Faculty Publications in the Biological Sciences   Papers in the Biological Sciences

7-2016

The maintenance of phenotypic divergence through sexual selection: An experimental study in barn swallows *Hirundo rustica*

Rebecca Safran
University of Colorado, Boulder, rebecca.safran@colorado.edu

Yoni Vortman
Tel-Hai College, Israel

Brittany R. Jenkins
University of Colorado, Boulder, brittany.jenkins@colorado.edu

Joanna K. Hubbard
University of Nebraska-Lincoln, jhubbard@truman.edu

Matt Wilkins
University of Nebraska-Lincoln, m.r.wilkins06@gmail.com

See next page for additional authors
Follow this and additional works at: http://digitalcommons.unl.edu/bioscifacpub

Part of the Behavior and Ethology Commons, Biodiversity Commons, Biology Commons, Evolution Commons, Genetics Commons, Ornithology Commons, Other Ecology and Evolutionary Biology Commons, and the Population Biology Commons

Authors

Rebecca Safran, Yoni Vortman, Brittany R. Jenkins, Joanna K. Hubbard, Matt Wilkins, Rachel J. Bradley, and Arnon Lotem

The maintenance of phenotypic divergence through sexual selection: An experimental study in barn swallows *Hirundo rustica*

Rebecca J. Safran,[1] Yoni Vortman,[2,3] Brittany R. Jenkins,[1] Joanna K. Hubbard,[1,4]
Matthew R. Wilkins,[1,4] Rachel J. Bradley,[1] and Arnon Lotem[2]

1 Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, Colorado 80309
2 Department of Zoology, Faculty of Life Sciences, Tel-Aviv University 69978, Tel-Aviv, Israel
3 Hula Research Centre, Department of Animal Sciences, Tel-Hai College, Upper Galilee 12210, Israel
4 School of Biological Sciences, University of Nebraska–Lincoln, Lincoln, Nebraska 68588

R. Safran & Y. Vortman are co-first authors.

Corresponding author — R. Safran, email Rebecca.Safran@colorado.edu

Abstract

Previous studies have shown that sexual signals can rapidly diverge among closely related species. However, we lack experimental studies to demonstrate that differences in trait-associated reproductive performance maintain sexual trait differences between closely related populations, in support for a role of sexual selection in speciation. Populations of Northern Hemisphere distributed barn swallows *Hirundo rustica* are closely related, yet differ in two plumage-based traits: ventral color and length of the outermost tail feathers (streamers). Here we provide experimental evidence that manipulations of these traits result in different reproductive consequences in two subspecies of barn swallow: (*H. r. erythrogaster* in North America and *H. r. transitiva* in the East Mediterranean). Experimental results in Colorado, USA, demonstrate that males with (1) darkened ventral coloration and (2) shortened streamers gained paternity between two successive reproductive bouts. In contrast, exaggeration of both traits improved reproductive performance within *H. r. transitiva* in Israel: males with a combination treatment of darkened ventral coloration and elongated streamers gained paternity between two successive reproductive bouts. Collectively, these experimental results fill an important gap in our understanding for how divergent sexual selection maintains phenotype differentiation in closely related populations, an important aspect of the speciation process.

Keywords: Barn swallow, Hirundo rustica, mate selection, paternity, phenotype manipulation, sexual selection, sexual signal, speciation

While differences in natural selection pressures have long been at the forefront of hypothesized drivers of speciation (Darwin 1859; Schluter 2001, 2009; Nosil 2012; Langerhans & Riesch 2013), the role of sexual selection in speciation is still under debate (Ritchie 2007; Seddon et al. 2013). There has long been an interest in testing the role of sexual selection, or patterns of differential mating and fertilization success, in the evolution of phenotypic distinctions that may ultimately be responsible for reproductive isolation (Darwin 1871; West-Eberhard 1983; Sætre et al. 1997; Price 1998;

Uy and Borgia 2000; Boughman 2001; Irwin et al. 2001; Panhuis et al. 2001; Coyne & Orr 2004; Mendelson & Shaw, 2005; Boul et al. 2007; Ritchie 2007; Kraaijeveld et al. 2011; Seddon et al. 2013). Sexual selection may facilitate speciation when sexual signal traits and their associated preferences divergently co-evolve in closely related populations, leading to the evolution of premating isolation (West-Eberhard 1983; Price 1998; Panhuis et al. 2001; Ritchie 2007; Butlin et al. 2012; Safran et al. 2013). The potential strength of sexual selection in facilitating speciation is that such traits have the

potential to evolve much faster (Hoekstra et al. 2001; Kingsolver et al. 2001), and directly lead to prezygotic reproductive isolation before post-insemination reproductive isolation evolves (Coyne & Orr 1989).

Evidence suggests that sexual selection may be especially important in the early stages of speciation (West-Eberhard 1983; Panhuis et al. 2001; Kraaijeveld et al. 2011; Butlin et al. 2012; Safran et al. 2013). Results from comparative analyses (Gray & Cade 2000; Masta & Maddison 2002; Kraaijeveld et al. 2011; Seddon et al. 2013), theory (Lande 1981; Kirkpatrick & Ravigné 2002; van Doorn et al. 2009), and empirical studies (Irwin et al. 2001; Svensson et al. 2006; Rodríguez et al. 2013; Svensson & Waller 2013) demonstrate that variation in sexual signals, mate selection behaviors, or mate preferences can diverge among closely related taxa and lead to reproductive isolation and the evolution of independent lineages. Whereas previous studies have shown that sexual signals can rapidly diverge among closely related species (Irwin et al. 2001; Svensson et al. 2006; Rodríguez et al. 2013; Seddon et al. 2013; Svensson & Waller 2013), we lack studies that experimentally demonstrate evidence for associations between reproductive performance and trait divergence among closely related populations. Studies that have used playback experiments (Boul et al. 2007; Rodríguez et al. 2013), and manipulated male phenotypes (Svensson & Waller 2013) have suggested important roles for sexual selection in population divergence. However, while critical for examining trait and trait-preference divergence, previous research has not directly quantified whether variation in male signal traits is a function of divergence in reproductive performance associated with those traits. Thus, there are still fundamental gaps in our understanding of how sexual selection influences speciation.

While divergence of sexual signal phenotypes among closely related populations is arguably an important aspect of the process of speciation (Coyne & Orr 2004), the causes of population divergence are difficult to address long after species have become isolated in independent lineages because the process of divergence is confounded by further evolutionary changes speciation (Coyne & Orr 2004). Recently formed sister taxa or species complexes are advantageous for comparative studies of reproductive isolation, as they enable inferences about the maintenance of population-level differences by ongoing divergent selection. Accordingly, the strongest evidence for a role for divergent sexual selection in speciation would come from experimental studies in recently diverged taxa where it is possible to analyze whether the maintenance of phenotype differences is associated with variation in trait-associated reproductive performance (Panhuis et al. 2001, Safran et al. 2013).

The barn swallow species complex (*Hirundo rustica*: Family Hirundinidae) encompasses a widespread breeding distribution throughout the Northern Hemisphere. Recent phylogeographic and population genetic analyses suggest this group formed rapidly and, as such, is not strongly genetically differentiated, despite marked phenotypic and behavioral differentiation among populations (Dor et al. 2010, 2011; Scordato and Safran 2014, Safran et al. 2016). In particular, melanin-based feather color and the length of the outermost tail feathers (streamers; McGraw et al. 2004, 2005) of male barn swallows vary among closely related subspecies (Turner & Rose 1989; Turner 2006; Scordato and Safran 2014). The European barn swallow (*H. r. rustica*) has been the subject of intense research activity that has revealed sexual selection for elongated tail streamers in males (Møller 1988, 1989, 1994; Turner 2006; Saino et al. 2007). Research in North American populations (Safran et al. 2005) demonstrated a causal relationship between ventral coloration and reproductive success, whereas correlational and experimental studies in another closely related, yet geographically distant, subspecies in the east Mediterranean (*H. r. transitiva*) revealed that both elongated tail streamers and dark plumage color are associated with greater reproductive performance in males (Vortman et al. 2011, 2013). Correlational studies from Japan (*H. r. gutturalis*) indicated that sexual selection may act on one aspect of ventral color—the size and color of the male throat patch—with reduced sexual selection on streamer length compared to *H. r. rustica* males (Hasegawa et al. 2010, 2014). Further, previous work has shown that both streamer length and ventral color are heritable within and among different populations of barn swallows (reviewed in Scordato and Safran 2014; Vortman et al. 2015; Hubbard et al. 2015). The results of these studies collectively suggest the hypothesis that phenotypic differences in heritable male sexual traits among subspecies are indeed related to variation in trait-associated reproductive performance, but do not provide experimental evidence for divergent sexual selection.

