PHYLOGENY OF SOME MIDDLE AMERICAN PITVIPERS BASED ON A CLADISTIC ANALYSIS OF MITOCHONDRIAL 12S AND 16S DNA SEQUENCE INFORMATION

Steven D. Werman[1] (swerman@mesastate.edu), Brian I. Crother[2] (bcrother@selu.edu) and Mary E. White[2] (mwhite@selu.edu)

[1] Department of Biology, Mesa State College, Grand Junction, Colorado 81502 USA
[2] Department of Biological Sciences, Southeastern Louisiana University, Hammond, Louisiana 70402 USA

Abstract. The cladistic relationships of several Middle American pitvipers representing the genera *Bothrops* (*sensu stricto*), *Bothriechis*, *Cerrophidion*, *Lachesis* and *Porthidium* were determined using mitochondrial 12S and 16S DNA sequence information. Maximum parsimony analyses were performed using PAUP on aligned sequences that included published information for related taxa. Two sets of analyses were conducted: one disregarding gaps in the aligned matrix, and another with gaps treated as a fifth base. When gaps were excluded resolution declined, although the general arrangement of the taxa changed little. A consistent relationship was the grouping of ((*Porthidium, Bothriechis*) *Lachesis*). The placement of *Lachesis*, as nested within other bothropoid genera, is only partially supported by results of other authors. The arrangement of *Crotalus*, *Bothrops* and *Cerrophidion* was ambiguous when gaps were discounted. In both trees, *Agkistrodon* was basal to the New World forms. The remaining genera, *Trimeresurus* (*Protobothrops*), *Vipera*, *Azemiops*, and *Coluber*, were uniformly distant to the former taxa. Also of interest is the lack of close relationship, based on the DNA data here and elsewhere, between *Bothrops* and *Porthidium*. This is in striking contrast to results based on morphologic and allozymic analyses of previous studies. It is concluded that additional DNA sequence information from a larger sample of taxa will be necessary to better assess the phylogenetic relationships among Middle American and related pitvipers.

The bothropoid pitvipers comprise a diverse and widespread assemblage of venomous snakes distributed from southern Mexico to southern Argentina. In the last decade, progress from systematic studies of pitvipers has led to descriptions of new species and the recognition of several new generic arrangements for those species formerly assigned to *Bothrops* (*sensu lato*). Detailed species accounts can be found in Campbell and Lamar (1989), whereas the most recent generic arrangement of New World pitvipers can be found in Campbell and Brodie (1992), and summarized in Campbell and Lamar (1992).

The phylogenetic relationships among Neotropical pitvipers remain problematic. Although several studies using phenotypic character information (Crother et al., 1992; Werman, 1992; Gutberlet, 1998) and molecular data (Knight et al., 1992; Kraus et al., 1996; Cullings et al., 1997; Salomão et al., 1997; Vidal et al., 1997; Wüster et al., 1997) have been completed, many inconsistencies of phylogenetic inference persist (see Werman, 1998). Herein, we present a cladistic analysis of novel mitochondrial 12S and 16S DNA sequence information for some Middle American pitviper species, in conjunction with published sequences for related genera.

Of primary concern is the relationship of *Lachesis* to the other New World pitviper genera and the relationship of *Bothrops* (*sensu stricto*) to *Porthidium*. Among DNA studies that include *Lachesis* (Kraus et al., 1996; Cullings et al., 1997; Vidal et al., 1997) there is no clear agreement as to the position of this genus relative to other pitviper genera. *Lachesis* is either a somewhat basal lineage, or is found nested among other bothropoid genera (see Werman, 1998). Regarding *Bothrops* and *Porthidium* none of the present DNA analyses place these together as sister lineages. This is curious because Werman (1992), based on a cladistic analysis of phenotypic information, regarded them as a terminal clade of recent divergence among the New World pitvipers.

MATERIALS AND METHODS

Tissue samples of five pitviper species, representing several Middle American genera, were used as DNA sources. Total genomic DNA samples (that included mitochondrial DNA) were isolated and purified from liver and/or skeletal muscle tissue using standard digestion (SDS-Proteinase K) and extraction (phenol-chloroform) techniques. Microgram quantities of genomic DNAs were obtained for the following taxa: *Bothrops asper, Bothriechis rowleyi, Cerrophidion godmani, Lachesis muta,* and *Porthidium nasutum.* Locality data and voucher information are available on request from the authors.

