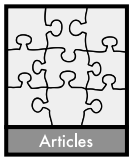

J. Avian Biol. 41: 128–138, 2010
doi: 10.1111/j.1600-048X.2009.04773.x

Received 12 December 2008, accepted 2 June 2009

Geographic variation in the plumage coloration of willow flycatchers *Empidonax traillii*

Eben H. Paxton, Mark K. Sogge, Thomas J. Koronkiewicz, Mary Anne McLeod and Tad C. Theimer

E. H. Paxton (correspondence), U.S. Geol. Survey, Southwest Biol. Sci. Center, Flagstaff, Arizona USA 86011, and Dept. of Biol. Sci., Northern Arizona Univ., Flagstaff, Arizona USA 86011. E-mail: eben_paxton@usgs.gov. – M. K. Sogge, U.S. Geol. Survey, Western Regional Office, Seattle, Washington USA 98104. – T. J. Koronkiewicz and M. A. McLeod, SWCA Environm. Consultants, Flagstaff, Arizona USA 86001. – T. C. Theimer, Dept. of Biol. Sci., Northern Arizona Univ., Flagstaff, Arizona USA 86011.

The ability to identify distinct taxonomic groups of birds (species, subspecies, geographic races) can advance ecological research efforts by determining connectivity between the non-breeding and breeding grounds for migrant species, identifying the origin of migrants, and helping to refine boundaries between subspecies or geographic races. Multiple methods are available to identify taxonomic groups (e.g., morphology, genetics), and one that has played an important role for avian taxonomists over the years is plumage coloration. With the advent of electronic devices that can quickly and accurately quantify plumage coloration, the potential of using coloration as an identifier for distinct taxonomic groups, even when differences are subtle, becomes possible. In this study, we evaluated the degree to which plumage coloration differs among the four subspecies of the willow flycatcher *Empidonax traillii*, evaluated sources of variation, and considered the utility of plumage coloration to assign subspecies membership for individuals of unknown origin. We used a colorimeter to measure plumage coloration of 374 adult willow flycatchers from 29 locations across their breeding range in 2004 and 2005. We found strong statistical differences among the mean plumage coloration values of the four subspecies; however, while individuals tended to group around their respective subspecies' mean color value, the dispersion of individuals around such means overlapped. Mean color values for each breeding site of the three western subspecies clustered together, but the eastern subspecies' color values were dispersed among the other subspecies, rather than distinctly clustered. Additionally, sites along boundaries showed evidence of intergradation and intermediate coloration patterns. We evaluated the predictive power of colorimeter measurements on flycatchers by constructing a canonical discriminant model to predict subspecies origin of migrants passing through the southwestern U.S. Considering only western subspecies, we found that individuals can be assigned with reasonable certainty. Applying the model to migrants sampled along the Colorado River in Mexico and the U.S. suggests different migration patterns for the three western subspecies. We believe that the use of plumage coloration, as measured by electronic devices, can provide a powerful tool to look at ecological questions in a wide range of avian species.

The wide range of colors and patterns displayed in the plumage of birds has historically played an important role in elucidating taxonomic groups, particularly in the identification of species, subspecies, and geographic races (Mayr 1963). With the advent of molecular methods for resolving taxonomic relationships, the use of morphological characteristics for taxonomic reconstructions waned. For avian subspecies, discordance between morphology and molecular genetic-based taxonomic groupings has led to confusion about the validity of subspecific taxonomy based on morphological characters alone (Zink 2004, Haig et al. 2006). One criticism of the use of morphological characteristics such as plumage coloration is that historically it was a qualitative trait, not easily quantified and difficult to test rigorously with statistical models. Additionally, the realization that birds see colors differently than humans, and across a larger range of wavelengths, raised questions about whether species-recognition color cues were being identified. However, with the development of electronic devices that precisely measure color (e.g., colorimeters, spectrometers), interest in the use of plumage coloration has increased, especially as a complementary trait to other sources of taxonomic information. For example, studies using electronic devices have revisited early taxonomic determinations that relied on qualitative measurements of plumage coloration (Johnson et al. 1998, Patten and Unitt 2002), combined plumage coloration with genetics to discern cryptic species and subspecies (Johnson and Jones 2001, Isler et al. 2002), and contrasted genetic patterns with plumage coloration patterns to make inferences about the demographic history of species (Johnsen et al. 2006).

The ability to quantify plumage coloration via electronic devices creates the potential for rapidly identifying the taxonomic or geographic origin of individuals of unknown

status, or for identifying boundary regions between taxonomic units and thereby helping to define taxonomic or geographic groups for conservation. For migratory birds, the ability to assign individuals to specific taxonomic subgroups is important for understanding connectivity between the non-breeding and breeding grounds (Webster et al. 2002), as well as identifying migration routes. Measurements of plumage coloration could be combined with other intrinsic markers, such as isotopes and genetics (Smith et al. 2005), to help refine the ability to assign individuals to their place of origin. Other established uses of plumage coloration include measurements of coloration for sexual selection studies (Endler 1980), identification of sex via differences in UV plumage reflectance (Eaton 2005), and identification of demographic parameters such as age and membership within particular breeding populations (Figuerola et al. 1999).

Efforts to identify taxonomic or geographic groups are particularly important for species of conservation concern. One species in which plumage coloration is used to identify subspecific groups, and for which there is a need for easy and rapid subspecies identification, is the willow flycatcher *Empidonax traillii*. The willow flycatcher is a Neotropical migrant that breeds across much of the conterminous U.S. and southern Canada, and winters from central Mexico south to northern South America. The willow flycatcher is a polytypic species (Unitt 1987), with four subspecies commonly recognized: *E. t. adastus*, ranging across the northern Rocky Mountains and Great Basin; *E. t. brewsteri*, found west of the Sierra Nevada and Cascade Mountains along the Pacific Slope; *E. t. extimus*, the southwestern willow flycatcher, which breeds across the southwestern U.S.; and *E. t. traillii*, ranging east of the northern Rocky Mountains (Fig. 1). In 1995 the southwestern willow flycatcher was declared an endangered species (USFWS 1995), raising interest in the ability to discriminate this subspecies from its conspecifics.

Morphological differences among the flycatcher subspecies, based largely on differences in plumage coloration (Unitt 1987), are subtle (Hubbard 1987, 1999) and identifying the willow flycatcher subspecies via museum specimens requires considerable skill and a complete set of voucher specimens (Hubbard 1999). Nonetheless, multiple taxonomists have evaluated and agreed on the general division and geographic distribution of the subspecies (Phillips 1948, Aldrich 1951, Hubbard 1987, Unitt 1987, Browning 1993). Our objective for this study was to evaluate whether quantitative differences in plumage coloration, as measured via an electronic device, could potentially yield a relatively rapid and inexpensive method that could be employed to identify subspecies affinity of individuals on both the non-breeding and breeding grounds and along migratory routes. Before utilizing such a tool, we first needed to assess quantitatively the variation in coloration across the range of breeding willow flycatchers, assess the degree to which the patterns are consistent with the established taxonomic relationships, and then determine the utility of quantitative plumage coloration measurements for identifying individuals and populations of unknown origin or affinity.