To directly test for a role of divergent sexual selection in phenotypic differentiation, we replicated the same phenotype manipulation experiment in two geographically isolated, closely related, but phenotypically distinct subspecies of barn swallows (*H. r. erythrogaster* in North America and *H. r. transitiva* in Israel). Males in North America have shorter tail streamers but darker ventral coloration compared to males in Israel. Like most songbirds, extra-pair mating is prevalent in barn swallows (Saino et al. 1997, Kleven et al. 2005; Safran et al. 2005; Neuman et al. 2007, Vortman et al. 2011). We thus designed an experiment to analyze changes in a male's paternity in two successive breeding bouts, before and after an experimental manipulation of a male's phenotype. The results from the *H. r. transitiva* study in Israel are published elsewhere (Vortman et al. 2013). Here, we (1) report experimental details and results for a population of *H. r. erythrogaster* in Colorado, USA, and then (2) directly compare results obtained from the same experiment conducted separately in each subspecies to (3) clearly assess differences in reproductive outcomes associated with divergent mate preferences.

Methods

Barn swallows in the USA were studied at 15 breeding sites in Boulder and Jefferson Counties, Colorado from April–July 2009. Each breeding location is a discrete location (barn, garage, bridge) in which the number of breeding pairs ranged from 1 to 45. Barn swallows in Israel were studied at two breeding locations over two consecutive breeding seasons from November 2008 through July 2010 (additional details can be found in Vortman et al. 2013). The exact same experimental protocol, described below, was applied to breeding populations in Colorado and Israel.

Phenotype Data Collection

In each study location we captured using mist nets, uniquely marked with color bands, and collected a blood and feather sample from all males and females before they initiated their first clutch of eggs. Individuals were observed at their nests to identify social breeding pairs. We measured or collected the following from each individual: (1) a sample of ventral feathers for objective color quantification, (2) length of tail streamers, and (3) blood samples as a DNA source from each bird (approx. 50 µl, stored in 2% SDS lysis buffer).

Color measurement

We collected 5–10 feathers from the breast (upper ventral region) and stored them in small envelopes in a dark, dry environment prior to measurement (*following* Safran and McGraw 2004). We assessed the color of these samples by measuring plumage brightness using an Ocean Optics USB4000 spectrometer (Dunedin, FL). Reflectance data were generated relative to a white standard (Ocean Optics WS-1) and a dark standard (all light excluded), and spectra were recorded with the SpectraSuite software package (ver. 2.0.125; Ocean Optics, Inc.). We used average brightness, which was calculated from three separate measurements of the collected breast feathers, as a representative metric of overall ventral plumage color because the three traditional axes of color (hue, chroma, and brightness) were found previously to be highly correlated across the ventral region of individual barn swallows (Safran and McGraw 2004, Mc-Graw et al. 2005; J. K. Hubbard, unpubl. data), and brightness is the most variable dimension of color in this region (J. K. Hubbard, unpubl. data). Lower brightness scores (% reflectance) indicate plumage color that appears darker, redder, and more saturated when compared to feathers with higher brightness scores.

Experimental Setup

We analyzed reproductive outcomes associated with the same set of phenotype manipulations in Colorado and Israel and then compared these results to test hypotheses about divergent sexual selection. After social pairs were identified, we (1) allowed a female to settle with a mate to lay a complete clutch of eggs, and (2) experimentally removed this first clutch of eggs one week to 10 days into the incubation period to simulate a nest predation event. On the same day as the egg removal, we (3) recaptured each male to manipulate his phenotype using six treatment groups (see text below and Table S1). Clutch removal induced each female to lay a replacement clutch after the phenotype of her mate was experimentally manipulated, enabling us to (4) directly test differences in a male's paternity before and after phenotypic manipulation. The use of within-individual comparisons is a powerful means for removing potential confounding factors that are otherwise difficult to control (e.g., aspects of the pre-manipulation phenotype: age, experience, arrival date, nest site, identity, and phenotype of mate).

Phenotype manipulations

In total, 67 males were manipulated in Israel, 66 in Colorado. On day 10–11 of the incubation period, when the eggs from first clutches were removed, males were recaptured and randomly assigned to one of six treatment groups: "dark", "short", "long", "dark + short", "dark + long", and "control." For (1) "dark" males ($n = 8$ Israel, $n = 7$ Colorado), all of the feathers along the entire ventral portion of each male were darkened using a Prismacolor nontoxic art marker (shade: light walnut); color resulting from this manipulation falls within the natural range of variation (Safran et al. 2005, 2008; Vitousek et al. 2013; Vortman et al. 2013). For (2) "short" ($n = 6$ Israel, $n = 7$ Colorado) and (3) "long" males ($n = 6$ Israel, $n = 6$ Colorado), tail streamers were shortened or elongated by reducing or increasing the length of each streamer by 1.5 standard deviations from the population mean (i.e., 13 mm in Israel [Vortman et al. 2013] and 9.3 mm in Colorado). For (4) "dark + short" ($n = 6$ Israel, $n = 8$ Colorado) and (5) "dark + long" males ($n = 6$ Israel, $n = 9$ Colorado), combinations of streamer reduction or elongation were paired with ventral feather darkening. Finally, as a (6) "control" ($n = 7$ Israel, $n = 9$ Colorado), we included an unmanipulated group of males. Sham control treatments for both the tail streamer and color manipulations have previously been shown to not affect paternity outcomes in barn swallows (Saino et al.1997; Safran et al. 2005). It is difficult to remove melanin-based color from birds without abrading the microstructure of their feathers, and we have thus far not found a suitable, nontoxic way of doing this. Accordingly, our treatments do not include a "light" group. Note that each individual was captured temporarily and released immediately after the phenotype manipulation. See Table S1 for a complete description of each experimental treatment. The post-manipulation data for Israel is reported in Vortman et al. (2013). Below, we report the post-manipulation phenotype information for each treatment for the Colorado population.

Sample size

Our initial sample included 66 study pairs, of which 59 pairs (89%) initiated a replacement clutch within one week after clutch removal. Of those 59 pairs, 11 failed due to nest predation ($n = 6$), a falling nest ($n = 2$), or the failure of eggs to hatch ($n = 3$). Two additional pairs were removed from our analysis because of genotyping error. The number of nest failures did not differ by treatment group (likelihood ratio test, chi-square = 5.085, $P > 0.40$, see Table S2). Treatments were randomly assigned across 15 breeding sites with additional consideration of minimizing the replication of each treatment. There was no replication of any one of the six treatments at smaller sites consisting of eight or fewer pairs; only the largest breeding sites had more than one replicate of each treatment.

Changes in tail streamer length

Males randomly assigned to the tail streamer "short" or "long" groups had their streamers shortened by 1.5 standard deviations from the population mean, relative to their original streamer length (approximately 9.3 mm). For all individuals, tail streamers were cut 15 mm from the base of attachment and shorter or longer streamers were reattached by connecting each piece of streamer together with fishing line in the interior portion of each feather to add structure and with super glue to maintain the connection (Bro-Jørgensen et al. 2007). The end of the reattached streamer was then trimmed such that the manipulation added or reduced the original streamer length by 9.3 mm. Males in the "short" group had an original streamer length of (mean ± SD) 90.97 ± 7.21 mm, whereas the post-manipulation average was 80.80 ± 6.34 mm; their streamers were shortened by an average of 8.97 ± 6.82 mm (paired t-test: post-manipulation, minus pre-manipulation streamer length, $t = -5.87$, df = 19, $P < 0.001$). Males in the "long" group had an original streamer length of (mean ± SD) 91.10 ± 5.92 mm, whereas the post-manipulation streamer length average was 101.30 ± 5.38 mm; their streamers were lengthened by an average of 10.29 ± 2.68 mm (paired t-test: within-individual difference in post-manipulation minus pre-manipulation streamer length differs from zero, $t = 16$, df = 17, $P < 0.001$). After the streamer was trimmed to the appropriate length, we also rounded the edges to mimic the natural shape of this structure.

Changes in color

Males in the three color treatment groups ("dark", "dark + long", and "dark + short") had an original breast brightness measure of (mean ± SD) 25.93 ± 6.49% brightness, whereas the post-manipulation average was 20.20 ± 4.26% brightness; their feather color was darkened by reducing average brightness by (paired t-test: post-manipulation minus pre-manipulation color, $t = 23.18$, df = 23, $P < 0.001$).

Genetic paternity analyses

We used microsatellite-based analyses to determine the paternity of offspring in first versus replacement broods in order to examine changes in paternity allocation in response to experimental manipulation of male sexual signals (Safran et al. 2005; Vortman et al. 2013). DNA samples from each embryo in the first clutch and from each nestling in the replacement brood, as well as their parents, were genotyped at six variable microsatellite loci (see Supporting Information for details). We then compared paternity allocation to the same male across broods as a function of treatment group (Safran et al. 2005; Vortman et al. 2013).

Statistical Analyses

All analyses were conducted using SAS version 9.2. We measured reproductive performance using three metrics to characterize a male's paternity of the offspring in his nest: number of own young, number of extra-pair young, and proportion of own young (number of own young/total number of offspring). Paternity outcomes were compared for each breeding bout (before and after phenotype manipulation) to directly assess the impact of the phenotype manipulation. We used these three metrics to describe paternity because the proportional and count data can offer different insights into the outcome of our experiment. First, proportional data are sensitive to variation in clutch and brood size and thus the complement of providing count data is important. Second, proportional data can provide insights into how females allocate resources toward particular males, whereas information about the number of eggs and offspring (count data) can provide insights about female investment. Because these three paternity measures are correlated with one another, we assess the significance of treatment effects using bootstrap analyses when $P < 0.05$. For models where the effect of 'treatment' is significantly different from zero, we analyze the significance of multiple comparisons among treatment groups using confidence intervals around parameter estimates (Nakagawa 2004).