Amplifications were carried out in a Thermolyne Amplitron thermal cycler utilizing established protocols with modifications specific for pitvipers (Knight and Mindell, 1993). Symmetrical amplification of the 12S sequences was accomplished with the L strand primer, 5'-AAACTGGGATTAGATACCCCACTAT-3', and the H strand primer, 5'-GTACACTTACCTTGTTACGACTT-3'. The 16S sequences were amplified with the L

strand primer, 5'-CGCCTGTTTATCAAAAACAT-3', and the H strand primer, 5'-CCGGTCTGAACTCAGATCACGT-3' (Knight and Mindell, 1993). The cycle parameters for the 12S and 16S sequences were: 85º C, 5 min, followed by 30 cycles of 94º C, 35 sec; 50º C, 35 sec; 72º C, 1 min; with an extension on the last cycle of 72º C, 5 min, then 4º C, dwell/soak. The amplifications resulted in approximately 960 bp for the combined 12S and 16S gene fragments.

Amplification products were either purified or ligated directly into *Invitrogen* pCR II TA cloning vectors following *Invitrogen* specific protocols. Ligation products were transformed into *Invitrogen* "One Shot" competent cells and plated on LB/Amp/X-gal media for colony selection. Plasmid DNAs were isolated from positive colonies using *5 Prime-3Prime* "Perfect prep" miniprep kits. Purified plasmid DNAs were digested with EcoRI, to release the inserts, and size fractionated on 2% agarose gels. Plasmids positive for 12S and 16S inserts were collected and stored at -70º C.

Sequencing effort included standard dideoxynucleotide termination methods (Sanger et al., 1977; Hillis et al., 1996) using CircumVent thermal cycle DNA sequencing kits (New England Biolabs) in combination with chemiluminescent Phototope detection kits (NEB). In addition, sequences were also determined with an automated Applied Biosystems Inc. (ABI) Prism DNA sequencer. In both cases, standard M13 forward and reverse sequencing primers were used. Sequences were determined by comparing both forward and reverse sequencing reactions. Approximately 410 nucleotide positions of the 12S gene and 550 nucleotide positions for the 16S gene were scored (e.g., 410 and 548 for *P. nasutum* 12S and 16S, respectively).

Partial 12S and 16S DNA sequences for *Coluber constrictor*, *Vipera ammodytes*, *Azemiops feae* and *Agkistrodon bilineatus* were obtained from Knight and Mindell (1993). 12S sequence information for *Trimeresurus* (*Protobothrops*) *mucrosquamatus* was obtained from Genbank, accession # D31613 (Eguchi, unpublished, 1994). 12S and 16S sequences for *Crotalus aquilus* were also obtained from Genbank under the accession numbers L14373 and L14374, respectively (Knight et al., 1993). These six taxa were used for outgroup comparisons and rooting purposes in the cladistic analyses.

Sequence alignments were accomplished by eye using a color coding scheme specific for each nucleotide. Insertions and deletions in the sequence matrix generated 975 total characters. Aligned sequences were subjected to phylogenetic analysis using PAUP, version 3.1.1 (Swofford, 1993). Maximum parsimony (MP) analyses were performed using random addition of sequences, tree-bisection-reconnection branch swapping, the MULPARS option with ACCTRAN optimization. Gaps were either treated as a fifth base or as missing information in the data matrix. Characters were not weighted. Unknown nucleotide information was designated as "n" and treated as missing information. Strict consensus trees were retained for multiple MP solutions. Bremer support values (Bremer, 1994), that identify how many extra steps on a particular branch are necessary to collapse the branch in a consensus tree of proximate parsimonious solutions, were used to assess nodal stability.Two analyses were performed: the first with gaps

included as character information and a second where gaps were discounted. The rationale for these analyses was to determine the effect of gaps on cladogram resolution.

RESULTS

The MP analysis for all taxa, with gaps included, resulted in a single parsimonious solution ([Figure 1](#)) with a tree length of 403 steps and a consistency index (CI) of 0.556, homoplasy index (HI) of 0.444, retention index (RI) of 0.399 and a rescaled consistency index (RC) of 0.222. Two terminal sister clades included ((*Porthidium*, *Bothriechis*) *Lachesis*) and ((*Bothrops, Crotalus*) *Cerrophidion*) with proximal and distal lineages being *Agkistrodon* and *Trimeresurus*, respectively. These taxa are rooted proximally by a clade composed of *Azemiops* and *Vipera* and distally by *Coluber*. Bremer support values were strongest for the *Porthidium-Bothriechis* grouping (9) and the relationship of *Lachesis* to these two taxa (6).

When gaps were discounted as characters, an analysis of all taxa resulted in five equally parsimonious solutions all with 343 steps, a CI of 0.528, HI of 0.472, RI of 0.377 and a RC of 0.199. One of these solutions as well as a strict consensus tree of the five solutions is shown in [Figure 2](#). At the terminus of this tree is a clade composed of ((*Porthidium*, *Bothriechis*) *Lachesis*) with sister relations to a trichotomy that includes *Bothrops*, *Cerrophidion* and *Crotalus*. Below this group and proximal to a polychotomy of *Trimeresurus*, *Azemiops*, *Vipera* and *Coluber*, a lineage represented by *Agkistrodon* diverges. Bremer support values are strongest for the *Porthidium-Bothriechis* grouping (5) and the sister group relation of *Lachesis* (4) to these former taxa.

