FIRST CALIFORNIA RECORDS OF THE LITTLE STINT AND NAZCA BOOBY CONFIRMED THROUGH MOLECULAR ANALYSIS

DOU-SHUAN YANG, H. T. Harvey & Associates, Suite 1B, Building 83, California Polytechnic State University–San Luis Obispo, One Grand Avenue, San Luis Obispo, California 93407; dou.s.yang@gmail.com

STEPHEN C. ROTTENBORN and SCOTT TERRILL, H. T. Harvey & Associates, Building D, 983 University Avenue, Los Gatos, California 95032

ADAM JAMES SEARCY, Western Foundation of Vertebrate Zoology, 439 Calle San Pablo, Camarillo, California 93012

FRANCIS X. VILLABLANCA, Department of Biology, Building 33, California Polytechnic State University–San Luis Obispo, One Grand Avenue, San Luis Obispo, California 93407

ABSTRACT: We used molecular genetic analysis to assess the identification of two bird specimens otherwise difficult to identify with certainty. A *Calidris* sandpiper collected in 1974 had been identified at various times as the Little Stint (*C. minuta*) or as the Red-necked Stint (*C. ruficollis*). The remains of a booby salvaged in 2013 were identified as either the Nazca (*Sula granti*) or the Masked (*S. dactylatra*). We were able to identify the first as *C. minuta* and the second as *S. granti*, providing the first California records of each species.

Figure 1. Nazca Booby carcass *in situ*, McGrath State Beach, Ventura County, California, 23 July 2013.

Photo by Alexis Frangis

Accurate species identification is essential for surveys of wildlife diversity. Over the past few decades, species identification of collected specimens by genetic characters has become commonplace. Genetic identification is particularly useful when the phenotypic characters usually used to distinguish species are difficult to apply or when a salvaged specimen is so degraded that phenotypic characters cannot be used. Here we present results from the genetic identifications of a sandpiper and a booby that constitute first records for California.

McCaskie (1975) described encountering a likely Rufous-necked Sandpiper or Red-necked Stint (*Calidris ruficollis*) at the Salton Sea, California, on 17 August 1974. He collected the bird, in worn first alternate plumage (San Diego Natural History Museum 38887), and identified it as the first example of *C. ruficollis* recorded in California. Subsequently, the possibility was raised that the specimen might be instead an example of *C. minuta*, though the California Bird Records Committee (CBRC) could not reach a consensus on the bird's identification. Veit (1988), using additional morphological criteria, concluded that the specimen was indeed *C. ruficollis*. However, subsequent morphological comparisons of SDNHM 38887 with over 100 specimens of *C. ruficollis* and *C. minuta* by M. Ralph Browning and Claudia Wilds at the U.S. National Museum of Natural History were inconclusive (Erickson and Hamilton 2001).

On 23 July 2013, Alexis Frangis, a California State Parks biologist, salvaged the remains of a booby (*Sula*) from McGrath State Beach in Oxnard, Ventura County, California. The specimen comprised two feathered wings, a skinless torso and skull, tail feathers, and foot webbing (Figure 1). Nevertheless, at the Western Foundation of Vertebrate Zoology (WFVZ) Searcy and Steven Tucker were able to determine, on the basis of its underwing pattern, that the specimen (WFVZ 56922) was either a Masked Booby (*S. dactylatra*) or a Nazca Booby (*S. granti*). Extensively dark upperparts implied the booby was in its first plumage cycle. Reliable identification criteria for first-cycle Masked/Nazca boobies based on plumage characters have not yet been identified, and mensural characters were not useful because of the degradation of the specimen. At the time California had no specimen records of the Masked Booby, and the Nazca Booby had not been accepted to the main list of California birds. The only Nazca Booby recorded previously in California, a first-cycle individual that had ridden a boat from Mexican waters to San Diego, was identified on the basis of bill coloration and genetic analysis. The CBRC voted to relegate this record to its supplemental list because the bird had traveled to California by boat (Garrett and Wilson 2003).

Though identification of avian specimens to species by genetic techniques has become routine following the publication of the Avian Barcode of Life primers and associated database of sequences for species identification (Hebert et al. 2004), genetic identification of these specimens presented challenges. The *Calidris* was collected over 40 years prior to our genetic analysis, and DNA degradation in the interim did not allow us to use the Barcode of Life primers. Because *S. granti* and *S. dactylatra* are closely related, we also investigated the possibility that the specimen was a hybrid.