Analyzing baseline paternity before the phenotype manipulation

The proportional data (paternity of males across each treatment group at the start of the experiment) could be biased by differences in the clutch and brood sizes. Thus, we applied a logistic model using the binomial dependent variable (number of own young in clutch versus total number of young in clutch) and a logit-link function to analyze differences among treatment groups at the start of the experiment (following Safran et al. 2005). Further, to analyze a male's pre-manipulation paternity from his social mate, we used both parametric (ANOVA) and nonparametric (Kruskal–Wallis) analyses, given the nonnormal distribution for two of these three paternity measures.

Analyzing changes in a male's paternity as a function of the phenotype manipulation

Next, we analyzed the differences in these three paternity variables between the two successive breeding attempts to directly assess the consequences of phenotypic manipulations. Differences in paternity measures (calculated as the difference in paternity before and after the manipulation), were all normally distributed, with statistically equivalent variances, within each treatment group (Fig. S1). Accordingly, we were able to use F-tests (one-way ANOVAs) to describe the results of our experiment with respect to the three measures of paternity.

Hypothesis tests

In each population, we manipulated both ventral coloration and tail streamer length using the exact same methods. The replication of the same experiment in two closely related, yet geographically isolated populations enabled us to test the following hypothesis: divergent sexual selection maintains heritable phenotypic differences between these young subspecies. Specifically, we predicted that *H. r erythrogaster* and *H. r. rustica* males differ in reproductive outcomes associated with the same set of phenotype manipulations. We analyzed whether the same phenotype treatments had different reproductive consequences for males in each population by testing whether the statistical interaction between breeding location (Colorado and Israel) and phenotype treatment differed from zero. Support for this hypothesis would provide causal evidence for a role of divergent sexual selection in maintaining phenotype differences among these closely related populations. Further, our experimental design enabled us to directly analyze and compare which traits are causally related to reproductive performance within and among each population of barn swallows.

We adopted two complementary hypothesis-testing approaches to examine the effects of the phenotype manipulations. Each of these approaches assessed the significance of "treatment" (Colorado) or "treatment × country" (Colorado vs. Israel). Note that although the data from Israel are based on two years of data collection, previous analyses within Israel tested for and did not find an influence of "year" on paternity outcomes (Vortman et al. 2011, 2013).

For all models with a significant "treatment" or "treatment × country" effect, we applied post hoc *t*-tests to compare pairwise differences among treatment groups. For the Colorado population, we analyzed treatment outcomes relative to outcomes in the control group using a specialized multiple comparison approach (Dunnett's test; Dunnett 1955). The null hypothesis for this analysis is that there are no differences in a male's paternity as a function of the experimental treatment when compared to changes in males within the control group. For the comparisons of Colorado to Israel, we used adjusted treatment effects, accounting for differences in control group outcomes (see Table S5), to analyze the significance of the "treatment × country" interaction.

For the comparative study of treatment outcomes in Israel and Colorado, we tested the hypothesis of divergent sexual selection by analyzing whether treatment outcomes differ among populations with the null hypothesis stated as no effect of a "treatment × country" interaction. In cases where we failed to find support for the null hypothesis (the effect of the "treatment × country" interaction was significantly different from zero), we analyzed post hoc comparisons of the same treatment outcome using *t*-tests. We further analyzed differences in treatment outcomes relative to outcomes for males in the control group.

Based on previous correlational work, we predicted that males in Colorado in our "dark" treatment would have the greatest paternity gains (Safran and McGraw 2004; Neuman et al. 2007), as was shown in a previous manipulative experiment of this trait alone (Safran et al. 2005). Further, based on previous correlational work in the North American population, we did not *a priori* predict that males in either of our streamer length treatments would have strong paternity gains (Safran and McGraw 2004; Neuman et al. 2007). In contrast, previous experimental work within *H. r. transitiva* has shown that males whose streamers were elongated and ventral plumage darkened had the greatest paternity gains (Vortman et al. 2013).

Significance testing

We analyzed the significance of multiple comparisons among treatment groups using confidence intervals around parameter estimates (Nakagawa 2004). We further applied bootstrapping analyses, a resampling method used for analyzing the precision and accuracy of significance tests (Whitlock & Schluter 2009). For all significance tests related to our experimental design (full models), we conducted 5000 bootstraps from which we report the mean and 95% confidence intervals for each P-value estimate.

Results

Trait Differences

Among the six widespread subspecies of barn swallow, both the North American and Israeli populations are considered relatively dark in ventral color (Scordato and Safran 2014). Between these two populations, North American males have shorter tail streamers (mean ± SD, Colorado: 91.02 mm ± 7.25, Israel: 102.81 mm ± 10.20, $t = 8.28$, $P < 0.001$, df = 71.92) and slightly darker ventral plumage (mean breast % brightness ± SD, Colorado: 23.09 ± 6.43, Israel: 29.51 ± 7.84, $t = 3.02$, $P < 0.001$, df = 74.15) compared to males in Israel.

Phenotype Experiment

Experimental details

There were no differences in the color (ANOVA, breast brightness: $F_{5, 25} = 0.46$, $P > 0.80$) or tail-streamer lengths (ANOVA, tail streamer length, $F_{5, 45} = 1.18$, $P > 0.33$) of males that were randomly assigned to the six treatment groups at the start of the experiment. Moreover, because our experimental manipulations were highly synchronized, there were no differences in the clutch-initiation dates of the first breeding attempts among the six treatment groups (ANOVA: $F_{5, 46} = 0.96$, $P > 0.45$). Finally, there were no overall differences in the clutch sizes within nests associated with males randomly assigned to our experimental treatments in the first (ANOVA: $F_{5, 46} = 0.65$, $P > 0.66$) and second, replacement clutches (ANOVA: $F_{5, 46} = 1.90$, $P > 0.11$). Whereas there is typically a seasonal reduction in reproductive output in passerine birds, we did not find a difference in the clutch size between the first and replacement breeding attempts: average clutch size for first attempt = 4.72, SD = 0.64, average clutch size for replacement attempt = 4.69, SD = 0.73, two-sample t-test $t = 0.80$, $P > 0.80$, df = 120.

Baseline paternity in first clutches

In the clutches laid before phenotype was manipulated, there were no initial differences in paternity across treatments (number of own young in clutch versus total number of young in clutch): logistic model, chi-square = 5.90, $P > 0.31$, number of a male's own young in nest: Kruskal–Wallis test, chi-square = 5.32, df = 5, $P > 0.37$; number of extra-pair young in male's nest (Kruskal–Wallis test, chi-square = 5.55, df = 5, $P > 0.35$ (Fig. 1).

Changes in paternity as a function of the phenotype manipulation

Relative to a male's paternity in the first breeding attempt, there was a significant effect of the plumage color and streamer manipulations on two of the three measures of changes in a male's paternity (differences in proportion of paternity: ANOVA, $F_{5,45} = 2.52$, $P < 0.02$, bootstrap analysis: mean P-value < 0.009, lower and upper 95% CI = 0.007–0.009; differences in number of a male's own young in nest: ANOVA, $F_{5,45} = 1.48$, $P > 0.21$; bootstrap analysis: $P < 0.06$, CI = 0.050–0.057; changes in the number of extra-pair young in male's nest: ANOVA, $F_{5,45} = 3.92$, $P < 0.01$, bootstrap analysis: $P < 0.008$, CI = 0.006–0.007; Fig. 2).

We applied post hoc pairwise comparisons (t-tests) for the analysis of changes in proportion of paternity (Fig. 2a) and differences in the number of extra-pair young (Fig. 2c). Males in the "dark" and "short" groups gained paternity and had reduced levels of extra-pair young in their second, post-manipulation, reproductive bouts compared to their first (Table S3).
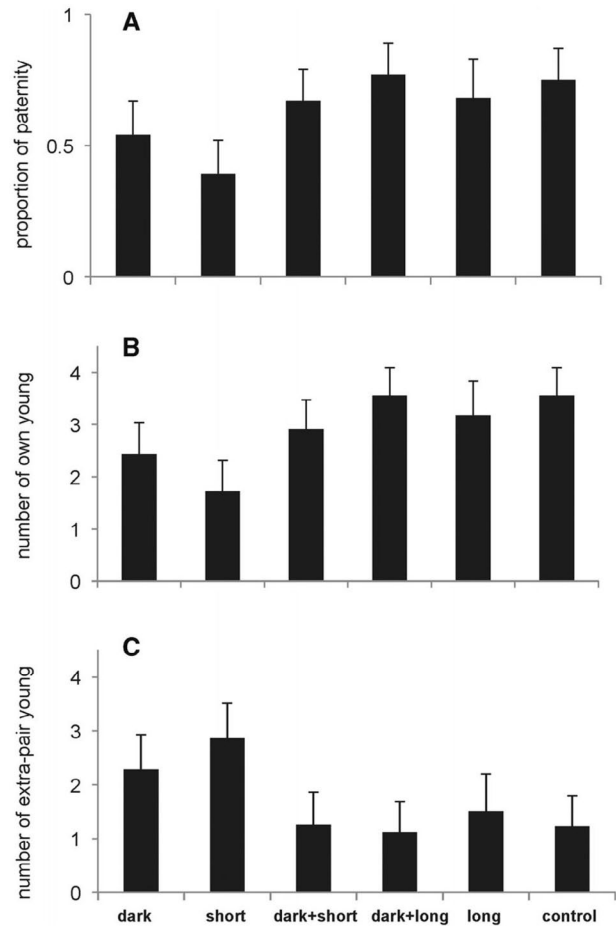
Figure 1. (a) The proportion of within-pair young sired by males in three treatment groups did not differ at the start of the experiment. (b) The number of offspring in the nest that were sired by the focal male did not differ across treatments at the start of the experiment. (c) The number of extra-pair offspring (a measure of cuckoldry) in the nest of the focal male that were sired by other males in the population did not differ across treatments at the start of the experiment. In all three figures, mean ± SE is shown.