DISCUSSION

In spite of several recent attempts to ascertain the phylogenetic relationships of pitvipers many species relationships remain unresolved (reviewed in Werman, 1998). Character information that includes morphology (Werman, 1992; Crother et al., 1992; Gutberlet, 1998), allozymes (Werman, 1992; Crother et al., 1992) and several mitochondrial gene sequences (Kraus et al., 1996; Cullings et al., 1997; Salomão et al., 1997), exists for many taxa, but a stable phylogeny for pitvipers remains elusive (Werman, 1998). In spite of the general lack of congruence among diverse studies conducted previously, some patterns are emerging to which our present analyses offer support.

It is clear from [Figure 1](#), [Figure 2](#), and Knight and Mindell (1993) that *Coluber*, *Vipera* and *Azemiops* are basal to the pitvipers. However, in our analysis, when gaps are included as character information, *Azemiops* clusters

with *Vipera,* and *Coluber* represents the earliest divergent lineage. This arrangement is not in complete agreement with that of Knight and Mindell (1993). They suggested that *Azemiops* is a sister lineage to *Agkistrodon ,* relative to *Vipera* and *Coluber*, and thus may well represent the sister group to all pitvipers. Kraus et al. (1996) found *Azemiops* to be basal to the pitvipers, although in some cases *Causus* was a closer sister group to the crotalines. In the present study, when gaps are removed from consideration, these genera are not resolved based on a consensus for equally parsimonious solutions (Figure 2). In addition, there remains little doubt that *Trimeresurus* (*Protobothrops*) is outside the New World pitviper radiation. This conclusion is supported by several other studies (e.g., Kraus et al., 1996; Salomão et al., 1997; Vidal et al., 1997). Furthermore, our study (Figure 1 and Figure 2) indicates that New World *Agkistrodon* is a basal lineage among the New World genera. This was also reported by Werman (1992) although conflicting interpretations based on DNA evidence exist (Kraus et al., 1996; Vidal et al., 1997).

Within the Middle American genera, the 12S and 16S sequence information suggests that *Porthidium* and *Bothriechis* are sister taxa (Figure 1 and Figure 2). This arrangement, based on Bremer support values, is robust. In addition, *Lachesis* emerges as a sister lineage to the former two genera. This is a curious result because no other gene sequence analysis to date supports a close relationship between *Porthidium* and *Bothriechis* ; also *Lachesis* has rarely been placed close to the former two taxa. For example, Werman (1992) showed a monophyletic relationship among *Bothriechis*, *Porthidium* and *Bothrops* , but depicted *Lachesis* as a lineage of early divergence in relation to Middle American pitvipers. Consequently, the relationship of *Bothriechis* to *Porthidium* represents yet another systematic hypothesis to be tested by future study. The relationship of *Lachesis* to other New World pitvipers remains problematic and both previous and present DNA analyses offer little to resolve the position of this interesting genus.

The relationships of *Bothrops* , *Cerrophidion* and *Crotalus* are unclear, although they appear basal in divergence among New World pitvipers. They either comprise a sister group to the above pitviper genera (Figure 1) or an unresolved trichotomy when gaps are treated as missing data (Figure 2). The basal placement of *Cerrophidion* is supported by Knight et al. (1992) and Werman (1992), whereas a basal placement for *Crotalus* is supported by Werman (1992) and Salomão et al. (1997).

Werman (1992) found a close affinity between *Bothrops* and *Porthidium* in a cladistic analysis of morphology and allozymes. He hypothesized that these taxa constituted a clade of recent divergence, relative to other New World pitviper genera with similar and presumably derived cranial features (Werman, 1992, 1998) and allozymes (Werman, 1992, 1997). This sister relationship was further supported by an in-depth phylogenetic analysis of morphological characters of many bothropoid genera by Gutberlet (1998). The close relationship of *Bothrops* and *Porthidium* is not supported by the present analysis or other DNA studies (e.g., Kraus et al., 1996). Although addressed in detail in Werman (1998), it is apparent that the analyses of morphologic and allozymic information are discordant with respect to the DNA analyses. Either the phenotypic data

are strongly convergent or the DNA sequence data are not rigorous nor voluminous enough to allow for accurate phylogenetic reconstructions. Surely, more DNA sequence information is necessary to resolve the relationship
between *Bothrops* and *Porthidium,* as well as the relationships among other New World pitviper genera. However, we suspect that the incongruences among data sets may simply reflect the disjoint evolutionary nature of mitochondrial DNA with respect to the nuclear genome. Perhaps an exploration of the latter will allow for further resolution of the relationships among these interesting snakes.