METHODS

From each specimen, we sampled a small 5-mm section of foot skin, following the sampling methods of Ellegren (1991) and Mundy et al. (1997). We then extracted DNA from each sample, following the protocol of a standard Qiagen DNEasy Blood & Tissue kit (Valencia, CA), with the addition of a 36-hour digestion with 40 mL (the *Sula* specimen) or 60 mL (the *Calidris* specimen) of proteinase K (Crowe et al. 1991). We then amplified DNA regions containing species-diagnostic genetic variation by PCR (the polymerase chain reaction).

Attempts to amplify the *Calidris* sample with the Barcode of Life primers (BirdF1 and Bird R1; Hebert et al. 2004) were unsuccessful, likely because of the large size of the fragment (750 base pairs) targeted by these primers. Additionally, despite the availability of multiple reference sequences of cytochrome oxidase I for each *Calidris* species, we were unable to design short (<150 base pairs) primers for this region that encompassed species-diagnostic variation and also amplified successfully. Instead, we acquired sequence data for both *C. ruficollis* and *C. minuta* from the mitochondrial DNA (mtDNA) ND2 region from Genbank (www.ncbi.nlm.nih.gov/genbank/) and designed primers to amplify a fragment of 115 base pairs that encompassed DNA polymorphisms distinguishing each species (*Calidris* ND2-F: 5′-CCC ACC AAA CTC CAC AAA CC-3′; *Calidris* ND2-R: 5′-GTG ATT ATG GGG GAG AGT GGT-3′). For the *Sula* specimen, we used the b3 and b4 primers designed by Morris-Pocock et al. (2010) to amplify a 450-base-pair fragment of the mtDNA cytochrome-b region capable of distinguishing *S. granti* from *S. dactylatra*. Additionally, to investigate the possibility that the specimen was a hybrid, we also amplified a nuclear DNA region capable of distinguishing *S. granti* from *S. dactylatra* (intron 8 of the nuclear gene for α-enolase) by using primers (F: 5′-TGGACCTTCAAATCCCCCGATGATCCCCAGC-3′, R: 5′-CCAGGCACCCCAGTCTACCTGGTCAAA-3′) described by Patterson et al. (2011).

Our PCR reaction mixes and thermocycling regime varied by primer set. For the *Calidris* sample we amplified 3 µL of DNA extract in a total volume of 25 µL with 2.5 µL 10X JumpStart RED*Taq* buffer (Sigma-Aldrich), 1 mM $MgCl_2$, 200 µM of each dNTP, 0.3 µM primer concentration, and 0.5 units of Jumpstart RED *Taq* (Sigma-Aldrich). Thermocycling consisted of an initial denaturation at 94°C for 4 minutes, then 40 cycles of 94°C for 15 seconds, 56°C for 1 minute, and 72°C for 30 seconds, followed by a final extension at 72°C for 7 minutes. For the *Sula* sample we amplified 1 µL of DNA extract in a total volume of 20 µL with 2 µL 10X Jumpstart RED*Taq* buffer, 200 mM of each dNTP, 0.2 µM primer concentration, and 0.4 units of Jumpstart RED *Taq*. In this case thermocycling comprised an initial denaturation at 94°C for 5 minutes, then 35 cycles of 94°C for 15 seconds, 48°C (65°C for α-enolase) for 30 seconds, and 72°C for 30 seconds, followed by a final extension at 72°C for 7 minutes. We tested the PCR for success in amplifying the DNA by agarose gel electrophoresis, then sent successful amplifications to Sequetech Corp. (Mountain View, CA) for EXO-SAP PCR cleanup, Sanger sequencing in the forward and reverse directions, sequencing cleanup, and capillary electrophoresis to generate

sequence data. We trimmed and aligned the forward and reverse sequence data to generate consensus sequences for the *Calidris* specimen and each gene analyzed for the *Sula* specimen in Geneious (Biomatters Ltd.).

We aligned the *Calidris* consensus sequence alongside reference sequences of *C. minuta* and *C. ruficollis* from Gibson and Baker (2012) acquired from Genbank (accessions KC 969101 and 969105, respectively). To place our results in a larger taxonomic context, we also aligned the *Calidris* sequence with reference sequences of 19 species of *Calidris* (including *C. minuta* and *C. ruficollis*) also from Gibson and Baker (2012; Genbank accessions KC969093–KC969108). We then visually inspected these alignments for characters diagnostic of species within *Calidris*. For each of these two *Calidris* alignments, we also used Geneious to generate a neighbor-joining phylogenetic tree with a Jukes–Cantor model of substitution (Jukes and Cantor 1969) to assign the *Calidris* specimen to its most closely related species. We used 1000 bootstrap replicates of each data set to assess statistical confidence in the relationships inferred within each phylogenetic tree.