Comparisons of changes in paternity relative to males assigned to the control group indicate that males in the "dark" and "short" groups were significantly different from the control groups in terms of changes in the proportion of paternity. In contrast, only males in the "dark" group had changes in the number of extra-pair young between two successive breeding attempts (Table S3).

Phenotype Experiments: Comparing Colorado And Israel

For all three measures of paternity, we found a significant effect of the interaction of country and treatment (Table 1, Fig. 3). Post hoc pairwise comparisons of the same treatment outcomes (Fig. 3, Table S4) indicate that population differences are most pronounced for males in the "dark + long" and "dark + short" treatment groups. Although we did not

Table 1. Population differences in changes in paternity as a function of phenotype manipulation in 85 breeding pairs (*n* = 46 from Colorado and *n* = 39 from Israel)

Factor	*F* (df)	*P*	Bootstrap results (5000 permutations) Mean *P*-value/lower and upper 95% CI
Changes in a male's proportion of paternity in successive reproductive bouts			
Country	0.47 (1,73)	>0.50	0.42/ 0.42–0.44
Treatment	2.92 (5,73)	<0.02	0.02/0.02–0.03
Country × Treatment	3.31 (5,73)	<0.01	0.02/0.02–0.02
Changes in a male's number of own offspring in successive reproductive bouts			
Country	0.62 (1,73)	>0.43	0.40/0.36–0.41
Treatment	0.92 (5,73)	>0.47	0.29/0.22–0.23
Country × Treatment	2.59 (5,73)	<0.05	0.04/0.04–0.04
Changes in a male's number of extra-pair offspring in a focal male's nest in successive reproductive bouts			
Country	0.91 (1,73)	>0.34	0.37/ 0.37–0.38
Treatment	3.15 (5,73)	<0.02	0.03/0.02–0.03
Country × Treatment	4.17 (5,73)	<0.01	0.01/<0.01–0.01

find statistically significant differences in the control groups across the three different measures of paternity outcomes, Colorado males in the control treatment lost paternity, whereas males in Israel did not. We thus explored these potentially confounding differences. Controlling for differences in the paternity outcomes for the control groups in Israel and Colorado (Table S5) indicate that our results are not explained by population differences in control group reproductive performance (Table S6).

Discussion

Phenotype differentiation is inextricably involved in the speciation process, whether in relation to local adaptation to ecological or social context, or both (Maan and Seehausen 2011; Safran et al. 2013). Further, differences in phenotype have been shown to obstruct gene flow between closely related populations (e.g., Shafer and Wolf 2013), indicating the role of trait differentiation and the evolution and maintenance of genetic divergence and reproductive isolation. Whereas previous work has demonstrated the importance of ecological adaptation to the speciation process (e.g., Schluter 2001, 2009), empirical research based on population studies and phylogenetic comparative analyses has also provided evidence that sexual signal traits can rapidly diverge among closely related species (e.g., Irwin et al. 2001; Svensson et al. 2006; Rodríguez et al. 2013; Seddon et al. 2013; Svensson & Waller 2013). However, experimental evidence for divergence in trait-associated reproductive outcomes is needed to infer a role of sexual selection in the maintenance of sexual signal differences among closely related populations.

Here, we establish experimental evidence for a role of divergent sexual selection in the maintenance of phenotype differences in two closely related, geographically isolated

populations. Specifically, these results indicate that sexual selection favors different combinations of the same traits in recently diverged, yet geographically isolated populations of barn swallows. In Colorado, males gained paternity in association with both the "dark" and "short" treatments, suggesting that darker ventral plumage and shorter streamer length are associated with reproductive performance in this population. Males in Israel gained paternity in association with the combination "dark + long" treatment, indicating that darker ventral plumage and longer streamer length are associated with reproductive performance in this population (Vortman et al. 2013). Direct comparisons between the two populations indicate a significant effect of the interaction of "country × treatment" and post hoc analyses to compare paternity outcomes associated with the same phenotype treatment show the most divergent effects for the "dark + short" and "dark + long" treatments.

Further, the outcomes from experimental treatments mirror the quantitative trait variation among these populations. Both populations are fairly dark in ventral plumage compared to other subspecies of barn swallows, whereas males in Colorado and throughout North America have short tail streamers compared to other barn swallow populations and males in Israel have among the longest streamers (Turner 2006; Scordato and Safran 2014). Our comparative experiments therefore suggest that differences in sexual selection play an important role in maintaining phenotypic differences in these closely related populations of barn swallows.

Colorado Paternity Experiment

Males with experimentally darkened ventral color gained paternity benefits from their social mates, corresponding to the results of a similar experiment conducted in a population
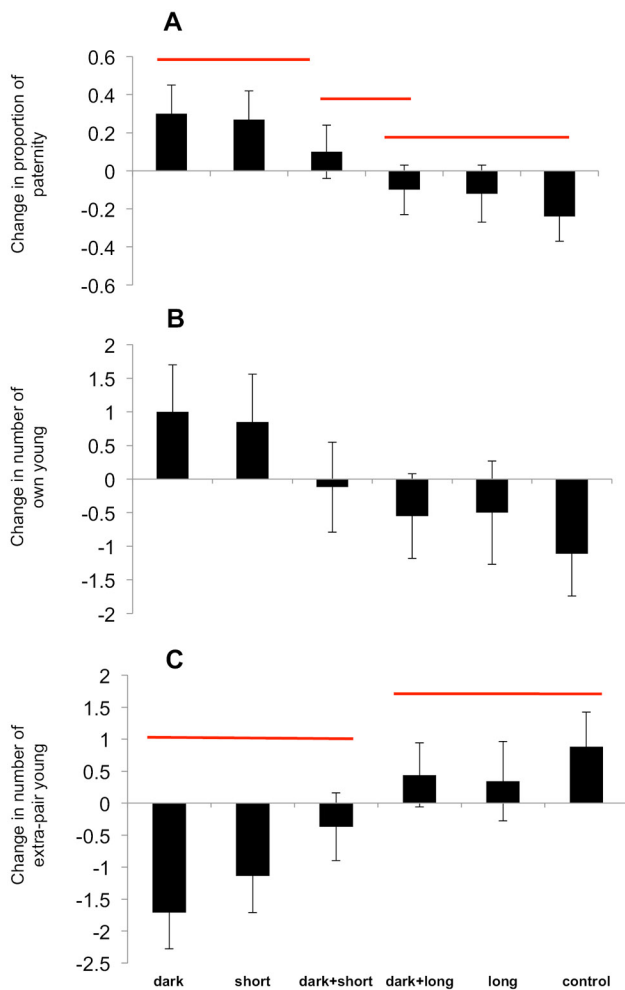
Figure 2. Changes in a male's paternity between two successive breeding events: the post-manipulation event minus the first breeding event. (a) Differences in proportion of paternity for the replacement broods minus the proportion of paternity for the first breeding attempts; a positive change indicates that males gained paternity from their social mate between breeding events; (b) Differences in the number of young sired by the focal male indicate no overall changes in the number of offspring a male sired in his nest as a function of phenotype manipulation. (c) Differences in the number of extra-pair young sired by other males in the population. A negative change in the number of extra-pair young (c) indicates that males cared for fewer extra-pair offspring in their second breeding event, compared to their first. In all three figures, mean ± SE is shown. In (a) and (c), where we found a significant effect of "treatment," we illustrate post hoc differences in treatment groups as follows: treatment groups connected by a line indicate that the paternity outcomes did not differ significantly among these treatments. Treatment groups not connected by a line indicate significant differences among those groups.

of North American barn swallows in New York State (Safran et al. 2005). A lack of clear benefits associated with longer streamers corresponds to several previous correlational (Safran and McGraw 2004; Neuman et al. 2007) and

experimental (Smith & Montgomerie 1991; Smith et al. 1991) results from barn swallow populations in the northeastern United States, with one exception. A correlational study of streamer length and paternity outcomes in a Canadian population of barn swallows (Kleven et al. 2006) indicated that males with longer streamers do not have paternity advantages within their social pair bond, but rather outside of it. Our experiment in Colorado was designed to analyze within-pair changes in a male's paternity as a function of the phenotype manipulation. However, data from a smaller breeding populations in Israel suggest that manipulated males of the "dark + long" treatment group also increased their number of extra-pair young outside of their social pair bond (Vortman et al. 2013), suggesting that these traits further influence males' reproductive success gains via extra-pair matings. Although the possibility of alternative reproductive strategies involving within- versus extra-pair reproductive success merits further study, the fact that our experimental results concur with phenotype differences among these closely related populations suggests we are not missing important and potentially surprising relationships between male phenotype and reproductive performance.

