PARACOTALPA URSINA SPECIES COMPLEX REVEALED: THE TRUE BIODIVERSITY OF THE CALIFORNIA BEAR SCARABS (COLEOPTERA: SCARABAEIDAE)

A Thesis by

Oliver Keller

Bachelor of Science, Saginaw Valley State University, 2012

Submitted to the Department of Biological Sciences
and the faculty of the Graduate School of
Wichita State University
in partial fulfillment of
the requirements for the degree of
Master of Science

May 2014

PARACOTALPA URSINA SPECIES COMPLEX REVEALED: THE TRUE BIODIVERSITY OF THE CALIFORNIA BEAR SCARABS (COLEOPTERA: SCARABAEIDAE)

The following faculty members have examined the final copy of this thesis for form and content, and recommend that it be accepted in partial fulfillment of the requirement for the degree of Master of Science with a major in Biological Sciences.

____
Mary Liz Jameson, Committee Chair

____
James Beck, Committee Member

____
Peer Moore-Jansen, Committee Member

DEDICATION

To my family and my two best friends:
Danke für eure Geduld, Zeit, und Hilfsbereitschaft

“The rise of genetics during the first thirty years of this century had a rather unfortunate effect on the prestige of systematics. The spectacular success of experimental work in unraveling the principles of inheritance and the obvious applicability of these results in explaining evolution have tended to push systematics into the background. There was a tendency among laboratory workers to think rather contemptuously of the museum man, who spent his time counting hairs or drawing bristles, and whose final aim seemed to be merely the correct naming of his specimens.”
– Ernst Mayr (1942:3)

Acknowledgments

I thank the institutions and private collections that provided insect loans for this work: Brett Ratcliffe and MJ Paulsen (UNSM: University of Nebraska State Museum, Lincoln, NE, USA), Dave Furth and Floyd Shockley (USNM: U.S. National Museum, Washington, D.C., USA, currently housed at UNSM), David Kavanaugh (California Academy of Sciences, San Francisco, CA, USA), Peter Oboyski (Essig Museum of Entomology, University of California-Berkeley, CA, USA), James Boone (Field Museum of Natural History, Chicago, IL, USA), Shawn Clark (Monte L. Bean Life Science Museum, Brigham Young University, Provo, UT, USA), Brian Brown (Natural History Museum of Los Angeles County, Los Angeles, CA, USA), Christopher Marshall (Oregon State Arthropod Collection, Oregon State University, Corvallis, OR, USA), Jim Berrian (San Diego Natural History Museum; San Diego, CA, USA), Frank Merickel (W.F. Barr Entomology Museum, University of Idaho, Moscow, ID, USA), Zach Falin (Snow Entomological Museum, University of Kansas, Lawrence, KS, USA), and Richard Zack (M.T. James Museum, Washington State University, Pullman, WA, USA).

Many thanks to David Hawks for sharing his knowledge about the genus *Paracotalpa*, his invaluable information about collecting sites, and the DNA-grade specimens without which this research would have not been possible.

I am grateful to the following Coleoptera experts for their guidance throughout my research: Jim Hogue and William Warner. Many thanks to the Jameson Biodiversity Laboratory (Wichita State University, Wichita, KS): Cristian Beza-Beza, Audrey McTaggart, Matthew Moore, and David Wickell. I thank Dr. Mary Liz Jameson and Dr. James Beck for their advice and guidance during the development of this research, and Dr. Peer Moore-Jansen for his invaluable help.

Funding for this research was provided by the Department of Biological Sciences, Wichita State University; the Graduate School, Wichita State University; the Southwestern Association of Naturalists Howard McCarley Research Award; and NSF DBI 0743783 to M.L. Jameson and collaborators.

ABSTRACT

The complex geology of southern California is an important mechanism that drives lineage isolation and promotes high levels of inter- and intraspecific variation. This, in combination with anthropogenic habitat fragmentation, potentially threatens a host of endemic animals and plants in this biodiversity hotspot region. The endemic California "bear scarabs" offer a model for examining highly variable lineages, allowing us to elucidate evolutionary mechanisms that generate variation and understand threats to biodiversity. The phenotypically variable *Paracotalpa ursina* species complex is assessed using a combination of DNA data (nuclear AFLP and mitochondrial sequences), morphology, and distribution modeling. I use molecular data to examine discontinuities between populations and investigate the role of geography and urban centers in genetic isolation among *P. ursina* morphotypes. All data sets suggest that the Transverse Mountain Ranges in California form a biogeographic barrier isolating populations and acting, in part, as a mechanism that promotes variation. Based on these results, two groups in the *P. ursina* species complex are identified: *Paracotalpa ursina* (Horn, 1867) (northern group) and *Paracotalpa rotunda* (Casey, 1915) (southern group). As a result, the following are considered synonyms of *P. ursina* (Horn): *Paracotalpa seriata* (Casey, 1915) and *P. ursina piceola* Saylor, 1940. The following names are considered synonyms of *P. rotunda* (Casey): *Paracotalpa brevis* (Casey, 1915), *P. laevicauda* (Casey, 1915), *P. rubripennis* (Casey, 1915), *P. nigripennis* (Casey, 1915), and *P. leonina* (Fall, 1932). A key to the five species of *Paracotalpa* is presented and based on non-variable species characteristics. Results demonstrate that AFLP genotyping assists in revealing cryptic species in beetles.

TABLE OF CONTENTS

Chapter Page

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF FIGURES

LIST OF PLATES

LIST OF ABBREVIATIONS

BIOCLIM	World-Clim global climate data
CO1	Mitochondrial Cytochrome Oxidase subunit I
MAXENT	Maximum Entropy
M	Molar
ml	Milliliter
PCR	Polymerase Chain Reaction
µM	Micromolar
µl	Microliter
TMR	Transverse Mountain Ranges

CHAPTER 1

INTRODUCTION

About 44% of the planet's vascular plants and 35% of four vertebrate groups (birds, mammals, amphibians, and reptiles) occur in biodiversity hotspots (Myers et al. (2000). These hotspots are defined as areas with high endemic species concentrations and continuing habitat loss (Myers et al. 2000, Mittermeier et al. 2005). The California Floristic Province (CFP) is one of these 34 designated biodiversity hotspots (Mittermeier et al. 2005). About one third of California's native plants and animals are found in this region that encompasses almost 70% of California and parts of Oregon and Baja California (Mittermeier et al. 2005, Rissler et al. 2006, Chatzimanolis & Caterino 2007, Vandergast et al. 2007). The CFP originally comprised over 290,000 km^2 but only about 25% (~73,000 km^2) of undisturbed habitat remains (Mittermeier et al. 2005), primarily due to the increasing human population density of California (15 million in 1960; 37 million in 2010 [U. S. Census Bureau 1961, 2010, respectively]). The vast majority of coastal sage scrub habitats, coastal prairies, and vernal pools have been lost to the increasing population growth of southern coastal California (Vandergast et al. 2007). Increased urbanization has transformed the region by shifting land use for housing and recreation, reducing air and water quality, and increasing the frequency of fire (Mittermeier et al. 2005, Chatzimanolis & Caterino 2007). The resulting habitat loss threatens species and could lead to a severe decline or even extinction of these taxa (Brooks et al. 2002, Chatzimanolis & Caterino 2007). In southern California, anthropogenic barriers may lead to the formation of genetic reservoirs for many formerly widespread species (Vandergast et al. 2007). Organisms including carnivores, birds, and lizards demonstrate this loss of genetic diversity in the region (Riley et al. 2006, Delaney et al. 2010). However, as is true of most biodiversity hotspots, population-level research and phylogeographic studies on insects in the CFP (which is home to more than 30,000 species of insects) are few or lacking (Chatzimanolis & Caterino 2007).

The complicated topography of California along with plate tectonics of the Pacific and North American plates has impacted the evolution of the modern fauna in the region (Chatzimanolis & Caterino 2007). Uplift and formation of the main mountain ranges during the last 10 million years have caused vicariance events that influence the distribution and diversification of species in California (Wake 1997, Kuchta & Tan 2006, Chatzimanolis & Caterino 2007). The Sierra Nevada ranges were formed 10 million years ago (mya) to present, and the Coastal, Transverse, and Peninsular mountain ranges were formed about 5-3 mya to present. Geological studies suggest uplift of more than 1000 m (1000+ m for the Sierra Nevada, 2000+ m for the San Bernardino and San Jacinto Mountains in the Transverse Mountain Range) within the last 3 million years (Pliocene and Pleistocene) creating geomorphological events that have the potential to influence the evolution of species (Jacobs et al. 2004, Kuchta & Tan 2006). The Transverse Mountain Range (TMR) is of special interest for phylogenetic and geographic studies due to its unique west to east orientation, and many studies have identified it as a factor that can disrupt gene flow (Maldonado et al. 2001, Calsbeek et al. 2003, Forister et al. 2004, Chatzimanolis & Caterino 2007).

The *Paracotalpa ursina* (Horn 1867) species complex (the “bear scarabs”) is a morphologically highly variable group that is endemic to the California Floristic Province, the Mojave Desert, and the Basin and Range region. They are associated with the North American Deserts and the Mediterranean California ecoregions, and have been found on plants of the Asteraceae and Rosaceae families (Arnett 2000, pers. observations). The widespread shrub *Artemisia* (Asteraceae) has been reported as a host plant for the larvae and adult stages of

Paracotalpa (Ritcher 1948, pers. observations). The dense setosity of these beetles provides evidence that they are likely pollinators of their host plants. Members of this group have been described as nine separate species due to high intraspecific variation including coloration of the clypeus, head, and thorax either green or blue; reddish, brown, or castaneous coloration of the elytra; size, density, and distribution of punctures on the thorax and elytra, and; setae that are long or short, white or yellow, and sparse or dense. Several subspecies were recognized in the last revision of the genus (Saylor 1940), but lack of consensus has resulted in nomenclatural tumult (Table 1), and populations are referred to either as species, subspecies, "morphotypes", or "variations". Most recently, all names were synonymized under *P. ursina* (Smith 2009). Descriptions of all nine taxa were based exclusively on morphological, phenotypic characters (Casey 1915, Saylor 1940, Hardy 1971), all of which appear to be highly variable within the species. Hardy (1971) stated that further investigation utilizing geography was needed to disentangle the species complex due to the sympatry of several morphotypes and subtle differences in variation and melanism. In these problematic cases molecular data can potentially provide information to unravel these morphologically cryptic lineages (Rissler et al. 2006, Chatzimanolis & Caterino 2007) and clarify the composition of the *P. ursina* species complex.

TABLE 1. Classification hypotheses for *Paracotalpa* based on Saylor (1940), Machatschke (1972), Krajcik (2008), and Smith (2009). The *P. ursina* species complex is outlined in red.

Saylor (1940)	**Machatschke (1972)**
Paracotalpa Ohaus, 1915 -syn. *Pocalta* Casey, 1915 ***Paracotalpa deserta*** Saylor, 1940 ***Paracotalpa granicollis*** (Haldeman, 1852) (*Cotalpa*) -syn. *Pocalta pubicollis* Casey, 1915 ***Paracotalpa puncticollis*** (LeConte, 1863) (*Cotalpa*) ***Paracotalpa ursina ursina*** (Horn, 1867) (*Cotalpa*) -syn. *Pocalta brevis* Casey, 1915 -syn. *Pocalta laevicauda* Casey, 1915 -syn. *Cotalpa leonina* Fall, 1932 ***Paracotalpa u. piceola*** Saylor, 1940 (*Cotalpa*) ***Paracotalpa u. rotunda*** Casey, 1915 (Pocalta) -syn. *Pocalta seriata* Casey, 1915 ***Paracotalpa u. nigripennis*** Casey, 1915 (*Pocalta*) ***Paracotalpa u. rubripennis*** Casey, 1915 (Pocalta)	***Paracotalpa*** Ohaus, 1915 -syn. *Pocalta* Casey, 1915 ***Paracotalpa deserta*** Saylor, 1940 ***Paracotalpa granicollis*** (Haldeman, 1852) (*Cotalpa*) -syn. *Pocalta pubicollis* Casey, 1915 ***Paracotalpa laevicauda*** Casey 1915 (*Pocalta*) ***Paracotalpa puncticollis*** (LeConte, 1863) (*Cotalpa*) ***Paracotalpa ursina*** (Horn, 1867) (*Cotalpa*) -syn. *Pocalta brevis* Casey, 1915 -syn. *Pocalta seriata* Casey, 1915 -syn. *Cotalpa leonina* Fall, 1932 -ssp. *Paracotalpa piceola* Saylor, 1940 -ssp. *Pocalta rubipennis* Casey, 1915 -ssp. *Pocalta nigripennis* Casey, 1915 -ssp. *Pocalta rotunda* Casey, 1915
Krajcik (2008)	**Smith (2009)**
Paracotalpa Ohaus, 1915 -syn. *Pocalta* Casey, 1915 ***Paracotalpa deserta*** Saylor, 1940 ***Paracotalpa granicollis*** (Haldeman, 1852) (*Cotalpa*) -syn. *Pocalta pubicollis* Casey, 1915 ***Paracotalpa laevicauda*** Casey 1915 (*Pocalta*) ***Paracotalpa puncticollis*** (LeConte, 1863) (*Cotalpa*) ***Paracotalpa ursina ursina*** (Horn, 1867) (*Cotalpa*) -syn. *Pocalta brevis* Casey, 1915 -syn. *Cotalpa leonina* Fall, 1932 ***Paracotalpa u. nigripennis*** Casey, 1915 (*Pocalta*) ***Paracotalpa u. piceola*** Saylor, 1940 ***Paracotalpa u. rotunda*** Casey, 1915 (Pocalta) -syn. *Paracotalpa rotundata* Machatschke, 1972 -syn. *Pocalta seriata* Casey, 1915 ***Paracotalpa u. rubipennis*** Casey, 1915 (Pocalta)	***Paracotalpa*** Ohaus, 1915 -syn. *Pocalta* Casey, 1915 ***Paracotalpa deserta*** Saylor, 1940 ***Paracotalpa granicollis*** (Haldeman, 1852) (*Cotalpa*) -syn. *Pocalta pubicollis* Casey, 1915 ***Paracotalpa puncticollis*** (LeConte, 1863) (*Cotalpa*) ***Paracotalpa ursina*** (Horn, 1867) (*Cotalpa*) -syn. *Pocalta brevis* Casey, 1915 -syn. *Pocalta laevicauda* Casey, 1915 -syn. *Pocalta nigripennis* Casey, 1915 -syn. *Pocalta rotunda* Casey, 1915 -syn. *Pocalta rubripennis* Casey, 1915 -syn. *Pocalta seriata* Casey, 1915 -syn. *Cotalpa leonina* Fall, 1932 -syn. *Paracotalpa piceola* Saylor, 1940

All nine taxa (hereafter referred to as "morphotypes") of the highly variable *P. ursina* species complex occur in the CFP, making it a perfect model for understanding the biogeographical and evolutionary mechanisms that have influenced this variation. To test if the *P. ursina* species complex comprises one highly variable species or several distinct lineages, I used a combination of phylogenetics, genotyping, geographical distribution, and morphology to examine populations within the species complex. The aims of this paper are: (1) to disentangle the *P. ursina* species complex using conventional methods (phylogenetics, biogeography, and morphology) in combination with AFLP genotyping to test the feasibility and utility of this method for cryptic species delimitation, (2) to provide potential conservation priorities for the *P. ursina* species complex due to anthropogenic habitat loss, and (3) to examine the influence of vicariance events on dispersal and gene flow within the *P. ursina* species complex.

CHAPTER 2

MATERIALS AND METHODS

Specimen selection

Twelve populations of the *Paracotalpa ursina* species complex were sampled to address the hypothesis that the complex topographies, especially the Transverse Mountain Ranges (TMR) in California, play a role in the genetic discontinuities of populations in the *P. ursina* species complex. Populations were collected from across the range of the species complex to include as much morphological variation and geographical range as possible. In particular, sampling from both north and south of the TMR was necessary to test the impact of the TMR as a genetically isolating barrier. Specimens of *Paracotalpa puncticollis* LeConte, 1863 were used as an outgroup. The validity of species within the *P. ursina* species complex was tested using the phylogenetic species concept (Wheeler & Platnick 2000).

Specimen acquisition

Twelve populations of the *P. ursina* species complex were collected or obtained with the assistance of David Hawks (UC-Riverside). Collecting was conducted in southern California in regions with prior collections. Conventional collecting methods for scarabs such as hand-collecting from vegetation, sweep netting, and aerial netting (Jorgensen 1963, Hardy 1972, Ratcliffe & Paulsen 2008, Kaufman & Jameson 2009) were used. Adults were transferred to vials with 95% ethanol and stored at cool temperatures. Locality data, including GPS data and elevation, were recorded at each collection locality using a Magellan Triton 2000 Handheld GPS Receiver. Voucher specimens are deposited at the Wichita State University Insect Collection

(WICH) (Appendix 1). Additional specimens were obtained from 13 museums and 1 private collection (Table 2).

TABLE 2. Private and institutional loan providers (curators/collections managers in parentheses).

Museum Acronym	**Museum Name**
BYU	Monte L. Bean Life Science Museum, Provo, Utah (Shawn Clark)
CASC	California Academy of Sciences, San Francisco, California (Dave Kavanaugh)
EMEC	Essig Museum of Entomology, Berkeley, California (Peter Oboyski)
FMNH	Field Museum of Natural History, Chicago, Illinois (Margaret Thayer)
MLJC	Mary Liz Jameson Collection, Wichita, Kansas
LACM	Natural History Museum of Los Angeles County, Los Angeles, California (Brian Brown)
OSAC	Oregon State University, Corvallis, Oregon (Christopher Marshall)
SDMC	San Diego Natural History Museum, San Diego, California (Michael Wall)
SEMC	Snow Entomological Museum, Lawrence, Kansas (Zach Falin)
UCRC	Entomology Research Museum, Riverside, California (Doug Yanega)
UNSM	University of Nebraska State Museum, Lincoln, Nebraska (Brett Ratcliffe)
USNM	U.S. National Museum of Natural History, Washington, D.C. (Currently housed at the University of Nebraska State Museum for off-site enhancement)
WFBM	W.F. Barr Entomological Collection, Moscow, Idaho (James Johnson)
WSU	W.T. James Entomological Collection, Pullman, Washington (Richard Zack)

DNA extraction

Tissue for extraction was obtained from the right meso- and/or metathoracic legs and thoracic muscle tissue. DNA was extracted using a Qiagen DNeasy® Blood and Tissue Kit (Qiagen, Germantown, Maryland, USA) (Appendix 2) and with the following modifications to Qiagen's "Purification of total DNA from Insects" protocol. Tissues were pulverized into a fine powder using a modified reciprocating saw method (Alexander et al. 2007) until the tissue resulted in a fine powder. After incubation in step 3, 4 µl of RNase (100 ng/ml) were added, and the sample was mixed by vortexing and incubated at room temperature before continuation with step 4. One elution with 100 µl Buffer AE was used to increase DNA concentration, and step 9 of the protocol was omitted.

DNA sequence amplification and phylogenetic analysis

Mitochondrial gene regions were used because they provide information at the inter- and intraspecific level (Zhang & Hewitt 1996, Hwang & Kim 1999). Two regions were sequenced, the 5' end of the small ribosomal subunit 12S rRNA and the complete cytochrome oxidase subunit 1 (CO1) region. For both CO1 (approximately 1250 bp) and 12S (approximately 400 bp), amplification was attempted for the entire gene region using primer sets as described in Simon et al. (1994) and Kambhampati & Smith (1995) (Table 3). Thirty-three individuals were included in the analysis (Appendix 3).

TABLE 3. Primer sets for CO1 and 12S. Modified primers indicated by asterisk (*).

Gene Region	Name	Primer Sequence 5’ to 3’	Reference
CO1	C1-J-1751	GGATCACCTGATATAGCATTCCC	Simon et al. 1994
	TL2-N-3014	TCCAATGCACTAATCTGCCATATTA	Simon et al. 1994
12S	12S 2F	TACTATGTTWMGACTTATCC	Kambhampati & Smith 1995*
	SR-N-14594	AAACTAGGATTAGATACCC	Kambhampati & Smith 1995

Twenty µl PCRs included 2 µl of 10x Klentaq (DNA Polymerase Technology, Inc., St. Louis, MO) reaction buffer (1x final concentration), 2 µl dNTPs (0.25 µM final concentration), 2 µl of each primer (1 µM per primer final concentration), 2 µl of DNA template, 0.1 µl of Klentaq LA, and 10 µl of DI water. Cycling conditions followed suggested protocols in the Klentaq documentation: 2 minutes initial denaturation at 94°C, followed by 1) denaturation at 94°C for 40 seconds, 2) annealing at variable temperatures and times (Table 4), and 3) elongation at 68°C for variable times (Table 4). Steps 1-3 were repeated for 35 cycles. The CO1 amplification program also included a final 2 minute elongation period at 68°C.

TABLE 4. Annealing temperatures and times, and elongation times for primer sets.

Gene region	Primer combination	Annealing temperature (°C)	Annealing times (Minutes:Seconds)	Elongation time at 68°C (Minutes:Seconds)
CO1	C1-J-1751/ TL2-N-3014	55	3:10	2:00
12S	12S 2F/SR-N-14594	45	0:40	1:30

All amplicons were sequenced with both forward and reverse primers at the Comprehensive Cancer Center at the University of Chicago. CLC Genomics Workbench v6.9 (CLC Inc, Aarhus, Denmark) was used to assemble forward and reverse chromatogram traces and construct consensus sequences. Sequences for each gene region were then aligned by eye using SeaView version 4.4.0 (Gouy et al. 2010) (Appendix 3).

The combined mitochondrial data (CO1 and 12S) were analyzed with both maximum parsimony (PAUP 4.0; Swofford 2002), and Bayesian (Mr. Bayes 3.2; Ronquist & Huelsenbeck 2003) optimality criteria. A heuristic maximum parsimony analysis was performed, and support calculated with 1,000 bootstrap replicates, 15,000 trees saved per replicate, and 100 random addition replicates. Clades were considered well supported if bootstrap support was higher than 80% (Baum & Smith 2013). Bayesian analysis assumed a GTR + I + G model of sequence evolution, and comprised two independent runs with 1,000,000 generations each, and a tree sampling every 1,000 generations. Convergence between runs was assessed using Excel 2010 (Microsoft) graphing likelihood (LnL) versus time (Gen) X-Y scatter plots. Both runs reached stationarity within the first 100,000 generations. The first 100 trees of each run were therefore discarded and a 50% majority-rule consensus tree was created using the remaining 901 trees. Clades with Bayesian posterior probabilities of 0.95 or higher are considered well supported (Baum & Smith 2013).

AFLP genotyping and analysis

Amplified Fragment Length Polymorphism data was obtained from multiple (between 2-4) individuals from each sampled population (Table 5). In order to assess the genotyping error rate, four randomly chosen samples (OK_P01_05, OK_P08_01, OK_P10_06, and OK_P12_01) were set up in duplicate for the PCR reactions. This number of replicates (4/32) was in the range suggested by Crawford et al. (2012). Restriction ligation was conducted at 37°C for 12 hours and incorporated the *Eco*RI and *Mse*I adaptors as noted in Vos et al. (1995).

TABLE 5. Subset of specimens used for AFLP genotyping. Duplicate specimens marked with an asterisk (*).

Population	**Voucher**	**Voucher**	**Voucher**	**Voucher**
1	OK_P01_05	OK_P01_05B*	OK_P01_06	
2	OK_P02_01	OK_P02_04	OK_P02_05	
3	OK_P03_01	OK_P03_04	OK_P03_05	
4	OK_P04_03	OK_P04_05		
5	OK_P05_01	OK_P05_03	OK_P05_05	
6	OK_P06_01	OK_P06_03	OK_P06_05	
7	OK_P07_05	OK_P07_06		
8	OK_P08_01	OK_P08_01B*	OK_P08_03	OK_P08_04
9	OK_P09_03	OK_P09_04	OK_P09_05	
10	OK_P10_06	OK_P10_06B*	OK_P10_07	
11	OK_P11_03	OK_P11_04	OK_P11_05	
12	OK_P12_01	OK_P12_01B*	OK_P12_03	OK_P12_04

Pre-amplification cycling conditions were those noted in Beck et al (2012). AFLP pre-amplification reactions were set up using the following components: 1µl GoTaq buffer (Promega), 0.6 µl $MgCl_2$ (25 mM), 1 µl dNTPs (2.5 mM each), 0.5 µl primer MseI+A (10 µM), 0.5 µl primer EcoRI+A (10 µM), 3.65 µl DI water, 0.25 µl GoTaq (5 U/µl) (Promega), and 2.5 µl diluted (1:10) restriction/ligation product for a total volume of 10 µl per reaction. The following thermocycler protocol was used: 1) 72°C for 2 minutes, 2) 94°C for 30 seconds, 3)

56°C for 30 seconds, 4) 72°C for 2 minutes, and 5) 72°C for 5 minutes. Steps 2-4 were repeated for 30 cycles. Pre-amplification products were then stored at -4°C.