We aligned the cytochrome-b sequence of the booby alongside five reference sequences of *S. dactylatra* and five of *S. granti* (Genbank accessions AY156695-156699 and JX569178-569182, respectively). To place our results in a larger taxonomic context we also aligned this dataset alongside reference sequences of *S. leucogaster, S. nebouxii, S. sula,* and *S. variegata* (Genbank accessions JX 416755, KC 577254, JX 416799, and KC 577252, respectively). We then visually inspected these alignments to identify species-specific variation and analyzed it phylogenetically as described for the *Calidris* dataset.

Bootstrap resampling assesses the likelihood of observing a given relationship in a new dataset generated by resampling the dataset with replacement, a "nodal support" value that is not directly analogous to statistical support based on deviation from a normal distribution of variation. For example, our *Calidris* data set was composed of 115 DNA base pairs ("sites") that can be numbered 1, 2, 3, etc., to 115. To create a bootstrap replicate, we randomly selected 114 of these sites with replacement (e.g. 1, 7, 7, 12, 23, 23, 50, 72, 72, 103, etc.) and analyzed this resampled data set in the same way. The bootstrap nodal support indicates the percentage of time across bootstrap replicates that two taxa (or groups of taxa) were found to be most closely related to one another. In this way, we could evaluate whether a phylogenetic relationship observed in the original data set was driven by few weakly informative points of variation, which are unlikely to be resampled consistently across bootstrap replicates, or multiple strongly informative points of variation, which should be resampled (at least in part) across the majority of bootstrap replicates. While bootstrap resampling of datasets is frequently used in phylogenetic analyses, the mathematics and methods of the approach are complex, as described by Felsenstein (1985).

RESULTS

The *Calidris* specimen (SDNHM 38887) possessed the variation characteristic of *C. minuta* at eight of eight diagnostic sites across the 115

Table 1 Variation at Eight Sites within the Mitochondrial Gene ND2 Distinguishing *Calidris minuta* from *C. ruficollis*

Position[a]	882	885	900	912	919	930	936	946
SDNHM 38887	C	G	T	A	T	C	G	C
C. minuta	C	G	T	A	C	C	G	C
C. ruficollis	T	A	C	T	C	T	A	T

[a]In alignment with the reference sequence of *C. minuta* (Genbank accession 969101).

base pairs surveyed (Table 1). The phylogenetic analyses also supported its identification as *C. minuta*, with 99% nodal support in a direct comparison with *C. minuta* and *C. ruficollis* (Figure 2) and 92% nodal support in a comparison of 19 species of *Calidris* combined (Figure 3). The 115-base-pair sequence is available via Genbank as accession KU310674.

The *Sula* specimen (WFVZ 56922) possessed variation characteristic of *S. granti* at four of four diagnostic sites across the 450 base pairs of mitochondrial DNA (mtDNA) surveyed (Table 2). The phylogenetic analysis of mtDNA also supported its identification as *S. granti* with 99% nodal support (Figure 4). Additionally, the nuclear gene-sequence data also placed WFVZ 56922 as *S. granti* (rather than as a hybrid of *S. granti* and *S. dactylatra*) at the single diagnostic site. The mitochondrial cytochrome b sequence for this specimen is available via Genbank as accession KU310675, and the nuclear α-enolase sequence is available as accession KU364145.

DISCUSSION

Despite the ever-increasing array of genetic technology available to study wild populations, poor quality and insufficient quantity of template DNA is a fundamental obstacle to genetic analysis. For SDNHM 38887, we initially attempted to acquire sequence data from the region that had been sequenced for the Barcode of Life (Hebert et al. 2004). By design, however, this region is relatively large in order to capture some genetic distinction between

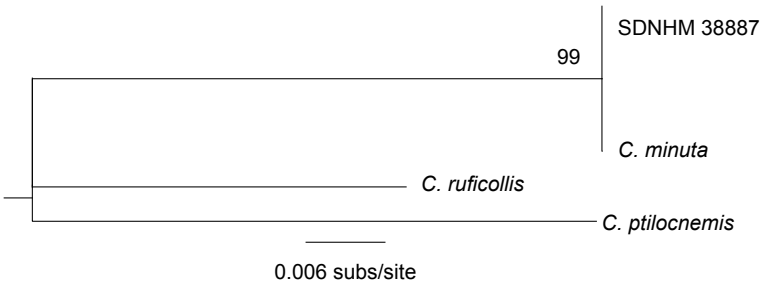

Figure 2. Neighbor-joining tree with 1000 bootstrap replicates of 122 base pairs of mitochondrial ND2 data comparing SDNHM 38887 to *C. minuta* and *C. ruficollis* only. Bootstrap support <90% not shown. Tree is rooted with a sequence of *C. ptilocnemis*, the Rock Sandpiper.

