Austin.

Morrison, R. B. 1965. Quaternary Geology of the Great Basin. pp. 265-286. *In* H. E. Wright and D. G. Frey (eds.). *The Quaternary of the United States*, Princeton University Press, Princeton, NJ.

Myers, G. and G. N. Pavlakis. 1992. Evolutionary Potential of Complex Retroviruses. pp. 51-105. *In* J. A. Levy (ed.). *The Retroviridae*, Plenum Press, New York.

Odendaal, F. J. 1989. Mature egg number influences the behavior of female *Battus philenor* butterflies. Journal of Insect Behavior 2: 15-25.

Odendaal, F. J., and M. D. Rausher. 1990. Egg load influences search intensity, host selectivity, and clutch size in *Battus philenor* butterflies. Journal of Insect Behavior 3: 183-193.

Palumbi, S. R. 1996. Nucleic acids II: The polymerase chain reaction, pp. 205-247. *In:* D. M. Hillis, C. Mortiz, and B. K. Mable (eds.), *Molecular Systematics*, 2nd ed. Sinauer Associates, Sunderland, Massachusetts.

Peacock, M. M. and A. T. Smith. 1997. Nonrandom mating in pikas *Ochotona princeps*: evidence for inbreeding between individuals of intermediate relatedness. Molecular Ecology 6: 801-811.

Pielou, E. C. 1992. *After the Ice Age*. University of Chicago Press, Chicago.

Pilson, D. and M. D. Rausher. 1988. Clutch size adjustment by a swallowtail butterfly. Nature 333: 361-63.

Porter, S. C., K. L. Pierce, and T. D. Hamilton. 1983. Late Wisconsin mountain glaciation in the western United States. pp. 71-114. *In* S. C. Porter (ed.), *Late-Quaternary Environments of the United States*, University of Minnesota Press, Minneapolis.

Power, M. D., P. A. Marx, M. L. Bryant, M. B. Gardner, P. J. Barr, and P. A. Luciw. 1986. Science 231: 1567-1572.

Prinsloo, P. and T. J. Robinson. 1992. Geographic mitochondrial DNA variation in the rock hyrax, *Procavia capensis*. Molecular Biology and Evolution 9: 447-456.

Prokopy, R. J., B. D. Roitberg, and R. I. Vargas. 1994. Effects of egg load on finding and acceptance of host fruit in *Ceratitis capitata* flies. Physiological Entomology 19: 124-132.

Querat, G., G. Audoly, P. Sonigo, and R. Vigne. 1990. Nucleotide sequence analysis of SA-OMVV, a Visna-related ovine lentivirus: phylogenetic history of lentiviruses. Virology 175: 434-447.

Riddle, B. R. 1996. The molecular phylogeographic bridge between deep and shallow history in continental biotas. Trends in Ecology and Evolution 11: 207-211.

Richmond, G. M. 1965. Glaciation of the Rocky Mountains. pp. 217-230. *In* H. E. Wright, Jr., and D. G. Frey (eds.) *The Quaternary of the United States*. Princeton University Press, Princeton, New Jersey.

Rosenheim, J. A., and D. Rosen. 1991. Foraging and oviposition decisions in the parasitoid *Aphytis lignanensis*: distinguishing the influence of egg load and experience. Journal of Animal Ecology 60: 873-893.

Scott, G. R. 1965. Nonglacial Quaternary geology of the southern and middle Rocky Mountains. pp. 243-254 *In* H. E. Wright and D. G. Frey (eds.). *The Quaternary of the United States*, Princeton University Press, Princeton, NJ.

Sharp, P. M., C. J. Burgess, B. H. Hahn. 1995. The Molecular Evolution of the Human Immunodeficiency Viruses. pp. 438-454. *In* A. Gibbs, C. H. Calisher, and F. Garcia-Arenal (eds.). *Molecular Basis of Virus Evolution*, Cambridge University Press.

Shields, G. F. and T. D. Kocher. 1991. Phylogenetic relationships of North American ursids based on analysis of mitochondrial DNA. Evolution 45: 218-221.

Singer, M. C. 1983. Determinants of multiple host use by a phytophagous insect population. Evolution 37: 389-403.

Singer, M. C., D. Vasco, C. Parmesan, C. D. Thomas, and D. Ng. 1992. Distinguishing between 'preference' and 'motivation' in food choice: an example from insect oviposition. Animal Behaviour 44: 463-71.

Smith, A. T. and M. L. Weston. 1990. *Ochotona princeps.* Mammalian Species 352: 1-8.

Smith, G. and F. A. Street-Perrott. 1983. Pluvial lakes of the western United States. pp. 190-214. *In* S. C. Porter (ed.), *Late-Quaternary Environments of the United States,* University of Minnesota Press, Minneapolis.

Sokal, R. R. and F. J. Rohlf. 1981. *Biometry*, 2nd edition. W. H. Freeman and Co., New York.