Three separate selective amplifications were conducted, incorporating one of following labeled *Eco*RI primers (*Eco*RI+ACG [6-FAM], *Eco*RI+AGA [6-FAM] or *Eco*RI+ATG [HEX]). Selective amplification cycling conditions were as described in Beck et al. (2012). Selective amplifications were set up using the following components: 2.6 µl DI water, 0.7 µl primer *Mse*I+ATC (10 µM), 0.7 µl primer *Eco*RI+AXX, 5 µl 2× Multiplex PCR Master Mix (Qiagen), and 1µl diluted (1:20) pre-amplification product for a total volume of 10 µl per reaction. The following thermocycler protocol was used: 1) 94°C for 15 minutes, 2) 94°C for 20 seconds, 3) 66°C for 30 seconds (-1°C per cycle), 4) 72°C for 2 minutes, 5) 94°C for 20 seconds, 6) 56°C for 30 seconds, 7) 72°C for 2 minutes, and 8) 60°C for 30 minutes. Steps 2-4 were repeated for 10 cycles followed by 20 cycles of steps 5-7. Selective amplification products were diluted 1:10, and genotyping was performed at the Comprehensive Cancer Center at the University of Chicago using the LIZ 500 size standard.

GeneMarker 1.9 (SoftGenetics, State College, Pennsylvania, USA) was used to determine initial AFLP presence-absence matrices for all three *Mse*I/*Eco*RI primer combinations. Only loci between 50 and 500 bp were called with a standard 1 bp bin width. Additional settings included smoothing, stutter-peak filter disabled, a 50 rfu peak height threshold, and a peak score setting of fail<1check<1pass as suggested in Holland et al. (2008). The three locus matrices were combined into a total allele matrix comprising 504 total loci. Error rates were calculated by dividing the total number of differences between a pair of replicates by the total number of loci (Bonin et al. 2004). Replicates were then excluded from further analysis. The 504 locus matrix was used to construct a pairwise genetic distance matrix in GenAlEx 6.501 (Peakall & Smouse

2012), and this matrix was subjected to a principal coordinates analysis (PCoA). The 504 locus matrix was also analyzed with modal clustering (PCO-MC) approach, as described in Reeves & Richards (2009; 2011). Briefly, PCO-MC identifies the most densely clustered groups in multidimensional space and assigns a “stability value” to each.

Distribution analysis

GPS data for populations 11 and 12 were obtained during the collecting events and recorded in Decimal Degree (DD) format. No GPS data were available for populations 1-10 and therefore had to be generated using Google Earth (Google Earth 2013). GPS data were generated with three decimals to account for a degree of uncertainty. Two different types of locality data were available: distances and directions from waypoints to localities, and exact localities. Missing GPS data were geo-referenced as described by Beza-Beza (2012) using standard methods.

Climatic data, including 19 climatic variables (Table 6), with a resolution of 2.5 arc-minutes (~ 5 x 5 km resolution) were obtained through the DIVA-GIS BIOCLIM specific package downloaded from the DIVA-GIS web site (Hijmans et al. 2005). Climatic data were restricted to include only southern California and the northern part of Baja California in Mexico for the distribution analyses. The subset was restricted to minimum X: W 124.583°, maximum X: 114.041°, minimum Y: N 29.875°, and maximum Y: N 42.041°.

Potential distribution models for three geographical groups, as identified by the molecular analyses, were generated using the Maximum Entropy Species Distribution Model (Philips et al. 2004) algorithm. MAXENT version 3.3.3k (http://www.cs.princeton.edu/~schapire/MAXENT/)

with default settings was used for the prediction of species distributions based on collection data points (Philips et al. 2004). Predicted species distributions are based on the ecological niche concept where the environment a species can occupy is limited by environmental conditions and negative interactions and expanded by positive interactions (Hutchinson 1957; Puliam 2000; Scheldeman & van Zonnefeld 2010).

TABLE 6. BIOCLIM climatic variables used to generate potential distribution with the Maximum Entropy Species Distribution Model (Philips et al. 2004) algorithm.

Name	**Variable**	**Calculation**
BIO1	Annual mean temperature	
BIO2	Mean diurnal range	Mean of monthly (max temp – min temp)
BIO3	Isothermality	BIO2/BIO7 * 100
BIO4	Temperature Seasonality	Standard deviation * 100
BIO5	Max temperature of warmest month	
BIO6	Min temperature of coldest month	
BIO7	Temperature annual range	BIO5 – BIO6
BIO8	Mean temperature of wettest quarter	
BIO9	Mean temperature of warmest quarter	
BIO10	Mean temperature of warmest quarter	
BIO11	Mean temperature of coldest quarter	
BIO12	Annual precipitation	
BIO13	Precipitation of wettest month	
BIO14	Precipitation of driest month	

TABLE 6 (continued)

Name	**Variable**	**Calculation**
BIO15	Precipitation of seasonality	Coefficient of variation
BIO16	Precipitation of wettest quarter	
BIO17	Precipitation of driest quarter	
BIO18	Precipitation of warmest quarter	
BIO19	Precipitation of coldest quarter	

Maps comparing the predicted distributions with a current land use cover for California (Homer et al. 2012) for all three groups were generated in ArcMap 10.1 (ESRI 2011). The following map layers were used: World Shaded Relief (ESRI 2013), USA States (ESRI 2013), USA Counties (ESRI 2013), and California Land Cover Data (Homer et al. 2012). The shape file used for the projection of occurrence points (realized distribution) was generated in Diva-GIS, and the GCH_WGS_1984 coordinate system was used to project both the potential distribution and occurrence points. Biogeographic information for occurrence points for each population was extracted from the following spatial data layers: USDA Level I and III ecogeographic regions (CEC 1997) and the California National Land Cover Dataset (Homer et al. 2012).

Morphology data

Morphological characters within the *P. ursina* species complex are highly variable and do not provide a consistent means of distinguishing between morphotypes and identification of morphotypes. Diagnoses have relied primarily on coloration (elytra, pronotum, head), setae (length, coloration, density), and punctation of the pronotum (size, density, distribution) (Horn

1867, Casey 1915, Fall 1932, Saylor 1940). Form of the male parameres, which is often used in scarab systematics, is not useful for identification because it is constant for all morphotypes (Saylor 1940, pers. observation). Results of the molecular and distribution analyses were used to inform my examination of morphological characters, specifically examining populations from north and south of the TMR.

To link populations with described species, I examined type specimens and original descriptions (Horn 1867, Casey 1915, Fall 1932, Saylor 1940). Type specimens for all species described by Casey (*Pocalta laevicauda, Pocalta brevis, Pocalta rotunda, Pocalta seriata, Pocalta rubripennis, and Pocalta nigripennis*) and Saylor (*Paracotalpa piceola*) were available for examination. The type specimen of *P. ursina* (Horn) was observed online (Harvard Museum or Comparative Zoology Insect Type Database; http://insects.oeb.harvard.edu/mcz/Species_record.php?id=7718). The type specimen of *P. leonina* (Fall) was not available for study.

Leica M80 microscopes and LED light sources were used to observe morphological characters, and images were taken with a Leica IC80 HD digital camera. Images were processed using the Leica Application Suite version 4.0.1, and modified using Photoshop CS 4 v. 11.0.2.

Taxonomy

Names for the northern and southern group were assigned based on results of the phylogenetic, AFLP, and biogeographic analyses and the subsequent discovery of the morphological character that distinguishes the two groups.

CHAPTER 3

RESULTS

Phylogenetic analysis

Analysis of the combined mitochondrial CO1 and 12S matrix (1706 aligned characters) produced 310 (18.2%) parsimony-informative and 442 (25.9%) variable characters. Each of the 1000 random addition replicates recovered the same most parsimonious tree (length = 731, CI = 0.724, RI = 0.883). This tree, along with bootstrap support (>80) (BS) and Bayesian posterior probabilities (>.95) (PP), is shown in Fig. 1. The three specimens of *P. puncticollis,* which were used as the outgroup, had strong (1 PP, 100 BS) Bayesian and bootstrap support. Two major clades correspond to populations north (1, 4, 5, 6, 7, 9, 10, and 11) and south (2, 3, 8, and 12) of the TMR were recovered within the *P. ursina* clade. Both exhibited maximum support (1 PP, 100 BS). Within the southern clade, two strongly supported clades (population 12; populations 2, 3, and 8 were also recovered with maximum support (1 PP, 100 BS).

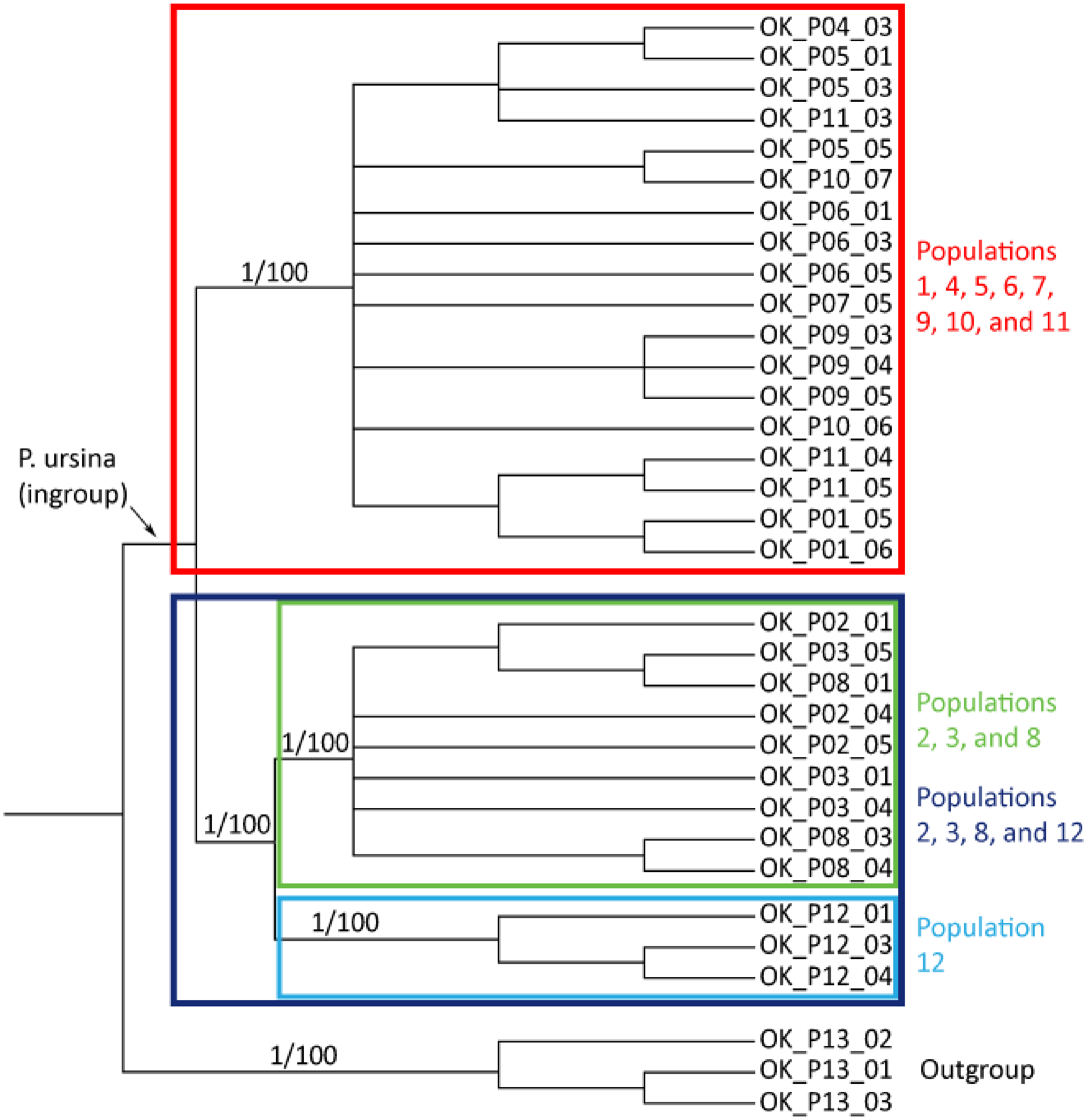


Fig. 1. The single most parsimonious tree inferred from the analysis of the combined CO1 and 12S mitochondrial data set using 30 specimens of *P. ursina* (ingroup) and 3 specimens of *P. puncticollis* (outgroup). Support values over 0.95 (Bayesian posterior probability) and over 80% (parsimony bootstrap) are shown on nodes. Clades outlined in color are discussed in the text.

AFLP genotyping

The three individual locus matrices included 504 total loci (*Eco*RI+ACG = 152 loci; *Eco*RI+AGA = 171; *Eco*RI+ATG = 181). Individual error percentages for all four duplicate samples were calculated for each individual AFLP matrix: *Eco*RI+ACG (Population 1: 19.1%; Population 8: 11.2%; Population 10: 8.6%; Population 12: 9.9%); *Eco*RI+AGA (26.9%; 6.4%; 7.6%; 8.2%); and *Eco*RI+ATG (26.5%; 6.6%; 14.9%; 12.2%); and the total allele matrix (24.4%; 7.9%; 10.5%; 10.1%). The overall error rate was 13.2%. Four groups with a stability value of >10% were recovered with PCO-MC, all of which were easily distinguishable on a plot of PCoA axes 1 and 2 (Fig. 2).

Northern populations (1, 4, 5, 6, 7, 9, 10, and 11) formed a coherent group (stability value 70.2%), corroborating results of the phylogenetic analysis (Fig. 1). The remaining four populations all south of the TMR also form a coherent group (55.3%). Results also indicate that population 12 is distinct (38.3%) from the remaining three southern populations (27.7%) as suggested by the phylogenetic analysis (Fig. 1). The two most distinct groupings in the data set are the northern (red) and the southern populations (blue) with respect to the TMR (Fig. 2). The higher the assigned stability value the more tightly clustered the groups is in multidimensional space.

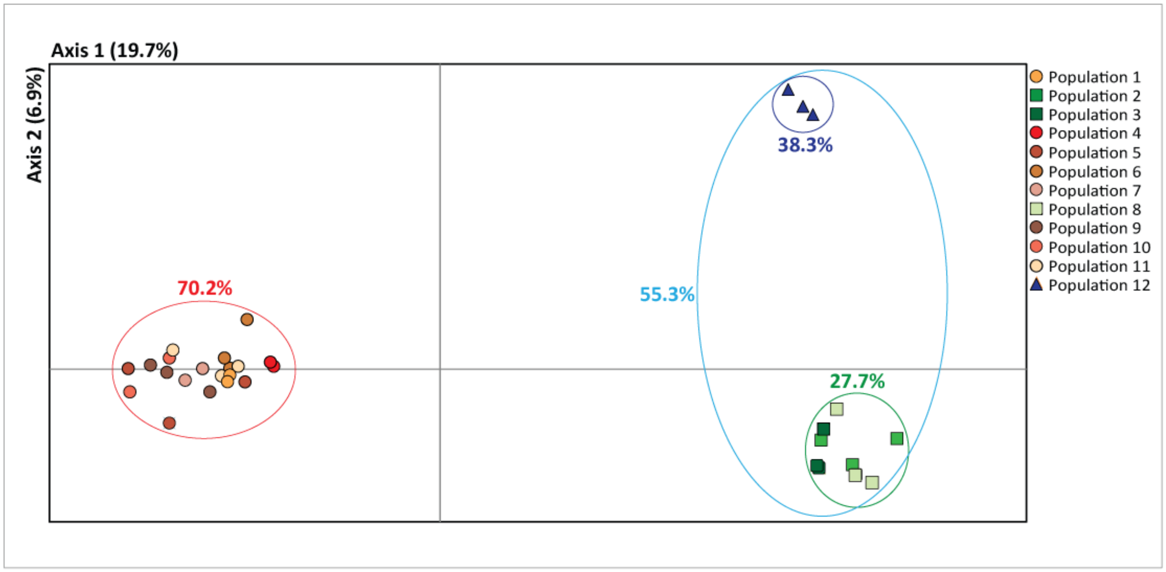


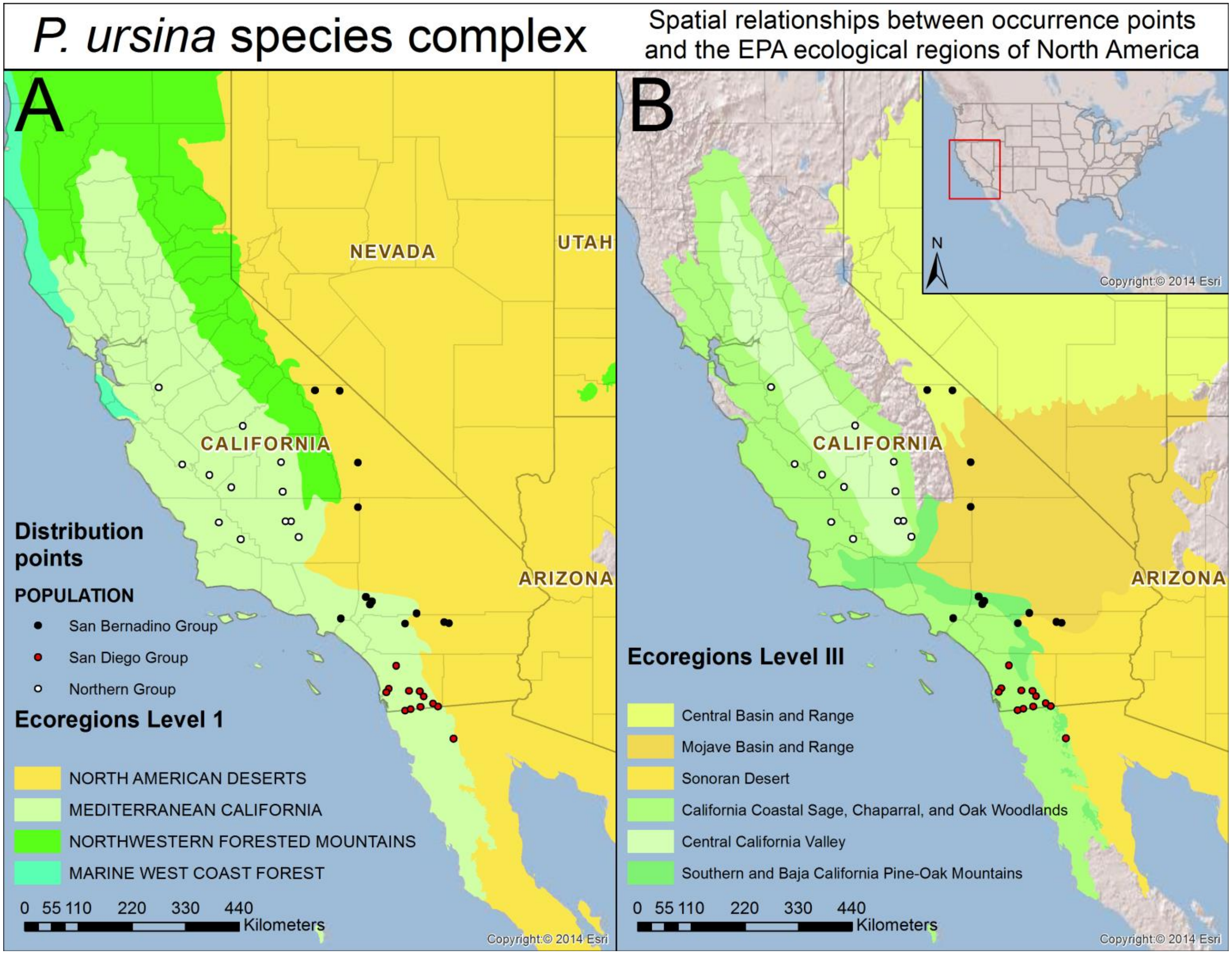
Fig. 2. Plot of the scores of 32 *P. ursina* individuals resulting from a principal coordinates analysis (PCoA) of the AFLP data. Colors of filled circles (see legend) indicate population membership. Colored outlines denote the four most stable groups identified by Principal Coordinates with Modal Clustering (PCO-MC), all of which are annotated with their stability values.

Distribution analysis

Based on results of the phylogenetic and AFLP analyses, localities for 36 specimens including museum material within the *P. ursina* species complex were geo-referenced: 12 for a northern group (populations 1, 4, 5, 6, 7, 9, 10, and 11; north of the TMR), 12 for a San Bernardino group (populations 2, 3, and 8; south of the TMR), and 12 for a San Diego group (population 12; south of the TMR). All localities were either in California, USA or Baja California, Mexico. Specimens of the northern group were geo-referenced from the following counties: Fresno, Kern, Kings, Monterey, San Luis Obispo, Stanislaus, and Tulare. Specimens from the San Bernardino group were geo-referenced from the following counties: Inyo, Kern, Los Angeles, Riverside, and San Bernardino. For the San Diego group, specimens were geo-referenced from California (San Diego County) and Mexico (Tecate and Tijuana Municipalities).

All 36 specimens were distributed within two Level I Ecoregions of North America (Plate 1A). All northern and San Diego group specimens occurred within the Mediterranean California ecoregion, while San Bernardino group specimens occurred in the North American Deserts and Mediterranean California ecoregions. Based on the Level III Ecoregions of North America (Plate 1B), the northern group specimens are found in the California Coastal Sage, Chaparral, and Oak Woodlands and Central California Valley ecoregions. The San Bernardino group is found in the Central Basin and Range, Mojave Basin and Range, California Coastal Sage, Chaparral, and Oak Woodlands, and Southern and Baja California Pine-Oak Mountains ecoregions. The San Diego group specimens are only found in the California Coastal Sage, Chaparral, and Oak Woodlands ecoregion.

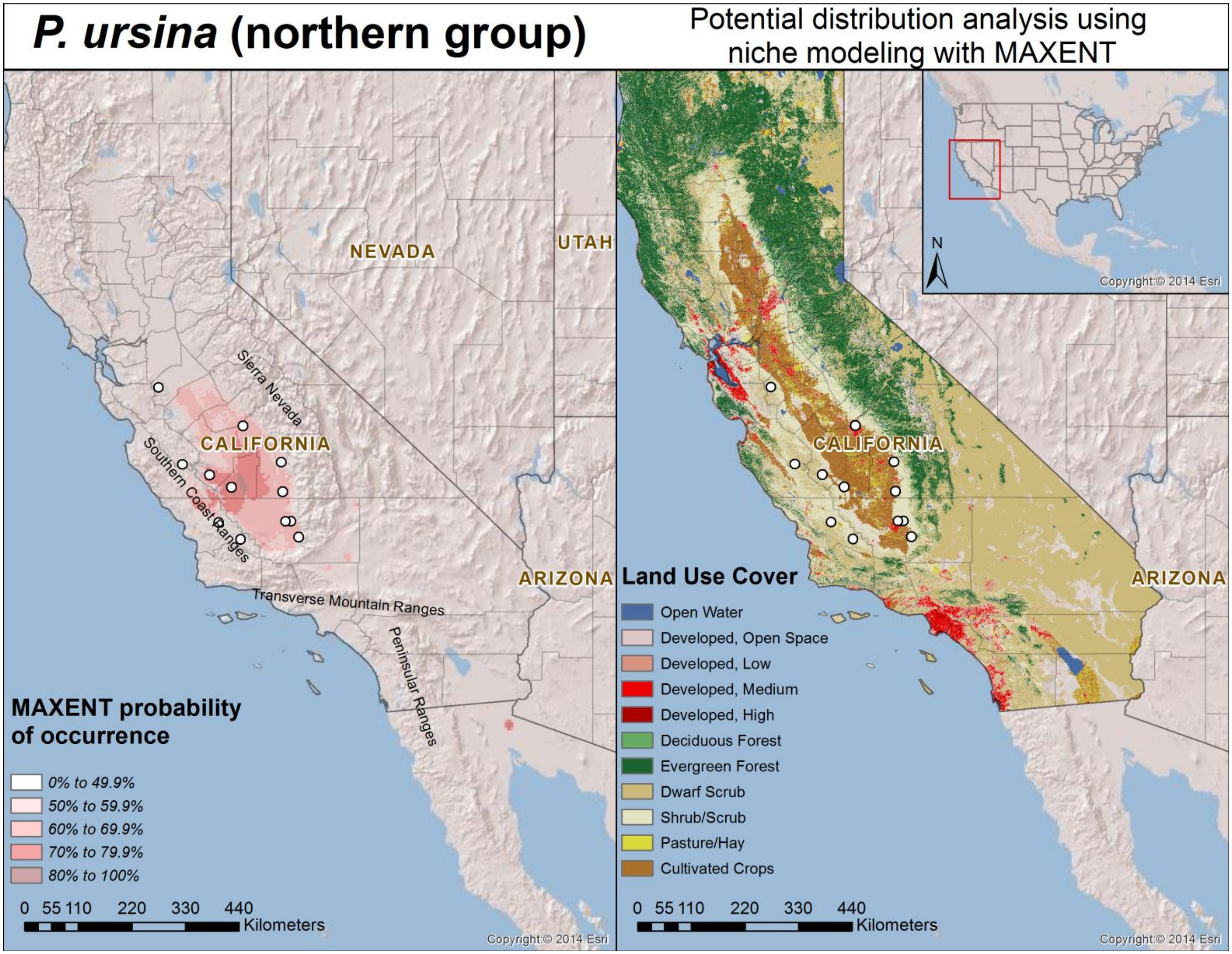
Plate 1. EPA ecological regions of North America and occurrence points and the spatial relationship between them.