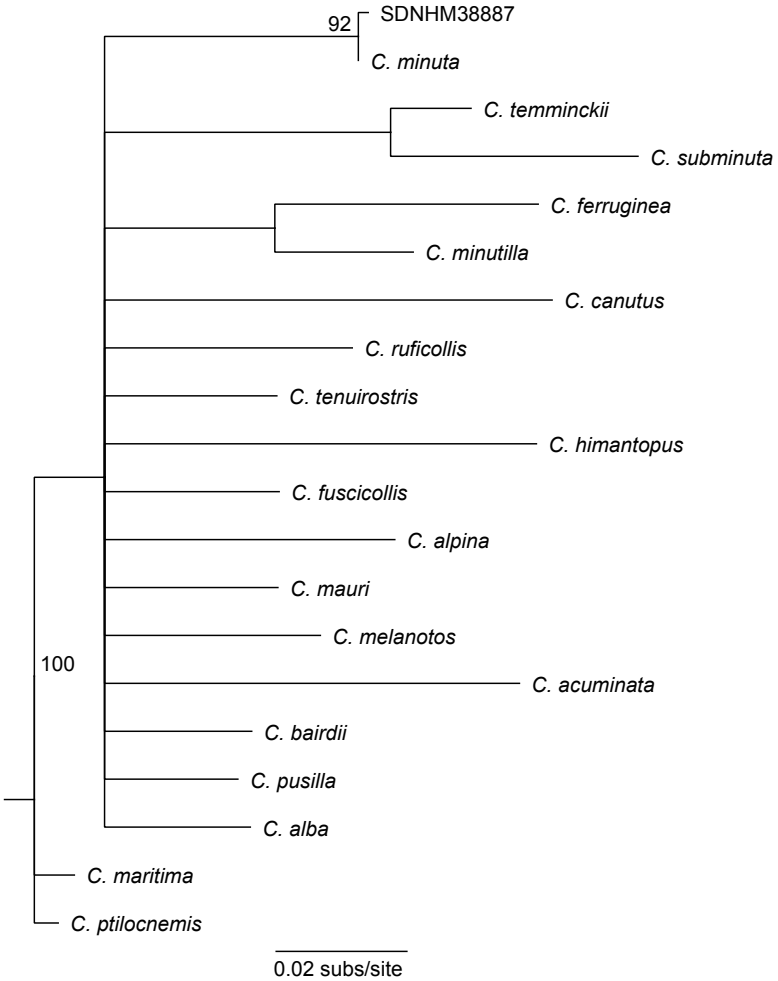

Figure 3. Unrooted neighbor-joining tree with 1000 bootstrap replicates of 122 base pairs of mitochondrial ND2 data comparing SDNHM 38887 to reference sequences from 19 species of *Calidris*. Bootstrap support <90% not shown.

every avian species. The DNA we extracted from SDNHM 38887 was too degraded to allow amplification of the 750-base-pair segment sequenced for the Barcode of Life, so we focused on smaller regions of <200 base pairs in order to amplify DNA successfully. Data analysis with the Barcode of Life region would have been extremely straightforward, with a simple analysis matching the sequence of SDNHM 38887 to the existing database. Only one reference sequence of the mitochondrial gene ND2 was available for

Table 2 Variation at Four sites within the Mitochondrial Gene ND2 Distinguishing *Sula granti* from *S. dactylatra*

Position[a]	75	192	405	429
WFVZ 56922	G	C	T	A
S. granti	G	C	T	A
S. dactylatra	A	T	C	C

[a]In alignment with reference sequences of *S. dactylatra* (Genbank accessions AY156695–156699).

the two candidate species of *Calidris*, so we cannot assess the placement of SDNHM 38887 with respect to the full range of genetic variation within each species. Nevertheless, SDNHM 38887 matched *C. minuta* at all eight diagnostic sites, a pattern unlikely to arise in an individual of *C. ruficollis* by chance alone. Thus SDNHM 38887 represents the earliest record of *C. minuta* in California, by over nine years, preceding one photographed at Bolinas, Marin County, 14–22 September 1983 (Roberson 1986).