Somers, P. 1973. Dialects in southern Rocky Mountain pikas, *Ochotona princeps* (Lagomorpha). Animal Behavior 21: 124-137.

Stanek, E. J., S. R. Diehl, N. Dgetluck, M. E. Stokes, and R. J. Prokopy. 1987. Statistical methods for analyzing discrete responses of insects tested repeatedly. Environmental Entomology 16:320-326.

Swofford, D. L. 1998. PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Swofford, D. L., G. J. Osen, P. J. Waddell, and D. M. Hillis. 1996. Phylogenetic Inference. pp. 407-514. *In* D. M. Hillis, C. Moritz, and B. K. Mable (eds.). *Molecular Systematics.* Sinauer Associates, Inc., Sunderland, Massachusetts.

Temin, H. M. 1992. Origin and General Nature of Retroviruses. pp. 1-18. *In* J. A. Levy (ed.). *The Retroviridae*, Plenum Press, New York.

Thomas, W. K., S. Paabo, F. X. Villablanca, and A. C. Wilson. 1990. Spatial and temporal continuity of kangaroo rat populations shown by sequencing mitochondrial DNA from museum specimens. Journal of Molecular Evolution 31: 101-112.

Tolliver, K. K., M. H. Smith, P. E. Jones, and M. W. Smith. 1985. Low levels of genetic variability in pikas from Colorado. Canadian Journal of Zoology 63: 1735-1737.

United States Geological Survey. 1984. Map of Quaternary Lake Bonneville. 1:800,00 Map, Pleistocene Series 3. Atlas Publishing.

United States Geological Survey. 1991. Topological map series of the states. 1:500,00 Map. (Colorado, Nevada, Utah, and Wyoming).

Wagaman, P. C., C. S. Hassalkus-Light, M. Henson, D. L. Lerner, T. R. Phillips, and J. H. Elder. 1993. Molecular cloning and characterization of deoxyuridine triphosphatase from Feline Immunodeficiency Virus (FIV). Virology 196: 451-457.

Waitt, R. B., Jr., and R. M. Thorson. 1983. The Cordilleran Ice Sheet in Washington, Idaho, and Montana. pp. 53-70. *In* S. C. Porter (ed.), *Late-Quaternary Environments of the United States,* University of Minnesota Press, Minneapolis.

Wasserman, S. S. and D. J. Futuyma. 1981. Evolution of host plant utilization in laboratory populations of the southern cowpea weevil, *Callosobruchus maculatus* Fabricius (Coleoptera: Bruchidae). Evolution 35: 605-617.

Watt, W. B., F. S. Chew, L. R. G. Snyder, A. G. Watt, and D. E. Rothschild. 1977. Population structure of Pierid butterflies: I. Numbers and movements of some montane *Colias* species. Oecologia (Berlin) 27:1-22.

Whitworth, M. R. and C. H. Southwick. 1984. Sex difference in the ontogeny of social behavior in pikas: possible relationships to dispersal and territoriality. Behavior, Ecology and Sociobiology 15: 175-182.

Wolfram, Stephen. 1998. MATHEMATICA. Version 4.0. Wolfram Research, Inc.

Yang, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. Journal of Molecular Evolution 39:306-314.

Yang, Z., N. Goldman, and A. Friday. 1994. Comparison of models for nucleotide substitution used in maximum-likelihood phylogenetic estimation. Molecular Biology and Evolution 11: 316-324.

Zalucki, M. P. 1981. The effects of age and weather on egg laying in *Danaus plexippus* L. (Lepidoptera: Danaidae). Researches in Population Ecology 23: 318-27.

Zamudio, K. R., K. B. Jones, and R. H. Ward. 1997. Molecular systematics of short-horned lizards: biogeography and taxonomy of a widespread species complex. Systematic Biology 46:284-305.

Vita

Kelly Kathleen Agnew was born June 19, 1970 in Lawton, Oklahoma. She was adopted in August 1970 by Brad and Sue Agnew of Tahlequah, Oklahoma. She attended Tahlequah public schools and published her first paper (on the 1872 Goingsnake Affair in the Cherokee Nation, Indian Territory) when she was sixteen. After graduating from Tahlequah High School in 1988, she attended Hendrix College, in Conway, Arkansas, where she majored in Biology. She spent a term at Birkbeck College, University of London, where she studied English literature and theater. She graduated from Hendrix in 1992, and was awarded the Spessard Biology Award and the Hendrix College President's Medal. She was a student at the Rocky Mountain Biological Laboratory in Gothic, Colorado, and has spent several field seasons in both the Colorado Rockies and the California Sierra Nevada. While a graduate student at the University of Texas at Austin, she taught many courses, including Biology for Liberal Arts Majors, Human Physiology, Genetics, Genetics Laboratory, and Evolution. In her free time, she enjoys playing with her dog, equestrian sports, and gardening.

Permanent address: 1212 Hunter Conway, AR 72032 USA

This dissertation was typed by the author.