Paracotalpa groups and their distributions

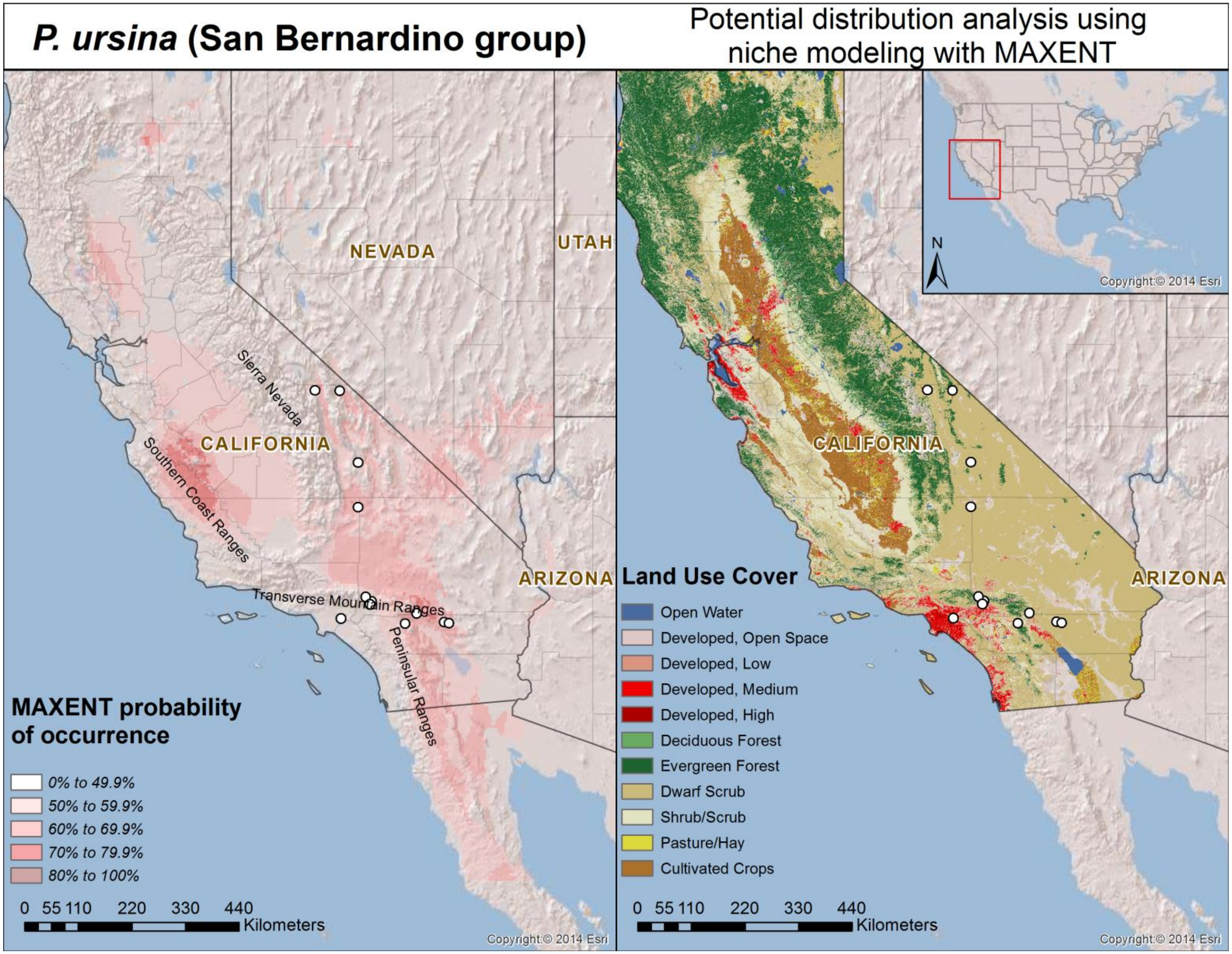
Paracotalpa ursina northern group: Specimens belonging to this group are distributed north of the TMR and within the Central Valley of California. They are limited to the area in between the Sierra Nevada Mountains (east), the Southern Coast Ranges (west), and the TMR (south). Their distribution islimited to the foothills of the mountains. Mean temperature for their habitat is 16.7°C, and the annual precipitation is about 306 mm. According to MAXENT, suitable habitat (probability >60%) is located in the middle of the Central Valley (Plate 2), but due to the agricultural usage of the area, the realized habitat is on the periphery of the predicted range. Predicted habitats occur within the Mojave Basin and the Sonoran Desert, but these are not contiguous with the main distribution (thus accounting for lack of specimen occurrence). Both the Sierra Nevada and the TMR could act as dispersal barriers that isolate this group to the north of the TMR. The realized distribution (as shown by occurrence points and the main predicted distribution) coincides with the predicted distribution, thus limiting dispersal of the northern group to the Central Valley.

Plate 2. *P. ursina* (northern group) potential distribution analysis using niche modeling with MAXENT.



Paracotalpa ursina San Bernardino group: Specimens in the San Bernardino group are distributed within five counties to the south of the TMR: Inyo, Kern, Los Angeles, Riverside, and San Bernardino counties. Mean temperature for their habitat is 15.5°C, and the annual precipitation is about 328 mm. All specimens occur either south of the TMR or east of the Sierra Nevada (Plate 3). According to MAXENT, suitable habitat (probability >60%) is present throughout most of southern California, the Southern Coast Ranges, and the Central Valley region, Nevada and Arizona, and Baja California in Mexico. Suitable habitat is mainly restricted to south and east of the TMR with small patches in the Mojave Basin and areas east of the Southern Coast Ranges. The most suitable habitat (>80%) is split into two regions: the Southern Coast Ranges and the San Bernardino Mountains. However, no specimen occurrences are found within the Southern Coast Ranges region. Occurrence points in this group suggest biogeographic barriers to the north (TMR), to the south (Peninsular Range), and to the west (Sierra Nevada), as well as anthropogenic barriers (in Riverside and San Bernardino counties) that limit westward dispersal.

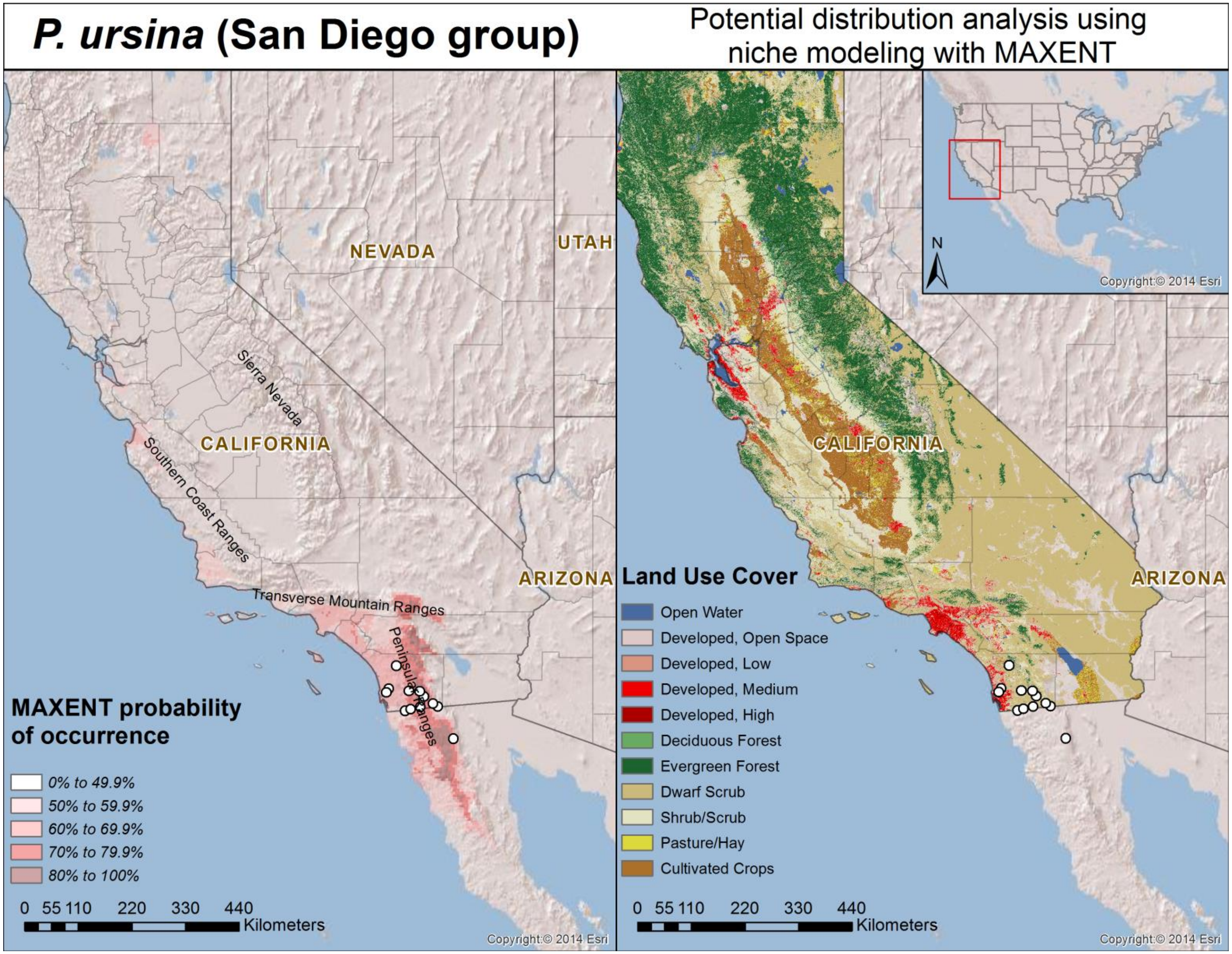
Plate 3. *P. ursina* (San Bernardino group) potential distribution analysis using niche modeling with MAXENT.



Paracotalpa ursina San Diego group

San Diego group specimens were only found in San Diego County and Baja California in Mexico. Mean temperature for their habitat is 15.0°C, and the annual precipitation is about 407 mm. Most (10 of 12) occurrence points were found within or on the periphery of the Peninsular Ranges, however two specimens were collected within small, undeveloped locales within the greater San Diego area. Most occurrence points coincided with patches of suitable habitat (>70% probability; Plate 4). Areas with the highest predicted probability of occurrence according to MAXENT are across the range of the Peninsular Ranges with a small patch alongside the east boundary of the TMR. Patches with lower probability (<60%) of occurrence are distributed along the Pacific coast as far north as the San Francisco Bay and one at the northernmost edge of the Central Valley. The realized range, according to occurrence data, is primarily restricted to the north by the highly populated southern California coastal region. Previous suitable habitat may have been eliminated due to human population growth, especially along the California coast.

Plate 4. *P. ursina* (San Diego group) potential distribution analysis using niche modeling with MAXENT.



Morphology data

Observation of all 12 populations used for the molecular analyses and additional museum specimens revealed one easily recognizable morphological character that allows separation and identification of northern and southern populations. All males from the northern populations possess a median projection on the inner median surface of the fifth protarsomere (Fig. 3A). The median projection is also present on the fifth meso- and metatarsomere of males, and it is present (but reduced) in females of the northern populations. Males and females from the southern populations lack the median projection (Fig. 3B) on all tarsomeres.

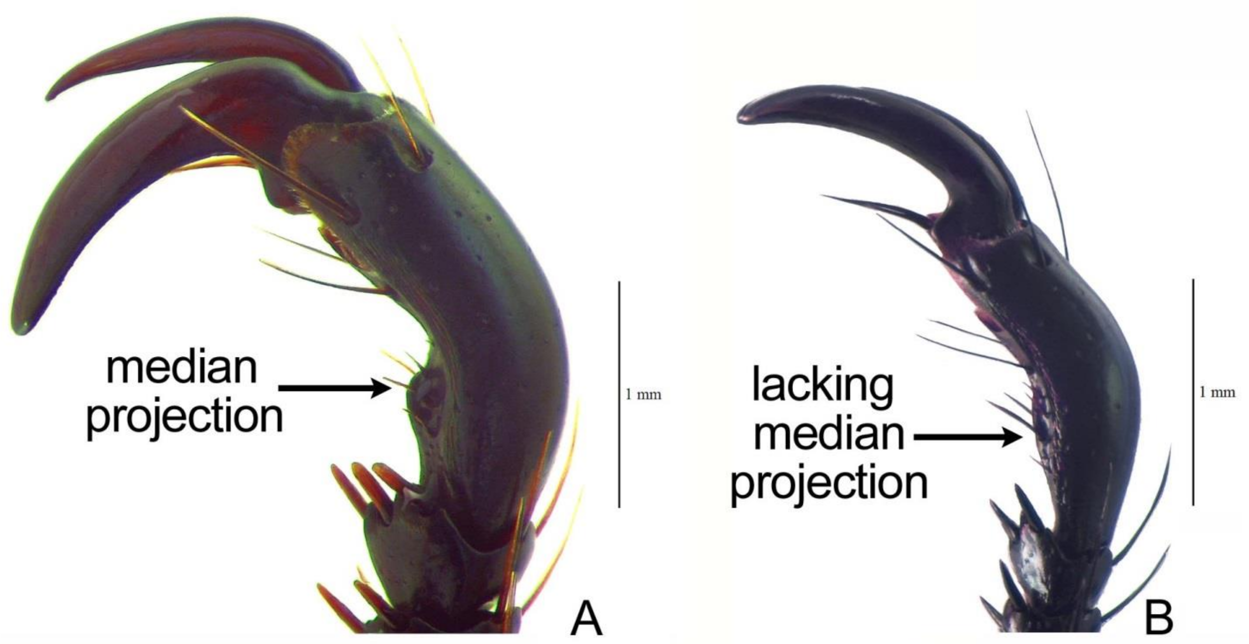


Fig. 3. Lateral view of protarsomeres 4 and 5 showing median projection on the inner median surface of fifth protarsomere of males present (A) or lacking (B).

Type specimens were examined for presence or absence of this morphological character. Examination of type material revealed that *P. ursina* possessed the median projection (= northern group). According to article 23.1 of the ICZN (1999) this name has nomenclatural priority over the following names that also possess this trait and that I consider synonyms of *P. ursina*: *P. seriata* (Casey, 1915:96) and *P. ursina piceola* Saylor, 1940:197. Examination of type specimens revealed that *P. rotunda* lacks of the median projection (=southern group). I elected to use this species name because type specimens included more precise locality data and the type series included four specimens (rather than one or two). Based on the absence of the median projection, I consider the following to be synonyms of *P. rotunda*: *Paracotalpa brevis* (Casey, 1915:95), *P. laevicauda* (Casey, 1915:94–95), *P. rubripennis* (Casey, 1915:96–97), *P. nigripennis* (Casey, 1915:97), and *P. leonina* (Fall, 1932:203–204). Although AFLP, molecular, and distribution analyses provide evidence for the San Bernardino group and the San Diego group, I did not discover morphological characters that allow confident separation of these two groups. I noted that these groups can be clearly distinguished by pronotal coloration: the San Bernardino group possesses a bluish pronotum whereas the San Diego group possesses a green pronotum. This character alone, however, does not confidently allow these groups to be assigned to a taxon (species or subspecies) name.

CHAPTER 4

DISCUSSION

Phylogenetic and population genetics analyses

Both genetic analyses strongly supported a northern group (populations 1, 4, 5, 6, 7, 9, 10, and 11) and a southern group (populations 2, 3, 8, and 12) (Figs. 1; 2). All populations of the northern group were isolated from the southern populations by the Southern Coast Ranges, the TMR, and the Sierra Nevada. Northern populations were distributed within the Central Valley, a region where dispersal and gene flow seems to be limited by the biogeographic barriers of the Southern Coast Ranges (west), the central Transverse Ranges and Tehachapi Mountains (TMR; south), and of the Sierra Nevada (east). Southern populations were distributed in areas south of the TMR, the Mojave Desert, or east of the Sierra Nevada. Their distribution also reaches into the northern half of Baja California.

Within the southern group, both analyses provided strong support for two groups. These results were corroborated based on distribution analyses, showing strong support for the San Bernardino group (populations 2, 3, and 8) and the San Diego group (population 12) (Figs. 1; 2). The San Bernardino group (three populations) was found in the area between the San Gabriel and San Bernardino mountain ranges, while the San Diego group (population 12) was found in San Diego County. Separation of these groups could be attributed to the Peninsular Ranges that form a biogeographic barrier in combination with the ongoing regional human population growth that forms an anthropogenic barrier thus inhibiting gene flow.

Distribution analysis

The potential distributions (predicted and realized ranges) for populations of *P. ursina* showed results quite similar to the genetic analyses. For the northern group, distribution modeling predicted high occurrence probabilities within the Central Valley (Plate 2). Small patches with high predicted probabilities also occur in the Mojave Desert and northern Mexico, however these are not contiguous with the habitat within the Central Valley. The twelve distribution data points for populations used to conduct the potential distribution modeling were all selected based on the presence of the median projection. The comparison of the predicted distribution model with a land use cover map of California (Plate 2) shows that all the data points are on the periphery of the predicted distribution. The central core of the predicted distribution coincides with a region of intense agriculture and high population density within the central Valley. Occurrence points from the past nine decades are distributed in the foothill regions of the Southern Coast Ranges (west), the central Transverse Ranges and Tehachapi Mountains (TMR; south), and of the Sierra Nevada (east). These populations correspond to the following ecoregions: California Coastal Sage, Chaparral, and Oak Woodlands, and Central California Valley (Map Plate 1). Much of the Central California Valley ecoregion is currently used for agriculture and urbanization. Almost 47% (27,114 km^2 of 58,000 km^2) of the area is used for agriculture, and the human population is predicted to grow from 4 million in 1992 to over 12 million in 2040 (American Farmland Trust 1995). Additionally, the majority of the plant communities are endangered: coastal sage scrub (90% eliminated) and coastal prairie (99% eliminated) habitats have been lost to urbanization in coastal southern California (McCaull 1994, Mattoni & Longcore 1997, Vandergast et al. 2007). Host plant associations for *P. ursina* include *Artemisia* spp. which is widespread within the California Coastal Sage, Chaparral, and Oak

Woodlands, and Central California Valley ecoregions. Larvae have also been found feeding on the roots of *Artemisia* (Ritcher 1948).

For the San Bernardino group, high occurrence probabilities were calculated for the Mojave Desert region north into the most eastern parts of the Sierra Nevada, south throughout the Peninsular Ranges and into the Baja California region, and in the western Central Valley and the Southern Coast Ranges (Plate 3). Predicted habitats in the Central Valley and Southern Coast Ranges are isolated from the other habitats by the Sierra Nevada and the TMR. Also, despite high probabilities throughout the Peninsular Ranges, the mountain range might act as a barrier limiting southward dispersal. Another southern barrier limiting dispersal might be the greater Los Angeles/Ontario/San Bernardino/Riverside Metropolitan area with a population of more than 17 million (U.S. Census Bureau 2010). Distribution data points were primarily from along the southern side of the eastern TMR and from the Mojave Desert. All but one population (from Los Angeles County) were also collected from elevations above 700m. This coincides with the low probability of occurrence alongside the coastal areas of California due to lower elevation. This group is also found in four ecoregions: Central Basin and Range, Mojave Basin and Range, California Coastal habitats (sage, chaparral, and oak woodlands), and Southern and Baja California Pine-Oak Mountains (Plate 1). This demonstrates a widespread distribution for the San Bernardino group and may indicate a wide range of host plants.

Probabilities of occurrence for the San Diego group were high in the eastern TMR, throughout the Peninsular Ranges, and ranging into Baja California. Coastal habitats are also predicted for this group ranging from the Los Angeles Basin southward into the northern third of Baja California in Mexico. Several disconnected smaller patches are also found north of Los Angeles along the coast. Distribution data points were mainly from the Peninsular Ranges, and all specimens were found in San Diego County (USA) or Tecate and Tijuana Municipalities (Mexico). Habitats throughout these coastal areas are being fragmented or destroyed by increasing urbanization (Riley et al. 2003). Urbanization in this area may also limit dispersal to the north. The San Diego group is found only in the California Coastal Sage, Chaparral, and Oak Woodlands ecoregions (Plate 1) and, therefore, much of its preferred habitat is endangered (McCaull 1994, Mattoni & Longcore 1997, Vandergast et al. 2007).

Morphology

The median projection on the fifth tarsomeres is a previously unobserved morphological character that can be used to distinguish between the northern and southern groups of the *P. ursina* species complex (Fig. 3). Characters previously used (e.g., coloration, setae, punctation) (Horn 1867, Casey 1915, Fall 1932, Saylor 1940) are quite variable and do not aid in distinguishing morphotypes. No new morphological character was found to distinguish between the San Bernardino group and the San Diego group. However, these groups can be distinguished by coloration of the pronotum, scutellum, and pygidium: the San Bernardino group is bluish whereas the San Diego group is green. Color characters, however, can be very subjective and are often criticized (Hardy 1971, Mawdsley 2001). By examining new character suites, it is possible that another distinguishing character may be discovered. Within this study, however, only the northern and southern groups are differentiated by external morphology.

Taxonomy

Based on molecular and biogeographic analyses, in combination with the morphological character, I propose that the *P. ursina* species complex includes two valid species: *P. ursina* (Horn, 1867; the northern group) and *P. rotunda* (Casey, 1915; the southern group). Despite the evidence from molecular and distribution analyses for the San Bernardino group and San Diego group, I decided to conservatively consider these morphotypes as *P. rotunda*. Additional analyses with more specimens and examination of new suites of characters may show that this morphotype (which can be associated with the type specimens of *P. rubripennis* and *P. nigripennis* based on the green pronotal color) is a valid species as previously suggested by Fall (1932) and Hardy (1971).

Key to the species of *Paracotalpa*

1.	Frontoclypeal suture weakly impressed. Clypeal apex quadrate. Pronotum with anterior margin distinctly bisinuate.	*Paracotalpa deserta* Saylor 1940
1’.	Frontoclypeal suture strongly impressed. Clypeal apex broadly rounded. Pronotum with anterior margin not bisinuate.	2
2.	Mentum deeply emarginate at apex, emargination surpassing insertion of labial palps.	3
2’.	Mentum not deeply emarginated at apex, emargination not surpassing insertion of labial palps.	4
3.	Pronotum densely rugopunctate.	*Paracotalpa granicollis* (Haldeman, 1852)
3’.	Pronotum sparsely to densely punctate (not rugopunctate).	*Paracotalpa puncticollis* (LeConte, 1863)

4. Fifth tarsomeres (all legs, male and female) with median projection on the inner median surface (Fig. 3A) — *Paracotalpa ursina* (Horn, 1867)

4'. Fifth tarsomeres (all legs, male and female) lacking median projection on the inner median surface (Fig. 3B) — Paracotalpa rotunda (Casey, 1915)

Conservation implications

High human population growth in southern California, the increasing prevalence of agriculture in the Central Valley, and the precipitous decline of unfragmented habitat within the California Floristic Province provide the impetus for increased conservation efforts in this biodiversity hotspot. Increased fragmentation due to anthropogenic activity leads to isolated populations and the loss of genetic diversity (Vandergast et al. 2007). Of special interest are cryptic species complexes, such as the *P. ursina* species complex, that may require different conservation approaches because they contain more than one species, each with its own habitat and host plant requirements. Populations of *P. ursina*, a species endemic to the Central Valley, are being pushed into the foothill regions by expanding agriculture and urbanization. Additionally, *P. rotunda,* a cryptic species recognized based on this research, has lost much of its predicted habitat due to the increasing population growth in southern California and loss of habitats such as the coastal sage scrub and coastal prairies (McCaull 1994, Mattoni & Longcore 1997, Vandergast et al. 2007). For both species, habitat associated with many historical data points has been lost. Moving forward, systematists and ecologists must work together to unravel cryptic species complexes within biodiversity hotpots so that efforts can be made to accommodate conservation strategies for each unique species (Bickford et al. 2006).