Materials and methods

Study sites and field methods

In 2004 and 2005, we captured and measured plumage coloration of 374 willow flycatchers from 29 breeding sites across the species' breeding range within the U.S. (Table 1,

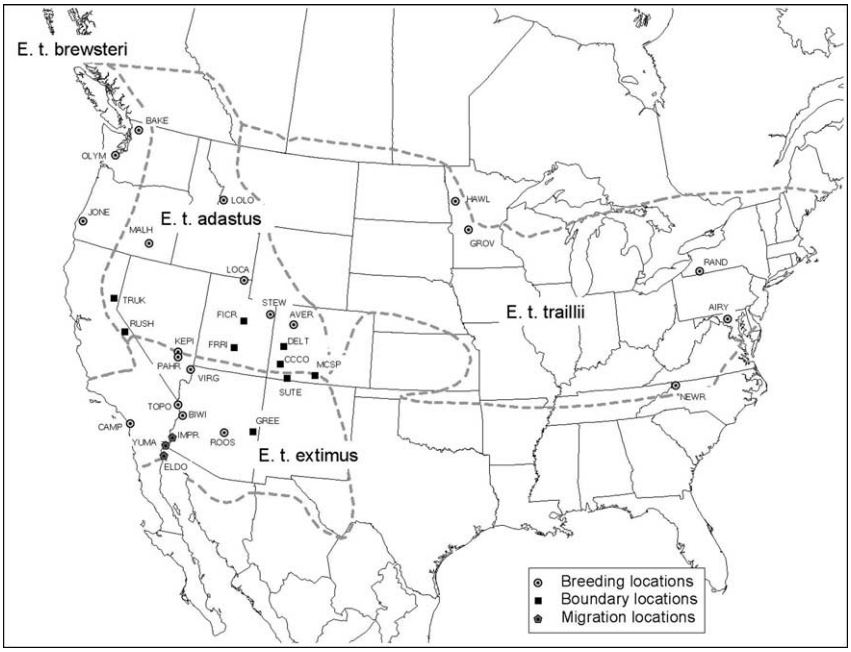

Figure 1. Locations of willow flycatcher sample sites collected across the breeding range of the four subspecies (see Table 1). Sites were designated as breeding sites, boundary sites if near a boundary known for intergradation between subspecies or of unknown affiliation, and migration stopover sites. Dotted line represents the approximate boundary of the four subspecies based on morphological studies as per Unitt (1987).

Table 1. Names, sample sizes, and locations of willow flycatcher breeding and migration sample sites. Sites were sampled during the breeding season across the range of the four willow flycatcher subspecies in 2004 and 2005. Sites near boundaries with known intergradation among subspecies and those of unknown status were designated boundary sites. In addition, three migration stopover sites along the Colorado River in Mexico and Arizona were sampled.

Site	Name	State	n	Latitude (N)	Longitude (W)	Year
E. t. adastus						
AVER	Avery Lake	CO	10	39.97	107.64	2005
LOCA	Logan Canyon	UT	7	41.92	111.56	2005
LOLO	Lolo	MT	15	45.95	114.13	2005
MALH	Malheur NWR	OR	25	42.83	118.86	2004
STEW	Stewart Lake	UT	12	40.34	109.35	2005
E. t. brewsteri						
BAKE	Baker Lake Rd	WA	20	48.56	121.82	2004
JONE	Big Creek	OR	13	43.03	123.97	2005
OLYM	Olympia Capitol Forest	WA	5	46.92	123.06	2004
E. t. extimus						
BIWI	Bill Williams River NWR	AZ	7	34.28	114.07	2005
CAMP	Camp Pendleton	CA	10	33.30	117.31	2005
KEPI	Key Pittman	NV	9	37.57	115.22	2005
PAHR	Pahranagat NWR	NV	34	37.31	115.12	2004, 2005
ROOS	Roosevelt Lake	AZ	52	33.76	111.24	2004, 2005
TOPO	Topock Marsh	AZ	11	34.81	114.52	2004, 2005
VIRG	Virgin River	NV	28	36.77	114.14	2005
E. t. traillii						
AIRY	Mount Airy	MD	5	39.28	77.27	2005
GROV	Grove Lake Township	MN	4	45.57	95.24	2005
HAWL	Hawley	MN	13	46.83	96.33	2005
NEWR	New River	NC	4	36.41	81.40	2005
RAND	Randolph	NY	13	42.15	78.98	2005
Boundary						
CCCO	Clear Creek	CO	4	37.79	108.23	2005
DELT	Escalante SWA	CO	8	38.75	108.15	2005
FICR	Fish Creek	UT	14	39.77	111.20	2005
FRRI	Fremont River	UT	3	38.30	111.51	2005
GREE	Greer	AZ	3	34.03	109.43	2005
MCSP	McIntire Springs NWR	CO	13	37.28	105.81	2005
RUSH	Rush Creek	CA	9	37.93	119.06	2005
SUTE	Southern Ute	CO	6	37.11	107.63	2004
TRUK	Truckee River	CA	17	39.47	120.38	2005
Migration						
ELDO	Colorado River Delta	MX	49	32.03	115.03	2007
IMPR	Imperial NWR	AZ	6	32.98	114.46	2004, 2005
YUMA	Yuma	AZ	90	32.57	114.79	2005–2007

Fig. 1). All flycatchers used in this study were adults, captured in mist nets either passively or via target netting (Sogge et al. 2001), and all were banded to ensure no duplication of individuals. Sample sizes ranged from 3 to 52 (median = 10) individuals per site, with samples collected from 11 May to 6 August across two years. In addition, 145 spring migrants were captured and measured at three locations where breeding has not been documented along the Colorado River in Baja del Norte, Mexico, and southwestern Arizona, U.S., from 5 May to 20 June (Table 1, Fig. 1). Our general sampling strategy was to collect a relatively large sample (10–15 individuals) from each of 2 to 3 breeding sites within the core range of each subspecies, and to sample individuals from as many additional sites as possible to capture geographic variation across the core range of each subspecies. However, many breeding sites were composed of a small number of breeders and sample sizes in some cases were constrained by the number of birds available, so breeding sites with small (≤ 4) samples were grouped with other sites if geographically adjacent to one another. Because *E. t. adastus* and *E. t. brewsteri* migrate through the breeding range of *E. t. extimus*, flycatchers sampled in the Southwest were recorded as resident only if sampled during the typical non-migratory period (15 June–20 July; Unitt 1987) at known breeding locations or if individuals were determined to be residents by other means, such as evidence of breeding. Sampling dates for the other three subspecies tended to be in the core breeding period, with *E. t. adastus* sampled from 8 June to 30 July, *E. t. brewsteri* from 14 June to 19 July, and *E. t. traillii* from 14 May to 5 Aug. Finally, long-term monitoring studies conducted at several *E. t. extimus* breeding sites allowed us to recapture and resample 10 individuals in different years to directly compare changes in plumage coloration from one year to the next.

