The Journal of Wildlife Management 82(4):723–733; 2018; DOI: 10.1002/jwmg.21421

Research Article

Distinguishing Between Northern Salt Marsh and Western Harvest Mice

DIEGO SUSTAITA,[1,2] *California Department of Fish and Wildlife, Bay Delta Region, 7329 Silverado Trail, Napa, CA 94558, USA*

SARAH K. BROWN,[3] *Biological Sciences Department, California Polytechnic State University, San Luis Obispo, CA 93407, USA*

LAUREEN BARTHMAN-THOMPSON, *California Department of Fish and Wildlife, Bay Delta Region, 7329 Silverado Trail, Napa, CA 94558, USA*

SARAH ESTRELLA, *California Department of Fish and Wildlife, Bay Delta Region, 7329 Silverado Trail, Napa, CA 94558, USA*

LAURA PATTERSON,[4] *California Department of Water Resources, Division of Environmental Services, 3500 Industrial Boulevard, West Sacramento, CA 95691, USA*

PATTY Q. FINFROCK, *California Department of Water Resources, Integrated Regional Water Management, 901 P Street Room 411A, Sacramento, CA 95814, USA*

FRANCIS X. VILLABLANCA, *Biological Sciences Department, California Polytechnic State University, San Luis Obispo, CA 93407, USA*

ABSTRACT The northern subspecies of the salt marsh harvest mouse (*Reithrodontomys raviventris halicoetes*) is morphologically similar to the western harvest mouse (*R. megalotis*) with which it co-occurs in the Suisun Marsh, California, USA, and therefore they are difficult to distinguish in the field. The salt marsh harvest mouse is a federal and California state-listed endangered species, whereas the western harvest mouse has no special status. Thus, our objective was to identify the most effective field metrics that distinguish the species. First we identified a barcode of life and restriction fragment length polymorphism approach for genetically distinguishing between the species. Then we performed univariate tests to examine variation in standard external morphological traits within and between species, and found that differences between species were confounded by sex and age. We then used discriminant function analysis and multiple logistic regression (MLR) to find combinations of characters that resulted in the highest percentages of correct classification based on a data set of individuals with genetically verified species identity. The best model (MLR) correctly classified $90.1 \pm 3.5\%$ ($\bar{x} \pm \mathrm{SD}$) of individuals, though all approaches performed relatively poorly with smaller, ostensibly younger, mice. Therefore, tail length, body length, and tail diameter, if treated in a comprehensive multivariate context, can yield substantial accuracy for distinguishing between coexisting northern salt marsh and western harvest mice. © 2018 The Wildlife Society.

KEY WORDS discriminant function analysis, field identification, multiple logistic regression, *Reithrodontomys*, salt marsh harvest mouse, Suisun Marsh, western harvest mouse.

Correctly identifying animals becomes important when managing endangered species (Bickford et al. 2007) because of the ecological and political implications of accurately establishing their range limits, population trends, and presence on public and private lands. The northern subspecies of the salt marsh harvest mouse (*Reithrodontomys raviventris halicoetes*) inhabits marshes bordering the northern extension of San Francisco Bay, San Pablo Bay, and Suisun Bay, California, USA, including the Suisun Marsh. The southern subspecies (*R. r. raviventris*) primarily occupies the southern portion of the San Francisco Bay (Shellhammer 1982). The salt marsh harvest mouse is geographically restricted to these North- and South-Bay salt marshes, and despite the relatively short geographic distances between them, there is significant genetic (Statham et al. 2016) and morphological (Fisler 1965) differentiation between subspecies. In addition, both are federal and California state-listed endangered species. The congeneric western harvest mouse (*R. megalotis*) is geographically widespread and inhabits most of California and contiguous states. The western harvest mouse occurs in a diversity of vegetation types and in San Francisco and associated bays, it overlaps the range of, and coexists with, both salt marsh harvest mouse subspecies (Fisler 1965, Shellhammer 1982,

Received: 3 March 2017; Accepted: 24 October 2017

[1]*E-mail: dsustaita@csusm.edu*
[2]*Current affiliation: Department of Biological Sciences, California State University San Marcos, 333 S. Twin Oaks Valley Rd., San Marcos, CA 92096*
[3]*Current affiliation: Washington Department of Fish and Wildlife, 600 Capitol Way N., Olympia, WA 98501*
[4]*Current affiliation: California Department of Fish and Wildlife, Nongame Wildlife Program, 1812 Ninth Street, Sacramento, CA 95811*

Webster and Jones 1982, Bias and Morrison 2006, Sustaita et al. 2011). This presents a problem for identifying and monitoring salt marsh harvest mouse populations in the field because of difficulties distinguishing between some ambiguous individuals with intermediate morphology (Hooper 1944, Fisler 1965, Bias and Morrison 2006, Statham et al. 2016).

At one point these more ambiguous forms were thought to be hybrids, considering the presumed close phylogenetic relation of the congeners (Hooper 1944, Fisler 1965). However, despite their morphological similarity and sympatry, salt marsh and western harvest mice are not sister taxa, and in fact, these 2 species are inferred to have diverged approximately 7 million years ago (Irwin et al. 1991). Chromosomal (Hood et al. 1984), allozyme (Nelson et al. 1984), and molecular (Bell et al. 2001; Brown 2003; Arellano et al. 2005, 2006) studies have demonstrated that the salt marsh harvest mouse is most closely related to the plains harvest mouse (*R. montanus*). Fisler (1965) reported no evidence that salt marsh and western harvest mice interbreed in captivity, and both mitochondrial (Brown 2003) and nuclear (Statham et al. 2016) DNA evidence do not support the existence of hybrids.

Despite the presence of ambiguous individuals, differences in morphological characteristics serve to distinguish salt marsh from western harvest mice in most cases. According to Fisler (1965), the pelage of salt marsh harvest mice is typically thicker than that of western harvest mice, and their tails are indistinctly bicolored and blunt tipped, as opposed to distinctly bicolored (brown or gray hairs on dorsal side, gray or white hairs on ventral side) and pointed. Venter color of salt marsh harvest mice varies from cinnamon-colored (southern subspecies) to much whiter (northern subspecies) compared to the typically dull gray-colored venters of western harvest mice (Fisler 1965, Shellhammer 1982). There are also additional qualitative and quantitative characteristics that help to distinguish between these species. Western harvest mice tend to have shorter head, body, tail, and hind foot lengths; lower tail:body length ratios; and lower body masses than salt marsh harvest mice (Fisler 1965, Shellhammer 1984). However, the magnitude of these interspecific differences depends on the subspecies of salt marsh harvest mouse, such that the southern subspecies overlaps with the western harvest mouse in these measurements substantially (Fisler 1965). There also appears to be pronounced differences in aspects of their reproductive biology, sociality, and temperament (Fisler 1965), but these are often difficult to distinguish in the short time mice are usually observed. In addition, field identification is complicated by such factors as observer experience (Shellhammer 1984), intraspecific variation in morphology (Fisler 1965), and allometric scaling of the most discriminatory characteristics, such as tail length (this study).