ACKNOWLEDGMENTS

LITERATURE CITED

BREMER, K.
        1994. Branch support and tree stability. Cladistics 10: 295-304.

CAMPBELL, J.A. AND E. D. BRODIE, JR.
        1992. Biology of the pitvipers. Selva, Tyler, Texas.

CAMPBELL, J. A. AND W. W. LAMAR.
        1989. The venomous reptiles of Latin America. Comstock Publ. Assoc. Cornell Univ. Press, Ithaca, New York.
        1992. Taxonomic status of miscellaneous Neotropical viperids, with the description of a new genus. Occasional Papers Museum Texas Tech University 153: 1-31.

CROTHER, B. I., J. A. CAMPBELL AND D. M. HILLIS.
        1992. Phylogeny and historical biogeography of the palm-pitvipers, genus *Bothriechis*: Biochemical and morphological evidence. Pages 1-20 in Biology of the pitvipers. (Campbell, J. A. and E. D. Brodie, Jr. eds). Selva, Tyler, Texas.

CULLINGS, K.W., D.J. MORAFKA, J. HERNENDEZ AND J.W. ROBERTS.
       1997. Reassessment of phylogenetic relationships among pitviper genera based
       on mitochondrial cytochrome b gene sequences. Copeia 1997: 429-432.

GUTBERLET, JR., R.L.
       1998. The phylogenetic position of the Mexican black-tailed pitviper (Squamata:
       Viperidae: Crotalinae). Herpetologica 54: 184-206.

HILLIS, D.M., B. K MABLE, A. LARSON, S. K. DAVIS AND E. A. ZIMMER.
       1996. Nucleic acids IV: Sequencing and cloning. Pages 321-381 in Molecular
       systematics. (Hillis, D. M., Moritz, C. and B. K. Mable, eds). 2nd ed. Sinauer
       Associates Inc. Sunderland, MA, USA.

KNIGHT, A., L. D. DENSMORE III. AND E. D. RAEL.
       1992. Molecular systematics of the *Agkistrodon* complex. Pages 49-70 in Biology
       of the pitvipers. (Campbell, J. A. and E. D. Brodie, Jr., eds). Selva, Tyler, Texas.

_____ AND D. P. MINDELL.
       1993. Substitution bias, weighting of DNA sequence evolution, and the
       phylogenetic position of Fea's viper. Syst. Biol. 42: 18-31.

_____, STYER, D. S., S. PELIKAN, J. A. CAMPBELL, L. D. DENSMORE III AND D. P.
MINDELL.
       1993. Choosing among hypotheses of rattlesnake phylogeny: A best-fit rate test
       for DNA sequence data. Systematic Biology 42: 356-367.

KRAUS, F., D. G. MINK AND W. M. BROWN.
       1996. Crotaline intergeneric relationships based on mitochondrial DNA sequence
       data. Copeia 1996: 763-773.

SALOMÃO, M. G., W. WÜSTER, R. S. THORPE, J-M. TOUZET   AND BBBSP.
       1997. DNA evolution of South American pitvipers of the genus *Bothrops* (Reptilia:
       Serpentes: Viperidae). Pages 89-98 in Venomous snakes: Ecology, evolution
       and snakebite. (R. S., Thorpe, Wüster, W. and A. Malhotra, eds).
       Symposia Zoological Society of London No. 70., Oxford University Press.

SANGER, F. NICKLEN AND A. R. COULSON.
       1977. DNA sequencing with chain-termination inhibitors. Proceedings of the
       National Academy of Sciences U.S.A. 74: 5463-5467.

SWOFFORD D. L.
       1993. PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1. Formerly
       distributed by Illinois Natural History Survey, Champaign, Illinois.

VIDAL, N., G. LECOINTRE, J.C. VIE AND J.P GASC.

1997. Molecular systematics of pitvipers: Paraphyly of the *Bothrops* complex. C.R. Acad. Sci. Paris, Sciences de la vie/Life Sciences 320: 95-101.

WERMAN, S.D.
1992. Phylogenetic relationships of Central and South American pitvipers of the genus *Bothrops* (*sensu lato*): Cladistic analyses of biochemical and anatomical characters. Pages 21-40 in Biology of the pitvipers (Campbell, J. A. and E. D. Brodie, Jr., eds). Selva, Tyler, Texas.
1997. Systematic implications of lactate dehydrogenase isozyme phenotypes in Neotropical pitvipers (Viperidae: Crotalinae). Pages 79-88 in Venomous snakes: Ecology, evolution and snakebite. (R. S., Thorpe, Wüster, W. and A. Malhotra, eds). Symposia Zoological Society of London No.70, Oxford University Press.
1998. Molecular phylogenetics and morphological evolution in Neotropical pitvipers: An evaluation of mitochondrial DNA sequence information and the comparative morphology of the cranium and palatomaxillary arch. Kaupia, Darmstadter Beitrage zur Naturgeschichte (in press).