One drawback of species identification by mtDNA is that mtDNA is inherited exclusively through the maternal line. The lack of paternal contribution to mtDNA variation can lead to difficulties in identifying patterns of gene flow (Yang and Kenagy 2009). Therefore, mtDNA data are not capable of identifying offspring resulting from interspecific hybridization. Nevertheless, we believe that SDNHM 38887 is unlikely to be a hybrid *C. minuta* × *C. ruficollis* origin, given that *C. ruficollis* is not known to hybridize (Pyle 2008).

In the case of *S. granti* and *S. dactylatra*, however, mixed pairs and apparent hybrid offspring have been observed (Howell and Webb 1990, Pitman and Jehl 1998). Thus we could not rule out the possibility of a hybrid origin for WFVZ 56922 by the mtDNA analysis alone. We addressed this uncertainty by sequencing the nuclear gene for α-enolase, in which *S. granti* and *S. dactylatra* differ (Patterson et al. 2011). In this marker, WFVZ 56922 possessed the sequence characteristic of *S. granti* with no indication of heterozygosity. Patterson et al. (2011) examined a total of five nuclear genes, but they found that among these only intron 8 of α-enolase consistently distinguished *S. dactylatra* from *S. granti*. They generated their reference sequences from two individuals of each species, sampled from opposite ends of each species' geographic range, increasing their ability to detect intraspecific polymorphism and thereby also increasing the likelihood that any genetic differences reflect interspecific rather than intraspecific differentiation. The low level of genetic differentiation between these two species can be attributed to their relatively recent evolutionary divergence (Friesen et al. 2002, Patterson et al. 2011).

As both mitochondrial and nuclear DNA results support the identification of WFVZ 56922 as *S. granti*, on this basis the CBRC added the Nazca Booby to the main list of California birds, the record from Ventura County being the northernmost of the species.

Our results reflect the utility of genetic analysis for species identification. For this purpose, the discrete nature of genetic data can serve as a valuable supplement to the more continuous nature of morphometric data traditionally used for species identification. In the vast majority of cases, morphometric

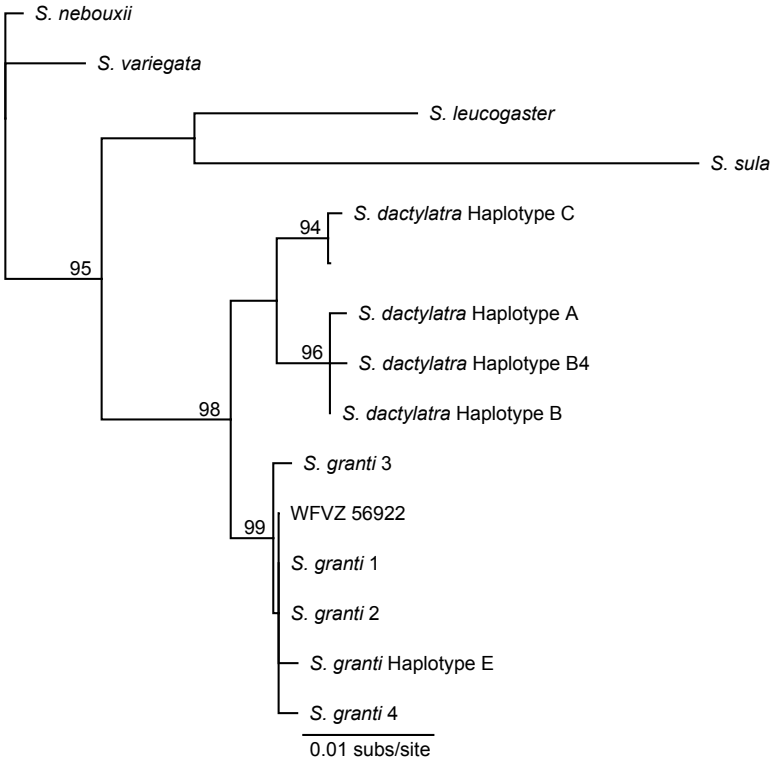

Figure 4. Neighbor-joining phylogenetic tree comparing sequences of the mitochondrial gene for cytochrome b from WFVZ 56922 with those reference sequences of the Masked, Nazca, Blue-footed (*Sula nebouxii*), Peruvian (*S. variegata*), Brown (*S. leucogaster*), and Red-footed (*S. sula*) boobies. The data strongly support the placement of WFVZ 56922 with *S. granti*. Bootstrap support <90% not shown.

data are more than sufficient for unambiguous species identification. In the examples we have addressed here, however, where diagnostic morphological characters were degraded, genetic analysis was invaluable. Additionally, in cases where species are distinguished mainly by absolute size rather by than morphological ratios or meristic characters, genetic analysis provides a valuable method to ensure that exceptionally large or small individuals are identified accurately.