As shown in an earlier ventral plumage manipulation experiment (Safran et al. 2005), Colorado males with darkened plumage color gained paternity in replacement clutches, demonstrating that ventral color is causally related to reproductive performance in two different populations of North American barn swallows. The additional insights gained from the tail streamer manipulations in this North American population are surprising: counter to tail elongation experiments in European populations (Saino et al. 1997) and in Israel (Vortman et al. 2013), we did not find evidence of paternity gains associated with streamer elongation. Instead, our results indicate paternity gains for males with shortened streamers. Surprisingly, the paternity outcomes associated with the combination treatment "dark + short" did not differ from zero, suggesting a lack of response to the presentation of both phenotype manipulations within an individual. The lack of paternity gains associated with the "dark + short" treatment is interesting given that ventral color and streamer length is not correlated in North American populations of barn swallows (Wilkins et al. 2015); Safran and McGraw 2004.

Divergent Sexual Selection

Accumulating case studies have documented a coupled divergence in sexual signaling traits and reproductive performance, which indicates sexual selection is playing a role in speciation (e.g., reviewed in Maan and Seehausen 2011). However, little is known about why populations diverge in signaling traits related to reproductive performance. Variation among populations in sexually selected traits may have evolved due to drift or stochastic events in each population, later amplified in a Fisherian "run-away" process, or due to environmental differences which impose different costs on
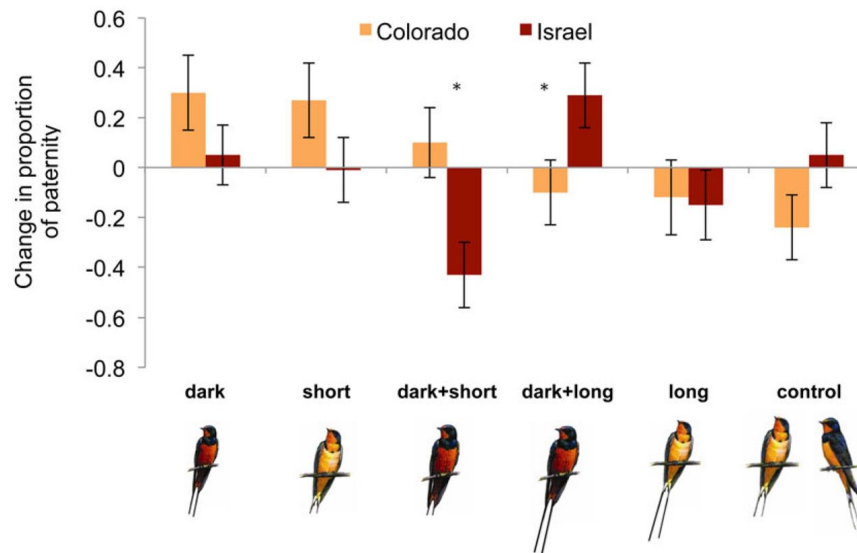
Figure 3. Differences in a male's proportion of paternity (post-manipulation proportion minus pre-manipulation proportion) across the same treatment groups in two different populations. Treatment mean (±SE) differences are shown to contrast treatment outcomes for males in Israel and Colorado. Statistically significant differences in least square means for the interaction term "country × treatment" ($P < 0.05$) are indicated by asterisk. See Table S1 for additional details about treatment differences for other measures of paternity.

signal maintenance or detectability (Schluter and Price 1993; Seehausen et al. 2008; Maan & Seehausen 2011; Wagner et al. 2012). Recent theory (van Doorn et al. 2009), synthesis (Maan and Seehausen 2011), and a classic model (Schluter and Price 1993) propose that sexual selection can drive speciation as organisms adapt to new environmental and social contexts. Here, a general hypothesis is that signal divergence is underlain by a pattern in which females prefer the most informative male signal in a given environmental and social context (Schluter and Price 1993). Although populations of barn swallows throughout their widespread breeding distribution are fairly similar in social behavior and ecological variables related to breeding (nest site, aerial insectivory), individuals breeding in Israel and Colorado are exposed to different ecological contexts in terms of predators, parasites, and climate that might be relevant to the development or costs of sexual signals (i.e., aerodynamic restrictions; Bro-Jørgensen et al. 2007). Additionally, interesting life history differences are apparent between these closely related populations, as barn swallows in Israel are considered resident, whereas barn swallows in North American populations are long-distance migrants. These differences in large annual movements may influence traits related to flight, including wing length and the extent of tail streamer elongation. Moreover, the *transitiva* population is located between two other distinct subspecies, *H. r. rustica* and *H. r. savignii*, which may affect the evolution of its sexual signals through gene flow or reinforcement (see Vortman et al. 2013), while the North American population is allopatric with respect to other subspecies of barn swallow (Scordato and Safran 2014). Further analyses are needed to determine whether and which specific ecological variables influence sexual signal

development or expression in order to test hypotheses about interactions of sexual and natural selection on morphological trait divergence (Safran et al. 2013). One interesting possibility rests on differences in the parasite communities that both adults and nestlings are exposed to in Israel and North America, as these are known to influence the development of plumage color in the Colorado population (A. K. Hund and J. K. Hubbard, unpubl. data). Finally, nonadaptive hypotheses as to why phenotypes differ between these populations in the first place must also be considered, including nonrandom associations between phenotype and founder effects, as barn swallows radiated from Africa to the entire northern hemisphere. Yet, the concurrence between reproductive outcomes of the phenotype manipulation and actual phenotype divergence between these two populations suggests some adaptive function to differences in male phenotypes of traits in Israel and Colorado.

Collectively, these experimental results fill an important gap in our understanding for how divergent sexual selection explains phenotype differentiation in closely related populations, an important aspect of the speciation process. While the exact causes for differences in sexual selection between the two populations are yet to be clarified, the presence of such differences that are consistent with the hypothesis that differences in trait-associated reproductive performance maintains phenotypic differences between the two closely related populations.

Data Archiving

The doi for our data is 10.5061/dryad.5865f.

Literature Cited

Boughman, J. W. 2001. Divergent sexual selection enhances reproductive isolation in sticklebacks. Nature 411:944–948.

Boul, K. E.,W. C. Funk, C. R. Darst, D. C. Cannatella, and M. J. Ryan. 2007. Sexual selection drives speciation in an Amazonian frog. Proc. Biol. Sci. 274:399–406.

Bro-Jørgensen, J., R. A. Johnstone, and M. R. Evans. 2007. Uninformative exaggeration of male sexual ornaments in barn swallows. Curr. Biol. 17:850–855.

Butlin, R., A. Debelle, C. Kerth, R. R. Snook, L. W. Beukeboom, Castillo Cajas, R. F. W., Diao, M.E. Maan, S. Paolucci, F.J. Weissing, et al. (2012). What do we need to know about speciation? Trends Ecol. Evol. 27:27–39.

Coyne, J. A., and H. A. Orr. 1989. Patterns of speciation in Drosophila. Evolution 43:362–381.

———. 2004. Speciation. Sinauer Associates, Sunderland, MA.

Darwin, C. 1859. On the origin of species by means of natural selection. 1st ed. John Murray, London.

———. 1871. The descent of man in relation to sex. John Murray, London.

Dor, R., R. J. Safran, F. H. Sheldon, D. W. Winkler, and I. J. Lovette. 2010. Phylogeny of the genus Hirundo and the Barn Swallow subspecies complex. Mol. Phylogenet. Evol. 56:409–418.

Dor, R., R. J. Safran, Y. Vortman, A. Lotem, A. McGowan, M. R. Evans, and I. J. Lovette. (2011). Population genetics and morphological comparisons of migratory European (*Hirundo rustica rustica*) and sedentary East-Mediterranean (*Hirundo rustica transitiva*) barn swallows. J. Hered. 103:55–63.

Dunnett, C.W. 1955. A multiple comparison procedure for comparing several treatments with a control. J. Am. Statist. Assoc. 50:1096–1121.

Gray, D. A., and W. H. Cade. 2000. Sexual selection and speciation in field crickets. Proc. Natl. Acad. Sci. USA 97:14449–14454.

Hasegawa, M., E. Arai, M.Watanabe, andM. Nakamura. 2010. Mating advantage of multiple male ornaments in the Barn Swallow *Hirundo rustica gutturalis*. Ornithol. Sci. 9:141–148.

