

Bill color, not badge size, indicates testosterone-related information in house sparrows

Silke Laucht · Bart Kempenaers · James Dale

Received: 8 January 2010 / Revised: 4 March 2010 / Accepted: 14 April 2010 / Published online: 29 May 2010
© Springer-Verlag 2010

Abstract The honesty of ornamental signals of quality is often argued to be enforced via costs associated with testosterone. It is still poorly understood, however, how seasonal variation of testosterone within individuals is related to the timing and extent of ornament development. Here, we studied inter- and intra-individual variability of plasma testosterone levels in a population of 150 captive male house sparrows (*Passer domesticus*) through the course of a full year. We further analyzed the relationship between plasma testosterone levels and two sexually dimorphic ornaments: badge size and bill coloration. Also, because of a known negative relation between molt and circulating testosterone levels, we analyzed the relationship between ornamentation and molt status during the fall. We found that testosterone levels increased towards the breeding season and decreased before the onset of annual molt. However, within individuals, relative testosterone titers demonstrated low repeatability between seasons. Plasma testosterone levels were not correlated with badge size in any season but were correlated strongly with bill coloration during all periods, except the breeding season when variation in bill color was low. Finally, we found that bill coloration strongly correlated with molt status during

fall. Our results indicate that bill coloration, not badge size, is the best ornamental indicator of a “running average” of male testosterone in house sparrows and therefore the best potential indicator of qualities and/or behavioral strategies associated with testosterone.

Keywords Badge size · Bill color · Honest signaling · Melanin · Strategy signals · Testosterone

Introduction

Current interest in the relationship between the steroid hormone testosterone (T) and sexually selected ornamentation is high because testosterone is well known to be responsible for the elaboration of many sexually selected ornaments (reviewed in, e.g., Candolin 2003; Roberts et al. 2004). The honesty of ornamental traits which specifically signal quality is argued to be enforced through inescapable costs associated with signaling (Grafen 1990; Johnstone and Norris 1993; Veiga 1993; Buchanan et al. 2003; Peters et al. 2004). Potential costs hypothesized to be associated with quality signals are testosterone-related suppression of the immune system (e.g., the Immunocompetence Handicap Hypothesis (ICHH) (Folstad and Karter 1992)) or testosterone-related depression of resistance to oxidative stress (e.g., the Oxidation Handicap Hypothesis (OHH) (Alonso-Alvarez et al. 2007)). Another form of signaling cost (often suggested for melanin-based ornaments) is social enforcement via costs created by aggressive interactions with, or testing by, conspecifics (i.e., the badge of status hypothesis (Maynard Smith and Harper 1988; Jawor and Breitwisch 2003; Tibbetts and Dale 2004)). Testosterone-dependent ornaments can thus be expected to provide information to conspecifics about qualities associ-

Communicated by: K. McGraw

S. Laucht (✉) · B. Kempenaers · J. Dale
Department of Behavioural Ecology and Evolutionary Genetics,
Max Planck Institute for Ornithology,
Eberhard-Gwinner-Straße, Haus 5,
82319 Seewiesen, Germany
e-mail: laucht@orn.mpg.de

J. Dale
Institute of Natural Sciences, Massey University,
Private Bag 102 904,
North Shore 0745 Auckland, New Zealand

ated with either physiological characteristics such as immunosuppression and/or oxidative stress or behavioral characteristics such as aggression (Folstad and Karter 1992; Johnstone and Norris 1993; Alonso-Alvarez et al. 2007; McGraw and Ardia 2007).

Since testosterone dramatically affects both behavior and physiology (reviewed in Wingfield et al. 1990; Ketterson and Nolan 1992; Adkins-Regan 2005; Hau 2007), maintaining elevated testosterone levels all year round can be costly (e.g., through increased risk of aggression-related injury or through increased metabolic costs (see Wingfield et al. 2001)). Temperate-zone bird species, for example, show dramatic fluctuations in plasma testosterone levels over the course of the year. They rise towards the breeding season to levels needed for the physiological changes and behaviors associated with breeding and drop to baseline levels with the onset of prebasic (or post-nuptial) molt in early fall when birds are more vulnerable and less aggressive (Humphrey and Parkes 1959; Wingfield et al. 1990; Hahn et al. 1992). Even during periods of elevated testosterone, T levels are generally kept at a certain breeding baseline and are then modulated on a short-term basis as a result of social interactions (e.g., aggression associated with reproductive behavior (Wingfield 1984, 1985; Wingfield et al. 1990)). Overall, it is generally argued that testosterone levels are kept at an optimum that is well balanced with the various behavioral and physiological costs of maintaining them (Folstad and Karter 1992; Wingfield et al. 2001; Adkins-Regan 2005; Hau 2007).