WÜSTER, W., M. G. SALOMÃO, R. S. THORPE, G. PUORTO, M. F. D. FURTADO, S.A. HOGE, R. D. G. THEAKSTON AND D. A. WARRELL.
1997. Systematics of the *Bothrops atrox* complex: New insights from multivariate analysis and mitochondrial DNA sequence information. Pages 99-113 in Venomous snakes: Ecology, evolution and snakebite. (R. S., Thorpe, Wüster, W. and A. Malhotra, eds). Symposia Zoological Society of London No. 70, Oxford University Press.

APPENDIX

Character matrix of aligned (L-strand) mitochondrial 12S and 16S DNA sequence information. The following abbreviations are: C- *Coluber constrictor*, AZ- *Azemiops feae*, V- *Vipera ammodytes*, AG- *Agkistrodon bilineatus*, N- *Porthidium nasutum*, R- *Bothriechis rowleyi*, L- *Lachesis muta*, CE- *Cerrophidion godmani*, AS- *Bothrops asper*, CR- *Crotalus aquilus*, T- *Trimeresurus mucrosquamatus*. Asterisks indicate the start of the 12S and 16S sequences. Missing information is designated by "n", gaps by "-".

```
    *12S                                                50
C   nnnnnnnnnnnnnnnnncaatcaaattaccaattgttngnnaaataactac
AZ  nnnnnnnnnnnnnnnnncaattaaaccaccaattgttcgccaaacaactac
V   nnnnnnnnnnnnnnnnncaattaaattactaattgttcgccaaacaactac
AG  nnnnnnnnnnnnnnnnncaattaaattaccaattgttcgccaaacaactac
N   gcctagccacaacaaacaattaaactaccaattgttcgccaaacaactac
R   gcccaaccgtaacacacaattaaattaccaattgttcgccaaacaactac
L   gcctagccgtaacaggcaattaaaataccaattgttcgccaaacaactac
CE  gcctaaccgtaacacgcaattaaaccaccaattgttcgccaaacaactac
AS  gcctagc-aaaacacgcaattaaactaccaattgcccgccaaacaactac
```

```
CR nnnnnnnnnnnnnnnncaattaaaccaccaattgtccgccaaacaactac
T  nnnnnnnnnnnnnnnnnnnnnnnnnttcacaattgtccaacaaacaactac
                                                    100

C  gagtaaagacttaaaatttaaaagacttgacggtacctcataacaaccta
AZ gagtaataacttaaaacttaaaagacttgacggtgcttcatcacaaccta
V  gagtagtaacttaaaacttaaaagacttgacggtgcttcatcacagccta
AG gagta-ta-cttaaaaacttaaaagacttgacggtgcttcaccacgccta
N  gagccata-cttaaaaacttaaaagacttgacggtacttcatcacaccta
R  gagtccca-cttaaaaacttaaaagacttgacggtacttcaccacgccta
L  gagtcaca-cttaaaaacttaaaagacttgacggtacttcaccacgccta
CE gagtcata-cttaaaaacttaaaagacttgacggtacttcaccacgccta
AS gagttaaa-cttaaaaacttaaaagacttgacggtacttcaccacaccta
CR gagttaca-cttaaaaacttaaaagacttgacggtgcttcaccacgccta
T  gagttaca-cttaaaaactgaaaagacttgacggtacttcaccacaaccta
                                                    150

C  gaggagcctgtcaaataaccgatactccacgattaacccaacctccccta
AZ gaggagcctgtctaacaaccgacaacccacgattaacccaaccctttctt
V  gaggagcctgtctaacaaccgacaacccacgtttaaccccgccccttctt
AG gaggagcctgtccaacaaccgataatccacgattaacccaacccccctctt
N  gaggagcctgtctaataaccgacaacccacgattaacccaacccctccctt
R  gaggagcctgtctaacaaccgacaacccacgattaacccaaccccttctt
L  gaggagcctgtccaataaccgacaacccacgattaacccgacccccttcta
CE gaggagcctgtctattaaccgacaacccacgattaacccagcctcccctc
AS gaggagcctgtccaataaccgacgacccacgattaacccagcccccccctt
CR gaggagcctgtctagtaaccgataatccacgattaacccagcccccctctg
T  gaggagcctgtctaataaccgataatccacgattaatctcaccctttctt
                                                    200

C  gccttaacagtctatataccgccgtcgccagcctaccttgtgaaagaaac
AZ gcc--aacagtctatataccgccgtcgccagcttaccttgtaaaagaaat
V  gcc--aacagtctatataccaccgtcgcaagcttaccttgtaagagtttt
AG gcc--aacagtctatataccgccgtcgccagcttaccttgtaaaagaaat
N  gccccaacagcctatataccgccgtcgccagcctaccttgtaaaagaaat
R  gcctg-acagtctatataccgccgtcgccagcctaccttgtaagagaaaa
L  gccc-aacagtctatataccgccgtcgccagcttaccttgtaaaagaaat
CE gccc-aacagtctatataccgccgtcgccagcctaccttgtaaaagaagt
AS gcc--aacagtctatataccgccgtcgccagcctacctcctaagagaaat