———. 2014. Colourful males hold high quality territories but exhibit reduced paternal care in barn swallows. Behaviour 591–612.

Hoekstra, H. E., J. M. Hoekstra, D. Berrigan, S. N. Vignieri, A. Hoang, C. E. Hill, A. Hoang, P. Beerli, and J.G. Kingsolver. 2001. Strength and tempo of directional selection in the wild. Proc. Natl. Acad. Sci. USA 98:9157–9160.

Irwin, D. E., S. Bensch, and T. D. Price. 2001. Speciation in a ring. Nature 409:333–337.

Kingsolver, J. G., H. E. Hoekstra, J.M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, et al. 2001. The strength of phenotypic selection in natural populations. Am. Nat. 157:245–261.

Kirkpatrick, M., and V. Ravign´e. 2002. Speciation by natural and sexual selection: Models and experiments. Am. Nat. 159:S22–S35.

Kleven, O., F. Jacobsen, R. J. Robertson, and J. T. Lifjeld. 2005. Extrapair mating between relatives in the barn swallow: A role for kin selection? Biol. Lett. 1:389–392.

Kleven, O., F. Jacobsen, R. Izadnegahdar, R. J. Robertson, and J. T. Lifjeld. 2006. Male tail streamer length predicts fertilization success in the North American barn swallow (*Hirundo rustica erythrogaster*). Behav. Ecol. Sociobiol. 59:412–418.

Kraaijeveld, K., F. J. L. Kraaijeveld-Smit, and M. E. Maan. 2011. Sexual selection and speciation: the comparative evidence revisited. Biol. Rev. Camb. Philos. Soc. 86:367–377.

Lande, R. 1981. Models of speciation by sexual selection on polygenic traits. Proc. Natl. Acad. Sci. 78:3721–3725.

Langerhans, R. B., and R. Riesch. 2013. Speciation by selection: A framework for understanding ecology's role in speciation. Curr. Zool. 59:31–52.

Maan, M. E., and O. Seehausen. 2011. Ecology, sexual selection and speciation. Ecol. Lett. 14:591–602.

Masta, S. E., and W. P. Maddison. 2002. Sexual selection driving diversification in jumping spiders. Proc. Natl. Acad. Sci. USA 99:4442–4447.

McGraw, K. J., R. J. Safran, M. R. Evans, and K. Wakamatsu. 2004. European barn swallows use melanin pigments to color their feathers brown. Behav. Ecol. 15:889–891.

McGraw, K. J., R. J. Safran, and K. Wakamatsu. 2005. How feather colour reflects its melanin content. Funct. Ecol. 19:816–821.

Mendelson, T. C., and K. L. Shaw. 2005. Rapid speciation in an arthropod. Nature 433:375–376.

Møller, A. P. 1988. Female choice selects for male sexual tail ornaments in the monogamous swallow. Nature 332:640–642.

———. 1989. Viability costs of male tail ornaments in a swallow. Nature 339:132–135.

———. 1994. Sexual selection and the barn swallow. Oxford Univ. Press, New York.

Nakagawa, S. 2004. A farewell to Bonferonni: The problem of low statistical power and publication bias. Behav. Ecol. 15:1044–1045.

Neuman, C. N., R. J. Safran, and I. J. Lovette. 2007. Male tail streamer length does not predict apparent or genetic reproductive success in North American barn swallows *Hirundo rustica erythrogaster*. J. Avian Biol. 38:28–36.

Nosil, P. 2012. Ecological speciation. Oxford Univ. Press, New York.

Panhuis, T. M., R. Butlin, M. Zuk, and T. Tregenza. 2001. Sexual selection and speciation. Trends Ecol. Evol. 16:364–371.

Price, T. 1998. Sexual selection and natural selection in bird speciation. Philos. Trans. R. Soc. B Biol. Sci. 353:251–260.

Ritchie, M. G. 2007. Sexual selection and speciation. Annu. Rev. Ecol. Evol. Syst. 38:79–102.

Rodríguez, R. L., J. W. Boughman, D. A. Gray, E. A. Hebets, G. Höbel, and L. B. Symes. 2013. Diversification under sexual selection: the relative roles of mate preference strength and the degree of divergence in mate preferences. Ecol. Lett. 16:964–974.

Sætre, G., T. Moum, S. Bureš, M. Král, M. Adamjank, and J. Moreno. 1997. A sexually selected character displacement in flycatchers reinforces premating isolation. Nature 387:589–592.

Safran, R. J., and K. J. McGraw. 2004. Plumage coloration, not length or symmetry of tail-streamers, is a sexually selected trait in North American barn swallows. Behav. Ecol. 15:455–461.

Safran, R. J., C. R. Neuman, K. J. McGraw, and I. J. Lovette. 2005. Dynamic paternity allocation as a function of male plumage color in barn swallows. Science 309:2210–2212.

Safran, R. J., J. S. Adelman, K. J. McGraw, and M. Hau. 2008. Sexual signal exaggeration affects physiological state in male barn swallows. Curr. Biol. 18:461–462.

Safran, R. J., E. S. C. Scordato, L. B. Symes, R. L. Rodríguez, and T. C. Mendelson. 2013. Contributions of natural and sexual selection to the evolution of premating reproductive isolation: A research agenda. Trends Ecol. Evol. 28:643–650.

Safran R. J., E. S. C. Scordato, M. R. Wilkins, J. K. Hubbard, B. R. Jenkins, T. Albrecht, S. M. Flaxman, H. Karaardiç, Y. Vortman, A. Lotem, et al. 2016, Genome-wide differentiation in closely related populations: the roles of selection and geographic isolation. Molecular Ecology 25:3865–3883. doi: 10.1111/mec.13740

Saino, N., C. R. Primmer, H. Ellegren, and A. P. Møller. 1997. An experimental study of paternity and tail ornamentation in the barn swallow (*Hirundo rustica*). Evolution 51:562–570.

Saino, N., R. Martinelli, and M. Romano. 2007. Ecological and phenological covariates of offspring sex ratio in barn swallows. Evol. Ecol. 22(5):659–674.

Schluter, D. 2001. Ecology and the origin of species. Trends Ecol. Evol. 16:372–380.

———. 2009. Evidence for ecological speciation and its alternative. Science 323:737–741.

Schluter, D., and T. D. Price. 1993. Honesty, perception, and population divergence in sexually selected traits. Proc. R. Soc. Lond. B Biol. Sci. 253:117–122.

Scordato, E. S., and R. J. Safran. 2014. Geographic variation in sexual selection and implications for speciation in the Barn Swallow. Avian Res. 5:8–21.

Seddon, N., C. A. B., Botero, P. O., Dunn, H., MacGregor, D. R., Rubenstein, J. A., Tobias, J. A. C. Uy, J.,Weir, L.W., Whittingham, and R. J, Safran. 2013. Sexual selection accelerates signal evolution during speciation in birds. Proc. Biol. Sci. 280:20131065.

Seehausen, O., Y. Terai, I. S. Magalhaes, K. L. Carleton, H. D. J. Mrosso, R. Miyagi, et al. 2008. Speciation through sensory drive in cichlid fish. Nature 455:620–626.

Shafer, A. B. A., and J. B. W. Wolf. 2013. Widespread evidence for incipient ecological speciation: A meta-analysis of isolation-by-ecology. Ecol. Lett. 16:940–950.

Smith, H. G., and R. Montgomerie. 1991. Sexual selection and the tail ornaments of North American barn swallows. Behav. Ecol. Sociobiol. 28:195–201.

Smith, H. G., R. Montgomerie, T. Pdldmaa, B. N. White, and P. T. Boag. 1991. DNA fingerprinting reveals relation between tail ornaments and cuckoldry in barn swallows, *Hirundo rustica*. Behav. Ecol. 2:90–98.

Svensson, E. I., and J. T.Waller. 2013. Ecology and sexual selection: evolution of wing pigmentation in calopterygid damselflies in relation to latitude, sexual dimorphism, and speciation. Am. Nat. 182:E174–E195.

Svensson, E. I., F. Eroukhmanoff, and M. Friberg. 2006. Effects of natural and sexual selection on adaptive population divergence and premating isolation in a damselfly. Evolution 60:1242–1253.

Turner, A. 2006. The Barn Swallow. Poyser, London.

Turner, A., and C. Rose. 1989. Swallows and martins: An identification guide and handbook. Houghton Mifflin, Boston, MA.

Uy, J. A. C., and G. Borgia. 2000. Sexual selection drives rapid divergence in bowerbird display traits. Evolution 54:273–278.

van Doorn, G. S., P. Edelaar, and F. J.Weissing. 2009. On the origin of species by natural and sexual selection. Science 326:1704–1707.

Vitousek, M. N., R. A. Stewart, and R. J. Safran. 2013. Female plumage colour influences seasonal oxidative damage and testosterone profiles in a songbird. Biol. Lett. 9:20130539.

Vortman, Y., A. Lotem, R. Dor, I. J. Lovette, and R. J. Safran. 2011. The sexual signals of the East-Mediterranean barn swallow: A different swallowtale. Behav. Ecol. 22:1344–1352.