Grinnell (1910) emphasized that the value of museum specimens may not be realized until the lapse of many years. In the case of the specimen of the Little Stint, McCaskie's (1975) foresight allowed us the opportunity to revisit this specimen with biological techniques not available at the time it was collected. As shown here, archived natural history collections have been and will continue to be a valuable resource for biologists, often in unexpected ways (e.g., Mundy et al. 1997, Yang et al. 2011).

LITERATURE CITED

Crowe, J. S., Cooper, H. J., Smith, M. A., Sims, M. J., Parked, D., and Gewert, D. 1991. Improved cloning efficiency of polymerase chain reaction (PCR) products after proteinase K digestion. Nucleic Acids Res. 19:184.

Ellegren, H. 1991. DNA typing of museum birds. Nature 354:113.

Erickson, R., and Hamilton, R. 2001. Report of the California Records Committee: 1998 records. W. Birds 32:13–49.

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

Friesen, V., Anderson, D., Steeves, T., Jones, H., Schreiber, E. 2002. Molecular support for species status of the Nazca Booby (*Sula granti*). Auk 119:820–826.

Garrett, K. L., and Wilson, J. C. 2003. Report of the California Bird Records Committee: 2001 records. W. Birds 34:15–41.

Gibson, R., and Baker, A. 2012. Multiple gene sequences resolve phylogenetic relationships in the shorebird suborder Scolopaci (Aves: Charadriiformes). Molec. Phylogen. Evol. 64: 66–72.

Grinnell, J. 1910. The uses and methods of a research museum. Popular Sci. Monthly 77:163–169

Hebert, P., Stoeckle, M., Zemlak, T., Francis, C. 2004. Identification of birds through DNA barcodes. PLoS Biology 2(10):e312.

Howell, S. N. G., Lewington, I., and Russell, W. 2014. Rare Birds of North America. Princeton Univ. Press, Princeton, NJ.

Howell, S. N. G., and Webb, S. 1990. The seabirds of Las Islas Revillagigedo, Mexico. Wilson Bull. 102:140–146.

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

McCaskie, G. 1975. A Rufous-necked Sandpiper in southern California. W. Birds 6:111–113.

Morris-Pocock, J., Taylor, S., Birt, T., and Friesen, V. 2010. Concerted evolution of duplicated mitochondrial control regions in three related seabird species. BMC Evol. Biol. 10:14.

Mundy, N., Unitt, P., and Woodruff, D. 1997. Skin from feet of museum specimens as a non-destructive source of DNA for avian genotyping. Auk 114:126–129.

Patterson, S. A., Morris-Pocock, J. A., and Friesen, V. L. 2011. A multilocus phylogeny of the Sulidae (Aves: Pelecaniformes). Molec. Phylogen. Evol. 58: 181–191.

Pitman, R. L., and Jehl, J. R. Jr. 1998. Geographic variation and reassessment of species limits in the "Masked" boobies of the eastern Pacific Ocean. Wilson Bull. 110:155–170.

Pyle, P. 2008. Identification Guide to North American Birds, part II. Anatidae to Alcidae. Slate Creek Press, Point Reyes Station, CA.

Roberson, D. 1986. Ninth report of the California Bird Records Committee. W. Birds 17:49–77.

Roberson, D. 1998. Sulids unmasked: Which large booby reaches California? Natl. Audubon Soc. Field Notes 52:276–287.

Veit, R. 1988. Identification of the Salton Sea Rufous-necked Sandpiper. W. Birds 19:165–169.

Yang, D., Conroy, C., and Moritz, C. 2011. Contrasting responses of *Peromyscus* mice of Yosemite National Park to recent climate change. Global Change Biol. 17:2559–2566.

Yang, D., and Kenagy, G. 2009. Nuclear and mitochondrial DNA reveal contrasting evolutionary processes in populations of deer mice (*Peromyscus maniculatus*). Molec. Ecol. 18:5115–5125.

Accepted 22 November 2015