The prevailing method for distinguishing between salt marsh and western harvest mice is currently based largely on Shellhammer's (1984) dichotomous key, which was adapted from Fisler's (1965) work. However, this technique involves some subjective assessments of character states, such as the shape of the tail tip, ventral tail hair color, and the presence or absence of orange ear tufts (Shellhammer 1984). Furthermore, the dichotomous nature of the key presents difficulties for identifying individuals with ambiguous character states because it forces decisions to be made on character states that are neither entirely discrete nor binary. In a more general sense, because these keys (and the analyses upon which they were based) are operationally univariate, they do not account for potential differences between sex and age classes, allometry, or interactions among them when considering differences along key discriminatory metrics. Therefore, more probabilistic, statistical approaches based on morphological characteristics (Rich et al. 1996, Maldonado et al. 2004) are necessary, given that ambiguous or intermediate individuals occur in the field (Hooper 1944, Fisler 1965, Sustaita et al. 2011).

The problem of salt marsh harvest mouse field identification is not restricted to the northern subspecies (Statham et al. 2016). However, only the northern subspecies and western harvest mouse coexist throughout the Suisun Marsh study area, hence the imperative for distinguishing between these 2 taxa in particular. Considering the genetic and morphological differences between subspecies, and the comparatively small sample sizes of genetic and morphological data currently available for the southern subspecies (Statham et al. 2016), we included only the northern subspecies of the salt marsh harvest mouse in the interest of producing a more powerful predictive tool. Our primary objective was to characterize quantitative morphological and genetic differences between the endangered salt marsh harvest mouse and the western harvest mouse, and to produce a predictive model for assigning individuals to their most probable species.

STUDY AREA

We obtained the samples used in this study (Table S1, available online in Supporting Information; see also Sustaita et al. 2011) during long-term salt marsh harvest mouse population surveys, and targeted western harvest mouse trapping throughout the Suisun Marsh, Solano County, California (122° 0' W, 38° 10' N), from 1999–2004 (California Department of Water Resources 1984, 2000; Sustaita et al. 2011). The Suisun Marsh consisted of diked marshes managed for waterfowl and other wildlife, relatively unaltered tidal marsh, uplands, bays, sloughs, and other waterways (California Department of Water Resources [CDWR] 2000, Sustaita et al. 2011). The annual precipitation was 38–50 cm, with temperatures averaging 8° C in January and 22° C in July. Vegetation associations typically occurred in distinct elevational bands in the low-, mid-, and high tidal marshes, whereas in the diked marshes vegetation was either more patchily, or more homogenously, distributed in diked marshes (Sustaita et al. 2011). Trapping locations varied in plant species diversity, ranging from pickleweed (*Salicornia pacifica*)-dominated, to upland grass-dominated (e.g., Italian ryegrass [*Festuca perennis*] and bromes [*Bromus* spp.]), with most locations characterized by a mixture of halophytic species, such as Baltic rush (*Juncus*

balticus), Olney's threesquare bulrush (*Schoenoplectus americanus*), fathen (*Atriplex prostrata*), and saltgrass (*Distichlis spicata*). We selected trapping sites at random in some cases (Sustaita et al. 2011). In other cases, we used existing trapping grids established in particular locations for long-term monitoring purposes, and in some cases we haphazardly set traps in areas where one taxon or the other was known to occur in high abundances. We opened traps between sunset and sunrise, from March to October, when temperatures were 7–22° C.

METHODS

Field and Molecular Methods

We conducted salt marsh harvest mouse live-trapping under a California Department of Fish and Wildlife (CDFW) Memorandum of Understanding with the CDWR, a recovery permit between CDWR and the United States Fish and Wildlife Service (USFWS; no. TE835365-2), and a Cooperative Agreement between CDFW and USFWS. Upon capture, we recorded harvest mice as salt marsh, western, or unidentified; determined sex; and assessed them for reproductive condition (based on presence of inguinal testes in males and teats, pregnancy, or copulatory plugs in females (Adler and Wilson 1987, Skupski 1995). We collected standard mammal specimen preparation measurements (Nagorsen and Peterson 1980) including total length and tail length (from which we calculated body length). We measured tail diameter at approximately 20 mm caudal to the base of the tail to the nearest 0.05 mm with dial calipers (General, Secaucus, NJ, USA) and mass to the nearest 0.25 g with a 30-g Pesola scale (Pesola AG, Baar, Switzerland). Prior to their release, we individually marked all harvest mice using aluminum coded ear tags (size 0.1, 018 M National Band and Tag, Newport, KY, USA). Over the course of multiple trap nights and seasonal trapping sessions, mice were often recaptured by the same, or different, observers and re-measured 2–4 times. We subsequently used these additional measurements for repeatability analysis, but we used only the first measurement obtained for any given individual in the morphological analyses because we captured most individuals only once. Using sterile forceps, we collected tufts of hair, including follicles, from the rump of harvest mice. We stored hair samples in Eppendorf tubes at −18° C and delivered them to California Polytechnic State University, San Luis Obispo for molecular analysis.

We analyzed samples using a standard DNA barcode approach (Hebert et al. 2003) in 2002 and in 2004. We used samples from 2002 ($n = 64$) to initially create phylogenetic trees and determine restriction fragment length polymorphism cutting sites, which would distinguish between the species. Here we summarize the molecular techniques used for determining species identity (see Details of Molecular Methods in Supporting Information available online). We extracted DNA from hair samples using the Qiagen QiaAmp tissue kit (Qiagen, Valencia, CA, USA). We amplified a 450 base-pair fragment of the mitochondrial DNA cytochrome b via polymerase chain reaction (PCR) using MVZ 05 (Smith and Patton 1993) and MVZ 04R, a *Reithrodontomys*-specific primer (5′CCTCAGAATGATATTTGTCCTC 3′) based on MVZ 04 (Smith and Patton 1993) and sequenced on an ABI 377 sequencer (Applied Biosystems, Foster City, CA, USA). Maximum parsimony and bootstrap analyses were based on 329 nucleotides of the cytochrome b gene and included 2 salt marsh and 2 western harvest mouse reference sequences and Genbank sequences from the following outgroups: plains harvest mouse (*Reithrodontomys montanus*), Zacateca's harvest mouse (*R. zacatecae*), eastern harvest mouse (*R. humulis*), Sumichrast's harvest mouse (*R. sumichrasti*), fulvous harvest mouse (*R. fulvescens*), and Mexican harvest mouse (*R. mexicanus*; see Details of Molecular Methods in Supporting Information for accession numbers and further details). The mean percent sequence diversity within species was 1.03% for the salt marsh harvest mouse and 1.01% for the western harvest mouse. The mean between-species divergence was 12.03%. Monophyletic clades with 100% and 96% bootstrap values separated salt marsh and western harvest mice, respectively (Brown 2003). Thirty characters, out of 329, were identified as species-specific (synapomorphic) character state changes. Based on these results, Brown (2003) performed (APO-1 and EAR-1) restriction fragment length polymorphism analyses on samples taken in 2004 ($n = 192$) that cut the amplicons in a species-specific manner and distinguished salt marsh from western harvest mouse samples with greater efficiency. The enzyme APO-1 cuts at sites 137 and 175 in the western harvest mouse but does not cut salt marsh harvest mouse amplicons. The enzyme EAR-1 cuts at site 125 in salt marsh harvest mouse but does not cut western harvest mouse amplicons. We used two enzymes in all samples from 2004, which resulted in one species-specific diagnostic restriction pattern.