We measured plumage coloration of flycatchers with a Konica Minolta Chroma Meter CR-400 colorimeter. This instrument is highly portable and easy to use by a single person, and calculates a color value designed to mimic the human eye, a color space that was the foundation of the original subspecies taxonomy per Unitt (1987) and Browning (1993). The colorimeter measures differences in chromaticity and lightness, which are represented in CIELAB (Commission Internationale de L'Eclairages)

3-dimensional color space and denoted as values L*, a*, and b*. A color has three components: lightness, saturation, and hue. The value L* denotes how light or dark the color is (lightness), while a* and b* together indicate color directions (saturation and hue) in two-dimensional space. An increase in a* indicates more red, while a decrease indicates a movement to green; an increase in b* is an increase in yellow, while a decrease in b* indicates a shift toward blue color space. Colorimeters do not measure UV reflectance.

With an individual flycatcher held securely in the hand, the colorimeter was placed firmly against the bird on the crown and a measurement of the plumage coloration recorded. After each measurement, the colorimeter was lifted away from the bird and then placed back onto the same location for a total of eight separate measurements. This process was repeated to gather eight similar readings from the back (the interscapular region between the wings) of each individual. Crown and back measurements were chosen as the most informative locations to measure plumage coloration based on previous research (Unitt 1987, Browning 1993). A random sample of 25 individuals was selected to have a full set of measurements taken by two different observers within the same capture/handling period for the purpose of measuring observer variation. For samples with duplicates (observer or year), only the chronologically first reading was used for subsequent analysis.

Data analysis

Each individual flycatcher had 16 measurements taken: eight replicate measurements of the crown, and eight replicate measurements of the back, with each measurement producing three numerical values describing the 3-dimensional color space (L*, a*, b*). Several steps were taken to produce the final data set for analysis. We first removed any obvious single misreadings, which were defined as color values ≥ 4 standard deviations from the mean of all individual readings. Misreadings were rare ($<1\%$) and did not result in the exclusion of any individuals from the analysis. We then chose the four (out of eight) color measurements closest to the mean value of a particular individual to minimize the effects of minor misreadings in some of the repeated measurements; in retrospect, the high repeatability of measurements suggests only 4–6 measurements, simply averaged, are needed. The four readings for each body location (crown and back) were averaged for L*, a*, and b*, with those averages used in the subsequent analysis. Thus, for each flycatcher measured we had six variables: the mean of the three color values (L*, a*, b*) for both the crown and back.

We tested for observer measurement error, effects of feather color fade, and year to year variation to evaluate the extent that these sources of variation might confound geographic-based variation. To evaluate observer error, we used a paired t-test for each of the individuals that were measured twice in the same capture session, once each by two different observers. Willow flycatchers molt on their wintering grounds (Unitt 1987, Pyle 1997, USGS, unpubl. data), and after feathers are completely grown they are generally inert (Montgomerie 2006). However, feather coloration can gradually change due to mechanical wear, ultraviolet radiation fading, and possibly through biological degradation (Delhey et al. 2006). To evaluate whether feather coloration changes over time, we regressed each of the six color variables against the day of year the individual was measured for the two subspecies in which we obtained samples from across the breeding season (*E. t. adastus* and *E. t. extimus*). Regression slopes for the two subspecies were nearly identical, so we combined the two subspecies in the regression analysis to obtain estimates of average seasonal change and adjusted the color values for all four subspecies to account for seasonal fade by using the regression coefficient and date of capture. In addition to gradual change due to fading, environmental conditions such as diet and humidity, which can change from year to year, may influence the coloration at the time of feather development during molt on the non-breeding grounds. We tested for this by comparing color readings from individuals captured and measured in two different years using a paired t-test, and used a two-mean t-test to evaluate differences in the mean values of individuals from two breeding sites within the range of *E. t. extimus* in which we had a large sample from two consecutive years.

Because this study was intended to distinguish among established taxonomic units, we grouped breeding sites *a priori* into one of the four subspecies based on geographic ranges delineated via morphological studies (Unitt 1987, Browning 1993). Some sites were designated "boundary sites" either because they occur in areas of known intergradation between two subspecies (Paxton et al. 2008), or occur in an area where the approximate location of a boundary is uncertain. Boundary sites were not used in the evaluation of differences among subspecies, but were considered in subsequent analyses. After verifying that color values were normally distributed, we used a MANOVA to test for differences among subspecies considering all six variables simultaneously. The multivariate test was followed with one-way ANOVAs and α-corrected pairwise Tukey-Kramer tests for each of the six color values.

We used a Canonical Discriminant Analysis to represent the linear relationship of the plumage coloration by subspecies (grouping factor) in multivariate space using all six color values. A Discriminant Analysis is similar to a Principle Components Analysis, but with the principle axes rotated to maximize the ability to distinguish among groups (subspecies). We built the model using only large sample sites within the core area of each subspecies (2–3 sites per subspecies); the model was then applied to all individuals from all sites. The model provides Canonical axis values of each individual, and the mean value for each subspecies or site can be used to represent the relative distance of sites or subspecies from one another. The closer coordinates are to one another in ordination space, the closer their color values are. We visually portrayed the relative plumage coloration differences among individuals, sites, and subspecies by plotting their respective coordinates in ordination space using the first two canonical axes.

The discriminant function also provides likelihood-based probability estimates of an individual belonging to each of the groups (subspecies) considered, with an individual typically assigned to that group with the highest prediction probability. The level of confidence in a

prediction can be increased by setting a threshold of confidence. For example, only considering individuals predicted to belong to a particular subspecies with $\geq 60\%$ or $\geq 80\%$ probability thresholds provides higher accuracy, though this approach leads to the exclusion of those individuals that are not predicted to belong to any group with high probability. We constructed the canonical discriminant model using seven sites with large samples from the western three subspecies (2–3 sites per subspecies, $n = 187$), and tested the predictive ability of the model with an additional sample of 71 individuals from smaller sites in the west that were not included in the model building. We only considered the three western subspecies for this question because the eastern subspecies has intermediate coloration values that confound discrimination from the other subspecies (see Results) and the eastern subspecies is not known or likely to migrate through the southwestern U.S. west of the Rio Grande, New Mexico (Hubbard 1987, Unitt 1987). We then applied the model to spring migrants sampled along the Colorado River in southern Arizona and northern Mexico. Although the model predicts subspecies status for individuals, we adopted a conservative approach of describing results as proportion of each of the three subspecies, rather than actual numbers of individuals from each subspecies.

We used JMP v. 6.0 (SAS, Inc.) for all statistical analyses. All graphs were made in SigmaPlot (SPSS, Inc.), and the sample location map was constructed in ArcView 3.0 (ESRI, Inc.). Statistical significance was accepted at $P < 0.05$.