To date, researchers have resolved many examples of an endocrine basis to variation in color-based ornaments occurring in birds (reviewed in Hill and McGraw 2006), monkeys (e.g., Setchell et al. 2008; Clough et al. 2009; Lewis 2009), fish (e.g., Dijkstra et al. 2007; Kurtz et al. 2007), and lizards (e.g., Thompson and Moore 1991; Salvador et al. 1996; Calisi and Hews 2007; Huyghe et al. 2009) and comprised of all of the three major mechanisms of coloration—structural (e.g., Peters et al. 2006), carotenoid-based (e.g., Hill 2002), and melanin-based (reviewed in Jawor and Breitwisch 2003; Bókony et al. 2008)—in ornaments consisting of feathers (e.g., Evans et al. 2000; Safran and McGraw 2004), integuments (e.g., Verhulst et al. 1999; Mougeot et al. 2004; Blas et al. 2006), or bill (e.g., Keck 1933; Mundinger 1972; Murphy et al. 2009). Among these studies, the house sparrow (*Passer domesticus*) has become one of the model organisms for the study of testosterone and melanin-based ornaments.

Male house sparrows have at least two noteworthy sexually dimorphic ornaments: the black breast bib or the “badge” and the black bill. Many studies have demonstrated that the size of the badge relates to social status, age, and variation in sexual behavior (Møller 1987, 1990; Veiga 1993; Liker and Barta 2001; Vaclav and Hoi 2002;

McGraw et al. 2003; Nakagawa et al. 2007; Morrison et al. 2008) and is subject to natural and sexual selection (Jensen et al. 2008). Furthermore, three studies found that badge size correlated positively with plasma testosterone levels around annual (prebasic) molt (Evans et al. 2000; Buchanan et al. 2001; Gonzalez et al. 2001) when the new badge is formed. In contrast, the signaling function of the bill is still a mystery. Male bill color changes between seasons from a pale horn color in the non-breeding season to a blackish dark color in the breeding season (Witschi and Woods 1936), and the color change is due to pigmentation with eumelanins and is known to be mediated by male hormones (testosterone) (Keck 1933; Witschi 1936; Pfeiffer et al. 1944; Haase 1975; Donham et al. 1982). Although Václav (2006) suggested the possibility of an effect of condition on bill color, so far there have been no tests of this.

Even though there appear to be clear relationships between testosterone levels and these two ornaments in house sparrows, the relationship between within-individual variation in testosterone levels and the timing and extent of ornament development is still very poorly understood. For example, feather ornaments molted during times of lowest testosterone levels (such as the badge in the house sparrow) can only remain honest indicators of testosterone levels during the breeding season if testosterone levels during different periods of the year are correlated. This, however, is surprisingly poorly studied (Kempnaers et al. 2008). To the best of our knowledge, only one study has examined the relationship between breeding and post-breeding testosterone levels in house sparrows (Buchanan et al. 2003). To address this gap, here, we studied plasma testosterone levels, badge size, and bill color in a large captive population of male house sparrows over the course of 1 year. The objectives of the study were to test the following four non-exclusive predictions.

1. Within individuals, plasma testosterone levels should be correlated between seasons. This was found by the only study comparing two different seasons (Buchanan et al. 2003) despite it being a fundamental assumption of theories trying to explain the testosterone-related honesty of signals that are developed at different times than when they are actually used (e.g., the ICHH (Folstad and Karter 1992; Alonso-Alvarez et al. 2007) and the OHH (Alonso-Alvarez et al. 2007)).
2. Badge size should be positively correlated with testosterone levels. As previously stated, badge size signals social status and sexual behavior (e.g., communal displays), both of which are testosterone-related traits. Additionally, such a relation was previously found in correlational approaches and after artificial increases of testosterone levels during molt (Evans et al. 2000; Buchanan et al. 2001).