CR gcc-taacagtctatataccgccgtcgccagcttaccttgtaaaagaaat
T  gccc-aacagtctatataccgccgtcgccagcctaccttgtaaaagaaat
                                                    250

C  aaagtgagctaaatagccgccacactaacacgacaggtcgaggtgtaact
AZ atagtgagctaaatagtactg-cactaaaacgacaggtcgaggtgtaacc
V  aaagtgagctacaaagttgtcacacttaaacgacaggtcgaggtgtaacc
AG aaagtgagctaaatagtactaacactaaaacgacaggtcgaggtgtaact
N  aaagtagactaaacagtatcat-actaaaacgacaggtcgaggtgtagct
R  aaagtatgccaaatag---caccactaacacgacaggtcgaggtgtaact
L  aaagtaagccaaatagtatt-tcactaaaacgacaggtcgaggtgtaact
CE aaagtaggctaaacagta-ccacactaaaacgacaggtcgaggtgtaact
AS aaagtaggccaaacagtacccgcactaaaacgacaggtcgaggtgtaact
CR aaagtaagctaaatagca-ccacacttaaacgacaggtcgaggtgtaact
T  aaagtaagccaaacagtactg-cactaaaacgacaggtcgaggtgtaacc
                                                    300
```

```
C  tatgggaaggaccatgatgggctacattttctaaaacagaaaacacgaat
AZ aatgaaaaggatcaagatgggctacattctcctgacccgagaatacgaat
V  catgaggggggacaaagatgggctacattctc-taaccagagaacacgaat
AG catgaagggg-ctaagatgggctacattctcc--aaccgagaatacgaac
N  tatggaggggaccaagatgggctacattctcc--aactgagcacacgaac
R  tatggaggggcccaagatgggctacattctc---aaccgagcatacggac
L  tatgaagagg-tcaagatgggctacattctcc--aaccgagaatacgaac
CE catgggggaggactaagatgggctacattctcc--aaccaagaatacggat
AS catgaggcggacccagatgggctacactctcac-acccgagaatacggaa
CR aatgagagggactaagatgggctacattctctt-atccgagaatacgaac
T  aatgaaaggggctaagatgggctacattctcc--aaccgagaatacgaac
                                                    350

C  aa-actatgaaa-taagaaactgaaggcggatttagcagtaaactaagaa
AZ aacactatgaaaattagtgcttgaaggcggatttagcagtaagataagaa
V  ctcactatgaatatg-gtgactgaaggaggatttagcagtaagataagaa
AG aacactatgaaact-agtctttaaaggcggatttagcagtaagataagaa
N  aacactatgaaact-agtctttaaaggcggatttagcagtaagataggaa
R  aacactatgaa-cttagtctttaaaggcggatttagcagtaagataggaa
L  aacactacgaaact-agttttttaaaggcggatttagcagtaagctaggaa
CE aacactatgaaa-ttagtctttaaaggcggatttagcagtaagataagaa
AS aatactacgaaa-ttagtctttaaaggtggatttagcagtaaggtgggaa
CR aatactatgaaa-ttagtgtttaaaggcggatttagcagtaagataagaa
T  agcactatgaaa-caagtgcccgagggcggatttagcagtaagataagaa
                                                    400

C  ctaaatacctagtcgaaaccaatgcaatgaggtgcgtacacaccgcccgt
AZ taaaanacttaactgaanatatagcaatgaagtgcgtncacaccgcccgt
V  taaaatacctaactgaacattaagcaatgaagtgcgtacacaccgcccgt
AG taaaangcttaactgaacataacgcaatgaagtgcgtacacaccgcccgt
N  taaaacacctaactggacataacgcaatgaagtgcgcacacaccgcccgt
R  taaaatacctaactggacataacgcaatgaagtgtgtacacaccgcccgt
L  tagaatacctaactgaacatatcgcaatgaagtgcgtacacaccgcccgt
CE taaaatacctaactgaacataacgcaatgaagtgcgcacacaccgcccgt
AS taaaacacccaactgaacataacgcaatgaagtgcgtacacaccgcccgt
CR taaaanacttaactgaacataacgcaatgaagtgcgcacacaccgcccgt
T  taaaacacttaactgtacacaacnnnnnnnnnnnnnnnnnnnnnnnnnnnn

   *16S                                             450
C  catccctgtcnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
AZ catccctgccnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
V  catccctgccnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
AG catccctgccnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
N  catccctgctactacaactctttatctt-aa-taaacca-ctaacaaaaa
R  catcccttc-accacaaca-taa-cctt-aactaaaccaactattaaaat
L  catccctg--accacaatactaaaa-tt-aaataaactaaccaaaataa-
CE catccctgccaccacaacactaaacctt-aa-taaacc-attaaattaaa
AS catccctgctaacacaatactaaacctt-aa-taaac-aaccaaaataaa
CR caccccctgtnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
T  nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