Statistical Procedures

First, we examined the extent of overlap between genetic and field data sets for each morphological variable to determine if both were representative of the same population. Second, we tested the repeatability of each morphological measurement. We used the intraclass correlation (r_I) to test the null hypothesis of no correlation among (repeated) measurements (Lessells and Boag 1987, Zar 1999) within and among observers. Third, for each variable we conducted a 3-way analysis of variance (ANOVA) to determine differences between species, sexes, reproductive states, and their interactions using only the genetically verified cases to establish a more accurate baseline of inter- and intraspecific differences. Finally, we assessed tail length allometry by ordinary least-squares regression of $\log_{10}$ tail length on $\log_{10}$ body mass. We considered slopes to be allometric if they departed significantly (difference of slopes test; Sokal and Rohlf 1995) from $b = 0.33$ (i.e., the isometric expectation for a linear dimension with respect to a volumetric dimension; Schmidt-Nielsen 1984).

We performed 2 common forms of classification analysis–discriminant function analysis (DFA) and multiple logistic regression (MLR; Tabachnick and Fidell 2001, Quinn and

Keough 2002, Sustaita et al. 2014) and compared them to find the most robust predictive tool. To explicitly account for sex and reproductive condition in the framework of DFA, we conducted separate DFAs for each sex (M, F) and reproductive condition (reproductive, non-reproductive; see Table S2 for sample sizes). Reproductive condition has been regarded by some workers as a proxy of age, but we used the actual condition (reproductive, non-reproductive), rather than using the condition to imperfectly infer age. We ran the MLRs including sex and reproductive condition as additional categorical predictors to account for their potential effects and interactions with the quantitative variables.

Our ultimate objective was to determine the optimal combination of variables for discriminating between species and for classifying unidentified harvest mouse cases. Thus, our model selection procedure was not guided by specific *a priori* hypotheses but rather was influenced by long-run model classification accuracy. To this end, we used hierarchical and stepwise methods based on Akaike's Information Criterion (AIC; see Details of Model Selection Procedures in Supporting Information) in conjunction with

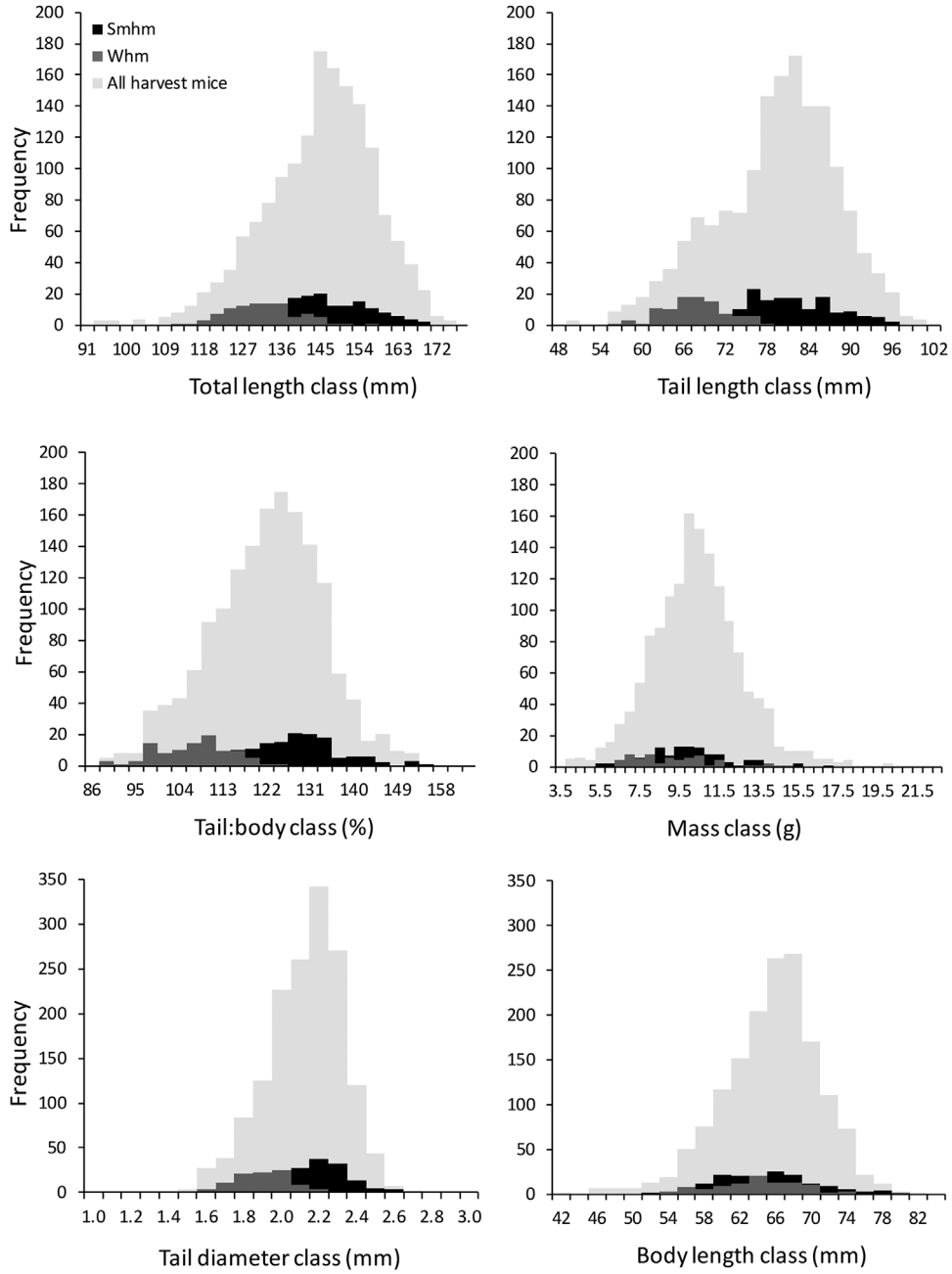

Figure 1. Distributions of field measurements for salt marsh (Smhm) and western (Whm) harvest mice captured in the Suisun Marsh, California, USA, 1999–2004. The lightest gray bars depict the distribution of all individuals (including genetically verified, field-identified, and unidentified cases; $n = 1{,}583$). The darker-gray and black bars depict the portions of the distributions that genetically verified Whm ($n = 97$) and Smhm ($n = 160$) comprise, respectively.