CHAPTER 5

CONCLUSION

The results of this study provide evidence for the utility of AFLP genotyping in combination with traditional phylogenetic, distribution modeling, and morphological techniques to disentangle a cryptic species complex. Two main groups were identified for the *P. ursina* species complex. Under the phylogenetic species concept (Wheeler & Platnick 2000), I recognize two species: *Paracotalpa ursina* (Horn, 1867) and *Paracotalpa rotunda* (Casey, 1915). This decision is based on and supported by molecular phylogenetics, AFLP DNA fingerprinting, morphology, and distribution analyses. Both *P. ursina* (northern group) and *P. rotunda* (southern group) are strongly supported based on phylogenetics and AFLP.

This study also showed that vicariance events such as the uplift of mountain ranges and the ongoing establishment of anthropogenic barriers limits dispersal and gene flow and can lead to fragmentation and isolation of populations. Lastly, this study revealed the clear need for conservation strategies that protect genetic diversity of species endemic to the California Floristic Province biodiversity hotspot.

REFERENCES

REFERENCES

Alexander, P. J., G. Rajanikanth, C. D. Bacon & C. D. Bailey. 2007. Recovery of plant DNA using a reciprocating saw and silica-based columns. *Molecular Ecology Notes* 7:5–9.

American Farmland Trust. 1995. Alternatives for future urban growth in California's Central Valley: the bottom line for agriculture and taxpayers. Washington D.C.: American Farmland Trust.

Arnett, R. H. 2000. American Insects: A Handbook of the Insects of America North of Mexico. Second Edition, CRC Press, Boca Raton, Florida. 1,000 pp.Baum, D. A. & S. D. Smith. 2013. *Tree-Thinking: An Introduction to Phylogenetic Biology*. Roberts & Co., Greenwood Village, Colorado, 476 pp.

Baum, D. A. & S. D. Smith. 2013. Tree Thinking: an introduction to phylogenetic biology. Roberts and Company Publishers, Greenwood Village, 476 pp.

Beck, J. B., J. R. Allison, K. M. Pryer & M. D. Windham. 2012. Identifying multiple origins of polyploidy taxa: a multilocus study of the hybrid cloak fern (*Astrolepis integerrima*; Pteridaceae). *American Journal of Botany* 99(11):1857–1865.

Beza-Beza, C. F. 2012. Cloud forest passalids: an evolutionary study of the genus *Yumtaax*. M.S. Thesis, Wichita State University, Wichita, Kansas. 137 pp.

Bickford, D., D. J. Lohman, N. S. Sodhi, P. K. L. Ng, R. Meier, K. Winkler, K. K. Ingram & I. Das. 2006. Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution* 22(3):148–155.

Bonin, A., E. Bellemain, P. Bronken Eidesen, F. Pompanon, C. Brochmann & P. Taberlet. 2004. How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* 13: 3261–3273.

Brooks, T. M., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca, A. B. Rylands, W. R. Konstant, P. Flick, J. Pilgrim, S. Oldfield, G. Magin & C. Hilton-Taylor. 2002. Habitat loss and extinction in the hotspots of biodiversity. *Conservation Biology* 16:909–923.

Calsbeek, R., J. N. Thompson & J. E. Richardson. 2003. Patterns of molecular evolution and diversification in a biodiversity hotspot: the California Floristic Province. *Molecular Ecology* 12:1021–1029.

Casey, T. L. 1915. A review of the American species of Rutelinae, Dynastinae and Cetoniinae. *Memoirs on the Coleoptera* 6:1–394.

CEC. 1997. Ecological regions of North America toward a common perspective. Commission for environmental cooperation: 71pp.

Chatzimanolis, S. & M. S. Caterino. 2007. Toward a better understanding of the "Transverse Range Break": lineage diversification in southern California. *Evolution* 61:2127–2141.

Crawford, L. A., D. Koscinski & N. Keyghobadi. 2012. A call for more transparent reporting of error rates: the quality of AFLP data in ecological and evolutionary research. *Molecular Ecology* 21:5911–5917.

Delaney, K. S., S. P. D. Riley & R. N. Fisher. 2010. A rapid, strong, and convergent genetic response to urban habitat fragmentation in four divergent and widespread vertebrates. PLoS One 5(9):e12767.

ESRI. 2011. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.

Fall, H. C. 1932. New North American Scarabaeidae, with remarks on known species. *Journal of the New York Entomological Society* 40:183–204.

Forister, M. L., J. A. Fordyce & A. M. Shapiro. 2004. Geological barriers and restricted gene flow in the holarctic skipper *Hesperia comma* (Hesperiidae). *Molecular Ecology* 13:3489–3499.

Gouy, M., S. Guindon & O. Gascuel. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution* 27(2):221–224.

Haldeman, S.S. 1852. Insects, pp. 366-378. *In*, Stansbury*, Exploration and Survey of the Valley of the Great Salt Lake of Utah, Including a Reconnoissance of a New Route through the Rocky Mountains*. Washington, D.C. 487 pp.

Hardy, A. R. 1971. The North American Areodina with a description of a new genus from California. *The Pan-Pacific Entomologist* 47:235–242.

Hardy, A.R. 1972. A note on *Paracotalpa deserta* Saylor (Coleoptera: Scarabaeidae: Rutelinae). *The Coleopterists Bulletin* 26:77–78.

Hijmans, R. J. S. E. Cameron, J. L. Parra, P. G. Jones & A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal for Climatology* 25:1965–1978.

Holland, B. R., A. C. Clarke & H. M. Meudt. 2008. Optimizing automated AFLP scoring parameters to improve phylogenetic resolution. *Systematic Biology* 57(3):347–366.

Homer, C.H., J. A. Fry & C. A. Barnes. 2012. The National Land Cover Database, U.S. Geological Survey Fact Sheet 2012-3020, 4 p.

Horn, G. H. 1867. Descriptions of new genera and species of western Scarabaeidae, with notes on others already known. *Transactions of the American Entomological Society* 1:163–170.

Hutchinson, G. E. 1957. Concluding remarks. *Cold Spring Harbor Symposia on Quantitative Biology* 22:415–427.

Hwang, U.-W. & W. Kim. 1999. General properties and phylogenetic utilities of nuclear ribosomal DNA and mitochondrial DNA commonly used in molecular systematics. *The Korean Journal of Parasitology* 37(4):215–228.

ICZN (International Commission on Zoological Nomenclature). 1999. *International Code of Zoological Nomenclature*. 4th Edition, International Trust for Zoological Nomenclature, London.

Jacobs, D. K., T. A. Haney & K. D. Louie. 2004. Genes, diversity, and geologic process on the Pacific coast. Annual Review of Earth and Planetary Sciences 32:601–652.

Jorgensen, C. D. 1963. Notes on the biology and distribution of *Paracotalpa granicollis* Haldeman. *The Pan-Pacific Entomologist* 39:154–156.

Kambhampati, S. & P. T. Smith. 1995. PCR primers for the amplification of four insect mitochondrial gene fragments. *Insect Molecular Biology* 4:233–236.

Kaufman, P. & M. L. Jameson. 2009. Biological observations and a new state record of *Paracotalpa granicollis* Haldeman (Coleoptera: Scarabaeidae: Rutelinae) in New Mexico. *The Coleopterists Bulletin* 63:513–515.

Krajcik M. 2008. Rutelinae of the World. Checklist of Scarabaeoidea of the World.2. Rutelinae (Coleoptera: Scarabaeidae: Rutelinae). Animma.x, suppl. 4. Published by the author, Plzeň, 142 pp.

Kuchta, S. R. & A. M. Tan. 2006. Lineage diversification on an evolving landscape: phylogeography of the California newt, Taricha torosa (Caudata: Salamandridae). *Biological Journal of the Linnean Society* 89:213–239.

LeConte, J. L. 1863. New species of North American Coleoptera. Prepared for the Smithsonian Institution. *Smithsonian Miscellaneous Collection* 167:1–86.

Machatschke, J. W. 1972. *Coleopterorum Catalous Supplementa, Part 66, Fascicle 1, Scarabaeoidea: Melolonthidae, Rutelinae*. Dr. W. Junk, Gravenhage. 361 pp.

Maldonado, J. E., C. Vila & R. K. Wayne. 2001. Tripartite genetic subdivisions in the ornate shrew (*Sorex ornatus*). *Molecular Ecology* 10:127–147.

Mattoni, R. & T. Longcore. 1997. The Los Angeles coastal prairie, a vanished community. *Crossosoma* 23:71–102.

Mawdsley, J. R. 2001. A new key to the species of the genus Cradytes Casey (Insecta: Coleoptera: Melyridae). *Insecta Mundi* 15(3):181–184.

McCaull, J. 1994. Chapter 14. The natural community conservation planning program and the coastal sage scrub ecosystem of southern California, pp. 281–292. In: R. E. Grumbine (Eds.). *Environmental Policy and Biodiversity*. Island Press, Washington, D.C., 426 pp.

Mittermeier, R. A., P. R. Gil, M. Hoffman, J. Pilgrim, T. Brooks, C. G. Mittermeier, J. Lamoreux & G. A. B. da Fonseca. 2005. *Hotspots revisited: Earth's biologically richest and most endangered terrestrial ecoregions*. Conservation International, Washington, DC, 389 pp.

Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. da Fonseca & J. Kent. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403:853–858.

Peakall, R. & P. E. Smouse. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research – an update. *Bioinformatics* 28:2537–2539.

Phillips, S. J., M. Dudik & R. E. Schapire. 2004. A maximum entropy approach to species distribution modelling. In: Proceedings of the 21st International Conference on Machine Learning. ACM Press, New York.

Puliam, H. R. 2000. On the relationships between niche and distribution. *Ecology Letters* 3:349–361.

Ratcliffe, B. C. & M. J. Paulsen. 2008. *The Scarabaeoid Beetles of Nebraska*. Bulletin of the University of Nebraska State Museum 22: i-iv, 1–570.

Reeves, P. A. & C. M. Richards. 2009. Accurate inference of subtle population structure (and other genetic discontinuities) using principal coordinates. *PLoS One* 4:e4269.

Reeves, P. A. & C. M. Richards. 2011. Species delimitation under the general lineage concept: an empirical example using wild North American hops (Cannabaceae: *Humulus lupulus*). *Systematic Biology* 60:45–59.

Riley, S. P. D., R. M. Sauvajot, T. K. Fuller, E. C. York, D. A. Kamradt, C. Bromley & R. K. Wayne. 2003. Effects of urbanization and habitat fragmentation on bobcats and coyotes in southern California. Conservation Biology 17(2):566–576.

Riley, S. P. D., J. P. Pollinger, R. M. Sauvajot, E. C. York, C. Bromley, T. K. Fuller & R. K. Wayne. 2006. A southern California freeway is a physical and social barrier to gene flow in carnivores. Molecular Ecology 15:1733–1741.

Rissler, L. J., R. J. Hijmans, C. H. Graham, C. Moritz & D. B. Wake. 2006. Phylogeographic lineages and species comparisons in conservation analyses: a case study of California herpetofauna. *The American Naturalist* 167:655–666.

Ritcher, P. O. 1948. Descriptions of the larvae of some ruteline beetles with keyes to tribes and species (Scarabaeidae). *Annals of the Entomological Society of America* 41(2):206–212.

Ronquist, F. H. & J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.

Saylor, L. W. 1940. Synoptic revision of the beetle genera *Cotalpa* and *Paracotalpa* of the United States, with description of a new subgenus. *Proceedings of the Entomological Society of Washington* 42:190–200.

Scheldeman, X. & M. van Zonneveld. 2010. Training manual on spatial analysis of plant diversity and distribution. Biodiversity International, Rome, Italy.

Simon, C., F. Frati, A. Beckenbach., B. Crespi, H. Liu & P. Flock. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America* 87:651–701.

Smith, A. B. T. 2009. Checklist and Nomenclatural authority file of the Scarabaeoidea of the Nearctic Realm. Version 4. Electronically published, Ottawa, Canada. 97 pp.

Swoffod, D. L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

U. S. Census Bureau. 1960. 1961 U. S. Census Report. Available from http://www2.census.gov/prod2/statcomp/documents/1961-02.pdf (accessed 20 March 2014).

U. S. Census Bureau. 2010. 2010 Census interactive population search. Available from http://www.census.gov/2010census/data/ (accessed 20 March 2014).

Vandergast, A. G., A. J. Bohonak, D. B. Weissman & R. N. Fisher. 2007. Understanding the genetic effects of recent habitat fragmentation in the context of evolutionary history: phylogeography and landscape genetics of a southern California endemic Jerusalem cricket (Orthoptera: Stenopelmatidae: Stenopelmatus). *Molecular Ecology* 16:977–992.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper & M. Zabeau. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* 23(21):4407–4414.

Wake, D. B. 1997. Incipient species formation in salamanders of the *Ensatina* complex. *Proceeding of the National Academy of Science USA* 55:7761–7767.

Wheeler, Q. D. & N. I. Platnick. 2000. Chapter 5. The phylogenetic species concept, pp. 55–69. *In*: Q. D. Wheeler & R. Meier (Eds.). *Species concepts and phylogenetic theory: a debate*. Columbia University Press, New York, xii + 231 pp.

Zhang, D. X. & G. M. Hewitt. 1996. Assessment of the universality and utility of a set of conserved mitochondrial COI primers in insects. *Insect Molecular Biology* 6:143–150.

APPENDICES

APPENDIX 1

VOUCHER SPECIMENS FOR DNA SEQUENCE AND AFLP GENOTYPING ANALYSES

Voucher	Population	Depository	Preservation	Label Data
OK_P01_05	1	WICH	EtOH (-80°C)	USA: CA: Kings Co., Kettleman Hills, Hwy 41, 5.4 mi NE of Hwy 33. 35.953°N 119.993°W. 06-iv-2013. Coll. D. C. Hawks
OK_P01_06	1	WICH	EtOH (-80°C)	USA: CA: Kings Co., Kettleman Hills, Hwy 41, 5.4 mi NE of Hwy 33. 35.953°N 119.993°W. 06-iv-2013. Coll. D. C. Hawks
OK_P02_01	2	WICH	EtOH (-80°C)	USA: CA: San Bernardino Co., Devore. 34.236°N 117.408°W. 03-v-2013. Coll. D. C. Hawks
OK_P02_04	2	WICH	EtOH (-80°C)	USA: CA: San Bernardino Co., Devore. 34.236°N 117.408°W. 03-v-2013. Coll. D. C. Hawks
OK_P02_05	2	WICH	EtOH (-80°C)	USA: CA: San Bernardino Co., Devore. 34.236°N 117.408°W. 03-v-2013. Coll. D. C. Hawks
OK_P03_01	3	WICH	EtOH (-80°C)	USA: CA: San Bernardino Co., Lower Lone Pine Canyon. On chamise. 34.303°N 117.516°W. 29-iv-2013. Coll. D. C. Hawks
OK_P03_04	3	WICH	EtOH (-80°C)	USA: CA: San Bernardino Co., Lower Lone Pine Canyon. On chamise. 34.303°N 117.516°W. 29-iv-2013. Coll. D. C. Hawks
OK_P03_05	3	WICH	EtOH (-80°C)	USA: CA: San Bernardino Co., Lower Lone Pine Canyon. On chamise. 34.303°N 117.516°W. 29-iv-2013. Coll. D. C. Hawks
OK_P04_03	4	WICH	EtOH (-80°C)	USA: CA: San Luis Obispo Co., 8.6 mi E of Shell Creek Rd on Hwy 58. 35.425°N 120.228°W. 07-iv-2013
OK_P04_05	4	WICH	EtOH (-80°C)	USA: CA: San Luis Obispo Co., 8.6 mi E of Shell Creek Rd on Hwy 58. 35.425°N 120.228°W. 07-iv-2013

APPENDIX 1 (continued)

Voucher	Population	Depository	Preservation	Label Data
OK_P05_01	5	WICH	EtOH (-80°C)	USA: CA: Kern Co., 0.5-1.1 mi W of Camp Okihi on Round Mountain Rd. 35.443°N 118.895°W. 08-iv-2013. Coll. D. C. Hawks
OK_P05_03	5	WICH	EtOH (-80°C)	USA: CA: Kern Co., 0.5-1.1 mi W of Camp Okihi on Round Mountain Rd. 35.443°N 118.895°W. 08-iv-2013. Coll. D. C. Hawks
OK_P05_05	5	WICH	EtOH (-80°C)	USA: CA: Kern Co., 0.5-1.1 mi W of Camp Okihi on Round Mountain Rd. 35.443°N 118.895°W. 08-iv-2013. Coll. D. C. Hawks
OK_P06_01	6	WICH	EtOH (-80°C)	USA: CA: Tulare Co., Hwy 198 & Yokohl Dr. 36.324°N 119.076°W. 06-iv-2013. Coll. D. C. Hawks
OK_P06_03	6	WICH	EtOH (-80°C)	USA: CA: Tulare Co., Hwy 198 & Yokohl Dr. 36.324°N 119.076°W. 06-iv-2013. Coll. D. C. Hawks
OK_P06_05	6	WICH	EtOH (-80°C)	USA: CA: Tulare Co., Hwy 198 & Yokohl Dr. 36.324°N 119.076°W. 06-iv-2013. Coll. D. C. Hawks
OK_P07_05	7	WICH	EtOH (-80°C)	USA: CA: Kern Co., 2 mi NE of Oildale, North Side of Kern River. 35.442°N 118.996°W. 08-iv-2013
OK_P07_06	7	WICH	EtOH (-80°C)	USA: CA: Kern Co., 2 mi NE of Oildale, North Side of Kern River. 35.442°N 118.996°W. 08-iv-2013
OK_P08_01	8	WICH	EtOH (-80°C)	USA: CA: San Bernardino Co., Lytle Creek Rd 0.1-0.5 mi N of Glen Helen Parkway. 34.192°N 117.442°W. 03-v-2013. Coll. D. C. Hawks

APPENDIX 1 (continued)

Voucher	Population	Depository	Preservation	Label Data
OK_P08_03	8	WICH	EtOH (-80°C)	USA: CA: San Bernardino Co., Lytle Creek Rd 0.1-0.5 mi N of Glen Helen Parkway. 34.192°N 117.442°W. 03-v-2013. Coll. D. C. Hawks
OK_P08_04	8	WICH	EtOH (-80°C)	USA: CA: San Bernardino Co., Lytle Creek Rd 0.1-0.5 mi N of Glen Helen Parkway. 34.192°N 117.442°W. 03-v-2013. Coll. D. C. Hawks
OK_P09_03	9	WICH	EtOH (-80°C)	USA: CA: Fresno Co., Woodward Park. 18-iv-2013. 36.866°N 119.785°W. Coll. D. C. Hawks & Bruyen
OK_P09_04	9	WICH	EtOH (-80°C)	USA: CA: Fresno Co., Woodward Park. 18-iv-2013. 36.866°N 119.785°W. Coll. D. C. Hawks & Bruyen
OK_P09_05	9	WICH	EtOH (-80°C)	USA: CA: Fresno Co., Woodward Park. 18-iv-2013. 36.866°N 119.785°W. Coll. D. C. Hawks & Bruyen
OK_P10_06	10	WICH	EtOH (-80°C)	USA: CA: Kern Co., 4 mi E of Arvin. 35.209°N 118.753°W. 18-iv-2013. Coll. D. C. Hawks
OK_P10_07	10	WICH	EtOH (-80°C)	USA: CA: Kern Co., 4 mi E of Arvin. 35.209°N 118.753°W. 18-iv-2013. Coll. D. C. Hawks
OK_P11_03	11	WICH	EtOH (-80°C)	USA: CA: San Luis Obispo Co., Route 166 Carrizo Plain National Monument. 35.06595°N 119.63167°W. Elevation 652 m. 20-iii-2013. Coll. O. Keller
OK_P11_04	11	WICH	EtOH (-80°C)	USA: CA: San Luis Obispo Co., Route 166 Carrizo Plain National Monument. 35.06595°N 119.63167°W. Elevation 652 m. 20-iii-2013. Coll. O. Keller
OK_P11_05	11	WICH	EtOH (-80°C)	USA: CA: San Luis Obispo Co., Route 166 Carrizo Plain National Monument. 35.06595°N 119.63167°W. Elevation 652 m. 20-iii-2013. Coll. O. Keller

APPENDIX 1 (continued)

Voucher	Population	Depository	Preservation	Label Data
OK_P12_01	12	WICH	EtOH (-80°C)	USA: CA: San Diego Co., Alliant International University. 32.89476°N 117.09292°W. Elevation 200 m. 21-iii-2013. Coll. O. Keller
OK_P12_03	12	WICH	EtOH (-80°C)	USA: CA: San Diego Co., Alliant International University. 32.89476°N 117.09292°W. Elevation 200 m. 21-iii-2013. Coll. O. Keller
OK_P12_04	12	WICH	EtOH (-80°C)	USA: CA: San Diego Co., Alliant International University. 32.89476°N 117.09292°W. Elevation 200 m. 21-iii-2013. Coll. O. Keller
OK_P13_01	*Paracotalpa puncticollis*	WICH	EtOH (-80°C)	USA: CA: Riverside Co., Pines to Palms Hwy, Route 74 near Bighorn Drive. 33.62106°N 116.41193°W. Elevation 958 m. 19-iii-2013. Coll. O. Keller & J. N. Hogue
OK_P13_02	*Paracotalpa puncticollis*	WICH	EtOH (-80°C)	USA: CA: Riverside Co., Pines to Palms Hwy, Route 74 near Bighorn Drive. 33.62106°N 116.41193°W. Elevation 958 m. 19-iii-2013. Coll. O. Keller & J. N. Hogue
OK_P13_03	*Paracotalpa puncticollis*	WICH	EtOH (-80°C)	USA: CA: Riverside Co., Pines to Palms Hwy, Route 74 near Bighorn Drive. 33.62106°N 116.41193°W. Elevation 958 m. 19-iii-2013. Coll. O. Keller & J. N. Hogue

APPENDIX 2

DNA EXTRACTION METHOD ALTERED FROM QIAGEN'S "PURIFICATION OF TOTAL DNA FROM INSECTS" SUPPLEMENTARY PROTOCOL

Remove sample tissue (leg and/or thoracic muscle) from specimen and place in a 2.0 ml pyramid microcentrifuge tube (Phenix Research Products). Let tissue dry over night until all ethanol residues are evaporated. After placing two metal ball bearings into each tube, use modified reciprocating saw for two minutes to pulverize tissue into fine powder. Add 180 µl Buffer ATL and 20 µl proteinase K to each tube, mix by vortexing, and incubate for 2 hours at 56°C. Vortex tubes every 20 minutes to disperse samples. Following incubation add 4 µl RNase A (100 mg/ml) (Applied Biosystems), vortex, and incubate for 2 minutes at room temperature. Vortex for 15 seconds, add 200 µl Buffer AL, vortex again, add 200 µl 100 % ethanol, and vortex again. Load DNeasy Mini spin column placed in a 2.0 ml collection tube with mixture from previous step (including precipitate) and centrifuge for 1 minute at 8,000 rpm. Place DNeasy Mini spin column into a new 2.0 ml collection tube and discard the old collection tube including flow-through. Add 500 µl Buffer AW1 to DNeasy Mini spin column and centrifuge for 1 minute at 8,000 rpm. Discard flow-through and collection tube again and place DNeasy Mini spin column into new 2.0 ml collection tube. Add 500 µl Buffer AW2, centrifuge for 3 minutes at 14,000 rpm drying the DNeasy membrane completely. Carefully remove spin column from collection tube to ensure no carryover of flow-through to new 1.5 ml microcentrifuge tube. Add 100 µl Buffer AE directly onto the DNeasy membrane and incubate for 1 minute at room temperature. Centrifuge for 1 minute at 8,000 rpm to elute sample, label microcentrifuge tube and store at -20°C for downstream applications.