Results

Sources of variation

To test for biases in color values as a function of the observer, 25 flycatchers were measured by two different people during the same capture period. Paired t-tests of the difference between observers indicated no difference in color values (smallest P-value was 0.37) and a high correlation (mean correlation = 0.85).

However, we did detect a significant effect of date on color values, with five out of the six color values showing significant change when regressed against day of year (L*_crown: $\beta = -0.0264$, $R^2 = 0.07$, $P = 0.01$; a*_crown: $\beta = 0.0097$, $R^2 = 0.14$, $P < 0.001$; b*_crown: $\beta = -0.0137$, $R^2 = 0.11$, $P = 0.002$; L*_back: $\beta = -0.0089$, $R^2 = 0.01$, $P = 0.46$; a*_back: $\beta = 0.0109$, $R^2 = 0.14$, $P < 0.001$; b*_back: $\beta = -0.0414$, $R^2 = 0.25$, $P < 0.001$). Because different sites and subspecies were sampled at different times over the breeding season, we corrected for the gradual change in color by multiplying the regression coefficient (β) by the date (day of year) starting from the earliest capture date (May 9th, day of year = 129). These seasonally adjusted color values were used for all subsequent analyses.

The analysis of 10 individuals that were captured and measured in two separate years suggested that individual color values did not significantly change from year to year (smallest P-value = 0.32), but sample size was small and the correlation among years was not strong (average correlation = 0.31). Comparisons of the mean values from two consecutive years at Virgin River, NV and Roosevelt Lake, AZ suggested no significant differences for any of the color values at the Virgin River, and only one color value showed significant differences between years at Roosevelt Lake (L*_crown, $t = -2.05$, $df = 50$, $P = 0.046$). Thus, while there is some indication of changes in color values from year to year, the overall weight of evidence suggests such changes are small.

Differences among subspecies

Our analysis using a MANOVA indicated highly significant differences in the mean color values of the subspecies (Wilks' Lambda = 0.358, approx. $F_{18,886} = 21.56$, $P < 0.001$). We explored the sources of the multivariate differences among subspecies by running one-way ANOVAs and Tukey-Kramer pairwise comparisons for each of the six color values (Table 2, Fig. 2). All six color values showed strong statistical differences ($P < 0.001$) among the subspecies, although the differences among the subspecies varied with different color values (Table 2). However, a canonical discriminant function model, representing the multidimensional color values with subspecies as the grouping factor, indicated that while individuals tended to cluster near their group mean, dispersion around each subspecies' mean color space overlapped (Fig. 3).

While there is overlap in the color values of individual flycatchers from the different subspecies, mean values for the breeding sites tended to cluster together by subspecies (Fig. 4). In general, the breeding sites from each of the three western subspecies clustered closest to one another, while the eastern subspecies was dispersed among the three western subspecies color values. The mean coloration of the eastern subspecies, which fell midway between the mean values of the three western subspecies (Fig. 3), is misleading as the actual distribution of the breeding sites are not clustered around that location in ordination space (Fig. 4). Similarly, the mean values of boundary sites, when plotted in ordination space, show intermediate locations between the three western subspecies (Fig. 5). This could be due to mixing of genetic traits or the fact that they are geographically mid-way between different regions. To evaluate whether plumage coloration changed as a function of geographic distance, we regressed canonical axis 1 against latitude and longitude for the two subspecies that had samples widely distributed across their range. We found no relationship between the color variables across the range of *E. t. adastus* (latitude, range = 8.2°, $R^2 = 0.098$, $P = 0.32$; longitude, range = 11.3°, $R^2 = 0.01$, $P = 0.77$), or *E. t. extimus* (latitude, range = 6.0°, $R^2 = 0.030$, $P = 0.61$; longitude, range = 8.2°, $R^2 = 0.07$, $P = 0.44$). However, across the shared boundary of these two subspecies, including boundary sites, there was a strong relationship in the canonical axis 1 value as a function of latitude across this boundary (range = 8.2°, $R^2 = 0.59$, $P < 0.001$).

Assigning migrating individuals to a particular subspecies

To evaluate which subspecies were migrating through the southwestern U.S. along the Colorado River, we used the

Table 2. Mean plumage coloration values for willow flycatchers and their associated error, 95% CI and statistical differences.

Color value	Subspecies	n	Mean	SE	95% CI	F-ratio (df = 3,318)	P-value	Pairwise differences*
Crown								
L*	*E. t. adastus*	82	25.95	0.18	25.59–26.30	32.18	<0.001	A
	E. t. brewsteri	42	23.90	0.26	23.37–24.44			B
	E. t. extimus	156	26.74	0.14	26.46–27.03			C
	E. t. traillii	42	25.45	0.247	24.97–25.93			A
A*	*E. t. adastus*	82	1.34	0.05	1.243–1.434	15.33	<0.001	A
	E. t. brewsteri	42	1.71	0.08	1.553–1.873			B
	E. t. extimus	156	1.76	0.04	1.671–1.842			B
	E. t. traillii	42	1.42	0.06	1.297–1.546			A
B*	*E. t. adastus*	82	9.91	0.08	9.75–10.06	57.50	<0.001	A
	E. t. brewsteri	42	10.57	0.13	10.31–10.83			B
	E. t. extimus	156	11.31	0.06	11.18–11.44			C
	E. t. traillii	42	10.84	0.14	10.56–11.13			B
Back								
L*	*E. t. adastus*	82	31.84	0.20	31.45–32.23	29.30	<0.001	A
	E. t. brewsteri	42	30.06	0.24	29.58–30.54			B
	E. t. extimus	156	32.74	0.13	32.49–32.98			C
	E. t. traillii	42	31.96	0.30	31.36–32.56			A
A*	*E. t. adastus*	82	0.90	0.05	0.79–1.00	17.93	<0.001	A
	E. t. brewsteri	42	1.19	0.06	1.06–1.32			B
	E. t. extimus	156	1.24	0.04	1.17–1.31			B
	E. t. traillii	42	0.80	0.05	0.69–0.91			A
B*	*E. t. adastus*	82	13.02	0.15	12.73–13.32	9.09	<0.001	A
	E. t. brewsteri	42	13.21	0.22	12.77–13.65			A,B
	E. t. extimus	156	13.95	0.12	13.71–14.18			C
	E. t. traillii	42	13.92	0.21	13.49–14.35			B,C

* Tukey-Kramer α-corrected significant difference among pairwise comparisons of means.