Vortman, Y., A. Lotem, R. Dor, I. Lovette, and R. J. Safran. 2013. Multiple sexual signals and behavioral reproductive isolation in a diverging population. Am. Nat. 182:514–523.

Vortman, Y., R. J. Safran, T. R. Brodetzki, R. Dor, and A. Lotem. 2015. Expression of multiple sexual signals by fathers and sons in the East-Mediterranean barn swallow: Are advertising strategies heritable? PLOS One 10:e0118054.

Wagner, C. E., L. J. Harmon, and O. Seehausen. 2012. Ecological opportunity and sexual selection together predict adaptive radiation. Nature 487:366–369.

West-Eberhard, M. J. 1983. Sexual selection, social competition, and speciation. Q. Rev. Biol. 58:155–183.

Whitlock, M. C., and D. Schluter. 2009. The analysis of biological data. Roberts and Company Publishers, Greenwood Village, CO.

Wilkins, M. R., M. B. Joseph, J. K. Hubbard, and R. J. Safran. 2015. Multimodal signaling in the North American barn swallow: Influences of intra- and intersexual selection on signal evolution. Proc. R Soc. Lond. Biol. Sci. 282:20151574.

Supplemental Information

Phenotypic evolution through divergent sexual selection

Safran, Rebecca J; Vortman, Yoni; Jenkins, Brittany R;
Hubbard, Joanna K; Wilkins, Matthew R ; Bradley, Rachel J ; Lotem, Arnon

Genetic paternity analyses. DNA was extracted from embryos, and blood samples collected from nestlings and adults. All blood samples were stored in lysis buffer and then DNA was extracted using DNeasy Blood & Tissue Extraction kits (Qiagen, Maryland, U.S.A). Polymerase chain reaction (PCR) was utilized to amplify six previously developed microsatellite loci (Hanotte et al., 1994; McDonald & Potts,1994, Bensch, Price, & Kohn, 1997; Tsyusko et al., 2007). Reaction conditions for pooled Escu6, Ltr6, Hir20, and Hir11 primers consisted of a 10 µl solution with 50-100 ng DNA, 0.12 mM of each labeled forward primer, 0.12 mM of each reverse primer, 200 µM each dNTP, 3.25 mM $MgCl_2$, 1x PCR Buffer, 0.15 units Taq polymerase (New England Biolabs, Massachusetts, U.S.A.), and were amplified with the following protocol: initial denaturation step of 94°C for 1 minute, followed by 10 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 45 s, with an additional 25 cycles starting at 87°C for 30 s instead of 94°C, and completed with a final extension at 72 °C for 3 min. The Pocc6 reaction was modified from the above conditions by using 1.25 mM $MgCl_2$, and the above conditions were altered for the Hir19 reaction with 3 mM $MgCl_2$, and 0.2 mM each forward and reverse primer. The PCR amplification protocol for Pocc6 and Hir19 was similar to the pooled loci protocol with the exception that 60°C was used for the annealing temperature. Amplified PCR products containing the fluorescently-labeled forward primer were detected using an ABI3730 DNA analyzer (ABI, Inc.).

Using GeneMapper software (v4.0, Applied Biosystems), we assigned genotypes for 265 nestlings, 292 embryos and 46 pairs of social parents at all six loci (Colorado) and for 157 nestlings, 183 embryos and 39 pairs of social parents at all six loci plus an additional locus (Vortman, Lotem, Dor, Lovette, & Safran, 2013). Genotypes from adults and offspring were incorporated into a paternity analysis using CERVUS software (v2.0) to calculate exclusion probabilities for assessing parentage, with a combined first-parent exclusion probability of 0.922 for all six loci. Paternity exclusion was conducted using similar parameters described in Neuman et al (Neuman et al., 2007). Briefly, we considered young as extra-pair if we detected two or more alleles across all loci examined that did not match the social father.

We used three complementary measures of paternity to assess a male's reproductive performance before and after the phenotype manipulation: number of own young (number of young a male sired in his nest), proportion of own young (number of own young a male sired in his nest relative to the total number of offspring in his nest), and number of extra-pair young (number of extra-pair young a male was caring for in his own nest).

Table S1. Description of six phenotype treatments applied to males in two populations. In all cases, a male's phenotype was manipulated after a simulated predation event (egg removal) during the first breeding bout such that paternity outcomes could be examined in two successive bouts within one breeding season: before and after the phenotype manipulation.

Treatment	Description	Israel (n)	US (n)
dark	Color darkened within the natural range of variation using the application of a non-toxic permanent ink marker	8	7
short	Streamer length reduced 1.5 SD of population mean	6	7
long	Streamer length elongated by 1.5 SD of the population mean	6	6
dark + short	Color darkened and streamer reduced	6	8
dark + long	Color darkened and streamer elongated	6	9
control	Phenotype not manipulated	7	9

Table S2: Description of nest outcomes as a function of experimental treatment, Colorado population.

Treatment	Original sample size	Final sample size	Failure due to				
			Predation on nestlings	Nest fell	Eggs did not hatch	Genotype error	Did not reinitiate replacement clutch
dark	12	7			1	1	3
short	10	7			1		2
long	11	6	2	1			2
dark + short	12	8		1	1	1	1
dark + long	10	9	1				
control	11	9	2				

Table S3. Post-hoc comparisons following a significant treatment effect (Colorado experiment).

Colorado: changes in proportion of young between broods

T-test differences in treatment groups. Confidence limits (CL) are used to assess the significance of differences given multiple comparisons.

Level	Compared to	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
dark	**control**	**0.54**	**0.18**	**0.17**	**0.91**	**0.01**
short	**control**	**0.51**	**0.18**	**0.19**	**0.88**	**0.01**
dark	**long**	**0.43**	**0.20**	**0.02**	**0.83**	**0.03**
dark	**dark + long**	**0.41**	**0.18**	**0.04**	**0.77**	**0.02**
short	**long**	**0.39**	**0.20**	**-0.01**	**0.80**	**0.05**
short	**dark + long**	**0.38**	**0.18**	**0.01**	**0.75**	**0.04**
dark	dark + short	0.28	0.18	-0.08	0.66	0.13
short	dark + short	0.26	0.183	-0.12	0.64	0.17
dark + short	control	0.25	0.17	-0.09	0.61	0.15
dark + short	long	0.14	0.19	-0.25	0.53	0.47
dark + long	control	0.13	0.17	-0.21	0.47	0.43
dark + short	dark + long	0.12	0.17	-0.23	0.47	0.49
long	control	0.11	0.19	-0.26	0.50	0.54
dark	short	0.03	0.19	-0.36	0.42	0.88
dark + long	long	0.01	0.199	-0.36	0.40	0.92

Dunnett's test: treatment comparisons to control group.

Level	Compared to	Difference	Std Err Dif	Z	p-Value
dark	**control**	**20.79**	**6.68**	**3.11**	**0.01**
short	**control**	**18.07**	**6.68**	**2.70**	**0.03**
dark + short	control	11.35	6.44	1.76	0.39
dark + long	control	7.33	6.25	1.17	1.00
long	control	6.55	6.99	0.94	1.00

Colorado: changes in number of extrapair young between broods

T-test differences in treatment groups. Confidence limits (CL) are used to assess the significance of differences given multiple comparisons.

Level	Compared to	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
control	**dark**	**2.60**	**0.76**	**1.05**	**4.15**	**0.01**
dark + long	**dark**	**2.15**	**0.76**	**0.60**	**3.70**	**0.01**
long	**dark**	**2.04**	**0.84**	**0.33**	**3.75**	**0.02**
control	**short**	**2.03**	**0.76**	**0.48**	**3.58**	**0.01**
dark + long	**short**	**1.58**	**0.76**	**0.03**	**3.13**	**0.04**
long	short	1.47	0.84	-0.23	3.18	0.08
dark + short	dark	1.33	0.78	-0.25	2.93	0.09
control	dark + short	1.26	0.73	-0.23	2.75	0.09
dark + long	dark + short	0.81	0.73	-0.67	2.31	0.27
dark + short	short	0.76	0.78	-0.82	2.35	0.33
long	dark + short	0.70	0.82	-0.95	2.36	0.39
short	dark	0.57	0.81	-1.07	2.21	0.48
control	long	0.55	0.80	-1.06	2.17	0.49
control	dark + long	0.44	0.71	-1.00	1.89	0.53
dark + long	long	0.11	0.80	-1.50	1.73	0.89

Dunnett's test: treatment comparisons to control group.

Level	Compared to	Difference	Std Err Dif	Z	p-Value
dark +long	control	-4.28	6.07	-0.70	1.00
long	control	-4.33	6.79	-0.63	1.00
dark + short	control	-9.62	6.26	-1.53	0.62
short	control	-15.28	6.49	-2.35	0.09
dark	**control**	**-20.07**	**6.49**	**-3.08**	**0.01**

Figure S1. Frequency distributions of paternity outcomes for comparative experiment.