Table 1. Repeatability of morphological measurements of recaptured harvest mice in the Suisun Marsh, California, USA during a given 4-night trap session, 1999–2004. The r_I is the intraclass correlation (Zar 1999) computed for each of 4 observers separately (within-observer r_I) and all observers collectively (among-observer r_I). Reported are the ranges of r_I, k (i.e., number of individual mice for which 2–4 repeated measurements were taken by a given observer), and P calculated for varying numbers of observers (Obs) who measured ≥8 recaptured mice for a given variable. We computed among-observer repeatability based on successive measurements of recaptured mice performed by different observers each time.

	Within-observer				Among-observer		
Measurements	r_I	k	P	**Obs**	r_I	$F_{1,2}$	P
Tail length (mm)	0.91–0.97	27–68	<0.001	4	0.81	$12.2_{426,\ 663}$	<0.001
Body length (mm)	0.30–0.78	24–68	≤0.070	4	0.61	$5.0_{415,644}$	<0.001
Tail diameter (mm)	0.38–0.80	8–13	0.008–0.066	2	0.10	$1.2_{141,168}$	0.163
Mass (g)	0.91–0.92	8–13	<0.001	2	0.85	$12.8_{109,129}$	<0.001
Total length (mm)[a]	0.89–0.95	26–67	<0.001	4	0.90	$24.1_{427,660}$	<0.001

[a] Not included in discriminant function/multiple logistic regression analyses.

classification performance (i.e., ability to correctly classify known cases) to select the best combination of variables for discriminating between species and classifying new cases. A substantial portion of the cases did not contain data on mass, so we split the data into 2 sets: one composed of all $n = 256$ cases and one composed of $n = 170$ cases that included mass. We used all cases in DFA and MLR analyses, save for the cross-validation procedures, in which we held out varying numbers of cases ($n = 1$–64) for testing model classification. We employed 2 types of cross-validation procedures to determine the percentage of correctly classified cases for each of the best DFA and MLR models. The first procedure involved randomly designating 25% of the genetically verified cases for salt marsh and western harvest mice as unselected cases that were not involved in generating model coefficients but rather were classified by them. We then repeated these analyses for 100 randomly permuted cross-validated results. The second approach (used for DFAs only) consisted of a jackknife (leave-one-out cross-validation) to successively remove each case and classify it, based on models derived using the remaining cases (Tabachnick and Fidell 2001, Quinn and Keough 2002). We then tested for differences in model classification performance within and between data sets that contained or excluded body mass and between the best DFA and MLR models using independent samples and paired t-tests and a chi-square goodness-of-fit test (see Details of Model Selection Procedures, available in Supporting Information).

We screened the data for outliers and distributional problems prior to analyses. We excluded 3 cases that were missing measurements from the classification analyses. We $\log_{10}$-transformed all (quantitative) variables prior to analyses to help improve normality and homoscedasticity but used raw data for the repeatability analyses. We used SPSS 14.0 (SPSS, Chicago, IL, USA) for all statistical tests except for the intraclass correlations (Zar 1999), linear regressions, and t-tests, which we computed in Excel (Microsoft, Redmond, WA, USA). For the DFAs we interpreted Box's M to assess the assumption of homogeneity of between-groups covariance matrices (Tabachnick and Fidell 2001). We used all valid cases in the analyses (despite the unequal group sample sizes that resulted) to ensure the range of variability in harvest mouse dimensions was adequately represented. We assessed significance levels at $\alpha = 0.05$.

RESULTS

Morphological Variation and Measurement Repeatability

The individuals used for the genetic analysis (Table S1; Brown 2003) adequately represented the variation in our

Table 2. Tail and body length, tail diameter, and mass measurements, by sex and reproductive (repro) condition, of genetically verified salt marsh and western harvest mice in the Suisun Marsh, California, USA, 1999–2004 used in the discriminant function and multiple logistic regression analyses. Superscripts indicate significant ($P < 0.05$) effects of species (a), sex (b), reproductive condition (c), and sex × reproductive condition (d), based on a 3-way analysis of variance for each measurement (all other interactions were not significant).

			Tail length (mm)[a,b,c,d]		Body length (mm)[d]		Tail diameter (mm)[a,b,c]		Mass (g)[b,c]	
Species	**Sex**	**Repro status**	$\bar{x}$	**SD (*n*)**	$\bar{x}$	**SD (*n*)**	$\bar{x}$	**SD (*n*)**	$\bar{x}$	**SD (*n*)**
Salt marsh harvest mouse	M	Repro	82.3	6.6 (58)	65.6	4.4 (58)	2.23	0.16 (58)	10.1	0.15 (42)
		Non-repro	81.9	7.6 (58)	63.7	4.6 (24)	2.20	0.18 (24)	8.5	0.43 (13)
	F	Repro	81.4	5.7 (40)	66.9	5.6 (40)	2.16	0.14 (40)	11.8	0.39 (29)
		Non-repro	74.6	6.7 (34)	60.0	4.9 (34)	2.00	0.18 (34)	8.0	0.30 (26)
	Pooled over sex and repro condition		80.3	7.2 (158)	64.4	5.5 (156)	2.17	0.18 (156)	9.8	0.20 (110)
Western harvest mouse	M	Repro	63.3	4.7 (27)	64.1	3.3 (27)	1.98	0.12 (27)	9.1	0.30 (9)
		Non-repro	65.9	3.8 (26)	60.7	4.8 (26)	1.88	0.18 (26)	7.6	0.27 (20)
	F	Repro	67.8	4.8 (31)	66.6	4.5 (31)	1.89	0.10 (31)	11.2	0.47 (20)
		Non-repro	65.0	3.5 (14)	61.0	4.7 (14)	1.85	0.22 (14)	8.1	0.37 (11)
	Pooled over sex and repro condition		67.1	4.5 (98)	63.6	4.5 (98)	1.91	0.15 (98)	9.1	0.28 (60)

larger data set, as indicated by the similarity of their distributions along each morphological variable (Fig. 1). Repeatability was variable but relatively high and significant for measurements taken among different observers, with the exception of tail diameter, which was not significantly repeatable (Table 1).

Salt marsh harvest mice had longer tails with larger diameters than did western harvest mice, when accounting for sex and reproductive condition. In both species, tail diameters were larger in males than in females and in reproductive individuals than in non-reproductive individuals (Table 2). However, differences in tail length between sexes and reproductive states were complicated by a significant sex × reproductive condition interaction (Table 2). Body length and mass did not differ appreciably between species, although there were significant differences (and interactions) between sexes and reproductive states (Table 2). Tail length increased with body mass in both species but at a rate slower than expected from geometric scaling (salt marsh harvest mouse: $r^2 = 0.13$, $b = 0.13$, $F = 16.50$, $P < 0.001$, $n = 106$; $t_{b=0.33} = 6.26$, $P < 0.001$; western harvest mouse: $r^2 = 0.22$, $b = 0.14$, $F = 18.01$, $P < 0.001$, $n = 61$; $t_{b=0.33} = 6.08$, $P < 0.001$). In other words, mass-specific tail length decreased with increasing body mass in both species (Fig. 2).