canonical discriminant model to predict likely membership within a particular subspecies. The model had an overall accuracy rate of 80%, based on individuals used to construct the model (n = 187). To test whether a more stringent criterion might increase the accuracy of the model, we re-evaluated the model's accuracy by excluding individuals that were predicted to be one or another subspecies at < 0.6 and < 0.8. Dropping individuals at the 0.6 threshold resulted in an increase in accuracy to 91%, with 46 (25%) individuals excluded. At the 0.8 level, accuracy increased to 95%, but with 85 (45%) of the total individuals excluded. We broadened our evaluation of this model by including eight additional sites (n = 71) that were not used to build the model. Results were similar, though misclassification rates increased, with 77% accuracy considering all individuals, 85% excluding those below a threshold of 0.6 (22% individuals excluded), and 92% at a threshold of 0.8 (47% excluded). The model tended to over-predict *E. t. brewsteri*, with *E. t. adastus* and *E. t. extimus* more likely to be misclassified as *E. t. brewsteri* than one another. This may be an artifact of feather fade, where individuals with greater fade than we accounted for with the day of year correction (above) would move toward the coloration of *E. t. brewsteri*. While the model does reasonably well at higher thresholds, the cost of the higher accuracy is the exclusion of many individuals. This could potentially bias the estimates by excluding certain subspecies. To evaluate this, we tested for differences in the proportion of actual individuals per subspecies versus the proportion of those predicted. Results indicate no inherent bias, particularly at higher thresholds (χ^2 tests (df = 2): no threshold, n = 558, $\chi^2 = 5.72$, P = 0.06; 60% threshold, n = 436, $\chi^2 = 4.51$, P = 0.11; 80% threshold, n = 294, $\chi^2 = 0.85$, P = 0.65).

We applied the three subspecies discriminant model to 145 individuals captured and measured during migration at three stopover sites along the Colorado River corridor in southern Arizona and northern Mexico. The results suggest that all three western subspecies migrate through the river corridor, but in differing proportions (Table 3). At the 80% threshold, 52% of the migrants sampled were identified as *E. t. adastus*, 47% as *E. t. extimus*, and 1% as *E. t. brewsteri*. However, the proportions among sites were not uniform, with the Arizona sites (Yuma and Imperial) having most of the predicted *E. t. extimus* individuals, while *E. t. adastus* was more predominant at the Mexico site (Table 3).

Discussion

Plumage coloration as measured with a colorimeter was significantly different among the willow flycatcher subspecies, agreeing with taxonomic studies that largely relied on qualitative comparison of museum specimen plumage coloration (Unitt 1987, Browning 1993). Generally, *E. t. brewsteri* showed the darkest plumage coloration, while *E. t. extimus* had the lightest, and the other two subspecies were intermediate. *E. t. brewsteri* and *E. t. extimus* were less green than *E. t. adastus* and *E. t. traillii*, while *E. t. extimus* and *E. t. traillii* tended toward a more yellowish coloration.

In agreement with the overall mean color values of the subspecies, breeding sites from each of the three western subspecies grouped closer to one another than other subspecies breeding sites. However, breeding sites of the eastern subspecies did not show cohesion but rather were interspersed in canonical space between the breeding sites

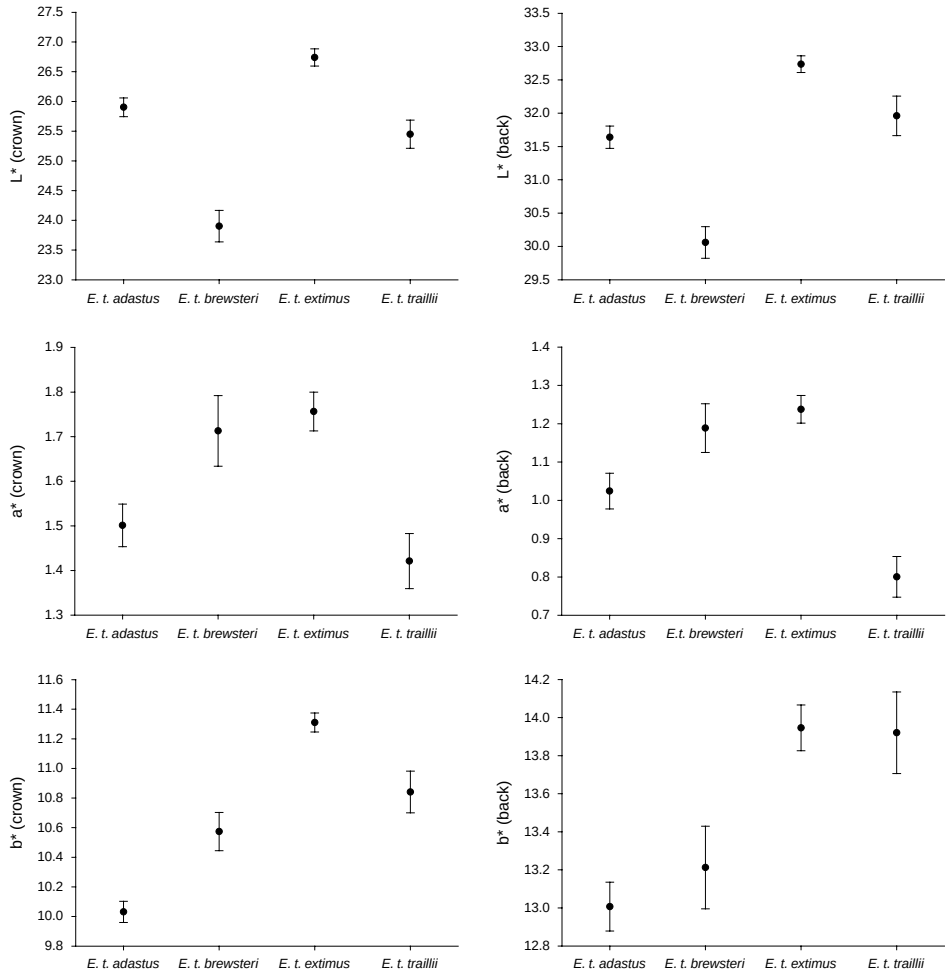

Figure 2. Mean measurement values (+/−1 SE) for the three color values obtained from the crown and back of the four willow flycatcher subspecies. Color values are represented in CIELAB 3-dimensional color space (L*, a*, and b*), with the value L* representing lightness and a* and b* together indicate color directions (saturation and hue). An increase in a* indicates a shift toward red, while a decrease indicates a shift toward green; an increase in b* represents an increase in yellow, while a decrease in b* indicates a shift toward blue. Strong statistical differences among subspecies were observed for all six color values (see Table 2).

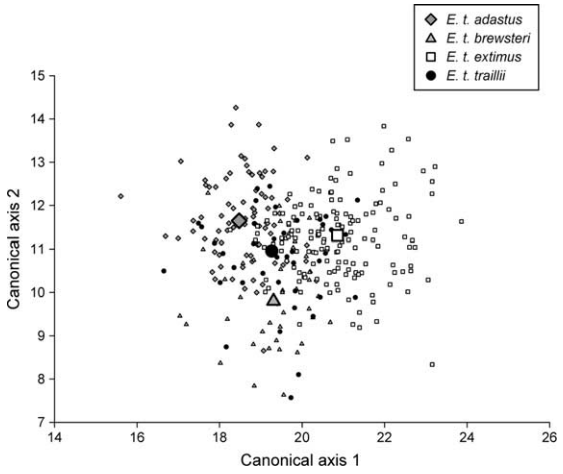

Figure 3. Relationship of the mean values of the subspecies (large symbols) and individuals assigned to each subspecies (small symbols). Individuals generally clustered near the mean value of their associated subspecies in ordination space (indicating similar coloration), but there was overlap among the subspecies.