APPENDIX 3

ALIGNED DNA DATA

12S

OK_P01_05

```
--ACTTATCCTGCTTTAGAGCAGGAGCGACGGGCGATATGTGCATATTATAGAGCTTAAG
TCATATAGTCTAA-CTAAACTATATTACAATCAAATCCACTTTAAAAAATAGTGTTAGCT
ACTTTAACCATATAAATAAATTTATTGTAACCCATCTCTCCTTACCTATCCGCTGTATCT
TGATCTGATTTTCTCTTT----AACCTAATATCTTGAACATTCCTTCTCCTTACAAAACA
TTCAACCTACGACGATATACAAA-CCTTTAAA--GCAAGTC-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAACTACTAATACTCAAGCATTTATATTTACNTTTTCAATNNTAGGGNANNTAA
N--------------
```

OK_P01_06

```
AGACTTATCCTGCTTTAGAGCAGGAGNGACGGGCGATATGTGCATATTATAGAGCTTAAG
TCATATAGTCTAA-CTAAACTATATTACAATCAAATCCACTTTAAAAAATAGTGTTAGCT
ACTTTAACCATATAAATAAATTTATTGTAACCCATCTCTCCTTACCTATCCGCTGTATCT
TGATCTGATTTTCTCTTT----AACCTAATATCTTGAACATTCCTTCTCCTTACAAAACA
TTCAACCTACGACGATATACAAA-CCTTTAAA--GCAAGTC-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAACTACTAATACTCAAGCATTTATATTTNCNTTTTCAATAATAGGGTATCTAA
---------------
```

OK_P02_01

```
AGACTTATCCTGCTTTAGAGCAGGAGCGACGGGCGATATGTGCATATTATAGAGCTTCAG
TCATATAGTTTAA-CTAAACTATATTACAATCAAATCCACTTTAAAAAAAAGAGGAAACT
ATATCAAAAACATAAAAAAAATTATAGTAACCCATCTCTCCTTGTCTATACCCTCTATCT
TGAGATCAGATCCCCTCTAT--AACCTAATATCTTGAATATTCCTCTTCCTTTCAAAACA
TTCAACCTACGACGATATACAAA-CCTTTAAAA--CAAGTA-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAGCTACTAATACTCAAGCATTTATATTTACATTTTCAATAATAGGGNATCNAA
A--------------
```

OK_P02_04

```
NGACTTATCCTGCTTTAGAGCAGGAGCGACGGGCGAAAAGGGCAAATTATAGAGCTTCAG
TCATATAGTTTAA-CGAAACTATATTACAATCAAATCCACTTTAAAAAATAGTGTTAACT
ATTTCAACCATATAAATAAATTTATTGTAACCCATCTCTCCTTGTCTATCCGCTGTATCT
TGATCTGATTTCCTCTT--T--AACCTAATATCTTGAATATTCCTCTTCCTTTCAAAACA
TTCAACCTACGACGATATACAAA-CCTTTAAAA--CAAGTA-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAGCTACTAATACTCAAGCATTTATATTTACATTTTCAATAATAGGGTATCTAA
T--------------
```

APPENDIX 3 (continued)

```
OK_P02_05
AGACTTATCCTGCTTTAGAGCAGGAGCGACGGGCGATATGTGCATATTATAGAGCTTCAG
TCATATAGTTTAA-CTAAACTATATTACAATCAAATCCACTTTAAAAAATAGTGTTAACT
ATTTCAACCATATAAATAAATTTATTGTAACCCATCTCTCCTTGTCTATCCGCTGTATCT
TGATCTGATTTCCTCTTT----AACCTAATATCTTGAATATTCCTCTTCCTTTCAAAACA
TTCAACCTACGACGATATACAAA-CCTTTAAAA--CAAGTA-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAGCTACTAATACTCAAGCATTTNTATTTACATTTTCAATAATAGGGTAT----
---------------
```

```
OK_P03_01
AGACTTATCCTGCTTTAGAGCAGGAGCGACGGGCGATATGTGCATATTATAGAGCTTCAA
TCAAATAGTTTAAAC-AAACTAAATTAAAATCAAATCCAACTTAAAAAATAGGATTAACT
ATTTCAACCATATAAATAAATTTATTGGAACCCATCTCTCCTTGTCTATCCGCTGTATCT
TGATCTGATTTCCTCTTT----AACCTAATATCTTGAATATTCCTCTTCCTTTCAAAACA
TTCAACCTACGACGATATACAAA-CCTTTAAAA--CAAGTA-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAGCTACTAATACTCAAGCATTTATATTTACATTTTCAATAATAGGGTCTCTAA
A--------------
```

```
OK_P03_04
AGACTTATCCTGCTTTAGAGCAGGAGCGACGGGCGATATGTGCATATGATAGAGCTTCAG
TCATATAGTTTAA-CTAAACTATATTACAATCAAATCCACTTTAAAAAATAGTGTTAACT
ATTTCAACCATATAAATAAATTTATTGTAACCCATCTCTCCTTGTCTATCCGCTGTATCT
TGATCTGATTTCCTCTTT----AACCTAATATCTTGAATATTCCTCTTCCTTTCAAAACA
TTCAACCTACGACGATATACAAA-CCTCTNTAAAACAAGTA-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCGGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAGCTACTAATACTCAAGCATTTATATTTACATTTTCAATAATAGGGTATCTAA
T--------------
```

```
OK_P03_05
AGACTTATCCTGCTTTAGAGCAGGAGCGACGGGCGATATGTGCATATTATAGACCGTCAA
GCATATAGATTAA-CTAAACCATATGAAAATCAAATCCACTCTAAAAAAGAAAGGGAACT
AAATCAACCAAATAAATAAATTCATTGTAACCCATCTCTACGGGTCTATACGCTGAATCT
CGAGATCAGATCCCCTCTAT--AACCTAATATCTTGAATATTCCTCTTCCTTTCAAAACA
TTCAACCTACGACGATATACAAA-CCTTTAAAA--CAAGTA-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAACTACTAATACTCAAGCATTTATATTTACATTTTCAATAATAGGGNATCTAA
T--------------
```

```
OK_P04_03
AGANTNATCNNGCNTTAGAGCAGGAGCGACGGGCGATATGTGCATATTATAGAGCTTCAG
TCATATAGTCTAA-CTAAACTATATTACAATCAAATCCACTTTAAAAAATAGTGTTAGCT
ACTTTAACCATATAAATAAATTTATTGTAACCCATCTCTCCTTACCTATCCGCTGTATCT
TGATCTGATTTTCTCTTT----AACCTAATATCTTGAACATTCCTTCTCCTTACAAAACA
```

APPENDIX 3 (continued)

TTCAACCTACGACGATATACAAA-CCTTTAAA--GCAAGTC-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAACTACTAATACTCAGGCATTTATNTTTACNTTTTCAATNATAGGGNATCTAA
N--------------

OK_P05_01
AGACTTATCCTGCTTTAGAGCAGGAGCGACGGGCGATATGTGCATATTATAGAGCTTCAG
TCATATAGTCTAA-CTAAACTATATTACAATCAAATCCACTTTAAAAAATAGTGTTAGCT
ACTTTAACCATATAAATAAATTTATTGTAACCCATCTCTCCTTACCTATCCGCTGTATCT
TGATCTGATTTTCTCTTT----AACCTAATATCTTGAACATTCCTTCTCCTTACAAAACA
TTCAACCTACGACGATATACAAA-CCTTTAAA--GCAAGTC-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAACTACTAATACTCAAGCATTTATATTTACNTTTTCAATNATAGGGNATCTAA
T--------------

OK_P05_03
AGNTTNANCTTGNNTNAGAGCAGGAGCGACGGGCTATAAGTCCCTANTATAGAGCTGCAG
GCAAANAGTCAAA-CAAANCTATATTACAATCAAATCCACTTTAAAAAATAGTGTTAGCT
ACTTTAACCATATAAATAAATTTATTGTAACCCATCTCTCCTTACCTATCCGCTGTATCT
TGATCTGATTTTCTCTTT----AACCTAANATCTTGAACATTCCTTCTCCTTACAAAACA
TTCAACCTACGACGATATACAAA-CCTTTAAAA-GCAAGTC-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCCTTAAATTTCA
AGAACATAACTACNAATACTCATGCATTTATATTTACNTTTTCAATNATAGGGNATCTAA
N--------------

OK_P05_05
AGACTTATCNTGCTTTAGAGCAGGAGCGACGGGCGATATGTGCATATTATAGAGCTTCAG
TCATATAGTCTAA-CTAAACTATATTACAATCAAATCCACTTTAAAAAATAGTGTTAGCT
ACTTTAACCATATAAATAAATTTATTGTAACCCATCTCTCCTTACCTATACGCTGTATCT
TGATNTGATTTTCTCTTT----AACCTAATATCTTGAACATTCCTTCTCCTTACAAAACA
TTCAACCTACGACGATATACAAA-CCTTTAAA--GCAAGTC-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAACTACTAATACTCAAGCATTTATATTTACNTTTTCAATNATAGGGNATCTAA
T--------------

OK_P06_01
AGACTTATCCTGCTTTAGAGCAGGAGCGACGGGCGATATGTGCATATTATAGAGCTTCAG
TCATATAGTCTAA-CTAAACTATATTACAATCAAATCCACTTTAAAAAATAGTGTTAGCT
ACTTTAACCATATAAATAAATTTATTGTAACCCATCTCTCCTTACCTATCCGCTGTATCT
TGATCTGATTTTCTCTTT----AACCTAATATCTTGAACATTCCTTCTCCTTACAAAACA
TTCAACCTACGACGATATACAAA-CCTTTAAA--GCAAGTC-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAACTACTAATACTCAAGCATTTATATTTACNTTTTCAATAATAGGGNATCTAA
T--------------

OK_P06_03

```
AGANTTATCNTGCCTNAGAGCAGGAGCGACGGGCGATATGTGCATATTATAGAGCTTCAG
TCATATAGTCTAA-CTAAACTATATTACAATCAAATCCACTTTAAAAAATAGTGTTAGCT
ACTTTAACCATATAAATAAATTTATTGTAACCCATCTCTCCTTACCTATCCGCTGTATCT
TGATCTGATTTTCTCTTT----AACCTAATATCTTGAACATTCCTTCTCCTTACAAAACA
TTCAACCTACGACGATATACAAA-CCTTTAAA--GCAAGTC-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGNTCCTCTGAATANACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAACTACTAATACTCANGCATTTATATTTACATTTTCAATAATAGGGTATCTAA
TCCAAGTTTAAAAAN
```

OK_P06_05

```
AGACTTATCNTGCTTTAGAGCAGGAGCGACGGGCGATATGTGCATATTATAGAGCTTCAG
TCATATAGTCTAA-CTAAACCATATTACAATCAAATCCACTTTAAAAAATAGTGTTAGCT
ACTTTAACCATATAAATAAATTTATTGTAACCCATCTCTCCTTACCTATCCGCTGTATCT
TGATNTGATTTTCTATCTAACCAACCTAATATCTTGAACATTCCTTCTCCTTACAAAACA
TTCAACCTACGACGATATACAAA-CCTTTAAA--GCAAGTCCAGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAACTACTAATACTCAAGCATT-AT-----------------------------
---------------
```

OK_P07_05

```
AGACTTATCCTGCTTTAGAGCAGGAGCGACGGGCGAGAAGGGCATATTATAGAGCTTCAG
TCATATAGTCTAA-CTAAACTATATTACAATCAAATCCACTTTAAAAAATAGTGTTAGCT
ACTTTAACCATATAAATAAATTTATTGTAACCCATCTCTCCTTACCTATCCGCTGTATCT
TGATCTGATTTTCTCTTT----AACCTAATATCTTGAACATTCCTTCTCCTTACAAAACA
TTCAACCTACGACGATATACAAA-CCTTTAAA--GCAAGTC-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAACTACTAATACTCAAGCATTTATATTTACATTTTCAATNATAGGGTATCTAA
T--------------
```

OK_P08_01

```
NGACTTATCCTGCTTTAGAGCAGGAGCGNCGGGCGATATGTGCAAAATATAGAGATGCAG
ACATATAGATTAA-CTAAACGATATGACAATCAAATCCACTCTAAAAAATAGAGGGAACT
ACTTCAACCAAATAAATAAATTCATTGTAACCCATCGCTACGGGTCTATACGCTGTATCT
CGATATCAGATCCCCTCTAT--AACCTAATATCTTGAATATTCCTCTTCCTTTCAAAACA
TTCAACCTACGACGATATACAAA-CCTTTAAAA--CAAGTA-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAGCTACTAATACTCAAGCATTTATATTTACATTTTCAATAATAGGG-------
---------------
```

OK_P08_03

```
NGACTTATCCTGCTTTAGAGCAGGAGCGACGGGCGAGATATGCATATTAAAGAGCTTCAG
TCATATAGTTTAA-CTAAACTATATTACAATCAAATCCACTTTAAAAAATAGTGTTAACT
ATTTCAACCATATAAATAAATTTATTGTAACCCATCTCTCCTTGTCTATCCGCTGTATCT
TGATCTGATTTCCTCTTT----AACCTAATATCTTGAATATTCCTCTTCCTTTCAAAACA
```

APPENDIX 3 (continued)

```
TTCAACCTACGACGATATACAAA-CCTTTAAAA--CAAGTA-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAGCTACTAATACTCAAGCATTTATATTTACATTTTCAATAATAGGGTATCTAA
N--------------


OK_P08_04
NGACTTATCCTGCTTTAGAGCAGGAGCGACGGGCGATAGGAGAAGAGTATAGAACTACAG
TCAGATAGAGTAA-CTAAAATATATAACAATAAAAACAACTTCAAAAAAAAGGGATAACT
ATTTCAACAATATAAATAAATTTATTGTAACCCATCTCTCCTTGTCTATCCGCTGTATCT
TGATCTGATTTCCTCTTT----AACCTAATATCTTGAATATTCCTCTTCCTTTCAAAACA
TTCAACCTACGACGATATACAAA-CCTCTNTAAAACAAGTA-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAGCTACTAATACTCAAGCATTTATATTTACATTTTCAATAATAGGGTATCNAA
N--------------

OK_P09_03
AGANTNANCNNNCNTTAGAGCAGGAGCGACGGGCGATATGTGCATATNATAGAGCCTCAA
TCAAAAAGTCTAAACAAAACAAAAT-AAAATCAAATCCACACTAAAAAATAGGGTTAGCA
ACTTTAACCATGTAAATAAATTTATTGTAACCCATCTCTCCTTACCTATCCGCTGTATCT
TGATCTGATTTTCTCTTTT----ACCTAATATCTTGAACATTCCTTCTCCTTACAAAACA
TTCAACCTACGACGATNTACAAA-CCTTTAAA--GCAAGTC-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACNATAATACCGCCAAATCCTTTAAATTTCN
AGAACATAACTACNAATACTCANGCNTTTATNTTTACTTTTTCAATAANAGGGTATCTAA
---------------

OK_P09_04
AGANTTATCCTGCTTTAGAGCAGGAGCGACGGGCGATAGGTGCATATCAAAAAACCTCAA
TCATATAGTCTAA-CTAAACTATATTACAATCAAATCCACTTTAAAAAATAGGGTTAGCT
ACTTTAACCATGTAAATAAATTTATTGTAACCCATCTCTCCTTACCTATCCGCTGTATCT
TGATCTGATTTTCTCTTTT----ACCTAATATCTTGAACATTCCTTCTCCTTACAAAACA
TTCAACCTACGACGATATACAAA-CCTTTAAA--GCAAGTC-AGANTTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAACTACTAATACTCAAGCATTTATNTTTACATTTTCAATAATAGGGTATCTAA
T--------------

OK_P09_05
-GANTTATCCTNCTTTAGAGCAGGAGCGANGGGCGATNTGTGCATATTATAGAGCTTCAG
TCATATAGTCTAA-CTAAACTATATTACAATCAAATCCACTTTAAAAAATAGGGTTAGCT
ACTTTAACCATGTAAATAAATTTATTGGAACCCATCTCTCCTTACCTATCCGCTGTATCT
TGATCTGATTTTCTCTTTT----ACCTAATATCTTGAACATTCCTTCTCCTTACAAAACA
TTCAACCTACGACGATATACAAA-CCTTTAAA--GCAAGTC-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATGTCA
AGAACATAACTACTAATACTCAAGCATTTATNTTTACTTTTTCAATAATAGGGTATC---
---------------
```

APPENDIX 3 (continued)

OK_P10_06
-GACTTATCCTGCTTTAGAGCAGGAGCGANGGGCGATATGTGCATATTATAGAGCTTCAG
TCATATAGTCTAA-CTAAACTATATTACAATCAAATCCACTTTAAAAAATAGTGTTAGCT
ACTTTAACCATATAAATAAATTTATTGTAACCCATCTCTCCTTACCTATCCGCTGTATCT
TGATCTGATTTTCTCTTT----AACCTAATATCTTGAACATTCCTTCTCCTTACAAAACA
TTCAACCTACGACGATATACAAA-CCTTTAAA--GCAAGTC-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAACTACTAATACTCAAGCATTTATATTTACATTTTCAATAATAGGGTATCTAA
T--------------

OK_P10_07
AGACTTATCCTGCTTTAGAGCAGGAGCGACGGGCGATATGTGCATATTATAGAGCTTCAG
TCATATAGTCTAA-CTAAACTATATTACAATCAAATCCACTTTAAAAAATAGTGTTAGCT
ACTTTAACCATATAAATAAATTTATTGTAACCCATCTCTCCTTACCTATACGCTGTATCT
TGATCTGATTTTCTCTTT----AACCTAATATCTTGAACATTCCTTCTCCTTACAAAACA
TTCAACCTACGACGATATACAAA-CCTTTAAA--GCAAGTC-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAACTACTAATACTCAAGCATTTATATTTACATTTTCAATAATAGGGNATCTAA
T--------------

OK_P11_03
AGNNTTATCCTGCTTTAGAGCAGGAGCGACGGGCGATATGTGCATATTATAGAGCTTCAG
TCATATAGTCTAA-CTAAACTATATTACAATCAAATCCACTATAAAAAATAGNGTTAGCT
ACTTTAACCATATAAATAAATTTATTGTAACCCATCTCTCCTTACCTANCCNCTGNATCT
TGATCTGATTTTCTCTTT----AACCTAANATCTTGAACATTCCTTCTCCTTACAAAACA
TTCAACCTACGACGATNTACAAA-CCTTTAAA--GCAAGTC-AGA-TTAATCGNGGATCA
TCANTTATNNGANAGGTTCCTCTGAATAGACNATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAACTACNAATACTCAAGCATTTATATTTACNTTTTCAATAATAGGGTATCTAA

OK_P11_04
AGACTTATCCTGCTTTAGAGCAGGAGCGACGGGCGATATGTGCATATTATAGAGCTTAAG
TCATATAGTCTAA-CTAAACTATATTACAATCAAATCCACTTTAAAAAATAGTGTTAGCT
ACTTTAACCATATAAATAAATTTATTGTAACCCATCTCTCCTTACCTATCCGCTGTATCT
TGATCTGATTTTCTCTTT----AACCTAATATCTTGAACATTCCTTCTCCTTACAAAACA
TTCAACCTACGACGATATACAAA-CCTTTAAA--GCAAGTC-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAACTACTAATACTCAAGCATTTATATTTACATTTTCAATAATAGGGTATCTAA
T--------------

OK_P11_05
AGACTTATCCTGCTTTAGAGCAGGAGCGACGGGCGATATGTGCATATTATAGAGCTTAAG
TCATATAGTCTAA-CTAAACTATATTACAATCAAATCCACTTTAAAAAATAGTGTTAGCT
ACTTTAACCATATAAATAAATTTATTGTAACCCATCTCTCCTTACCTATCCGCTGTATCT
TGATCTGATTTTCTCTTT----AACCTAATATCTTGAACATTCCTTCTCCTTACAAAACA

APPENDIX 3 (continued)

TTCAACCTACGACGATATACAAA-CCTTTAAA--GCAAGTC-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAACTACTAATACTCAAGCATTTATATTTACATTTTCAATAATAGGGTATCTAA
T--------------

OK_P12_01
AGACTTATCCTGCTTTAGAGCAGGAGCGACGGGCGATATGTGCATATTATAGAGCTTCAG
TCATATAGTTTAA-CTAAACTATATTACAATCAAATCCACTTTAAAAAATAGTGTTAGCT
ATTTCAACCATATAAATAAATTTATTGTAACCCATCTCTCCTTGTCTATATGCTGTATCT
TGATCTGATTTCCTCTCT----AACTTAATATCTTGAATATTCCTCTTCCTTTCAAAACA
TTCAACCTACGACGATATACAAA-CCTTTAAAA--CAAGTA-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAACTACTAATACTCAAGCATTTATATTTACATTTTCAATAATAGGGTATCTAA
T--------------

OK_P12_03
AGACTTATCCTGCTTTAGAGCAGGAGCGACGGGCGATATGTGCATATTATAGAGCTTCAG
TCATATAGTTTAA-CTAAACTATATTACAATCAAATCCACTTTAAAAAATAGTGTTAGCT
ATTTCAACCATATAAATAAATTTATTGTAACCCATCTCTCCTTGTCTATATGCTGTATCT
TGATCTGATTTCCTCTCT----AACTTAATATCTTGAATATTCCTCTTCCTTTCAAAACA
TTCAACCTACGACGATATACAAA-CCTTTAAAA--CAAGTA-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAACTACTAATACTCAAGCATTTATATTTACATTTTCAATAATAGGGTATNTAA
N--------------

OK_P12_04
AGACTTATCCTGCTTTAGAGCAGGAGCGACGGGCGATATGTGCATATTATAGAGCTTCAG
TCATATAGTTTAA-CTAAACTATATTACAATCAAATCCACTTTAAAAAATAGTGTTAGCT
ATTTCAACCATATAAATAAATTTATTGTAACCCATCTCTCCTTGTCTATATGCTGTATCT
TGATCTGATTTCCTCTCT----AACTTAATATCTTGAATATTCCTCTTCCTTTCAAAACA
TTCAACCTACGACGATATACAAA-CCTTTAAAA--CAAGTA-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAACTACTAATACTCAAGCATTTATATTTACNTTTTCAATAATNGGGTATCTAA
N--------------

OK_P13_01
AGACTTATCCTACTTTAGAGTAGGAGCGACGGGCGATATGTGCATATTATAGAGCTTCAG
TCATATAGTTTAA-CTAAACTATATTACAATCAAATCCACTTTAAAAGATAGTGTTAGCT
ACTTTAACCATATAAATAAATTTATTGTAACCCATCTCTCCTTGCCTATCCGCTGTATCT
TGATCTGATTTTCTCTCT----AATCTAATATCTTGAATATTCCTTTTCCTTACAAAACA
TTCAACCTACGACGATATACAAA-CCTTTTAA--GCAAGTA-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAGCTACTAATACTCAAGCATTTATATTTACATTTCCAATAATAGGGTATCTAA
A--------------

APPENDIX 3 (continued)

OK_P13_02

```
TGACTTATCCTACTTTAGAGTAGGAGCGACGGGCGATATGTGCATATTATAGAGCTTCAG
TCATATAGTTTAA-CTAAACTATATTACAATCAAATCCACTTTAAAAGATAGTGTTAGCT
ACTTTAACCATATAAATAAATTTATTGTAACCCATCTCTCCTTGCCTATCCGCTGTATCT
TGATCTGATTTTCTCTCT----AATCTAATATCTTGAATATTCCTTTTCCTTACAAAACA
TTCAACCTACGACGATATACAAA-CCTTTTAA--GCAAGTA-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAGCTACTAATACTCAAGCATTTATATTTACATTTCCAATAATAGGGTATCTAA
---------------
```

OK_P13_03

```
TGACTTATCCTACTTTAGAGTAGGAGCGACGGGCGATATGTGCATATTATAGAGCTTCAG
TCATATAGTTTAA-CTAAACTATATTACAATCAAATCCACTTTAAAAGATAGTGTTAGCT
ACTTTAACCATATAAATAAATTTATTGTAACCCATCTCTCCTTGCCTATCCGCTGTATCT
TGATCTGATTTTCTCTCT----AATCTAATATCTTGAATATTCCTTTTCCTTACAAAACA
TTCAACCTACGACGATATACAAA-CCTTTTAA--GCAAGTA-AGA-TTAATCGTGGATCA
TCAATTATAGGACAGGTTCCTCTGAATAGACTATAATACCGCCAAATCCTTTAAATTTCA
AGAACATAGCTACTAATACTCAAGCATTTATATTTACATTTCCAATAATAGGGTATCTAA
A--------------
```

CO1

OK_P01_05

```
-------------------------TCCCTAATCCTCCTATTGATAAGAAGATTAGTGGA
AAGAGGAGCTGGAACAGGTTGAACTGTTTATCCCCCCCTCTCATCTAATATTGCCCACAG
AGGAGCTTCTGTAGATTTAGCCATCTTTAGTCTCCACCTAGCTGGAGTCTCCTCGATCCT
AGGTGCAGTAAATTTTATTACTACAGTAATTAACATACGAGCAACAGGTATATCATTTGA
CCNGATACCCCTATTTGTCTGATCAGTTGTATTAACTGCCGTACTCCTGCTCCTGTCTCT
CCCAGTCCTTGCGGGAGCAATCACTATACTCCTTACAGATCGTAATATCAATACAACATT
CTTTGACCCAGCAGGTGGAGGAGATCC-TATCCTATACCAACACTTATTTTGG-TTCTTT
GGACACCCTGAAGTATACATTCTAA-TTCTCCCAGGATTTGGTATAATCTCTCATATCAT
TAGCCAAGAAAGAAGAAAAAAGGAAACTTTTGGAACCCTTGGTATAATCTACGCTATAAT
AGCAATTGGACTTCTAGGTTTTATTGTATGAGCTCACCACATATTTACAGTTGGTATAGA
CGTAGACACCCGGGCCTACTTCACTTCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCTACCCTCCATGGTTCACAACTAAACTATTCCCCCTCACT
ATTATGAGCACTAGGGTTCGTATTTCTCTTCACAGTAGGAGGTCTTACTGGAGTAATTCT
AGCTAACTCCTCAATTGATATTATTTTACACGACACTTATTATGTCGTTGCCCACTTCCA
CTATGTTCTTTCTATAGGAGCTGTGTTTGCAATTATAGGAGGATTTGTTCACTGGT-ATC
CTTTATTTACTGGTTTAACAATAAACCATAAATTCCTAAAAATCCAGTTCCTTACAATAT
TCATTGGAGTAAACACCACATTCTTCCCCCAACATT-TCTTAGGA-CTC--AGAGGAA-T
ACCTCGACGTTAC-TCTGACTATCCTGATGCGTATACTACATGAAATGTAATCTCTTCCA
TTGGATCCCTCATTTCCTTAATAAGAATTTTCATCTTTCTTTTTACTTTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTTGCACCTCTCAGACTACCCACATCCCTTGAATGATCTC
AACTTATACCTCCTGCCGAACACG------------------------------------
```