3. Bill color should be positively correlated with testosterone levels. As described above, seasonal changes in bill color are mediated by changes in testosterone levels, and therefore we predicted that the extent of coloration in this dynamic ornament should be related to current testosterone levels.
4. Bill color should be correlated with molt status. In other species, molt status is strongly negatively correlated with plasma testosterone levels (Schleussner 1990; Hahn et al. 1992; Nolan et al. 1992). Therefore, because we expected a positive relation between bill color and testosterone levels, we also predicted a relation between bill color and molt status.

Although our study is correlational and we do not test the signaling role of the bill directly, we nevertheless assume here that bill coloration is an evolved signal (i.e., that it can influence decision making in receivers), because it is a conspicuous, sexually dimorphic trait which has some apparent design in terms of the complex physiological processes responsible for its dynamic expression.

Materials and methods

Study population

We studied a population of 150 captive male house sparrows held at the Max Planck Institute for Ornithology, Seewiesen, Germany. All males were after-hatching year birds. They were either caught in rural areas in Bavaria, Germany (under license: permit nr. 55.1-8642.3-3-2006 of the “Regierung Oberbayern”, with several extensions) and held in captivity for at least 8 months ($n=136$) or born in captivity ($n=14$; exclusion did not qualitatively change the results). From July 2006 until July 2007, individuals were kept in all-male groups of five or six in aviaries of size $1.2 \times 2.0 \times 4.0$ m. After July 2007, we kept them in the same aviaries in groups of nine or ten (note that we did not measure testosterone in the males after we changed the group size). At all times, the birds had ad libitum access to food (wild seed mix for forest birds (Waldvogelfutter: RKW Sued, Universal Kraftfutterwerk, Kehl, Germany), sunflower seeds, crushed corn and wheat, oats, chicken starter, soybean meal extract, and mineral mix for birds), drinking and bathing water, and sand. The light–dark cycle and temperatures in the aviaries were close to natural conditions, as the aviaries were semi-outdoor with one side enclosed only by chicken wire. Although our study is on a captive population of house sparrows, we have confidence that our results reflect patterns present under wild populations because our sparrows were mostly wild-caught individuals that were housed in large semi-outdoor aviaries

that were grouped together in a barn-like building that was very similar to the actual barns where we caught the sparrows. Moreover, our house sparrows readily bred under these conditions, indicating that conditions were highly favorable for natural behavior. Finally, our study replicates research by other groups working on captive sparrows, and so our results are highly comparable to theirs.

During five periods throughout the course of the year (Oct./Nov. 2006, Jan. 2007, March 2007, June 2007, and Sept./Oct. 2007), we caught all individuals and took biometric measurements, standardized photographs of the bill and badge, and blood samples (except in Sept./Oct. 2007, when no blood samples were taken). In fall 2007, we additionally scored molt status. During each blood sampling period, conducted between 0700 and 1000 hours and between 1300 and 1500 hours, we took 150–200 μ l of blood from the wing vein within 15 min after first starting to catch the birds. The time passed since first starting to catch birds did not have an influence on T levels (linear mixed effect model (lme): $t=1.13$, $p=0.26$, $n=551$ with season, day time, and bird ID as random effects). We collected the blood in 75-mm Na-heparinized micro haematocrit capillaries and centrifuged it at 13,000 rpm for 3 min to separate the plasma. Plasma was stored at -80°C .