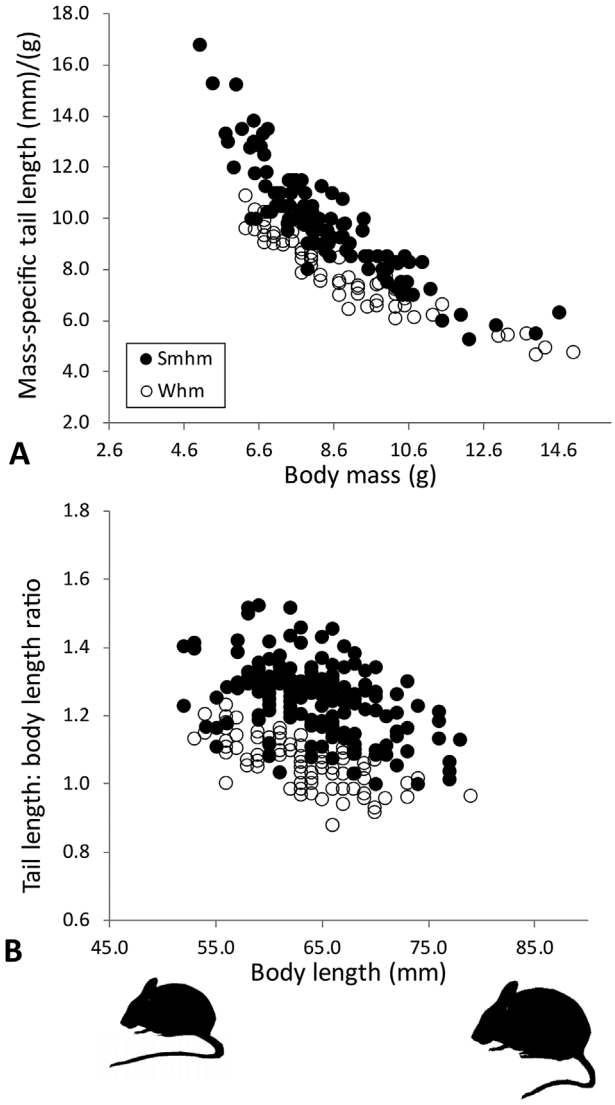

Figure 2. Scaling of tail length among genetically verified salt marsh (Smhm; filled circles) and western (Whm; open circles) harvest mice individuals captured throughout the Suisun Marsh, California, USA, 1999–2004. Raw mass-specific tail length (tail length/body mass) with respect to body mass is depicted here for clarity (A); note, however, that we computed the observed negative allometries from $\log_{10}$ tail $-$ $\log_{10}$ mass regressions. The same pattern holds with respect to body length (B). Although Smhm tend to have greater mass-specific tail lengths and tail:body ratios than do Whm for any given body size, these tend to decrease with increasing body size in both species (as depicted by the mouse silhouettes).

Classification Analyses

The DFA and MLR performed very well in terms of their overall percentages of correct classification. However, based on random permutations of cross-validation tests of the top model generated by each approach, the MLR yielded a higher correct-classification percentage than the DFA 60 out of 100 times ($\chi^2_1 = 4.0$, $P = 0.046$). Furthermore, the MLR showed a slightly higher average correct-classification percentage than the DFA ($90.1 \pm 3.5\%$ [SD] and $89.4 \pm 3.1\%$, respectively; paired t-test; $t_{50} = -2.28$, $P = 0.025$). The best DFA and MLR models agreed on the assignment of 91.2% of the cases (correct and incorrect) to harvest mouse species. When the 2 forms of analysis disagreed on the assignment of individual cases, the MLR was correct on 61.7% (vs. 38.3% for DFA) of the cases. Additionally, the DFA exhibited relatively greater difficulty correctly assigning salt marsh (85.4%) than western harvest mice (96.9%), whereas the MLR showed similar percentages of correct classification between groups (94.2% and 93.8%, respectively). As a result, here we detail the results of the MLR analysis and for the DFAs we refer readers to the Results of Discriminant Function Analyses, available in Supporting Information.

The best MLR model based on the dataset containing mass, selected by the stepwise procedure included 12 variables and interaction terms (-2Log likelihood $= 33.60$, AIC $= 57.6$; Table 3, Table S3). The percentage of cross-validated correctly classified unselected cases of known identity (from random subsets) was $90.1 \pm 4.4\%$. Three other models, containing 10–13 terms, selected by the stepwise procedure fell within 2 AIC units (Table 3). These models yielded similarly high rates of correct classification (95.2–97.6%). The best model derived from the hierarchical analysis trailed the stepwise models (ΔAIC $= 3.3$) and had 9 parameters and a correct-classification rate of 91.3%.

The best MLR model, based on the dataset excluding mass, selected by the stepwise procedure included 8 terms (-2Log likelihood $= 88.92$, AIC $= 104.9$; Table 3, Table S4; Fig. 3C). The percentage of correctly classified unselected cases of known identity (from random subsets) was $90.1 \pm 3.5\%$. Two other models, containing 7–9 terms, selected by the stepwise and hierarchical procedures fell within 2 AIC units (Table 3). These models yielded similarly high rates of correct classification for unselected cases (90.8–

Table 3. Top stepwise (S) and hierarchical (H) multiple logistic regression models used to classify salt marsh and western harvest mice in the Suisun Marsh, California, USA, 1999–2004, based on tail and body length [mm], tail diameter [taild; mm], body mass [g], sex, reproductive condition [repro], and interactions among these variables. The percentages of correctly classified cases (%CC) were based on those used to both generate and test model coefficients, whereas the percentages of correctly classified unselected cases of known identity (%CCrs) were based on the assignment of random subsets of 25% of cases held out of the model-building portion of the procedure. Also reported are the −2Log likelihood (−2Loglik) values for model fit, number of parameters (*K*), Akaike's Information Criterion (AIC), and ΔAIC values. We computed ΔAIC values from the full complement of 33 and 26 models, based on data including ($n = 168$) and excluding ($n = 254$) body mass, respectively.