                                                    500

C  nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
AZ nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
V  nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
AG nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
N  -ttaaccagggcaacctttagccaaacaagtattaaaggcgacgcctgcc
```

```
R  c--aaacagggcaaccttttagccaaacaagtattaaaggcaacgcctgcc
L  -tcaaccagggcaacctttagctaaacaagtattaacggcaacgcctgcc
CE gccaaccagggcaacctttagccaaacaagtattaaaggcaacgcctgcc
AS -ccaaacagggcaacctttagccaaccaagtattaaaggcgacgcctgcc
CR nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
T  nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
                                                   550

C  nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
AZ nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
V  nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
AG nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
N  cagtgaaaa-ttaaacggccgcggtaccctaaccgtgcaaaggtagcata
R  cagtgaaaa-ttaaacggccgcggtaccctaaccgtgcaaaggtagcata
L  cagtgaaaa-ttaaacggccgcggtatcctaaccgtgcaaaggtagcata
CE cagtgaaaaattaaacggccgcggtatcctaaccgtgcaaaggtagcata
AS cagtgaaaaattaaacggccgcggtatcctaaccgtgcaaaggtagcaca
CR nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
T  nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
                                                   600

C  nnnnnnnnnnnnnnnnnnnntagacctgtatgaaaggcaaaatgagggg-tt
AZ nnnnnnnnnnnnnnnnnnnntagacccgtatgaaaggcaaaatgagggg-tt
V  nnnnnnnnnnnnnnnnnnnntggactcgtatgaaaggcaaaatgagag-cc
AG nnnnnnnnnnnnnnnnnnnntagacctgtatgaaaggcaaaatgagag-tc
N  atcatttgtctattaattgtagacctgtatgaa-ggccaaatgagag-cc
R  atcatttgtctattaattgtagacctgaatgaa-agccaaatgagagccc
L  atcatttgtctattaattgtagacctgtatgaaaggcaaaatgagag-cc
CE atcatttgtctattaattgtagacctgtatgaaaggcaaaatgagag-cc
AS atcatttgtctattaattgtagacctctatgaaaggcaaaatgagag-cc
CR nnnnnnnnnnnnnnnnnnnntagaccactatgaaaggcaaaatgagag-cc
T  nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
                                                   650

C  tagctgtctcttataataaatc-aattaaact-gatctcctagtacaaaa
AZ caactgtctcttataatacatc-aattaaact-gatcttctagtacaaaa
V  tgactgtctcttacaactaatc-aattaaact-gatctcccagtacaaaa
AG caactgtctcttataacaaatc-aattaaact-gatctcctagtacaaaa
N  caactgtctcttataacaaat-caattaaacttgtcctcctagtacaaga
R  caactgtctcttataataaatccaattaaacttgatcttctagtncagaa
L  tgactgtctcttataacaaatc-aattaaact-gatctcttagtacaaaa
CE caactgtctcttataacaaatc-aattaaact-gatctcctagtacaaaa
AS caactgtctcttgtaacaaatc-aattaaact-gatctcctagtacaaaa
CR cagctgtctcttataataaatc-aattaaact-gatctcctagtacaaaa
T  nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
                                                   700

C  gctaggattccaacataagaccagaagaccctgtgaagcttaaactaaac
AZ gctagaatactaacataagaccagaagaccctgtgaagctttaactaaac
V  gctggaatacccccataagaccagaagaccctgtgaagctttaactaacc
AG gctagaatattaacataagaccagaagaccctgtgaagctttaactaacc
N  gctagaataataacataagaccagaagaccctgtgaagcttaacctaacc
R  gctagaataccaacataagaccagaagaccctgtgaagctttaactaacc
L  gctaagatactaacataagaccagaagaccctgtgaagctttaattaacc
CE gctagaatactaacataagaccagaagaccctgtgaagctttaactaacc
AS gctagaatactaacataagaccagaagaacctgtgaagctttaactaacc
```

```
CR gctagaatactaacataagaccagaagaccctgtgaagctttaactaacc
T  nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
                                                  750

C  tattaaaacatataatagctactttcggttggggcgaccttggaaaaaaa
AZ tattaaaccacttaataactactttcggttggggcgaccttggaaaaaaa