APPENDIX 3 (continued)

```
-----------

OK_P01_06
---------------------------------------------AAGAAGATTAGTGGA
AAGAGGAGCNGGAACAGGTTGAACTGTTTATCCCCCCCTCTCATCTAATATTGCCCACAG
AGGAGCTTCTGTAGATTTAGCCATCTTTAGTCTCCACCTAGCTGGAGTCTCCTCGATCCT
AGGTGCAGTAAATTTTATTACTACAGTAATTAACATACGAGCAACAGGTATATCATTTGA
CCAGATACCCCTATTTGTCTGATCAGTTGTATTAACTGCCGTACTCCTGCTCCTGTCTCT
CCCAGTCCTTGCGGGAGCAATCACTATACTCCTTACAGATCGTAATATCAATACAACATT
CTTTGACCCAGCAGGTGGAGGAGATCC-TATCCTATACCAACACTTATTTTGG-TTCTTT
GGACACCCTGAAGTATACATTCTAANTTCTCCCAGGATTTGGTATAATCTCTCATATCAT
TAGCCAAGAAAGAAGAAAAAAGGAAACTTTTGGAACCCTTGGTATAATCTACGCTATAAT
AGCAATTGGACTTCTAGGTTTTATTGTATGAGCTCACCACATATTTACAGTTGGTATAGA
CGTAGACACCCGGGCCTACTTCACTTCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCTACCCTCCATGGTTCACAACTAAACTATTCCCCCTCACT
ATTATGAGCACTAGGGTTCGTATTTCTCTTCACAGTAGGAGGTCTTACTGGAGTAATTCT
AGCTAACTCCTCAATTGATATTATTTTACACGACACTTATTATGTCGTTGCCCACTTCCA
CTATGTTCTTTCTATAGGAGCTGTGTTTGCAATTATAGGAGGATTTGTTCACTGAT-ATC
CTTTATTTACTGGTTTAACAATAAACCATAAATTCTTAAAAATCCAGTTCCTTACAATAT
TCATTGGAGTAAACACCACATTCTTCCCCCAACATT-TCTTAGGA-CTC--AGAGGAA-T
ACCTCGACGTTAC-TCTGACTATCCTGATGCGTATACTACATGAAATGTAATCTCTTCCA
TTGGATCCCTCATTTCCTTAATAAGAATTTTCATCTTTCTTTTTACTTTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTTGCACCTCTCAGACNACCCACATCCCTTGAATGATCTC
------------------------------------------------------------
-----------

OK_P02_01
-----GATTTTGGCTACTCCCCCCATCCCTAATCCTCCTTCTGATAAGAAGATTAGTGGA
AAGAGGAGCCGGAACAGGTTGAACTGTTTATCCCCCCCTCTCATCAAATATCGCCCATAG
AGGGGCTTCAGTAGACTTGGCAATCTTTAGACTCCATTTAGCTGGTGTCTCATCAATCCT
TGGTGCAGTAAACTTTATCACAACAGTAATCAATATACGAACAACTGGTATATCCTTTGA
TCAAATACCCCTATTTGTCTGATCAGTTGTTCTCACTGCTGTACTCCTCCTTCTATCCCT
TCCGGTCCTTGCGGGGGCAATCACTATACTCCTCACTGATCGTAATATCAACACTACATT
CTTTGATCCTGCAGGAGGAGGGGA-CCCTATCCTATATCAACACTTATTCTGA-TTCTTT
GGTCATCCCGAAGTTTATATTCTAA-TTCTTCCAGGATTTGGTATAATTTCCCACATCAT
CAGTCAAGAAAGAAGGAAAAAGGAAACCTTTGGAACTCTAGGAATAATCTACGCAATAAT
AGCAATTGGACTCCTAGGTTTTATTGTATGAGCCCACCATATATTCACAGTTGGTATAGA
TGTGGACACCCGGGCCTACTTTACATCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCTACCCTCCATGGTTCTCAACTAAATTATTCCCCATCATT
GTTATGAGCTCTAGGGTTCGTTTTCCTTTTTACAGTAGGGGGTCTTACTGGAGTTATTCT
GGCCAACTCCTCAATTGATATTATTTTACACGACACTTACTATGTTGTTGCTCACTTCCA
CTATGTGCTTTCTATAGGGGCTGTATTTGCAATTATAGGAGGATTCGTTCACTGAT-ATC
CCCTATTTACTGGCCTAACAATAAATCATAAATTCCTAAAAATCCAATTCCTTACTATAT
TCATTGGAGTAAACACTACATTCTTCCCCCAACATTTT-TTAGGA-TT--GAGAGGAA-T
ACCTCGACGTTAT-TCTGATTATCCTGATGCCTATACTACATGGAATGTAGTCTCCTCCA
```

```
TTGGATCCCTCATTTCCTTAATAAGAATTTTCATCTTTCTGTTTACTCTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTCGCACCTCTTAGATTATCCACTTCCCTTGAATGATCTC
AACTCATACCTCCTGCCGAACACAGATATGCN-AAATCCAATATG---------------
-----------
```

OK_P02_04

```
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-------------------------------------------AATATCAACACTACATT
CTTTGATCCTGCAGGAGGGGGAGA-CCCTATCCTATATCAACACTTATTCTGA-TTCTTT
GGTCATCCTGAAGTTTACATTCTAA-TTCTTCCAGGATTTGGTATAATTTCCCACATCAT
CAGTCAAGAAAGAAGAAAAAAGGAAACCTTTGGAACTCTAGGAATAATCTACGCAATAAT
AGCAATTGGACTCCTAGGTTTTATTGTATGAGCCCACCATATATTCACAGTTGGTATAGA
TGTGGACACCCGGGCCTACTTTACATCTGCCACAATAATTATTGCAGTTCCCACAGGAAT
TAAAATCTTCAGATGACTTGCTACCCTCCATGGTTCTCAACTAAATTATTCCCCATCATT
GTTATGAGCACTAGGGTTCGTTTTCCTTTTTACAGTAGGGGGTCTTACTGGAGTTATTCT
GGCCAACTCCTCAATTGATATTATTTTACACGACACTTACTATGTTGTTGCTCACTTCCA
CTATGTGCTTTCTATAGGGGCTGTATTTGCAATTATAGGAGGATTCGTTCACTGAT-ATC
CCCTATTTACTGGCCTAACAATAAACCATAAATTCCTAAAAATCCAATTCCTTACTATAT
TCATTGGAGTAAACACTACATTCTTCCCCCAACATTTT-TTAGGA-TT--GAGAGGAA-T
ACCTCGACGTTAT-TCTGACTATCCTGATGCCTATACTACATGGAATGTAGTCTCCTCCA
TTGGATCCCTCATCTCCTTAATAAGAATTTTCATCTTTCTGTTTACTCTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTTGCACCTCTTAGATTACCCACTTCCCTTGAATGATCTC
AACTCATACCTCCTGCCGAACAT-------------------------------------
-----------
```

OK_P02_05

```
-ATAAGATTTTGGCTACTCCCCCCATCCCTAATCCTCCTTCTGATAAGAAGATTAGTGGA
AAGAGGAGCCGGAACAGGTTGAACTGTTTATCCCCCCCTCTCATCAAATATCGCCCATAG
AGGGGCTTCAGTAGACTTGGCAATCTTTAGACTCCATTTAGCTGGTGTCTCATCAATCCT
TGGTGCAGTAAACTTTATCACAACAGTAATCAATATACGAGCAACTGGTATATCCTTTGA
TCAAATACCCCTATTTGTCTGATCAGTTGTTCTCACTGCTGTACTCCTCCTTCTATCCCT
TCCGGTCCTCGCGGGAGCAATCACTATACTCCTCACTGATCGTAATATCAACACTACATT
CTTTGATCCTGCAGGAGGAGGAGA-CCCTATCCTATATCAACACTTATTCTGA-TTCTTT
GGTCATCCTGAAGTTTATATTCTAA-TTCTTCCAGGATTTGGTATAATTTCCCACATCAT
CAGTCAAGAAAGAAGAAAAAAGGAAACCTTTGGAACTCTAGGAATAATCTACGCAATAAT
AGCAATTGGACTCCTAGGTTTTATTGTATGAGCCCACCATATATTCACAGTTGGTATAGA
TGTGGACACCCGGGCCTACTTTACATCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCTACCCTCCATGGTTCTCAACTAAATTATTCCCCATCATT
GTTATGAGCACTAGGGTTCGTTTTCCTTTTTACAGTAGGGGGTCTTACTGGAGTTATTCT
GGCCAACTCCTCAATTGATATTATTTTACACGACACTTACTATGTTGTTGCTCACTTCCA
CTATGTGCTTTCTATAGGGGCTGTATTTGCAATTATAGGAGGATTCGTTCACTGAT-ATC
```

APPENDIX 3 (continued)

```
CCCTATTTACTGGCCTAACAATAAACAAAAAATCCCAAAAAATCCAATTCCTTACTATAT
TCATTGGAGTAAACACTACATTCTTCCCCCAACATTTT-TTAGGA-TT--GAGAGGAA-T
ACCTCGACGTTAT-TCTGATTATCCTGATGCCTATACTACATGGAATGTAGTCTCCTCCA
TTGGATCCCTCATTTCCTTAATAAGAATTTTCATCTTTCTGTTTACTCTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTTGCACCTCTTAGATTATCCACTTCCCTTGAATGATCTC
AACTCATACCTCCTGCCGAACACAGATATGC-----------------------------
-----------

OK_P03_01
-------------------CCCNCATCCCTAATCCTCCTTCTGATAAGAAGATTAGTGGA
AAGAGGAGCCGGAACAGGTTGAACTGTTTATCCCCCCCTCTCATCAAATATCGCCCATAG
AGGGGCTTCAGTAGACTTGGCAATCTTTAGACTCCATTTAGCTGGTGTCTCATCAATCCT
TGGTGCAGTAAACTTTATCACAACAGTAATCAATATACGAACAACTGGTATATCCTTTGA
TCAAATACCCCTATTTGTCTGATCAGTTGTTCTCACTGCTGTACTCCTCCTTCTATCCCT
TCCGGTCCTTGCGGGGGCAATCACTATACTCCTCACTGATCGTAATATCAACACTACATT
CTTTGATCCTGCAGGAGGGGGGGA-CCCTATCCTATATCAACACTTATTCTGA-TTCTTT
GGTCATCCCGAAGTTTATATTCTAA-TTCTTCCAGGATTTGGTATAATTTCCCACATCAT
CAGTCAAGAAAGAAGGAAAAAGGAAACCTTTGGAACTCTAGGAATAATCTACGCAATAAT
AGCAATTGGACTCCTAGGTTTTATTGTATGAGCCCACCATATATTCACAGTTGGTATAGA
TGTGGACACCCGGGCCTACTTTACATCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCTACCCTCCATGGTTCTCAACTAAATTATTCCCCATCATT
GTTATGAGCTCTAGGGTTCGTTTTCCTTTTTACAGTAGGGGGTCTTACTGGAGTTATTCT
GGCCAACTCCTCAATTGATATTATTTTACACGACACTTACTATGTTGTTGCTCACTTCCA
CTATGTGCTTTCTATAGGGGCTGTATTTGCAATTATAGGAGGATTCGTTCACTGAT-ATC
CCCTATTTACTGGCCTAACAATAAATCATAAATTCCTAAAAATCCAATTCCTTACTATAT
TCATTGGAGTAAACACTACATTCTTCCCCCAACATTTT-TTAGGA-TT--GAGAGGAA-T
ACCTCGACGTTAT-TCTGATTATCCTGATGCCTATACTACATGGAATGTAGTCTCCTCCA
TTGGATCCCTCATTTCCTTAATAAGAATTTTCATCTTTCTGTTTACTCTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTCGCACCTCTTAGATTACCCACTTCCCTTGAATGATCTC
AACTCATACCTCCTGCCGAACACNGATATGC--A--------------------------
-----------

OK_P03_04
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------ATCCCT
TCCGGTCCTCGCGGGGGCAATCACTATACTCCTCACTGATCGTAATATCAACACTACATT
CTTTGATCCTGCAGGAGGAGGAGA-CCCTATCCTATATCAACACTTATTCTGA-TTCTTT
GGTCATCCTGAAGTTTATATTCTAA-TTCTTCCAGGATTTGGTATAATTTCCCACATCAT
CAGTCAAGAAAGAAGAAAAAAGGAAACCTTTGGAACTCTAGGAATAATCTACGCAATAAT
AGCAATTGGACTCCTAGGTTTTATTGTATGAGCCCACCATATATTCACAGTTGGTATAGA
TGTGGACACCCGGGCCTACTTTACATCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCTACCCTCCATGGTTCTCAACTAAATTATTCCCCATCATT
```

APPENDIX 3 (continued)

```
GTTATGAGCACTAGGGTTCGTTTTCCTTTTTACAGTAGGGGGTCTTACTGGAGTTATTCT
GGCCAACTCCTCAATTGATATTATTTTACACGACACTTACTATGTTGTTGCTCACTTCCA
CTATGTGCTTTCTATAGGGGCTGTATTTGCAATTATAGGAGGATTCGTTCACTGAT-ATC
CCCTATTTACTGGCCTAACAATAAATCATAAATTCCTAAAAATCCAATTCCTTACTATAT
TCATTGGAGTAAACACTACATTCTTCCCCCAACATTTT-TTAGGA-TT--GAGAGGAA-T
ACCTCGACGTTAT-TCTGATTATCCTGATGCCTATACTACATGGAATGTAGTCTCCTCCA
TTGGATCCCTCATTTCCTTAATAAGAATTTTCATCTTTCTGTTTACTCTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTCGCACCTCTTAGATTACCCACTTCCCTTGAATGATCTC
AACTCATNNNNCCTGNNGANCACT------------------------------------
-----------
```

OK_P03_05

```
------------GCTACTCCCCCCATCCCTAATCCTCCTTCTGATAAGAAGATTAGTGGA
AAGAGGAGCCGGAACAGGTTGAACTGTTTATCCCCCCCTCTCATCAAATATCGCCCATAG
AGGGGCTTCAGTAGACTTGGCAATCTTTAGACTCCATTTAGCTGGTGTCTCATCAATCCT
TGGTGCAGTAAACTTTATCACAACAGTAATCAATATACGAGCAACTGGTATATCCTTTGA
TCAAATACCCCTATTTNTCTGATCGGTTGTTCTCACTGCTGTACCCCCCCTTCTATCCCT
TCCGGTCCTCGCGGGGGCAATCACTATACTCCTCACTGATCGTAATATCAACACTACATT
CTTTGATCCTGCAGGAGGAGGAGA-CCCTATCCTATATCAACACTTATTCTGA-TTCTTT
GGTCATCCTGAAGTTTATATTCTAA-TTCTTCCAGGATTTGGTATAATTTCCCACATCAT
CAGTCAAGAAAGAAGAAAAAAGGAAACCTTTGGAACTCTAGGAATAATCTACGCAATAAT
AGCAATTGGACTCCTAGGTTTTATTGTATGAGCCCACCATATATTCACAGTTGGTATAGA
TGTAGACACCCGGGCCTACTTTACATCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCTACCCTCCATGGTTCTCAACTAAATTATTCCCCATCATT
GTTATGAGCACTAGGGTTCGTTTTCCTTTTTACAGTAGGGGGTCTTACTGGAGTTATTCT
GGCCAACTCCTCAATTGATATTATTTTACACGACACTTACTATGTTGTTGCTCACTTCCA
CTATGTGCTTTCTATAGGGGCTGTATTTGCAATTATAGGAGGATTCGTTCACTGAT-ATC
CCCTATTTACTGGCCTAACAATAAATCATAAATTCCTAAAAATCCAATTCCTTACTATAT
TCATTGGAGTAAACACTACATTCTTCCCCCAACATTTT-TTAGGA-TT--GAGAGGAA-T
ACCTCGACGTTAT-TCTGATTATCCTGATGCCTATACTACATGAAATGTAGTCTCCTCCA
TTGGATCCCTCATTTCCTTAATAAGAATTTTCATCTTTCTGTTTACTCTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTCGCACCTCTTAGATTACCCACTTCCCTTGAATGATCTC
AACTCATACCTCCTGCCGAACACAGATATAC-----------------------------
-----------
```

OK_P04_03

```
--------------GNCTCNCCNCATCCCTAATCCTCCTATTGATAAGAAGATTAGTGGA
AAGAGGAGCTGGAACAGGTTGAACTGTTTATCCCCCCCTCTCATCTAATATTGCCCACAG
AGGAGCTTCTGTAGATTTAGCCATCTTTAGTCTCCACCTAGCTGGAGTCTCCTCAATCCT
AGGTGCAGTAAATTTTATTACTACAGTAATTAACATACNAGCAACAGGCATATCCNTTGA
CCAGATNCCCCTATTTGTCTGATCAGTTGTATTAACCGCCGTACTCCTGCTCCTATCTCT
TCCAGTCCTTGCGGGAGCAATCACTATACTCCTTACAGATCGTAATATCAATACAACATT
CTTTGACCCAGCAGGTGGAGGAGATCC-TATCCTATACCAACACTTATTTTGG-TTCTTT
GGACACCCTGAAGTATACATTCTAA-TTCTCCCAGGATTTGGTATAATCTCTCATATCAT
TAGCCAAGAAAGAAGAAAAAAGGAAACTTTTGGGACCCTTGGTATAATCTATGCTATAAT
```

```
AGCAATTGGACTTCTAGGTTTTATTGTATGAGCTCACCACATATTTACAGTTGGTATAGA
CGTAGACACCCGGGCCTACTTCACTTCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCCACCCTCCATGGTTCACAATTAAACTATTCCCCCTCACT
ATTATGAGCACTAGGGTTTGTATTTCTATTCACAGTAGGAGGTCTTACTGGAGTAATTCT
AGCTAACTCCTCAATTGATATTATTTTACACGACACTTATTATGTCGTTGCCCACTTCCA
CTATGTTCTTTCTATAGGAGCTGTGTTTGCAATTATAGGAGGATTTGTTCACTGGT-ATC
CTTTATTTACTGGTTTAACAATAAACCATAAATTCCTAAAAATCCAGTTCCTTACAATAT
TCATTGGAGTAAACACCACATTCTTCCCCCAACATTT-CTTAGGA-CTC--AGAGGAA-T
ACCCCGACGTTAC-TCTGACTATCCTGATGCATATACTACATGAAATGTAGTCTCTTCCA
TTGGATCCCTCATTTCCTTAATAAGAATTTTCATCTTTCTTTTTACTTTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTTGCACCTCTCAGACTACCCACATCCCTTGAATGATCTC
AACTTATACCTCCTGCCGAACACAGNTACGCTAAATACCATNC-----------------
-----------
```

OK_P05_01

```
-------------------------TCCCTAATCCTCCTATTGATAAGAAGATTAGTGGA
AAGAGGAGCTGGAACAGGTTGAACTGTTTATCCTCCCCTCTCATCTAATATTGCCCACAG
AGGAGCTTCTGTAGATTTAGCCATCTTTAGTCTCCACCTAGCTGGAGTCTCCTCAATCCT
AGGTGCAGTAAATTTTATTACTACAGTAATTAACATACGAGCAACAGGCATATCCTTTGA
CCAGATACCCCTATTTGTCTGATCAGTTGTATTAACCGCCGTACTCCTGCTCCTATCTCT
TCCAGTCCTTGCGGGAGCAATCACTATACTCCTTACAGATCGTAATATCAATACAACATT
CTTTGACCCAGCAGGTGGAGGAGATCC-TATCCTATACCAACACTTATTTTGG-TTCTTT
GGACACCCTGAAGTATACATTCTAA-TTCTCCCAGGATTTGGTATAATCTCTCATATCAT
TAGCCAAGAAAGAAGAAAAAAGGAAACTTTTGGGACCCTTGGTATAATCTATGCTATAAT
AGCAATTGGACTTTTAGGTTTTATTGTATGAGCTCACCACATATTTACAGTTGGTATAGA
CGTAGACACCCGGGCCTACTTCACTTCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCCACCCTCCATGGTTCACAATTAAACTATTCCCCCTCACT
ATTATGAGCACTAGGGTTTGTATTTCTATTCACAGTAGGAGGTCTTACTGGAGTAATTCT
AGCTAACTCCTCAATTGATATTATTTTACACGACACTTATTATGTCGTTGCCCATTTCCA
CTATGTTCTTTCTATAGGAGCTGTGTTTGCAATTATAGGAGGATTTGTTCACTGGT-ATC
CTTTATTTACTGGTTTAACAATAAACCATAAATTCCTAAAAATCCAGTTCCTTACAATAT
TCATTGGAGTAAACACCACATTCTTCCCCCAACATTT-CTTAGGA-CTC--AGAGGAA-T
ACCCCGACGTTAC-TCTGACTATCCTGATGCATATACTACATGAAATGTAGTCTCTTCCA
TTGGATCCCTCATTTCCTTAATAAGAATTTTCATCTTTCTTTTTACTTTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTTGCACCTCTCAGACTACCCACATCCCTTGAATGATCTC
AACTTATACCTCCTGCCGAACACAGCTACGC-----------------------------
-----------
```

OK_P05_03

```
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
```

```
--------------------------------------CAACNTTTATTTTGA-TTTTTT
GGACACCCTGAAGTATACATTCTAA-TTCTCCCAGGATTTGGTATAATCTCTCATATCAT
TAGCCAAGAAAGAAGAAAAAAGGAAACTTTTGGGACCCTTGGTATAATCTATGCTATAAT
AGCAATTGGACTTCTAGGTTTTATTGTATGAGCTCACCACATATTTACAGTTGGTATAGA
CGTAGACACCCGGGCCTACTTCACTTCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCCACCCTCCATGGTTCACAATTAAACTATTCCCCCTCACT
ATTATGAGCACTAGGGTTTGTATTTCTATTCACAGTAGGAGGTCTTACTGGAGTAATTCT
AGCTAACTCCTCAATTGATATTATTTTACACGACACTTATTATGTCGTTGCCCACTTCCA
CTATGTTCTTTCTATAGGAGCTGTGTTTGCAATTATAGGAGGATTTGTTCACTGAT-ATC
CTTTATTTACTGGTTTAACAATAAACCATAAATTCCTAAAAATCCAGTTCCTTACAATAT
TCATTGGAGTAAACACCACATTCTTCCCCCAACATTT-CTTAGGA-CTC--AGAGGAA-T
ACCCCGACGTTAC-TCTGACTATCCTGATGCATATACTACATGAAATGTAGTCTCTTCCA
TTGGATCCCTCATTTCCTTAATAAGAATTTTCATCTTTCTTTTTACTTTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTTGCACCTCTCAGACTACCCACATCCCTTGAATGATCTC
AACTTATACCTCCTGCCGAACACAGCTACGACGAAATTCCTATAATTTCTAACTTCTAAT
ATGGCAGATTA
```