Model type	Model terms	%CC	%CCrs	−2Loglik	*K*	AIC	ΔAIC
Including mass[a]							
S	Tail + body + sex + tail×sex + body×sex + body×repro + body×repro×sex + taild×sex + taild×repro×sex + mass×sex + mass×repro + mass×repro×sex	95.2	90.1 ± 4.4[b]	33.60	12	57.6	0.0
S	Tail + body + sex + tail×sex + body×sex + body×repro + body×repro×sex + taild×sex + mass×sex + mass×repro + mass×repro×sex	95.2	95.2	36.24	11	58.2	0.6
S	Tail + body + sex + tail×sex + body×sex + body×repro + body×repro×sex + taild×sex + mass×repro + mass×repro×sex	95.2	95.2	38.75	10	58.7	1.1
S	Tail + body + sex + repro×sex + tail×sex + body×sex + body×repro + body×repro×sex + taild×sex + taild×repro×sex + mass×sex + mass×repro + mass×repro×sex	94.6	97.6	32.85	13	58.9	1.3
H	Repro×sex + mass×repro + mass×sex + taild×repro + taild×sex + tail×sex + sex + body + tail	94.0	91.3	42.89	9	60.9	3.3
Excluding mass[c]							
S	Tail + body + taild + sex + tail×sex + body×repro + taild×sex + taild×repro	94.1	90.1 ± 3.5[b]	88.92	8	104.9	0.0
H	Tail + body + taild + sex + repro + tail×sex + taild×repro	91.7	90.9	92.64	7	106.6	1.7
S	Tail + body + taild + sex + tail×sex + tail×repro×sex + body×repro + taild×sex + taild×repro	93.7	90.8	88.78	9	106.8	1.9

[a] −2Log likelihood values from constant-only model = 220.74.
[b] $\bar{x} \pm$ SD correct-classification percentage based on 100 random sub-selections of 25% of the genetically verified cases for each species.
[c] −2Log likelihood values from constant-only model = 340.70.

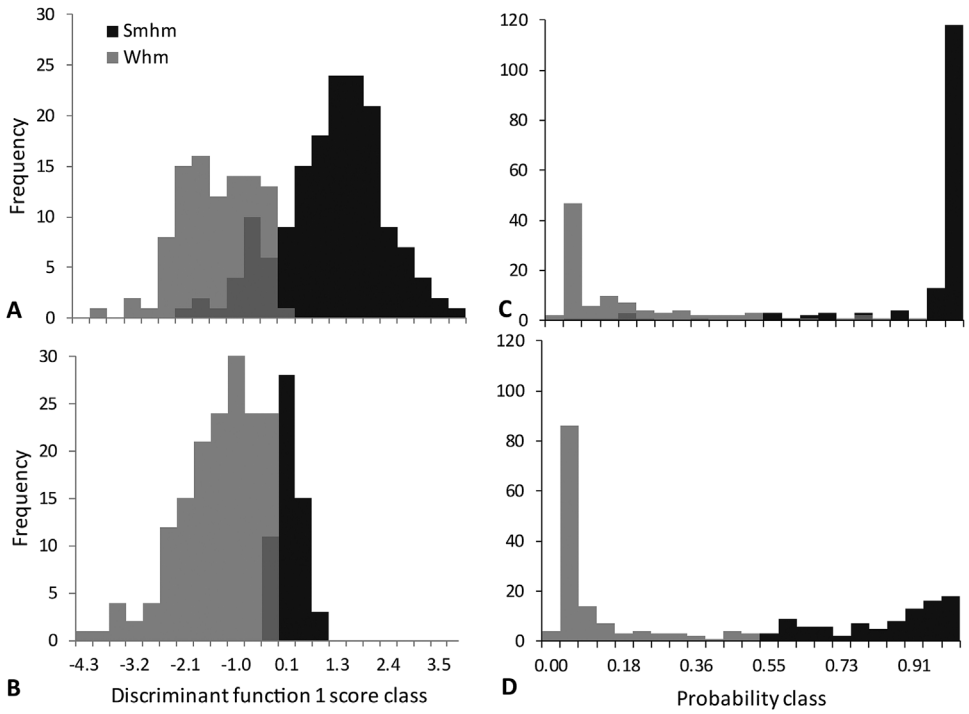

Figure 3. Relative frequency histograms of discriminant function scores (A, B) and multiple logistic regression posterior probabilities (C, D) generated from the top models based on the complete genetically verified data set (i.e., excluding mass; $n = 253$–255) of salt marsh and western harvest mice sampled throughout the Suisun Marsh, California, USA, 1999–2004. The upper graphs (A, C) show results for genetically verified salt marsh (Smhm; black) and western (Whm; gray) harvest mouse cases (note regions of overlapping distributions in darker gray). The lower graphs (B, D) show results for $n = 228$ cases of uncertain identity (i.e., those not easily diagnosable in the field and not genetically verified) assigned to Smhm or Whm by the models.

90.9%). Based on 100 random permutations of cross-validation test sets, there was no significant difference in correct-classification performance between the best MLR model with ($90.1 \pm 4.4\%$) and without ($90.1 \pm 3.5\%$) body mass (t-test assuming unequal variances; $t_{187} = -0.051$, $P = 0.959$).

DISCUSSION

Salt marsh and western harvest mice differ in key traits (e.g., tail length), and some of these differences depend on sex, reproductive condition, and body size. Therefore, these interactions (Table 2) should be taken into consideration in objective classification models. Our statistical models can assign individuals to their species with reasonable accuracy (correct classification of approximately 90%), despite considerable variation in within- and among-observer field measurement repeatability. Model selection and comparison identified advantages of MLR, over DFA, for classification analysis.

During this study we confronted the reality that when the identification of small mammals is based on external quantitative characters, the characters typically used describe adult individuals, and thus allow only for the identification of adult individuals (Hall 2001). We posit something that should be intuitive: reliance solely on adult characteristics may lead to misclassification of some individuals because of the effects or interactions of age, sex, and allometry. This limitation to taxonomy and diagnosis is likely common (Hall 2001). This is particularly problematic for a case like the federal and California state-listed endangered salt marsh harvest mouse that co-occurs with the western harvest mouse throughout much of its range (Fisler 1965, Shellhammer 1982), where currently both species are identified by adult characteristics without regard to allometry. Salt marsh and western harvest mice are very similar morphologically but differ primarily in the salt marsh harvest mouse's greater tail length (absolutely and relative to body length and mass) and tail diameter. These characters have been considered important quantitative metrics for distinguishing between species in the field (Shellhammer 1984, Brown 2003). Because smaller (in terms of body mass) mice of both species tend to have disproportionately longer tails than larger mice (Fig. 2) and salt marsh harvest mice tend to have longer tails than western harvest mice (Table 2), younger (hence, smaller) mice of both species are proportioned more like salt marsh harvest mice, which presents a problem for traditional univariate approaches and dichotomous keys.

Multivariate techniques that simultaneously account for body size, sex, and age effects are advantageous. Nevertheless, we found that even the best models still performed suboptimally when assigning the smallest (i.e., juvenile) mice, based on their equivocal probabilities (MLR) and posterior probability values (DFA). Considering the negative allometry observed, and given sexual size dimorphism (Table 2), females are likely to be misclassified more often. Likewise, individual salt marsh harvest mice with disproportionately long tails may correctly classify, whereas western harvest mice with disproportionately longer tails will more likely misclassify. Molecular tools may allow us to determine the direction of misclassification, and to more rigorously identify the mechanism (allometry, sexual dimorphism, age) resulting in misclassifications. It is possible that upon further investigation in other systems, we will find that not all ages and sexes contribute equally to misclassifications. Instead, interactions of age, sex, and allometry may define the segment of the population most prone to misclassification.