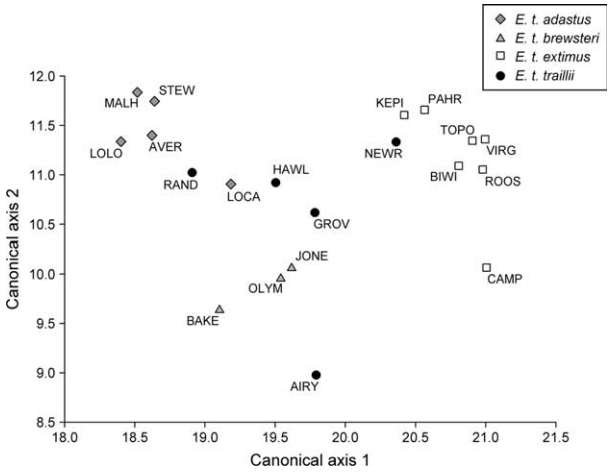

Figure 4. Mean values of breeding sites for each of the four subspecies. Breeding sites of each western subspecies generally clustered together, but the breeding sites of the eastern subspecies (*E. t. traillii*) were generally scattered among the three western subspecies clusters.

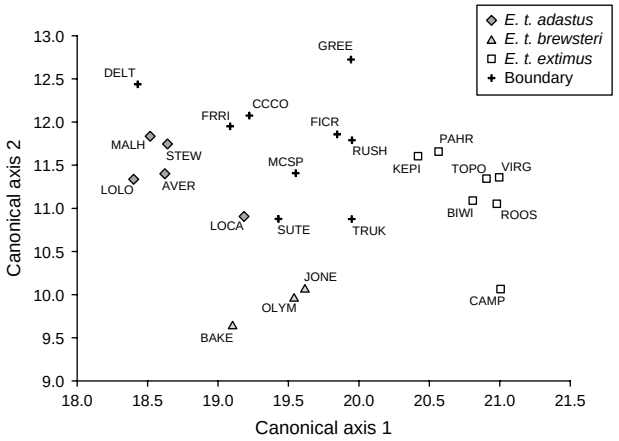

Figure 5. Location of breeding sites for the three western subspecies and boundary locations. In general, the boundary sites showed intermediate values between the core breeding sites of each subspecies.

of the three western subspecies. A possible explanation for the dispersed coloration values of the eastern subspecies may arise from its rapid expansion across large portions of the eastern U.S. as it exploits new habitat arising from anthropogenic changes in the landscape (Sedgwick 2000). This rapid geographic expansion could result in novel shifts in gene frequencies through founder effects, resulting in a broader range of coloration values than found in the other subspecies (Avise 1994). Additionally, the eastern subspecies occupies a large geographic area in which it is not uniformly distributed (Sedgwick 2000), possibly restricting gene flow among *E. t. traillii* breeding sites, and thus allowing for divergence among sites due to genetic drift or differing selection. This variability in *E. t. traillii* color values may be one reason that some investigators of willow flycatcher taxonomy believed that the eastern subspecies was actually two subspecies - *E. t. traillii* and *E. t. campestris* (Aldrich 1951, Browning 1993).

Boundary sites also showed intermediate color values between breeding sites within the core areas of the three western subspecies. These intermediate values presumably represent the result of intergradation among the subspecies, consistent with a molecular genetic study which found a broad zone of intergradation along the boundary between *E. t. extimus* and *E. t. adastus* (Paxton et al. 2008). The results of intermediate plumage coloration are consistent with coloration being a polygenic trait, in which hybridization would produce intermediate plumage coloration (Rohwer and Wood 1998). However, the relationship of the sites to one another based on differences in mean coloration values do not necessarily correspond to their distance from one another in geographic space, suggesting a complex pattern of transmission of genes and possible gene flow pathways. Likewise, the distance in ordination space of a boundary site from core breeding sites may not reflect their true genetic relationship to a particular subspecies; rather, sites with intergradation may have plumage coloration shifted in ways not easily translated into degrees of genetic relationship.

Divergence in plumage coloration among the subspecies could be due to selection, random genetic drift, or environmental interactions. Sexual selection, or adaptations to different environments, could quickly drive divergence between subspecies with limited gene flow (Avise 1994). For example, willow flycatchers follow Gloger's rule, with the darkest subspecies, *E. t. brewsteri*, breeding in the wettest region (the Pacific Northwest), and the lightest subspecies, *E. t. extimus*, breeding in the arid Southwest. Darker plumage may benefit populations in mesic regions by allowing for better absorption of radiant heat to promote the drying of feathers (Gill 1973). Similarly, plumage coloration pigments may have properties that resist degradation from biotics, such as bacteria, and therefore regionally different communities of bacteria could drive regional differences in coloration (Burtt and Ichida 2004). However, it is important to remember that flycatchers as long-distance migrants must balance selective pressures among their breeding, wintering, and migratory environments. Therefore, it may be difficult to tightly link an environmental quality at one particular location, such as rainfall on the breeding grounds, with the overall plumage coloration.

Alternatively, the differences among the plumage coloration of the subspecies are subtle and they may have arisen through random drift over thousands of years. If coloration is a polygenic trait, which is suggested by the intermediate coloration values at intergradation regions (Rohwer and Wood 1998), then small shifts in gene frequencies over time could lead to divergence in coloration. In addition, there is a substantial amount of variation among individuals of a given subspecies. Although the coloration of individuals tended to be closest to the overall mean of the subspecies of origin, coloration of individuals from different subspecies overlapped. If plumage coloration is weakly selected for, or

Table 3. Proportion of willow flycatcher western subspecies predicted at migration sites along the Colorado River using a Canonical Discriminant Function. The discriminant model produces likelihood-based probabilities of each individual originating from each of the three western subspecies. Using a threshold of confidence, only those predicted to belong to a particular subspecies at a 0.6 and 0.8 threshold increases the accuracy of the assignment, but decreases the number of individuals that are considered (see Methods).

Site	Subspecies	n	All count (%)	n	>60% count (%)	n	>80% count (%)
Colorado River Delta	*E. t. adastus*	49	36 (74%)	38	30 (79%)	25	24 (96%)
	E. t. brewsteri		8 (16%)		5 (13%)		1 (4%)
	E. t. extimus		5 (10%)		3 (8%)		0
Yuma and Imperial	*E. t. adastus*	96	38 (40%)	79	27 (34%)	58	19 (33%)
	E. t. brewsteri		4 (4%)		2 (3%)		0
	E. t. extimus		54 (56%)		50 (63%)		39 (67%)
Total	*E. t. adastus*	145	74 (51%)	117	57 (49%)	83	43 (52%)
	E. t. brewsteri		12 (8%)		7 (6%)		1 (1%)
	E. t. extimus		41 (41%)		53 (45%)		39 (47%)

differences among subspecies are due primarily to random genetic drift, then there may be little selection for a narrower distribution of color values. Additionally, there may be complex genetic and environmental interactions that result in individual variation. While the pigments or structural characteristics contributing to willow flycatcher plumage coloration are unknown, green feathers are thought to be a combination of carotenoid and melanin-based pigments. Carotenoid pigments, which would contribute the yellow coloration in flycatcher's feathers, can be strongly influenced by an individual's diet and condition (McGraw 2006), although all sources of plumage coloration are potentially subject to environmental influences (Paxton 2009). However, for the flycatchers in this study, yearly variation in plumage coloration appears to have contributed little to overall variation in plumage coloration.