OK_P05_05

```
---------CTGGCTACTCCCCCCATCCCTAATCCTCCTATTGATAAGAAGATTAGTAGA
AAGAGGAGCTGGAACAGGTTGAACTGTTTATCCCCCCCTCTCATCTAATATTGCCCACAG
AGGAGCTTCTGTAGATTTAGCCATCTTTAGTCTCCACCTAGCTGGAGTCTCCTCAATCCT
AGGTGCAGTAAATTTTATTACTACAGTAATTAACATACGAGCAACAGGTATAGCCTTTGA
CCAGATACCCCTATTTGTCTGATCAGTTGTATTAACCGCCGTACTCCTGCTCCTGTCTCT
CCCAGTCCTTGCGGGAGCAATCACTATACTCCTTACAGATCGTAATATCAACACAACATT
CTTTGACCCAGCAGGTGGAGGAGATCC-TATCCTATACCAACACTTATTTTGG-TTCTTT
GGACACCCTGAAGTATACATTTTAA-TTCTCCCAGGATTTGGTATAATCTCTCATATCAT
TAGCCAAGAAAGAAGAAAAAAGGAAACTTTTGGAACCCTTGGTATAATCTACGCTATAAT
AGCAATTGGACTTCTAGGTTTTATTGTATGAGCTCACCACATATTTACAGTTGGTATAGA
CGTAGACACCCGAGCCTACTTCACTTCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCTACCCTCCATGGTTCACAATTAAACTATTCCCCCTCACT
ATTATGAGCACTAGGGTTTGTATTTCTCTTCACAGTAGGGGGTCTTACTGGAGTAATTCT
AGCTAACTCCTCAATTGATATTATTTTACACGACACTTATTATGTCGTTGCCCACTTCCA
CTATGTTCTTTCTATAGGAGCTGTGTTTGCAATTATAGGAGGATTTGTTCACTGAT-ATC
CTTTATTTACTGGTTTAACAATAAACCATAAATTCCTAAAAATCCAGTTCCTTACAATAT
TCATTGGAGTAAACACCACATTCTTCCCCCAACATTT-CTTAGGA-CTC--AGAGGAA-T
ACCTCGACGTTAC-TCTGACTATCCTGATGCATATACTACATGAAATGTAGTCTCTTCCA
TTGGATCCCTCATTTCCTTAATAAGAATTTTTATCTTTCTTTTTACTTTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTTGCACCTCTCAGACTACCTACATCCCTTGAATGATCTC
AACTTATACCTCCTGCCGAACACAGCTATG------------------------------
-----------
```

OK_P06_01

```
-------------------------TCCCTAATCCTCCTATTGATAAGAAGATTAGTGGA
AAGAGGAGCTGGAACAGGTTGAACTGTTTATCCCCCCCTCTCATCTAATATTGCCCACAG
AGGAGCTTCTGTAGATTTAGCCATCTTTAGCCTCCACCTAGCTGGAGTCTCCTCAATCCT
```

APPENDIX 3 (continued)

```
AGGTGCAGTAAATTTTATTACTACAGTAATTAACATACGAGCAACAGGTATATCCTTTGA
CCAGATACCCCTATTTGTCTGATCAGTTGTATTAACCGCCGTACTCCTGCTCCTGTCTCT
CCCAGTCCTTGCGGGAGCAATCACTATACTCCTTACAGATCGTAATATCAATACAACATT
CTTTGACCCAGCAGGTGGAGGTGATCC-TATCCTATACCAACACTTATTTTGG-TTCTTT
GGGCACCCTGAAGTATACATTCTAA-TTCTCCCAGGATTTGGTATAATCTCTCATATCAT
TAGCCAAGAAAGAAGAAAAAAGGAAACTTTTGGAACCCTTGGTATAATCTACGCTATAAT
AGCAATTGGACTTCTAGGTTTTATTGTATGAGCTCACCACATATTTACAGTTGGTATAGA
CGTAGACACCCGGGCCTACTTCACTTCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCTACCCTCCATGGTTCACAATTAAACTATTCCCCCTCACT
ATTATGAGCACTAGGATTCGTGTTTCTCTTCACAGTAGGAGGTCTTACTGGAGTAATTCT
AGCTAACTCCTCAATTGATATTATTTTACACGACACTTATTATGTCGTTGCCCACTTCCA
CTATGTTCTTTCTATAGGAGCTGTGTTTGCAATTATAGGAGGATTTGTTCACTGGT-ATC
CTTTATTTACTGGTTTAACAATAAACCATAAATTCCTAAAAATCCAGTTCCTTACAATAT
TCATTGGAGTAAACACCACATTCTTCCCCCAACATTT-CTTAGGA-CTC--AGAGGAA-T
ACCTCGACGTTAC-TCTGACTATCCTGATGCGTATACTACATGAAATGTAGTCTCTTCCA
TTGGATCCCTCATTTCCTTAATAAGAATTTTCATCTTTCTTTTTACTTTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTTGTACCTCTCAGACTACCCACATCCCTTGAATGATCCC
AACTTATACCTCCTGCCGAACAC-------------------------------------
-----------
```

OK_P06_03

```
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
---TGCAGTAAATTTTATTACTACAGTAATTAACATACNAGCAANANGTATATCCTTTGA
CCAGATACCCCTATTTGTCTGATCAGTTGTATTAACCGCCGTACTCCNGCTCCNGTCTCT
CCCAGTCCTTGCGGGAGCAATCACTATACTCCTTACAGATCGTAATATCAATACAACATT
CTTTGACCCAGCAGGTGGGGGTGATCC-TATCCTATACCAACACTTATTTTGG-TTCTTT
GGACACCCTGAAGTATACATTCTAA-TTCTCCCAGGATTTGGTATAATCTCTCATATCAT
TAGCCAAGAAAGAAGAAAAAAGGAAACTTTTGGCACCCTTGGTATAATCTATGCTATAAT
AGCAATTGGACTTCTAGGTTTTATTGTATGAGCTCACCACATATTTACAGTTGGTATAGA
CGTAGACACCCGGGCCTACTTCACTTCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCTACCCTCCATGGTTCACAATTAAACTATTCCCCCTCACT
ATTATGAGCACTAGGGTTTGTATTTCTCTTCACAGTAGGAGGTCTTACTGGAGTAATTCT
AGCTAACTCCTCAATTGATATTATTTTACACGACACTTATTATGTCGTTGCCCACTTCCA
CTATGTTCTTTCTATAGGAGCTGTATTTGCAATTATAGGAGGATTTGTTCACTGAT-ATC
CTTTATTTACTGGTTTAACAATAAACCATAAATTCCTAAAAATCCAGTTCCTTACGATAT
TCATTGGAGTAAACACCACATTCTTCCCCCAACATTT-CTTAGGA-CTC--AGAGGAA-T
ACCTCGACGTTAC-TCTGACTATCCTGATGCNTATACTACATGAAATGTAGTCTCTTCCA
TTGGATCCCTCATTTCCTTAATAAGAATTTTCATCTTTCTTTTTACTTTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTTGCACCTCTCAGACTACCCACATCCCTTGAATGATCTC
AACTTATACCTCCTGCCGAACAC-------------------------------------
-----------
```

APPENDIX 3 (continued)

OK_P06_05

```
-------------------------TCCCTAATCCTCCTATTGATAAGAAGATTAGTGGA
AAGAGGAGCTGGAACAGGTTGAACTGTTTATCCCCCCCTCTCATCTAATATTGCCCACAG
AGGAGCTTCTGTAGATTTAGCCATCTTTAGTCTCCACCTAGCTGGGATCTCCTCAATCCT
AGGTGCAGTAAATTTTATTACTACAGTAATTAACATACGAGCAACAGGTATATCCCTTGA
ACAAAAAACCCCAATTGTCTGATCAGTTGTATTGACCGCCGTACTCCTGCTCCTGTCTCT
CCCAGTCCTTGCGGGAGCAATCACTATACTCCTTACAGATCGTAATATCAATACAACATT
CTTTGACCCAGCAGGTGGAGGAGATCC-TATCCTATACCAACACTTATTTTGG-TTCTTT
GGACACCCTGAAGTATACATTCTAA-TTCTCCCAGGATTTGGTATAATCTCTCATATCAT
TAGCCAAGAAAGAAGAAAAAAGGAAACTTTTGGAACCCTTGGTATAATCTACGCTATAAT
AGCAATTGGACTTCTAGGTTTTATTGTATGAGCTCACCACATATTTACAGTTGGTATAGA
CGTAGACACCCGGGCCTACTTCACTTCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCTACCCTCCATGGTTCACAATTAAACTATTCCCCCTCACT
ATTATGAGCACTAGGGTTCGTATTTCTCTTCACAGTAGGAGGTCTTACTGGAGTAATTCT
AGCTAACTCCTCAATTGATATTATTTTACACGACACTTATTATGTCGTTGCCCACTTCCA
CTATGTTCTTTCTATAGGGGCTGTGTTTGCAATTATAGGAGGATTTGTTCACTGGT-ATC
CTTTATTTACTGGTTTAACAATAAACCATAAATTCCTAAAAATCCAGTTCCTTACAATAT
TCATTGGAGTAAACACCACATTCTTCCCCCAACATTT-CTTAGGA-CTC--AGAGGAA-T
ACCTCGACGTTAC-TCTGACTATCCTGATGCGTATACTACATGAAATGTAGTCTCTTCCA
TTGGATCCCTCATTTCCTTAATAAGAATTTTCATCTTTCTTTTTACTTTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTTGCACCTCTCAGACTACCCACATCCCTTGAATGATCTC
AACTTATACCTCCTGCCGAACAC-------------------------------------
-----------
```

OK_P07_05

```
ANTAGGATTCTGGCTACTCCCCCCATCCCTAATCCTCCTATTGATAAGAAGATTAGTGGA
AAGAGGAGCTGGAACAGGTTGAACTGTTTATCCCCCCCTCTCATCTAATATTGCCCACAG
AGGAGCTTCTGTAGATTTAGCCATCTTTAGTCTCCACCTAGCTGGAGTCTCCTCAATCCT
AGGTGCAGTAAATTTTATTACTACAGTAATTAACATACGAGCAACAGGCATATCCTTTGA
CCAGATACCCCTATTTGTCTGATCAGTTGTATTAACCGCCGTACTCCTGCTCCTGTCTCT
CCCAGTCCTTGCGGGAGCAATCACTATACTCCTTACAGATCGTAATATCAATACAACATT
CTTTGATCCAGCAGGTGGGGGAGATCCC-ATCCTATACCAACACTTATTTTGG-TTCTTT
GGACACCCTGAAGTATACATTCTAA-TTCTCCCAGGATTTGGTATAATCTCTCATATCAT
TAGCCAAGAAAGAAGAAAAAAGGAAACTTTTGGAACCCTTGGTATAATCTACGCTATAAT
AGCAATTGGACTTCTAGGTTTTATTGTATGAGCTCACCACATATTTACAGTTGGTATAGA
CGTAGACACCCGGGCCTACTTTACTTCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCTACCCTCCATGGTTCACAATTAAACTATTCCCCCTCACT
ATTATGAGCACTAGGGTTTGTATTTCTCTTCACAGTAGGAGGTCTTACTGGAGTAATTCT
AGCTAACTCCTCAATTGATATTATCTTACACGACACTTATTATGTCGTTGCCCACTTCCA
CTATGTCCTTTCTATAGGAGCTGTGTTTGCAATTATAGGAGGATTTGTTCACTGGT-ATC
CTTTATTTACTGGTTTAACAATAAACCATAAATTCCTAAAAATCCAGTTCCTTACAATAT
TCATTGGAGTAAACACCACATTCTTCCCCCAACATTT-CTTAGGA-CTC--AGAGGAA-T
ACCTCGACGTTAC-TCTGACTATCCTGATGCATATACTACATGAAATGTAGTCTCTTCCA
TTGGATCCCTCATTTCCTTAATAAGAATTTTCATCTTTCTTTTTACTTTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTTGCACCTCTCAGACTACCCACATCCCTTGAATGATCTC
```

```
AACTTATACCTCCTGCCGAACACAGCTATG------------------------------
-----------

OK_P08_01
-------------------CCCNCGTCCCTAATCCTCCTTCTGATAAGAAGATTAGTGGA
AAGAGGAGCCGGAACAGGTTGAACTGTTTATCCCCCCCTCTCATCAAATATCGCCCATAG
AGGAGCTTCAGTAGACTTGGCAATCTTTAGACTCCATTTAGCTGGTGTCTCATCAATCCT
TGGTGCAGTAAACTTTATCACAACAGTAATCAACATACGAGCAACTGGTATATCCTTTGA
TCAAATACCCCTATTTGTCTGATCAGTTGTTCTCACTGCTGTACTCCTCCTTCTATCCCT
TCCGGTCCTCGCGGGGGCAATCACTATACTCCTCACTGATCGTAATATCAACACTACATT
CTTTGATCCTGCAGGAGGAGGAGA-CCCTATCCTATACCAACACTTATTCTGA-TTCTTT
GGTCATCCTGAAGTTTATATTCTAA-TTCTTCCAGGATTTGGTATAATTTCCCACATCAT
CAGTCAAGAAAGAAGAAAAAAGGAAACCTTTGGAACTCTTGGAATAATCTACGCAATAAT
AGCAATTGGACTCCTAGGTTTTATTGTATGAGCCCACCACATATTCACAGTTGGTATAGA
TGTGGACACCCGGGCCTACTTTACATCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCTACCCTCCATGGTTCTCAACTAAATTATTCCCCATCATT
GTTATGAGCACTGGGGTTCGTTTTCCTTTTTACAGTGGGGGGTCTTACTGGAGTTATTCT
GGCCAACTCCTCAATTGATATTATTTTACACGACACTTACTATGTTGTTGCTCACTTCCA
CTATGTACTCTCTATAGGGGCTGTATTTGCAATTATAGGAGGATTTGTTCACTGAT-ACC
CCCTATTTACTGGCCTAACAATAAATCATAAATTCCTAAAAATCCAATTCCTTACTATAT
TCATTGGAGTAAACACTACATTCTTCCCCCAACATTTT-TTAGGA---TTGAGAGGAA-T
ACCTCGACGTTAT-TCTGACTATCCTGATGCCTATACTACATGAAATGTAGTCTCCTCCA
TCGGATCCCTCATCTCCTTAATAAGAATTTTCATCTTTCTGTTTACTCTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTTGCACCTCTTAGATTACCCACTTCCCTTGAATGATCTC
AACTCATACCTCCTGCCGAACAC-------------------------------------
-----------

OK_P08_03
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-----------------GGGGAGA-CNCTATCCTATATCANNANTTANNCTNA-TNCTTT
GGTCATCCTGAAGTTTATATTTTAA-TTCTCCCAGGATTTGGTATAATTTCCCACATCAT
CAGTCAAGAAAGAAGAAAAAAGGAAACCTTTGGAACCCTAGGAATAATCTACGCAATAAT
AGCAATTGGACTCCTAGGTTTTATTGTATGAGCCCACCATATATTCACAGTNGGTATAGA
TGTAGACACCCGGGCCTACTTTACATCTGCCACAATAATTATNGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCTACCCTCCATGGTTCTCAACTAAATTATTCCCCATCATN
GTTATGAGCACTAGGGTTNGTTTTCCTTTTNACAGTAGGGGGTCNTACTGGAGTTATTNT
GGCCAACTCCTCAATTGATATTATTTTACACGACACTTANTATGTTGTTGCTCACTTNCA
CNATNTGCTTTCTATAGGGGCTGTATTTGCAATNATAGGAGGATTCGTTCNCTGAT-ACC
CCCTATTTACTGGCCTAACAATAAATCATAAATTCCTAAAAATCCAATTCCTTACTATAT
TCATCGGAGTAAACACTACATTCTTCCCCCAACATTTT-TTANGA---TTGAGAGGAA-T
```

```
ACCTNGCCGTTAT-TCTGATTATCCTGATGCCTATACTACATGGAATGTAGTCTCCTCCA
TNGGATCCCTCATCTCCTTAATAAGAATTTTCATCTTTCTGTTTACTNTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTTGCGCCTCTTAGATTACCCACTTCCCTTGAATGATCTC
AACTCATACCTCCTGCCGAACACAGATANGCT----------------------------
-----------

OK_P08_04
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
--------------------------------------------------------CCCT
TCCGGTCCTCGCGGGGGCAATCACTATACTCCTCACTGATNGTAATATCAACACTACATT
CTTTGATCCTGCAGGAGGGGGAGA-CCCTATCCTATATCAACACTTATTCTGA-TTCTTT
GGTCATCATGAAGTTTATATTTTAA-TTCTTCCAGGATTTGGTATAATTTCCCACATCAT
CAGTCAAGAAAGAAGAAAAAAGGAAACCTTTGGAACCCTAGGAATAATCTACGCAATAAT
AGCAATTGGACTCCTAGGTTTTATTGTATGAGCCCACCATATATTCACAGTTGGTATAGA
TGTGGACACCCGGGCCTACTTTACATCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCTACCCTCCATGGTTCTCAACTAAATTATTCCCCATCATT
GTTATGAGCACTAGGGTTCGTTTTCCTTTTTACAGTAGGGGGTCTTACTGGAGTTATTCT
GGCCAACTCCTCAATTGATATTATTTTACACGACACTTACTATGTTGTTGCTCACTTCCA
CTATGTGCTTTCTATGGGGGCTGTATTTGCAATTATAGGAGGATTCGTTCACTGAT-ACC
CCCTATTTACTGGCCTAACAATAAATCATAAATTCCTAAAAATCCAATTCCTTACTATAT
TCATTGGAGTAAACACTACATTCTTCCCCCAACATTTT-TTAGGA---TTGAGAGGAA-T
ACCTCGACGTTAT-TCTGATTATCCTGATGCCTATACTACATGGAATGTAGTCTCCTCCA
TTGGATCCCTCATCTCCTTAATAAGAATTTTCATCTTTCTGTTTACTCTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTTGCACCTCTTAGATTACCCACTTCCCTTGAATGATCTC
AACTCATNCCTCCTGCCGAACAG-------------------------------------
-----------

OK_P09_03
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
---------------------TGCAGGNATTAACATACNAGCAANGNGTATATNCNTTGA
CAGGATNCCCCTATTTGTCTGATCNGTTGTATNAAANGCNNTACTCNGGGNTCCG-CCCC
CCCAGTCCTTGCGGGAGCAATCACTATACTCCTTACAGATCGTAATATCAATACAACATT
CTTTGACCCAGCAGGTGGGGGAGATCC-TATCCTATACCAACACTTATTTTGGTTTCTTT
GGACACCCTGAAGTATACATTCTAA-TTCTTCCAGGATTTGGTATAATCTCTCAAATCAT
TAGCCAAGAAAGAAGAAAAAAGGAAACTTTTGGGACCCTTGGTATAATCTACGCTATAAT
AGCAATTGGACTTCTAGGTTTTATTGTATGAGCTCACCACATATTTACGGTTGGTATAGA
CGTAGACACTCGGGCCTACTTCACTTCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCTACCCTCCATGGTTCACAATTAAACTATTCCCCCTCACT
ATTATGAGCACTAGGGTTCGTATTTCTCTTCACAGTAGGAGGTCTTACTGGAGTAATTCT
AGCTAACTCCTCAATTGATATTATTTTACACGACACTTATTATGTCGTTGCCCACTTCCA
```

APPENDIX 3 (continued)

```
CTATGTTCTTTCTATAGGAGCTGTGTTTGCAATTATAGGAGGATTTGTTCACTGGT-ATC
CTTTATTTACTGGTTTAACAATAAACCATAAATTCCTAAAAATCCAGTTCCTTACAATAT
TCATTGGAGTAAACACCACATTCTTCCCCCAACATTTC-TTAGGG-CTC--AGAGGAA-T
ACCTCGACGTTAC-TCTGACTATCCTGATGCATATACTACATGAAATGTAGTCTCTTCCA
TTGGATCCCTCATTTCCTTAATAAGAATTTTCATCTTTCTTTTTACTTTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTTGCACCTCTCAGACTACCCTCATCCCTTGAATGATCTC
AACTTATACCTCCTGCCGAACACT------------------------------------
-----------

OK_P09_04
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
---TGCANTAAATTTTATTACTNCAGNAATTAACATACNAGCAACNGGTATATCCTTTGA
CCANATACCCCTATTTGTCTGATCAGTNGTATTAACCGCCGTACTCCGGCNCCNGTCTCT
CCCAGTCCTTGCGGGAGCAATCACTATACTCCTTACAGATCGTAATATCAATNCAACATT
CTTTGACCCAGCAGGTGGGGGAGATCC-TATCCTATACCAACACTTATTTTGG-TTCTTT
GGACACCCTGAAGTATACATTCTAA-TTCTTCCAGGATTTGGTATAATCTCTCATATCAT
TAGCCAAGAAAGAAGAAAAAAGGAAACTTTTGGGACCCTTGGTATAATCTACGCTATAAT
AGCAATTGGACTTCTAGGTTTTATTGTATGAGCTCACCACATATTTACGGTTGGTATAGA
CGTAGACACCCGAGCCTACTTCACTTCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCTACCCTCCATGGTTCACAATTAAACTATTCCCCCTCACT
ATTATGAGCACTAGGGTTCGTATTTCTCTTCACAGTAGGAGGTCTTACTGGAGTAATTCT
AGCTAACTCCTCAATTGATATTATTTTACACGACACTTATTATGTCGTTGCCCACTTCCA
CTATGTTCTTTCTATAGGAGCTGTGTTTGCAATTATAGGAGGATTTGTTCACTGGT-ATC
CTTTATTTACTGGTTTAACAATAAACCATAAATTCCTAAAAATCCAGTTCCTTACAATAT
TCATTGGAGTAAACACCACATTCTTCCCCCAACATTT-CTTAGGA-CTC--AGAGGAA-T
ACCTCGACGTTAC-TCTGACTATCCTGATGCATATACTACATGAAATGTAGTCTCTTCCA
TTGGATCCCTCATTTCCTTAATAAGAATTTTCATCTTTCTTTTTACTTTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTTGCACCTCTCAGACTACCCTCATCCCTTGAATGATNTC
AACTTATACCTCCTGCCGAACACT------------------------------------
-----------

OK_P09_05
-------------------------TCCCTAATCCTCCTATTGATAAGAAGATTAGTGGA
AAGAGGAGCTGGAACAGGTTGAACTGTTTATCCCCCCCTCTCATCTAATATTGCCCACAG
AGGAGCTTCTGTAGATTTAGCCATCTTTAGTCTCCACCTAGCTGGAGTCTCCTCAATCCT
AGGTGCAGTAAATTTTATTACTACAGTAATTAACATACGAGCAACAGGTATATCCTTTGA
CCAGATACCCCTATTTGTCTGATCAGTTGTATTAACCGCCGTACTCCTGCTCCTGTCTCT
CCCAGTCCTTGCGGGAGCAATCACTATACTCCTTACAGATCGTAATATCAATACAACATT
CTTTGACCCAGCAGGTGGGGGAGATCC-TATCCTATACCAACACTTATTTTGG-TTCTTT
GGACACCCTGAAGTATACATTCTAA-TTCTTCCAGGATTTGGTATAATCTCTCATATCAT
TAGCCAAGAAAGAAGAAAAAAGGAAACTTTTGGGACCCTTGGTATAATCTACGCTATAAT
AGCAATTGGACTTCTAGGTTTTATTGTATGAGCTCACCACATATTTACGGTTGGTATAGA
CGTAGACACTCGGGCCTACTTCACTTCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
```

APPENDIX 3 (continued)

```
TAAAATCTTCAGATGACTTGCTACCCTCCATGGTTCACAATTAAACTATTCCCCCTCACT
ATTATGAGCACTAGGGTTCGTATTTCTCTTCACAGTAGGAGGTCTTACTGGAGTAATTCT
AGCTAACTCCTCAATTGATATTATTTTACACGACACTTATTATGTCGTTGCCCACTTCCA
CTATGTTCTTTCTATAGGAGCTGTGTTTGCAATTATAGGAGGATTTGTTCACTGGT-ATC
CTTTATTTACTGGTTTAACAATAAACCATAAATTCCTAAAAATCCAGTTCCTTACAATAT
TCATTGGAGTAAACACCACATTCTTCCCCCAACATTT-CTTAGGG-CTC--AGAGGAA-T
ACCTCGACGTTAC-TCTGACTATCCTGATGCATATACTACATGAAATGTAGTCTCTTCCA
TTGGATCCCTCATTTCCTTAATAAGAATTTTCATCTTTCTTTTTACTTTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTTGCACCTCTCAGACTACCCTCATCCCTTGAATGATCTC
AACTTATACCTCCTGCCGAACACNGCTATGC-----------------------------
------------
```

OK_P10_06

```
-----------------TCCCCCCATCCCTAATCCTCCTATTGATAAGAAGATTAGTGGA
AAGAGGAGCTGGAACAGGTTGAACTGTTTATCCCCCCCTCTCATCTAATATTGCCCACAG
AGGAGCTTCTGTAGATTTAGCCATCTTTAGCCTCCACCTAGCTGGAGTCTCCTCAATCCT
AAGTCCAGTAAAGTAAATTACAACACTAAAGAACAAAAGACCAACAACAAGAACATCCGA
CAACAGACACCCAATTGTCTGATCAGTTGTATTAACCGCCTTACTCCTGCTCCTGTCTCT
CCCAGTCCTTGCGGGAGCAATCACTATACTCCTTACAGATCGTAATATCAATACAACATT
CTTTGACCCAGCAGGTGGAGGTGATCC-TATCCTATACCAACACTTATTTTGG-TTCTTT
GGACACCCTGAAGTATACATTCTAA-TTCTCCCAGGATTTGGTATAATCTCTCATATCAT
TAGCCAAGAAAGAAGAAAAAAGGAAACTTTTGGGACCCTTGGTATAATCTATGCTATAAT
AGCAATTGGACTTCTAGGTTTTATTGTATGAGCTCACCACATATTTACAGTTGGTATAGA
CGTAGACACCCGGGCCTACTTCACTTCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCTACCCTCCATGGTTCACAATTAAACTATTCCCCCTCACT
ATTATGAGCACTAGGGTTTGTATTTCTCTTCACAGTAGGAGGTCTTACTGGAGTAATTCT
AGCTAACTCCTCAATTGATATTATTTTACACGACACTTATTATGTCGTTGCCCACTTCCA
CTATGTTCTTTCTATAGGAGCTGTGTTTGCAATTATAGGAGGATTTGTTCACTGGT-ATC
CTTTATTTACTGGTTTAACAATAAACCATAAATTCCTAAAAATCCAGTTCCTTACAATAT
TCATTGGAGTAAACACCACATTCTTCCCCCAACATTT-CTTAGGA-CTC--AGAGGAA-T
ACCTCGACGTTAC-TCTGACTATCCTGATGCGTATACTACATGAAATGTAGTCTCTTCCA
TTGGATCCCTCATTTCCTTAATAAGAATTTTCATCTTTCTTTTTACTTTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTTGCACCTCTCAGACTACCCACATCCCTTGAATGATCTC
AACTTATACCTCCTGCCGAACACNGCTATGCTG---------------------------
------------
```