The advantages of these analyses over the existing dichotomous keys are 3-fold. First, they are based on a larger, Suisun Marsh-specific data set. Second, because they are multivariate procedures, they incorporate simultaneous effects and interactions among quantitative metrics and categorical states that vary differently between species. Third, these models can be used by observers with varying degrees of experience because they rely on objective field measurements that in some cases may not necessarily be highly repeatable. For example, tail diameter, a measurement that was not significantly repeatable among observers, actually improved model fit and correct-classification rates, perhaps because differences between species were on the order of approximately 0.2 mm, whereas our measurements were made to the nearest 0.05 mm. Body mass is another measure that could suffer from low repeatability, given that it is largely dependent on foraging success and body condition, and thus can vary widely within individuals. Nevertheless, mass demonstrated high repeatability in this study and was routinely included in the top models during model selection. However, there was no significant difference in long-run classification performance with or without mass, perhaps because the data set lacking mass was larger than the one that included it.

When we compared DFA and MLR, the difference in the average percentage of correctly classified cases was not substantial. However, the MLR was correct slightly more often when the 2 approaches disagreed on the assignment of an individual, and the MLR demonstrated less bias in its tendency to assign cases to one species or the other. Furthermore, the MLR has the advantage of explicitly incorporating categorical sex and reproductive condition states, which improved model fit and performance. When we applied both the MLR (Appendix A) and DFA (Table S5) to 228 cases of uncertain identity (i.e., those not easily diagnosable in the field and not genetically verified) from our 1999–2004 trapping records, we found that both approaches agreed on 81.5% of the cases. Notwithstanding the ostensibly superior performance of MLR (on average), it is virtually impossible to determine which analysis is accurate when applied blindly to cases of uncertain species identity. Thus, one conservative approach for dealing with the non-overlapping cases is to assess the relative magnitudes of their MLR assignment probabilities (Fig. 3D; see also Results of Discriminant Function Analyses, available in Supporting Information) on a case-by-case basis to help guide decisions of whether to assign them to species or simply leave them as unidentified. These models can be applied to new data from the same or other populations in Suisun Marsh that fall

within the ranges of data used to build them. We stress, however, that new models should be generated for mice from other geographic regions (e.g., the southern subspecies of the salt marsh harvest mouse) or those from the Suisun Marsh with measurements falling outside the ranges represented in this data set.

MANAGEMENT IMPLICATIONS

Our statistical approaches correctly classified approximately 90% of co-existing endangered salt marsh and non-listed western harvest mice in the Suisun Marsh, California, based on a few key field measurements. Because of its slightly better average correct-classification rate and flexibility for including categorical predictors, we recommend multiple logistic regression for building future classification models for other localities or subspecies of salt marsh harvest mice. Furthermore, we found that measurement repeatability varied within and among observers. Thus, we suggest that when multiple researchers are involved in measuring animals, and especially for the first few days of a new study, that each observer measure several of the same animals upon capture (or recapture), so that repeatability can be assessed, and ultimately improved.

These species assignments are probability-based estimates specific to the populations in the Suisun Marsh, which in some cases should not supplant more definitive molecular procedures for verifying species. We provide some prescriptions for their use above, but in general, their application should be mediated by the circumstances. For instance, if the intent is to demonstrate occupancy of a site by salt marsh harvest mice, then the multivariate approach could be appropriate. One instance of a high probability assignment to salt marsh harvest mouse would establish occupancy with confidence. If, on the other hand, every individual in a sample needs to be assigned to a species with ≥ 0.95 certainty, then a genetic approach (i.e., DNA barcoding) would be required for individuals that do not classify with high probability.

ACKNOWLEDGMENTS

We are grateful to G. S. Van Klompenburg, L. E. Briden, J. A. Niceswanger, and P. K. Garrison III, for vital assistance at various points throughout the field study, and to N. W. Clipperton for assistance with molecular analyses. We are also grateful to the California Department of Fish and Wildlife's Grizzly Island Wildlife Area and the Solano Land Trust's Rush Ranch for access. Funding was provided by the California Department of Water Resources and the United States Bureau of Reclamation. We thank J. Sanderlin, and 2 reviewers for their constructive comments toward improving the manuscript.

LITERATURE CITED

Adler, G. H., and M. L. Wilson. 1987. Demography of a habitat generalist, the white-footed mouse, in a heterogeneous environment. Ecology 68:1785–1796.

Arellano, E., F. X. González-Cozátl, and D. S. Rogers. 2005. Molecular systematics of Middle American harvest mice *Reithrodontomys* (Muridae), estimated from mitochondrial cytochrome b gene sequences. Molecular Phylogenetics and Evolution 37:529–540.

Arellano, E., D. S. Rogers, and F. X. González-Cozátl. 2006. Sistemática molecular del género *Reithrodontomys* (Rodentia: Muridae). Pages 27–35 *in* E. Vázquez-Domínguez and D. J. Hafner, editors. Genética y mamíferos mexicanos: presente y futuro. New Mexico Museum of Natural History and Science, Albuquerque, USA.

Bell, D. M., M. J. Hamilton, C. W. Edwards, L. E. Wiggins, R. M. Martinez, R. E. Strauss, R. D. Bradley, and R. J. Baker. 2001. Patterns of karyotypic megaevolution in *Reithrodontomys*: evidence from a cytochrome b phylogenetic hypothesis. Journal of Mammalogy 82:81–91.

Bias, M. A., and M. L. Morrison. 2006. Habitat selection of the salt marsh harvest mouse and sympatric rodent species. Journal of Wildlife Management 70:732–742.

Bickford, D., D. J. Lohman, N. S. Sodhi, P. K. L. Ng, R. Meier, K. Winker, K. K. Ingram, and I. Das 2007. Cryptic species as a window on diversity and conservation. Trends in Ecology and Evolution 22:148–155.

Brown, S. K. 2003. Development of a molecular diagnostic technique to Identify salt marsh harvest mice (*Reithrodontomys raviventris*) and western harvest mice (*Reithrodontomys megalotis*) in Suisun Marsh, California. Thesis, California Polytechnic University, San Luis Obispo, San Luis Obispo, USA.

California Department of Water Resources [CDWR]. 1984. Plan of protection for the Suisun Marsh including environmental impact report. California Department of Water Resources, Sacramento, USA.

California Department of Water Resources [CDWR]. 2000. Suisun Marsh Monitoring Program reference guide. The Environmental Services Office of the California Department of Water Resources of The Resources Agency of the State of California, Sacramento, USA.

Fisler, G. F. 1965. Adaptations and speciation in harvest mice of the marshes of San Francisco Bay. University of California Publications 77:1–108.

Hall, E. R. 2001. Mammals of North America. Blackburn Press, Caldwell, New Jersey, USA.