Frequency distributions for three paternity outcome variables for the entire sample of 85 individuals across six different treatments and two breeding locations. Each figure panel below portrays the frequency distributions for a) change in the number of own young, b) change in the number of extra-pair young, and c) change in proportion of paternity between two successive breeding bouts before and after the same set of six phenotype manipulation treatments.
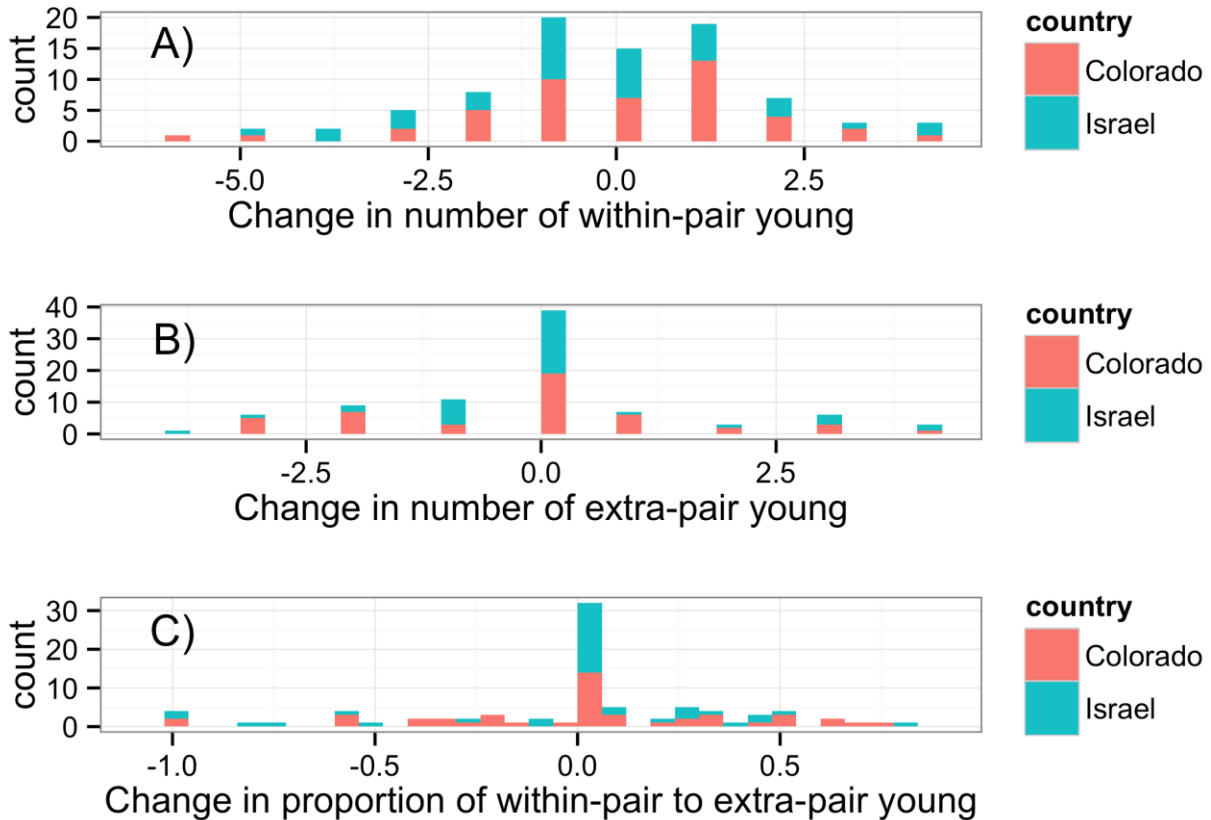
Table S4. Post-hoc comparisons following a significant treatment x country effect (Colorado compared to Israel).

Colorado x Israel: Changes in number of own young

T-test differences in treatment groups across countries. Confidence limits (CL) are used to assess the significance of differences given multiple comparisons.

Level	Compared to	Difference	SE	Lower CL	Upper CL	p-Value
short, Israel	short, Colorado	1.86	1.07	-0.27	3.99	0.08
dark + long, Israel	dark + long, Colorado,	1.72	1.02	-0.29	3.74	0.09
control, Israel	control, Colorado	1.68	0.97	-0.25	3.61	0.08
dark, Israel	dark, Colorado	1.50	0.99	-0.48	3.48	0.13
dark + short, Israel	dark + short, Colorado	1.37	1.04	-0.69	3.44	0.18
long, Israel	long, Colorado	0.67	1.11	-1.54	2.87	0.55

Colorado x Israel: Changes in proportion of own young

T-test differences in treatment groups across countries. Confidence limits (CL) are used to assess the significance of differences given multiple comparisons.

Level	Compared to	Difference	SE	Lower CL	Upper CL	p-Value
dark + short, Israel	**dark + short, Colorado**	**0.45**	**0.19**	**0.07**	**0.83**	**0.01**
dark + long, Israel	**dark + long, Colorado**	**0.39**	**0.18**	**0.03**	**0.76**	**0.03**
control,Israel	control,Colorado	0.29	0.18	-0.06	0.64	0.10
short,Israel	short, Colorado	0.27	0.19	-0.12	0.66	0.16
dark,Israel	dark,Colorado	0.25	0.18	-0.11	0.61	0.17
long,Israel	long,Colorado	0.02	0.20	-0.38	0.43	0.90

Colorado x Israel: Changes in number of extra pair young

T-test differences in treatment groups across countries. Confidence limits (CL) are used to assess the significance of differences given multiple comparisons.

Level	Compared to	Difference	SE	Lower CL	Upper CL	p-Value
dark + short, Israel	**dark + short, Colorado**	**2.21**	**0.80**	**0.61**	**3.80**	**0.01**
dark + long, Israel	**dark + long, Colorado**	**1.94**	**0.78**	**0.38**	**3.50**	**0.01**
dark,Israel	dark,Colorado	1.46	0.76	-0.06	2.99	0.05
short,Israel	short,Colorado	0.98	0.82	-0.66	2.62	0.23
control,Israel	control,Colorado	1.17	0.75	-0.31	2.66	0.11
long,Israel	long,Colorado	0.33	0.85	-1.37	2.04	0.69

Experimental outcomes controlling for differences in the control group between Israel and Colorado. For a conservative estimate of treatment differences by country, we adjusted each treatment outcome to a baseline, correcting for control group outcomes for each paternity measure (Table S4). The results (Table S5) indicate significant treatment x country interactions, as reported in Table 1.

Table S5. Differences in the control group outcomes between Colorado and Israel that were used to scale paternity outcomes within each treatment and each country.

Country	Control treatment group average changes in number young	Control treatment group average changes in proportion own young	Control treatment group average changes in number of extra pair young
Colorado	-1.11	-0.24	0.89
Israel	0.57	0.05	-0.28

Table S5. Population differences in changes in paternity as a function of phenotype manipulation in 85 breeding pairs, controlling for control group differences (n = 46 from Colorado and n = 39 from Israel).

Factor	*F* (df)	*P*
Changes in a male's proportion of paternity in successive reproductive bouts		
Country	20.89 (1,73)	< 0.01
Treatment	3.15 (5,73)	< 0.02
Country x Treatment	4.17 (5,73)	< 0.01
Changes in a male's number of own offspring in successive reproductive bouts		
Country	22.85 (1,73)	< 0.01
Treatment	0.92 (5,73)	> 0.47
Country x Treatment	2.59 (5,73)	< 0.03
Changes in a male's number of extra-pair offspring in a focal male's nest in successive reproductive bouts		
Country	19.81 (1,73)	< 0.01
Treatment	2.92 (5,73)	< 0.02
Country x Treatment	3.31 (5,73)	< 0.01

Supplemental References

Bensch, S., Price, T., & Kohn, J. (1997). Isolation and characterization of microsatellite loci in a Phylloscopus warbler. *Molecular Ecology*, *6*(1), 91–2.

Hanotte, O., Zanon, C., Pugh, A., Greig, C., Dixon, A., & Burke, T. (1994). Isolation and characterization of microsatellite loci in a passerine bird: the reed bunting Emberiza schoeniclus. *Molecular Ecology*, *3*(5), 529–530.

McDonald, D. B., & Potts, W. K. (1994). Cooperative display and relatedness among males in a lek-mating bird. *Science*, *266*(5187), 1030–2.

Neuman, C.R, Safran, R.J., & Lovette, I.J. (2007). Male tail streamer length does not predict apparent or genetic reproductive success in North American barn swallows Hirundo rustica erythrogaster. *Journal of Avian Biology*, *38*(1), 28–36. doi:10.1111/j.2007.0908-8857.03713.x

Tsyusko, O. V., Peters, M. B., Hagen, C., Tuberville, T. D., Mousseau, T. a., Møller, A. P., & Glenn, T. C. (2007). Microsatellite markers isolated from barn swallows (Hirundo rustica). *Molecular Ecology Notes*, *7*(5), 833–835. doi:10.1111/j.1471-8286.2007.01719.x

Vortman, Y., Lotem, A., Dor, R., Lovette, I., & Safran, R. J. (2013). Multiple sexual signals and behavioral reproductive isolation in a diverging population. *The American Naturalist*, *182*(4), 514–23. doi:10.1086/671908