In our study, feather wear and fading did contribute to overall plumage variation, as has been widely demonstrated in other bird species (Figuerola and Senar 2005, Delhey et al. 2006). For the willow flycatcher, wear and fade appear to be gradual and linear, affecting coloration (a* and b*) more than lightness (L*). Color values from both the crown and back indicate gradual increase in a* and decrease in b*, indicating a color moving away from green toward a more grayish color. This suggests the carotenoid-based yellow may be more susceptible to fading than the melanin-based pigments, consistent with the findings in great tits *Parus major* (Figuerola and Senar 2005) and green jays *Cyancorax yncas* (Johnson and Jones 1993). However, the effects are gradual and can be corrected for, at least for the period of time in which we can measure the degree of coloration change (i.e., the breeding season). Measurement error did not play a role in overall variation, suggesting a high accuracy of measurement once someone was trained to measure the flycatchers.

A key goal of this study was to exploit differences among the plumage coloration of subspecies to assign individuals of unknown status to a particular subspecies of origin, specifically migrants and wintering birds. Our results indicate that measurements of plumage coloration in flycatchers are a powerful tool for linking individuals to taxonomic groups. However, it is important to document sources of variation, and understand how they can place limits on the ability to address some questions. For example, using plumage coloration measurements on wintering willow flycatchers would require estimating the effects of wear and fade beyond the time period that we directly measured, either forward until the mid-winter molt, or backwards in time to a freshly molted plumage. Further, the inclusion of the eastern subspecies would confound results given their dispersed color values, at least in areas where they may co-winter with the western subspecies. As this technique is applied to other avian species, it will be important to fully measure sources of variation and understand what, if any, limitations they may place on the questions that can be addressed. However, combining other sources of information, such as stable isotopes and genetic markers, with plumage coloration may help overcome limitations for both flycatchers and other species.

For the flycatcher, plumage coloration provides crucial insight into migration strategies of the three western subspecies. Applying the discriminant model to spring migrants along the Colorado River in northern Mexico and southern U.S. suggests all three western subspecies use the river corridor, but in different proportions at different locations. Near the Colorado River Delta in Mexico, almost all (96%) of the migrants were estimated to be *E. t. adastus*, with some contribution from *E. t. brewsteri* and possibly *E. t. extimus*. Further north, near Yuma, AZ, approximately one third of the migrants were estimated to be *E. t. adastus*, but two thirds were assigned to *E. t. extimus*. Again, *E. t. brewsteri* was estimated to be absent or occurring in low numbers.

Documenting the differential occurrence of subspecies at stopover sites can help elucidate migration patterns and strategies, as different populations may have different migratory pathways (Paxton et al. 2007). Based on our results, it appears that the Colorado River is an important migration corridor for *E. t. adastus*, which was estimated to be the primary migrant at the Mexico stopover site and approximately one third of the migrants at the Arizona stopover sites. Contrary, *E. t. brewsteri* was estimated to contribute only a small proportion of the migrants sampled, perhaps indicating that a majority of *E. t. brewsteri* migrate west of the Colorado River to the Pacific coast before migrating north. Of particular interest was the estimate of *E. t. extimus* contributing a large proportion (67%) of migrants at the Arizona sites, which was unexpected because the endangered subspecies constitutes a small fraction ($<1\%$) of the entire species' total population size (Rich et al. 2004). Additionally, few to no *E. t. extimus* were detected in Mexico. This suggests that the Arizona stopover sites sampled are important for the endangered subspecies, at least during the period sampled. Sampling at the Arizona sites was conducted at the end of the migration period (7 June to 20 June), whereas the Mexico site was sampled throughout the migration period (5 May to 14 June). Thus, we can only estimate the relative contribution of migrants within the time frame they were sampled, and inference is limited to the two areas sampled. Clearly, additional effort is needed to sample multiple sites throughout the migration period to begin to understand which migration areas are important for the endangered subspecies. Given the evidence that migration can be a limiting period for populations (Sillett and Holmes 2002, Newton 2004), a better understanding of migration routes through the southwestern U.S. would benefit overall management of the endangered subspecies, and techniques developed in this study provide the tools to do so.

This study highlights that the use of a colorimeter can be an important tool for addressing evolutionary and ecological questions. We believe that the use of this technique holds great promise for a wide range of bird species, and can be particularly valuable for studies that seek to understand migratory connectivity among different regions. Museum specimens could be incorporated in such studies, for example to document coloration patterns in areas where a particular species is now extirpated, but care is needed as older museum specimens would need corrections for changes in coloration (McNett and Marchetti 2005). Other devices to measure coloration, such as a spectrometer, can provide additional information such as UV reflectance (Eaton 2005), and may answer additional questions. Additionally, for assigning individuals of

unknown origin, combinations of colorimeter measurements with other intrinsic markers, such as molecular genetic markers or stable isotopes, should enhance assignment accuracies (Smith et al. 2005).

Acknowledgements – This work was funded by the U.S. Geological Survey and the U.S. Bureau of Reclamation. Caroline Causey, Phil Unitt, and Mary Whitfield made important contributions to this research project. The hard work of many members of the SWCA and USGS banding crews made this study possible. Particular thanks to Amanda Bakian, Osvel Hinojosa-Huerta, Matt Johnson, Chris McCreedy, Gary Slater, and Scott Stoleson for assistance in sampling specific locations. This manuscript was greatly improved through comments made by Kristina Paxton and two anonymous reviewers. The use of product trade names in this paper does not constitute product endorsement by the USGS.

References

Aldrich, J. W. 1951. A review of the races of the Traill's flycatcher. – Wilson Bull. 63: 192–197.

Avise, J. C. 1994. Molecular markers, natural history and evolution. – Chapman and Hall, New York.

Browning, M. R. 1993. Comments on the taxonomy of *Empidonax traillii* (willow flycatcher). – Western Birds 24: 241–257.

Burtt, E. H. and Ichida, J. M. 2004. Gloger's rule, feather-degrading bacteria, and color variation among song sparrows. – Condor 106: 681–686.

Delhey, K., Peters, A., Johnsen, A. and Kempenaers, B. 2006. Seasonal changes in blue tit crown color: do they signal individual quality? – Behav. Ecol. 17: 790–798.

Eaton, M. D. 2005. Human vision fails to distinguish widespread sexual dichromatism among sexual ''monochromatic'' birds. – Proc. Nat. Acad. Sci. USA 102: 10942–10946.