Hebert, P. D. N., A. Cywinska, S. L. Ball, and J. R. DeWaard. 2003. Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences 270:313–321.

Hood, C. S., L. W. Robbins, R. J. Baker, and H. S. Shellhammer. 1984. Chromosomal studies and evolutionary relationships of an endangered species, *Reithrodontomys raviventris*. Journal of Mammalogy 65:655–667.

Hooper, E. T. 1944. San Francisco Bay as a factor influencing speciation in rodents. University of Michigan, Museum of Zoology, Ann Arbor, USA.

Irwin, D. M., T. D. Kocher, and A. C. Wilson. 1991. Evolution of the cytochrome b gene of mammals. Journal of Molecular Evolution 32:128–144.

Lessells, C. M., and P. T. Boag. 1987. Unrepeatable repeatabilities: a common mistake. Auk 104:116–121.

Maldonado, J. E., F. Hertel, and C. Vila. 2004. Discordant patterns of morphological variation in genetically divergent populations of ornate shrews (*Sorex ornatus*). Journal of Mammalogy 85:886–896.

Nagorsen, D. W., and R. L. Peterson. 1980. Mammal collectors' manual. A guide for collecting, documenting, and preparing mammal specimens for scientific research. Life Sciences Miscellaneous Publications, Royal Ontario Museum, Toronto, Canada.

Nelson, K. R., R. J. Baker, H. S. Shellhammer, and R. K. Chesser. 1984. Test of alternative hypotheses concerning the origin of *Reithrodontomys raviventris*: genetic analysis. Journal of Mammalogy 65:668–673.

Quinn, G. P., and M. J. Keough. 2002. Experimental design and data analysis for biologists. Cambridge University Press, Cambridge, United Kingdom.

Rich, S. M., C. W. Kilpatrick, and J. L. Shippee. 1996. Morphological differentiation and identification of *Peromyscus leucopus* and *P. maniculatus* in northeastern North America. Journal of Mammalogy 77:985–991.

Schmidt-Nielsen, K. 1984. Scaling, why is animal size so important? Cambridge University Press, New York, New York, USA.

Shellhammer H. 1982. *Reithrodontomys raviventris*. Mammalian species 169:1–3.

Shellhammer, H. S. 1984. Identification of salt marsh harvest mice *Reithrodontomys raviventris*, in the field and with cranial characteristics. California Fish and Game 70:113–120.

Skupski, M. 1995. Population ecology of the western harvest mouse, *Reithrodontomys megalotis*: a long-term perspective. Journal of Mammalogy 76:358–367.

Smith, M. F., and J. L. Patton. 1993. The diversification of South American murid rodents: evidence from mitochondrial DNA sequence data for the Akodontine Tribe. Biological Journal of the Linnaean Society 50:149–177.
Sokal, R. R., and F. J. Rohlf. 1995. Biometry. Third edition. W. H. Freeman and Company, New York, New York, USA.
Statham, M. J., S. Aamoth, L. Barthman-Thompson, S. Estrella, S. Fresquez, L. D. Hernandez, R. Tertes, and B. N. Sacks. 2016. Conservation genetics of the endangered San Francisco Bay endemic salt marsh harvest mouse (*Reithrodontomys raviventris*). Conservation Genetics 17:1055. https://doi.org/10.1007/s10592-016-0843-4
Sustaita, D., C. L. Owen, J. C. Villarreal, and M. A. Rubega. 2014. Morphometric tools for sexing loggerhead shrikes in California. Southwestern Naturalist 59:560–567.
Sustaita, D., P. F. Quickert, L. Patterson, L. Barthman-Thompson, and S. Estrella. 2011. Salt marsh harvest mouse demography and habitat use in Suisun Marsh, California. Journal of Wildlife Management 75:1498–1507.
Tabachnick, B. G., and L. S. Fidell. 2001. Using multivariate statistics. Allyn and Bacon, Boston, Massachusetts, USA.
Webster, W., and J. Jones. 1982. *Reithrodontomys megalotis*. Mammalian Species 167:1–5.
Zar, J. H. 1999. Biostatistical analysis. Prentice Hall, Upper Saddle River, New Jersey, USA.

Associate Editor: Jamie Sanderlin.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's website.

APPENDIX A. TOP MULTIPLE LOGISTIC REGRESSION MODEL PARAMETERS

Coefficients (B) from the top models used to classify harvest mice of uncertain species identity (i.e., those that are not easily diagnosable in the field and have not been genetically verified), captured throughout the Suisun Marsh, California, during 199–2004. Probability (P) $= e^X/1 + e^X$; $X = X_1B_1 + X_2B_2 + X_iB_i + X_0B_0$; where $P > 0.5$ = salt marsh harvest mouse and $P < 0.5$ = western harvest mouse.

	With tail diameter		Excluding tail diameter	
Independent variable X	***B***	**SE**	***B***	**SE**
Data set excludes body mass				
($\log_{10}$) tail length	134.89	27.66	107.90	22.91
($\log_{10}$) body length	−36.25	10.52	−38.01	9.07
($\log_{10}$) tail diameter	−26.65	19.10		
sex[a]	111.00	45.83	85.47	36.36
repro[b]			−86.86	46.03
($\log_{10}$) tail length × sex[a]	−64.99	26.73	−45.67	19.60
($\log_{10}$) tail length × repro[b]			46.00	24.75
($\log_{10}$) tail diameter × sex[a]	36.89	21.71		
($\log_{10}$) tail diameter × repro[b]	72.59	23.66		
($\log_{10}$) body length × repro[b]	−12.94	4.01		
Constant (B_0)	−176.68	43.29	−131.39	37.42
Data set includes body mass				
($\log_{10}$) tail length	157.23	50.99	219.35	66.85
($\log_{10}$) body length	−48.69	23.73	−168.26	50.16
($\log_{10}$) tail diameter				
($\log_{10}$) mass			51.68	17.51
sex[a]	154.35	77.25	145.54	91.90
repro[b]			−146.83	64.23
($\log_{10}$) tail length × sex[a]	−102.01	54.06	−163.33	69.85
($\log_{10}$) body length × repro[b]	1.37	6.47	104.75	40.30
($\log_{10}$) body length × sex[a]	12.08	27.83	102.01	47.41
($\log_{10}$) tail length × repro[b] × sex[a]			77.54	47.41
($\log_{10}$) body length × repro[b] × sex[a]	4.28	8.22	−77.73	48.27
($\log_{10}$) tail diameter × sex[a]	49.72	21.67		
($\log_{10}$) mass × sex[a]			−28.74	15.85
($\log_{10}$) mass × repro[b]	−3.27	11.87	−47.61	16.43
($\log_{10}$) mass × repro[b] × sex[a]	−8.24	14.77		
Constant (B_0)	−204.06	71.62	−151.14	81.04

[a]sex = 1 for female, 0 for male.
[b]repro = reproductive condition; 1 for yes, 0 for no.