OK_P10_07

```
ANTATGATTCTGGCTACTCCCCCCATCCCTAATCCTCCTATTGATAAGAAGATTAGTGGA
AAGAGGAGCTGGAACAGGTTGAACTGTTTATCCCCCCCTCTCATCTAATATTGCCCACAG
AGGAGCTTCTGTAGATTTAGCCATCTTTAGTCTCCACCTAGCTGGAGTCTCCTCAATCCT
AGGTGCAGTAAATTTTATTACTACCGTAATTAACATACGAGCAACAGGTATAGCCTTTGA
CCAGATACCCCTATTTGTCTGATCAGTTGTATTAACCGCCGTACTTCTGCTCCTGTCTCT
CCCAGTCCTTGCGGGAGCAATCACTATACTCCTTACAGATCGTAATATCAATACAACATT
CTTTGACCCAGCAGGTGGAGGAGATCC-TATCCTATACCAACACTTATTTTGG-TTCTTT
GGACACCCTGAAGTATACATTCTAA-TTCTCCCAGGATTTGGTATAATCTCTCATATCAT
```

APPENDIX 3 (continued)

```
TAGCCAAGAAAGAAGAAAAAAGGAAACTTTTGGAACCCTTGGTATAATCTACGCTATAAT
AGCAATTGGACTTCTAGGTTTTATTGTATGAGCTCACCACATATTTACAGTTGGTATAGA
CGTAGACACCCGAGCCTACTTCACCTCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCTACCCTCCATGGTTCACAATTAAACTATTCCCCCTCACT
ATTATGAGCACTAGGGTTTGTATTTCTCTTCACAGTAGGGGGTCTTACTGGAGTAATTCT
AGCTAACTCCTCAATTGATATTATTTTACACGACACTTATTATGTCGTTGCCCACTTCCA
CTATGTTCTTTCTATAGGAGCTGTGTTTGCAATTATAGGAGGATTTGTTCACTGAT-ATC
CTTTATTTACTGGTTTAACAATAAACCATAAATTCCTAAAAATCCAGTTCCTTACGATAT
TCATTGGAGTAAACACCACATTCTTCCCCCAACATTT-CTTAGGA-CTC--AGAGGAA-T
ACCTCGACGTTAC-TCTGACTATCCTGATGCATATACTACATGAAATGTAGTCTCTTCCA
TTGGATCCCTCATTTCCTTAATAAGAATTTTTATCTTTCTTTTTACTTTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTTGCACCTCTCAGACTACCTACATCCCTTGAATGATCTC
AACTTATACCTCCTGCCGAA----------------------------------------
-----------
```

OK_P11_03

```
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------NATGAG
CNGTGCAGTANNNTTTATNANTGNAGNNNTTAACATACTANNAANGNGTATNTCCNTTGA
CAGNANGCCCCTATTTGTCNGATCAGTNGTATTAACCGCNNTANNCCGGGANCCNTCTCT
TCCAGTCCTTGCGGGAGCAATCACTATACTCCTTACAGATCGTAATATCAATACAACATT
CTTTGACCCAGCAGGTGGAGGAGATCC-TATCCTATACCAACACTTATTTTGG-TTCTTT
GGACACCCTGAAGTATACATTCTAA-TTCTCCCAGGATTTGGTATAATCTCTCATATCAT
TAGCCAAGAAAGAAGAAAAAAGGAAACTTTTGGGACCCTTGGTATAATCTATGCTATAAT
AGCAATTGGACTTCTAGGTTTTATTGTATGAGCTCACCACATATTTACAGTTGGTATAGA
CGTAGACACCCGGGCCTACTTCACTTCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCCACCCTCCATGGTTCACAATTAAACTATTCCCCCTCACT
ATTATGAGCACTAGGGTTTGTATTTCTATTCACAGTAGGAGGTCTTACTGGAGTAATTCT
AGCTAACTCCTCAATTGATATTATTTTACACGACACTTATTATGTCGTTGCCCACTTCCA
CTATGTTCTTTCTATAGGAGCTGTGTTTGCAATTATAGGAGGATTTGTTCACTGGGAATC
CTTTATTTACTGGTTTAACAATAAACCATAAATTCCTAAAAATCCAGTTCCTTACAATAT
TCATTGGAGTAAACACCACATTCTTCCCCCAACATTT-CTTAGGA-CTC--AGAGGAAAT
ACCCCGACGTTACCTCTGACTATCCTGATGCATATACTACATGAAATGTAGTCTCTTCCA
TTGGATCCCTCATTTCCTTAATAAGAATTTTCATCTTTCTTTTTACTTTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTTGCACCTCTCAGACTACCCACATCCCTTGAATGATCTC
AACTTATACCTCCTGCCGAACACA------------------------------------
-----------
```

OK_P11_04

```
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
---TGCAGTAAATTTTATTACTACAGTAATTAACATACGAGCAACAGGTATATCATTTGA
CCAGATNCCCCTATTTGTCTGATCAGTTGTATTAACTGCCGTACTCCNGCNCCTGTCTCT
```

APPENDIX 3 (continued)

```
CCCAGTCCTTGCGGGAGCAATCACTATACTCCTTACAGATCGTAATATCAATACAACATT
CTTTGACCCAGCAGGTGGGGGAGATCC-TATCCTGTACCAACACTTATTTTGG-TTCTTT
GGACACCCTGAAGTATACATTCTAA-TTCTCCCAGGATTTGGTATAATCTCTCATATCAT
TAGCCAAGAAAGAAGAAAAAAGGAAACTTTTGGAACCCTTGGTATAATCTACGCTATAAT
AGCAATTGGACTTCTAGGTTTTATTGTATGAGCTCACCACATATTTACAGTTGGTATAGA
CGTAGACACCCGGGCCTACTTCACTTCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCTACCCTCCATGGTTCACAATTAAACTATTCCCCCTCACT
ATTATGAGCACTAGGATTCGTATTTCTCTTCACAGTAGGAGGTCTTACTGGAGTAATTCT
AGCTAACTCCTCAATTGATATTATTTTACATGACACTTATTATGTCGTTGCCCACTTCCA
CTATGTTCTTTCTATAGGAGCTGTGTTTGCAATTATAGGAGGATTTGTTCACTGGT-ATC
CTTTATTTACTGGTTTAACAATAAACCATAAATTCCTAAAAATCCAGTTCCTTACAATAT
TCATTGGAGTAAACACCACATTCTTCCCCCAACATTT-CTTAGGA-CTC--AGAGGAA-T
ACCTCGACGTTAC-TCTGACTATCCTGATGCGTATACTACATGAAATGTAGTCTCTTCCA
TTGGATCCCTCATTTCCTTAATAAGAATTTTCATCTTTCTTTTTACTTTATGAGAAAGTA
TAGTATCAATACGAAAAAGTCTTGCACCTCTCAGACTACCCACATCCCTTGAATGATCTC
AACTCATACNTCCTGCCGAACACT------------------------------------
-----------

OK_P11_05
-------------------------TCCCTAATCCTCCTATTGATAAGAAGATTAGTAGA
AAGAGGAGCTGGAACAGGTTGAACTGTTTATCCCCCCCTCTCATCTAATATTGCCCACAG
AGGAGCTTCTGTAGATTTAGCCATCTTTAGTCTCCACCTAGCTGGAGTCTCCTCGATCCT
AGGTGCAGTAAATTTTATTACTACAGTAATTAACATACGAGCAACAGGTATATCATTTGA
CCAGATACCCCTATTTGTCCGAACAAGTGGAATAAACGCCCGAACCCTGCTCCTGTCTCT
CCCAGTCCTTGCGGGAGCAATCACTATACTCCTTACAGATCGTAATATCAATACAACATT
CTTTGACCCAGCAGGTGGGGGAGATCC-TATCCTATACCAACACTTATTTTGG-TTCTTT
GGACACCCTGAAGTATACATTCTAA-TTCTCCCAGGATTTGGTATAATCTCTCATATCAT
TAGCCAAGAAAGAAGAAAAAAGGAAACTTTTGGAACCCTTGGTATAATCTACGCTATAAT
AGCAATTGGACTTCTAGGTTTTATTGTATGAGCTCACCACATATTTACAGTTGGTATAGA
CGTAGACACCCGGGCCTACTTCACTTCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCTACCCTCCATGGTTCACAATTAAACTATTCCCCCTCACT
ATTATGAGCACTAGGATTCGTATTTCTCTTCACAGTAGGAGGTCTTACTGGAGTAATTCT
AGCTAACTCCTCAATTGATATTATTTTACATGACACTTATTATGTCGTTGCCCACTTCCA
CTATGTTCTTTCTATAGGAGCTGTGTTTGCAATTATAGGAGGATTTGTTCACTGGT-ATC
CTTTATTTACTGGTTTAACAATAAACCATAAATTCCTAAAAATCCAGTTCCTTACAATAT
TCATTGGAGTAAACACCACATTCTTCCCCCAAAATTC-C--AGAA-CACACAGAAGAA-T
ACCTCGACGTTAC-TCTGACTATCCTGATGCGTATACTACATGAAATGTAGTCTCTTCCA
TTGGATCCCTCATTTCCTTAATAAGAATTTTCATCTTTCTTTTTACTTTATGGGAAAGTA
TAGTATCAATACGAAAAAGTCTTGCACCTCTCAGACTACCCACATCCCTTGAATGATCTC
AACTTATACCTCCTGCCGAACACNGCTATGC-----------------------------
-----------

OK_P12_01
-------------------------TCCCTAATCCTCCTTCTGATAAGAAGATTAGTGGA
AAGAGGAGCCGGAACAGGTTGAACTGTTTATCCCCCCCTCTCCTCAAATATTGCCCACAG
```

APPENDIX 3 (continued)

```
AGGAGCTTCAGTAGACCTGGCCATCTTTAGACTCCACTTAGCTGGTGTCTCTTCAATCCT
AGGTGCAGTAAATTTTATTACAACAGTAATCAACATACGAGCAACTGGCATATCCTTTGA
CCAAATACCCCTATTTGTCTGATCAGTTGTTCTTACTGCTGTACTCCTCCTCCTATCTCT
TCCAGTCCTCGCGGGAGCAATCACTATACTCCTTACTGATCGAAATATCAACACTACATT
CTTTGATCCCGCAGGTGGAGGAGACCCC-ATCCTGTACCAGCACCTATTCTGA-TTCTTT
GGTCATCCCGAAGTTTATATTCTAA-TTCTCCCAGGGTTTGGTATAATTTCCCACATCAT
TAGCCAAGAAAGAAGAAAAAAGGAAACCTTTGGGACTCTTGGAATAATCTACGCAATAAT
AGCAATTGGACTCCTAGGTTTTATTGTGTGAGCCCACCATATATTTACAGTTGGTATAGA
TGTAGACACCCGAGCCTACTTTACTTCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCCACCCTTCATGGTTCTCAACTAAATTATTCCCCGTCATT
GCTATGAGCACTAGGGTTCGTTTTTCTTTTTACAGTAGGGGGTCTTACCGGAGTTATTCT
GGCCAATTCCTCAATTGATATTATTTTACATGACACTTACTATGTTGTTGCCCACTTCCA
CTATGTACTTTCTATGGGGGCTGTATTTGCAATTATAGGAGGATTCGTCCACTGAT-ACC
CTTTATTTACTGGTTTAACAATAAACCATAAATTCCTAAAAATCCAATTCCTTACTATAT
TCATTGGAGTAAACACTACATTCTTCCCTCAACATTTTC-TAGGG---TTGAGAGGGA-T
ACCTCGACGTTAT-TCTGACTACCCTGATGCTTATACTACATGGAATGTAGTCTCCTCTA
TTGGATCTCTCATCTCCTTAATAAGAATTTTCATCTTTTTGTTTACTCTGTGAGAAAGCA
TAGTATCAATACGAAAAAGTCTTGCACCTCTTAGATTACCAACCTCCCTTGAATGATCCC
AGCTTATACCCCCTGCTGAACAC-------------------------------------
-----------

OK_P12_03
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
-----------------------------------ACTGCTGTANTCCGCCTNCCATCTCT
TCCAGTCCTCGCGGGAGCAATCACTATACTCCTTACTGATCGAAATATCAACACTACATT
CTTTGATCCCGCAGGTGGAGGAGA-CCCCATCCTGTACCAGCACCTATTCTGA-TTCTTT
GGTCACCCCGAAGTTTATATTCTAA-TTCTCCCAGGGTTTGGTATAATTTCCCACATCAT
TAGCCAAGAAAGAAGAAAAAAGGAAACCTTTGGGACTCTTGGAATAATCTACGCAATAAT
AGCAATTGGACTCCTAGGTTTTATTGTATGAGCCCACCATATATTTACAGTTGGTATAGA
TGTAGACACCCGAGCCTACTTTACTTCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCCACCCTTCATGGTTCTCAACTAAATTATTCCCCGTCATT
GCTATGAGCACTAGGGTTCGTTTTTCTTTTTACAGTAGGGGGTCTTACCGGAGTTATTCT
GGCCAATTCCTCAATTGATATTATTTTACATGACACTTACTATGTTGTTGCCCACTTCCA
CTATGTACTTTCTATGGGGGCTGTATTTGCAATTATAGGAGGATTCGCCCACTGAT-ACC
CTCTATTTACTGGTTTAACAATAAACCATAAATTCCTAAAAATCCAATTCCTTACTATAT
TCATTGGAGTAAACACTACATTCTTCCCTCAACATTTTC-TAGGG---TTGAGAGGAA-T
GCCTCGACGTTAT-TCTGACTACCCTGATGCTTATACTACATGGAATGTAGTCTCCTCTA
TTGGATCTCTCATCTCCTTAATAAGAATTTTCATCTTTTTGTTTACTCTGTGAGAAAGCA
TAGTATCAATACGAAAAAGTCTTGCACCTCTTAGATTACCAACCTCCCTTGAATGATCCC
AGCTTATACCCCCTGCTGAACACTG-----------------------------------
-----------
```

APPENDIX 3 (continued)

OK_P12_04

```
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------------------------------
------------------------------------TGCTGTACTCCGNCTCNTNTCTCT
TCCNGTCCTCGCGGGAGCAATCACTATACTCCTTACTGATCGAAATATCAACACTACATT
CTTTGATCCCGCAGGTGGAGGAGA-CCCCATCCTGTACCAGCACCTATTCTGA-TTCTTT
GGTCATCCCGAAGTTTATATTCTAA-TTCTCCCAGGGTTTGGTATAATTTCCCACATCAT
TAGCCAAGAAAGAAGAAAAAAGGAAACCTTTGGGACTCTTGGAATAATCTACGCAATAAT
AGCAATTGGACTCCTAGGTTTTATTGTATGAGCCCACCATATATTTACAGTTGGTATAGA
TGTAGACACCCGAGCCTACTTTACTTCTGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTCAGATGACTTGCCACCCTTCATGGTTCTCAACTAAATTATTCCCCGTCATT
GCTATGAGCACTAGGGTTCGTTTTTCTTTTTACAGTAGGGGGTCTTACCGGAGTTATTCT
GGCCAATTCCTCAATTGATATTATTTTACATGACACTTACTATGTTGTTGCCCACTTCCA
CTATGTACTTTCTATGGGGGCTGTATTTGCAATTATAGGAGGATTCGTCCACTGAT-ACC
CTCTATTTACTGGTTTAACAATAAACCATAAATTCCTAAAAATCCAATTCCTTACTATAT
TCATTGGAGTAAACACTACATTCTTCCCTCAACATTTTC-TAGGG---TTGAGAGGAA-T
GCCTCGACGTTAT-TCTGACTACCCTGATGCTTATACTACATGGAATGTAGTCTCCTCTA
TTGGATCTCTCATCTCCTTAATAAGAATTTTCATCTTTTTGTTTACTCTGTGAGAAAGCA
TAGTATCAATACGAAAAAGTCTTGCACCTCTTAGATTACCAACCTCCCTTGAATGATCCC
AGCTTATACCCCCTGCTGAACACG------------------------------------
-----------
```

OK_P13_01

```
------------------------CTCCCTGATCCTCCTATTGATAAGAAGATTAGTGGA
AAGAGGAGCCGGAACAGGTTGAACTGTCTACCCCCCTCTCTCATCCAATATCGCTCACAG
AGGGGCATCGGTAGACTTAGCCATCTTTAGTCTTCACTTAGCCGGGATCTCCTCAATCCT
AGGGGCAGTAAATTTCATTACTACAGTAATTAACATACGAGCAACAGGTATAGCCTTCGA
CCAGATACCTCTCTTTGTTTGGTCTGTCGTTCTCACTGCCGTACTCCTTCTCCTGTCTCT
TCCAGTCCTTGCAGGAGCAATCACTATACTCCTTACAGACCGTAACATCAACACCACATT
CTTTGACCCAGCAGGCGGGGGGGA-CCCCGTCCTATATCAACACTTATTCTGG-TTCTTT
GGCCACCCCGAAGTCTATATTCTAA-TCCTTCCAGGTTTTGGAATAATTTCTCATATCAT
CAGCCAAGAAAGAAGGAAAAAGGAAACTTTTGGAACTCTCGGAATAATCTATGCAATAAT
AGCAATCGGGCTCCTAGGTTTTATTGTATGAGCCCACCATATATTTACAGTAGGAATAGA
TGTAGACACCCGGGCCTACTTTACTTCCGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTTAGATGACTCGCCACCCTTCACGGTTCCCAATTAAATTATTCACCCTCATT
ACTATGGGCGCTAGGGTTCGTCTTTCTCTTTACAGTAGGAGGTCTTACGGGAGTCATTCT
AGCTAACTCCTCAATTGATATTATTTTACATGACACTTACTATGTTGTTGCCCACTTTCA
CTATGTCCTCTCCATGGGAGCTGTATTTGCAATTATAGGAGGATTTGTTCATTGAT-ACC
CTTTATTTACTGGCCTAACAATAAACCATAAATTTCTAAAAATTCAATTTCTTACGATAT
TCATTGGAGTAAACACCACATTCTTCCCTCAACATTT-CTTAGGA--CTT-AGAGGAA-T
GCCTCGCCGATAT-TCTGATTATCCCGATGCCTACACTACATGAAATGTAATCTCCTCCA
TCGGATCTCTCATCTCCCTGATAAGAATTTTCATCTTTCTCTTTATGTTGTGGGAAAGTA
TAGTATCAATACGAAAAAGCCTTGCACCTCTAAGGCTACCAACATCCCTTGAATGATCTC
```

APPENDIX 3 (continued)

```
AACTTATACCCCCTGCCGAGCACT------------------------------------
-----------
```

OK_P13_02

```
------------------------CTCCCTGATCCTCCTATTGATAAGAAGATTAGTGGA
AAGAGGAGCCGGAACAGGTTGAACTGTCTACCCCCCTCTCTCATCCAATATCGCTCACAG
AGGGGCATCGGTAGACTTAGCCATCTTTAGTCTTCACTTAGCCGGGATCTCCTCAATCCT
AGGGGCAGTAAATTTCATTACTACAGTAATTAACATACGAGCAACAGGTATAGCCTTCGA
CCAGATACCTCTCTTTGTTTGGTCTGTCGTTCTCACTGCCGTACTCCTTCTCCTGTCTCT
TCCAGTCCTTGCAGGAGCAATCACTATACTCCTTACAGACCGTAACATCAACACCACATT
CTTTGACCCAGCAGGCGGGGGGGA-CCCCGTCCTATATCAACACTTATTCTGG-TTCTTT
GGCCACCCCGAAGTCTATATTCTAA-TCCTTCCAGGTTTTGGAATAATTTCTCATATCAT
CAGCCAAGAAAGAAGGAAAAAGGAAACTTTTGGAACTCTCGGAATAATCTATGCAATAAT
AGCAATCGGGCTCCTAGGTTTTATTGTATGAGCCCACCATATATTTACAGTAGGAATAGA
TGTAGACACCCGGGCCTACTTTACTTCCGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTTAGATGACTCGCCACCCTTCACGGTTCCCAATTAAATTATTCACCCTCATT
ACTATGGGCGCTAGGGTTCGTCTTTCTCTTTACAGTAGGAGGTCTTACGGGAGTCATTCT
AGCTAACTCCTCAATTGATATTATTTTACATGACACTTACTATGTTGTTGCCCACTTTCA
CTATGTCCTCTCCATGGGAGCTGTATTTGCAATTATAGGAGGATTTGTTCATTGAT-ACC
CTTTATTTACTGGCCTAACAATAAACCATAAATTTCTAAAAATCAAATTCCTAACAAAAT
-CAATGGAAGAAAAAACAAAATCCTCCCCAAAACA---CTTAGGACGAAGAAGAAGAA-T
GCCTCGCCGATAT-TCTGATTATCCCGATGCCTACACTACATGAAATGTAATCTCCTCCA
TCGGATCTCTCATCTCCCTGATAAGAATTTTCATCTTTCTCTTTATGTTGTGGGAAAGTA
TAGTATCAATACGAAAAAGCCTTGCACCTCTAAGGCTACCAACATCCCTTGAATGATCTC
AACTTATACCCCCTGCCGAGCACT------------------------------------
-----------
```

OK_P13_03

```
------------------------ATCCCTGATCCTCCTATTGATAAGAAGATTAGTGGA
AAGAGGAGCCGGAACAGGTTGAACTGTTTACCCCCCTCTCTCATCCAATATCGCTCACAG
AGGGGCATCGGTAGACTTAGCCATCTTTAGTCTTCACTTAGCCGGGATCTCCTCAATCCT
AGGGGCAGTAAATTTCATTACTACAGTAATTAACATACGAGCAACAGGTATAGCCTTCGA
CCAGATACCTCTCTTTGTTTGGTCTGTCGTTCTCACTGCCGTACTCCTTCTCCTGTCTCT
TCCAGTCCTTGCAGGAGCAATCACTATACTCCTTACAGACCGTAACATCAACACCACATT
CTTTGACCCAGCAGGCGGGGGGGA-CCCCGTCCTATATCAACACTTATTCTGG-TTCTTT
GGCCACCCCGAAGTCTATATTCTAA-TCCTTCCAGGTTTTGGAATAATTTCTCATATCAT
CAGCCAAGAAAGAAGGAAAAAGGAAACTTTTGGAACTCTCGGAATAATCTATGCAATAAT
AGCAATCGGGCTCCTAGGTTTTATTGTATGAGCCCACCATATATTTACAGTAGGAATAGA
TGTAGACACCCGGGCCTACTTTACTTCCGCCACAATAATTATTGCAGTTCCTACAGGAAT
TAAAATCTTTAGATGACTCGCCACCCTTCACGGTTCCCAATTAAATTATTCACCCTCATT
ACTATGGGCGCTAGGGTTCGTCTTTCTCTTTACAGTAGGAGGTCTTACGGGAGTCATTCT
AGCTAACTCCTCAATTGATATTATTTTACATGACACTTACTATGTTGTTGCCCACTTTCA
CTATGTCCTCTCCATGGGAGCTGTATTTGCAATTATAGGAGGATTTGTTCATTGAT-ACC
CTTTATTTACTGGCCTAACAATAAACCATAAATTTCTAAAAATTCAATTTCTTACGATAT
TCATTGGAGTAAACACCACATTCTTCCCTCAACATTT-CTTAGGA--CTT-AGAGGAA-T
```

APPENDIX 3 (continued)

```
GCCTCGCCGATAT-TCTGATTATCCCGATGCCTACACTACATGAAATGTAATCTCCTCCA
TCGGATCTCTCATCTCCCTGATAAGAATTTTCATCTTTCTCTTTATGTTGTGGGAAAGTA
TAGTATCAATACGAAAAAGCCTTGCACCTCTAAGGCTACCAACATCCCTTGAATGATCTC
AACTTATACCCCCTGCCGAGCACT------------------------------------
-----------
```