Endler, J. A. 1980. Natural and sexual selection on color patterns in *Poecilia reticulate*. – Evolution 34: 76–91.

Figuerola, J. and Senar, J. C. 2005. Seasonal changes in carotenoid and melanin-based plumage coloration in the great tit *Parus major*. – Ibis 147: 797–802.

Figuerola, J., Senar, J. C. and Pascual, J. 1999. The use of a colorimeter in field studies of blue tit *Parus caeruleus* coloration. – Ardea 87: 269–275.

Gill, F. B. 1973. Intra-island variation in the Mascarene white-eye. – Ornithol. Monogr., vol. 12, Amer. Ornithol. Union.

Haig, S. M., Beever, E. A., Chambers, S. M., Draheim, H. M., Dugger, B. D., Dunham, S., Elliot-Smith, E., Fontaine, J. B., Kesler, D. C., Knaus, B. J., Lopes, I. F., Locshl, P., Mullins, T. D. and Sheffield, L. M. 2006. Taxonomic considerations in listing subspecies under the U.S. endangered species act. – Conserv. Biol. 20: 1584–1594.

Hubbard, J. P. 1987. The status of the willow flycatcher in New Mexico. Report to the New Mexico Dept. of game and fish. 29 pages.

Hubbard, J. P. 1999. A critique of Wang Yong and Finch's field-identifications of willow flycatcher subspecies in New Mexico. – Wilson Bull. 11: 585–588.

Isler, M. L., Alonso, J. A., Isler, P. R., Valqui, T., Begazo, A. and Whitney, B. M. 2002. Rediscovery of a cryptic species and description of a new subspecies in the *Myrmeciza hemimelaena* complex (Thamnophilidae) of the Neotropics. – Auk 119: 362–378.

Johnsen, A., Andersson, S., Fernandez, J. G., Kempenaers, B., Pavel, V., Questiau, S., Raess, M., Rindal, E. and Lifjeld, J. T. 2006. Molecular and phenotypic divergence in the bluethroat (*Luscinia svecica*) subspecies complex. – Mol. Ecol. 15: 4033–4047.

Johnson, N. K. and Jones, R. E. 1993. The green jay turns blue in Peru-interrelated aspects of the annual cycle in the arid tropical zone. – Wilson Bull. 105: 389–398.

Johnson, N. K. and Jones, R. E. 2001. A new species of tody-tyrant (Tyrannidae: *Poecilotriccus*) from northern Peru. – Auk 118: 334–341.

Johnson, N. K., Remsen, J. V. and Cicero, C. 1998. Refined colorimetery validates endangered species of the least tern. – Condor 100: 18–26.

Mayr, E. 1963. Animal Species and Evolution. – Harvard Univ. Press, Cambridge.

McGraw, K. J. 2006. Mechanics of carotenoid-based coloration. – In: Hill, G. E. and McGraw, K. J. (eds). Bird coloration: mechanisms and measurements. Vol. 1. Harvard Univ. Press, Cambridge.

McNett, G. D. and Marchetti, K. 2005. Ultraviolet degradation in carotenoid patches: live versus museum specimens of wood warblers (Parulidae). – Auk 122: 793–802.

Montgomerie, R. 2006. Cosmetic and adventitious colors. – In: Hill, G. E. and McGraw, K. J. (eds). Bird coloration: mechanisms and measurements. Vol. 1. Harvard Univ. Press, Cambridge.

Newton, I. 2004. Population limitations in migrants. – Ibis 146: 197–226.

Patten, M. A. and Unitt, P. 2002. Diagnosability versus mean difference of sage sparrow subspecies. – Auk 119: 26–35.

Paxton, E. H. 2009. The utility of plumage coloration for taxonomic and ecological studies. – The Open J. Ornithol. 2: 17–23.

Paxton, E. H., Sogge, M. K., Theimer, T. C., Girard, J. and Keim, P. 2008. Using molecular genetic markers to resolve a subspecies boundary: the northern boundary of the southwestern willow flycatcher in the four-corner states. U.S. – Geol. Surv. Open File Rep. 2008-1117.

Paxton, K. L., van Riper, C., Theimer, T. C. and Paxton, E. H. 2007. Spatial and temporal migration patterns of Wilson's warbler in the southwestern United States revealed by stable isotopes. – Auk 124: 162–175.

Phillips, A. R. 1948. Geographic variation in *Empidonax traillii*. – Auk 65: 507–514.

Pyle, P. 1997. Identification guide to North American birds. Part 1. – Slate Creek Press, Bolinas.

Rich, T. D., Beardmore, C. J., Berlanga, H., Blancher, P. J., Bradstreet, M. S. W., Butcher, G. S., Demarest, D. W., Dunn, E. H., Hunter, W. C., Iñigo-Elias, E. E., Kennedy, J. A., Martell, A. M., Panjabi, A. O., Pashley, D. N., Rosenberg, K. V., Rustay, C. M., Wendt, J. S. and Will, T. C. 2004. Partners in Flight North American landbird conservation plan. Cornell Lab of Ornithol. Ithaca. Partners in Flight website. http://www.partnersinflight.org/cont_plan/ (VERSION: March 2005).

Rohwer, S. and Wood, C. 1998. Three hybrid zones between hermit and Townsend's warblers in Washington and Oregon. – Auk 115: 284–310.

Sedgwick, J. A. 2000. Willow flycatcher (*Empidonax traillii*). – In: Poole, A. and Gill, F. (eds). The birds of NorthAmerica, no. 533. The birds of North America, Inc., Philadelphia.

Sillett, T. S. and Holmes, R. T. 2002. Variation in survivorship of a migratory songbird throughout its annual cycle. – J. Anim. Ecol. 71: 296–308.

Smith, T. S., Clegg, S. M., Kimura, M., Ruegg, K. C., Mila, B. and Lovette, I. J. 2005. Molecular genetic approaches to linking breeding and overwintering areas in five Neotropical migrant passerines. – In: Greenberg, R. and Marra, P. P. (eds). Birds of two worlds: The ecology and evolution of migration. The John Hopkins Univ. Press, Baltimore.

Sogge, M. K., Owen, J. C., Paxton, E. H. and Langridge, S. M. 2001. A targeted mist net capture technique for the willow flycatcher. – Western Birds 32: 167–172.
Unitt, P. 1987. *Empidonax traillii extimus*: an endangered subspecies. – Western Birds 18: 137–162.
United States Fish and Wildlife Service. 1995. Final rule determining endangered status for the southwestern willow flycatcher. – Fed. Reg. 60: 10694–10715.
Webster, M. S., Marra, P. P., Haig, S. M., Bensch, S. and Holmes, R. T. 2002. Links between worlds: unraveling migratory connectivity. – Trends Ecol. Evol. 17: 76–83.
Zink, R. M. 2004. The role of subspecies in obscuring avian biological diversity and misleading conservation policy. – Proc. R. Soc. B 271: 561–564.