Determination of plasma T levels

Frozen plasma samples were sent to the endocrine laboratory of the Leibniz Institute for Zoo and Wildlife Research in Berlin, Germany, where testosterone levels were determined by enzyme immunoassays (for details on the methods, see Roelants et al. 2002). The inter-assay coefficient of variation (CV) for the enzyme immunoassay was 12.3% and the intra-assay CV was 9.0%. Additionally, to calculate the true repeatability (intra-class correlation coefficient) of measuring serum T levels, we split the plasma samples of several males into duplicates right after centrifuging. Across the whole year, the repeatability of these plasma T estimates was $R=0.967 \pm 0.006$ (SE) ($p < 0.001$, $n=2 \times 122$). Note that this estimate includes additional non-assay sources of variation because of the immediate separation after centrifugation. It thus gives a conservative estimate of measurement error in plasma T levels. We assumed that all data points with a value of zero (29 out of 579) were actually below the detection limit (for each assay slightly different, but around 20 pg/ml) and thus assigned them the lowest value measured (15 pg/ml). Excluding those individuals entirely from the analysis did not qualitatively change the results. T levels are reported in picogram per milliliter. Testosterone concentrations were natural log-transformed to achieve a normal distribution, in order to fit standard least-squares models.

Determination of bill color

Immediately after blood samples were taken, we took two standardized photographs of each bird's bill. We used a Canon Power Shot S2 IS camera and took pictures at the highest resolution with flash. All males were held the same way (presenting the right side of the head and bill to the camera) in front of a gray card and color standard background at the same distance from the camera.

Digital photograph processing software written into R 2.4.0 (R Development Core Team 2006) was used to collect values of bill “brightness” as measured in the standard HSB (i.e., hue, saturation, and brightness) color space. Note that, in this color scoring scheme, “brightness” is an indicator of how dark or light a color is and correlates closely with mean total reflectance (Montgomerie 2006; also see Fig. 1). SL measured the brightness of the individual pixels located at five randomly chosen positions each on the upper bill, lower bill, and the gray background around the bill (used as a brightness standard). To standardize our bill brightness measurements between photos, we calculated the overall mean gray card brightness of all photos of each season, we then determined the deviation of gray card brightness of a focal photo from the overall mean, and subtracted this deviation from mean bill brightness for each picture. This standardization renders a bill brightness score for each male that both compensates for any minor differences in overall brightness between photos and that also keeps our brightness variable as an actual color measurement (rather than a difference). For analyses, we used the mean of the standardized upper and lower bill brightness from both pictures. As expected for a trait with considerable phenotypic variation, these measurements were highly repeatable within individuals (repeatability (Lessells and Boag 1987): $R=0.949\pm 0.004$ (SE), $p<0.001$, $n=2\times 581$ for two pic-

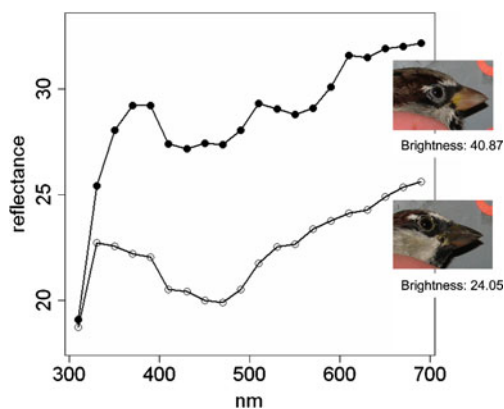


Fig. 1 Full spectrogram of two differently colored male house sparrow bills. Our measures of bill brightness as measured from digital photographs correspond closely with total reflectance as measured with spectrophotometry. *Open circles* represent a medium dark bill color, *closed circles* a pale bill color

tures). For repeatability measures between two consecutive years, we took additional photographs of the bills in January 2008.

To illustrate the correspondence between bill brightness (measured with digital photographs) and total reflectance, we measured the bills of two differently colored males with a hand-held spectrometer (Avantes, AvaSpec-2048, Eerbeek, The Netherlands) with a deuterium–halogen light source (Avantes, Ava-Light-D(H)-S). We measured five points each on the bill. In Fig. 1, we present averages over each 20-nm spectral range averaging also the five measurements and clearly demonstrate that total reflectance is dramatically different in bills with different brightness values.

Determination of badge size

We took four pictures of the birds' breast badges during each season. For each picture, we held the birds ventrally such that the throat and badge was stretched and presented to the camera. Between each photograph of the badge, the bird was arranged into a different position, before rearranging it into the badge-exposed position. SL measured the size of the badge from the photographs by encircling and measuring the area of the melanized badge in pixels using the program ImageJ 1.36b (Abramoff et al. 2004). For standardization, we divided this area by a standard area present in each photograph and measured in the same way and then converted the result into square centimeters. For analyses, we used the average of all four pictures for each bird. In males with white tips to their badge feathers (Møller and Erritzøe 1992), badge size was estimated by outlining as best as possible the area occupied by any apparent melanized feathers underlying the white tips. The measurements were highly repeatable within individuals ($R=0.943$, estimated according to Falconer and Mackay (1996) from repeatability of single pictures).