V  tattaaaccctataataactaatttcggttggggcgaccttggaaaaaaa
AG tattaaaccaaataatagctactttcggttggggcgaccttggaaaaaaa
N  tattaaacccaataataattactttcggttggggcgaccttggaataana
R  tattaaaccaaataatagttactttcggttgggccgaccttggaataaaa
L  tattaaacctaataataattactttcggttggcgcgaccttggaataaaa
CE tattaaacccaataataactactttcggttggggcgaccttggaataaaa
AS tactaaaccaaataatagctactttaggttggggcgaccttggaataaaa
CR tattaaacccaataataactacttttggttggggcgaccttggaataaaa
T  nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
                                                  800

C  aagaacttccaaacacaatgacttaa-accatactaaca-ggccaacaag
AZ aagaacttccaa--ata-tcaacatt-accattaaa----ggcagacaag
V  aagaacttccaa-cata-tgaattcc-tccatagaa---aggcaaacaag
AG aagaacttccaa-tata-tgaattcc-tccataaaaccaaggcgaacaag
N  aagaacttccaac-ata-tgactccc-tc-atagaagtaaggctaacaag
R  aagaacttccaacaaca-tgact-aa-tc-atagaagataggcaaacaag
L  aagaacttccaaca-ta-tgact-tc-tc-ata-aaataaggcaaacaag
CE aagaacttccaat-ata-tgacttcctc-ata-aaataaggccaacaag
AS aagaacttccaacc-ta-tgactttc-tc-ata--aataaggcaaacaag
CR aagaacttccaac-ata-tgactttc-tc-ata-aaatcaggcaaacaag
T  nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
                                                  850

C  ccactaaaagacccagtaaaactgataaatgaaccaagttactccaggga
AZ cccacataagacccagcatagctgacaattgaaccaagttactccaggga
V  cc-acattagacccagtatatctgattaatgaaccaagttactccaggga
AG ccaatattagacccagtaaagctgataatcgaaataagttactccaggga
N  ccaatactagatccagtatgactgataattgaaataagttactccaggga
R  ccaacaccagacccagtactactgataattgaaataagttactccaggga
L  ccaatattagacccagtactactgataattgaaataagttactccaggga
CE ccaacactagaccccgtataactgataattgaaataagttactccaggga
AS ccaacacaagacccagcactgctgataatcgaaacaagttactccaggga
CR ccaaaactagacccagcacagctgataattgaaacaagttactccaggga
T  nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
                                                  900

C  taacagcgcaatcttcttcaagagcccatatcaaaaagaaggtttacgac
AZ taacagcgctatcttcttcaagagcccatatcaaaaagaaggtttacgac
V  taacagcgctatcttcttcaagagcccatatcaaaaagaaggtttacgac
AG taacagcgctatcttctttaagagcccatatcaaaaagaaggtttacgac
N  taacagcgccatcttcttcaagagcccatatcaaaaagaaggtttacgac
R  taacagcgccatcttctncaagagcccatatc-agaagaaggtttgcgac
L  taacagcgccatcttctttatgagcccatatcaaaaaggaggtttacgac
CE taacagcgctatcttctttaagagcccatatcaaaaagaaggtttacgac
AS taacagcgctatcttctttaagagtccatatcaaaaagaaggtttacgac
CR gaacagcgctatcttcttcaagagcccatatcaaaaagaaggtttacgac
T  nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
                                                  950
```

```
C  ctcgatgttggatcaggacatcctaatggtgcagccgctattaagggttc
AZ ctcgatgttggatcaggacatcctagcaatgcaaccgttactaagggttc
V  ctcgatgttggatcaggacatcctagtaatgcaaccgttactgagggttc
AG ctcgatgttggatcaggacatccaagtaatgcagccgttactaaaggttc
N  ctcgatgttggatcaggacatcccagtaatgcaaccgttactaaaggttc
R  ctcgatgttggatcagggcatcccagtaatgcagccgttactaaaggttc
L  ctcgatgttggatcaggacaccccagtaatgcagccgttactaaaggctc
CE ctcgatgttggatcaggacatcccagtaatgcaaccgttactaaaggttc
AS ctcgatgttggatcaggacaccccagtaatgcaaccgctactaaaggttc
CR ctcgatgttggatcaggacatccaagtaatgcagccgttactaaaggttc
T  nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnn
                                                   975

C  gtttgttcaacgattaacagtccta
AZ gtttgttcaacgattaatagtccta
V  gtttgttcaacgattaacagtccta
AG gtttgttcaacgattaacagtccta
N  gtttgttcaacgattaatagccca
R  gtttgttcaacgattaacagccta
L  gtttgttcaacgattaatagtccta
CE gtttgttcaacgattaacagtccta
AS gtttgttcaacgattaatagtccta
CR gtttgttcaacgattaacagtccca
T  nnnnnnnnnnnnnnnnnnnnnnnnn
```

_____