Determination of molt stage

From September 4th to 14th, we scored the molt of 145 males. On October 9th and October 18th, we re-scored 30 males each. We determined molt status according to Ginn and Melville (2000, p. 28, Fig. 7b) for primaries. When wing feathers were asynchronously molted on the right wing, we used the scores of the left wing. We scored old feathers as 0, fully grown new feathers as 5, and growing feathers as 1 to 4 according to their length. For analyses, we added up all scores of the single primaries of one bird (hereafter BTO (British Trust for Ornithology) score). Additionally, we used a binary score, where a feather was simply determined as old (= 0) or molting and/or new (= 1). We then added up these scores (hereafter referred to as binary score). As the BTO score and the binary score

produced very similar results, we only present the results of the BTO score in the “Results” section.

Statistical analyses

We performed all statistical analyses using R 2.6.2 (R Development Core Team 2008; packages: ape, effects, nlme, RODBC, survival) at the significance level $\alpha=0.05$. For overall analyses, we used linear mixed effect models (lmes) using individual ID as a random factor to account for repeated sampling. When using data from single seasons where each individual was represented only once, we used linear models (lms). Before analysis, we checked for normal distribution; after analysis, we assessed whether the assumption of lms and of lmes, respectively, on the within-group errors and on the random effects were violated. In the analyses reported here, we did not remove any outliers (max. four per model) because we had large sample sizes and no a priori expectations of what constitutes normal ranges for our biological variables. In addition, all analyses yielded qualitatively similar results with or without outliers included.

For analyses of relationships between bill color and testosterone levels, we used the data of the four periods when plasma samples were taken. In addition, in the analysis of the relationship between badge size and testosterone or condition, we used the average of the March and June scores as a male’s badge phenotype because at these times the white feather edges have mostly abraded off (Møller and Erritzøe 1992). For analyses of molt status, we used the data of fall 2007. For analyses with body condition, we fitted three separate models: the first used the residuals of body mass regressed onto tarsus length, the second used body mass alone, and the third one used tarsus length alone. We calculated date as days passed since the beginning of measurements (Sept 25th ‘06 for analyses of testosterone, badge size, and bill color; Sept 4th ‘07 for analyses of molt and bill color). Therefore, date starts in fall (during molt) and ends in June (peak breeding season), and we expect close to linear relationships between date and T levels or bill color. We additionally accounted for day time in all models including T or used residuals of T in relation to date and day time (see indication in “Results”). Note that the differences in sample sizes are due to missing data points for single individuals because of sample loss during centrifugation or natural death of a few individuals.

Results

Dynamics in plasma testosterone levels

Plasma T levels changed dramatically throughout the course of the year (Fig. 2a). We found a significant positive

relationship between T and date (lme: $p<0.0001$, $t_{424}=24.67$, random effect—bird ID; after accounting for day time: $p<0.0001$, $t_{414}=23.58$). However, within individuals, T levels were only weakly correlated between seasons (see lines in Fig. 2a connecting measurements of the same individuals): T levels were significantly correlated only between Sept/Oct and January and between January and March (Table 1).

We found no significant relationship between T and body condition when accounting for date: Neither tarsus length nor the residuals of body mass on tarsus length were related to T (lmes for overall analyses, lms for single seasons; all $p>0.07$; body mass and tarsus length were positively correlated: $p<0.001$ overall and for all seasons separately when accounting for date). However, when January data were analyzed separately, T was positively related to body mass after accounting for date ($t_{136}=2.63$, $p=0.009$, variance explained by the model $R^2=0.21$; other seasons and overall, $p>0.06$). These results were very similar when using residuals of T in relation to date and day time instead of T except for the relation between T and body mass over all seasons ($t_{412}=2.78$, $p=0.006$).

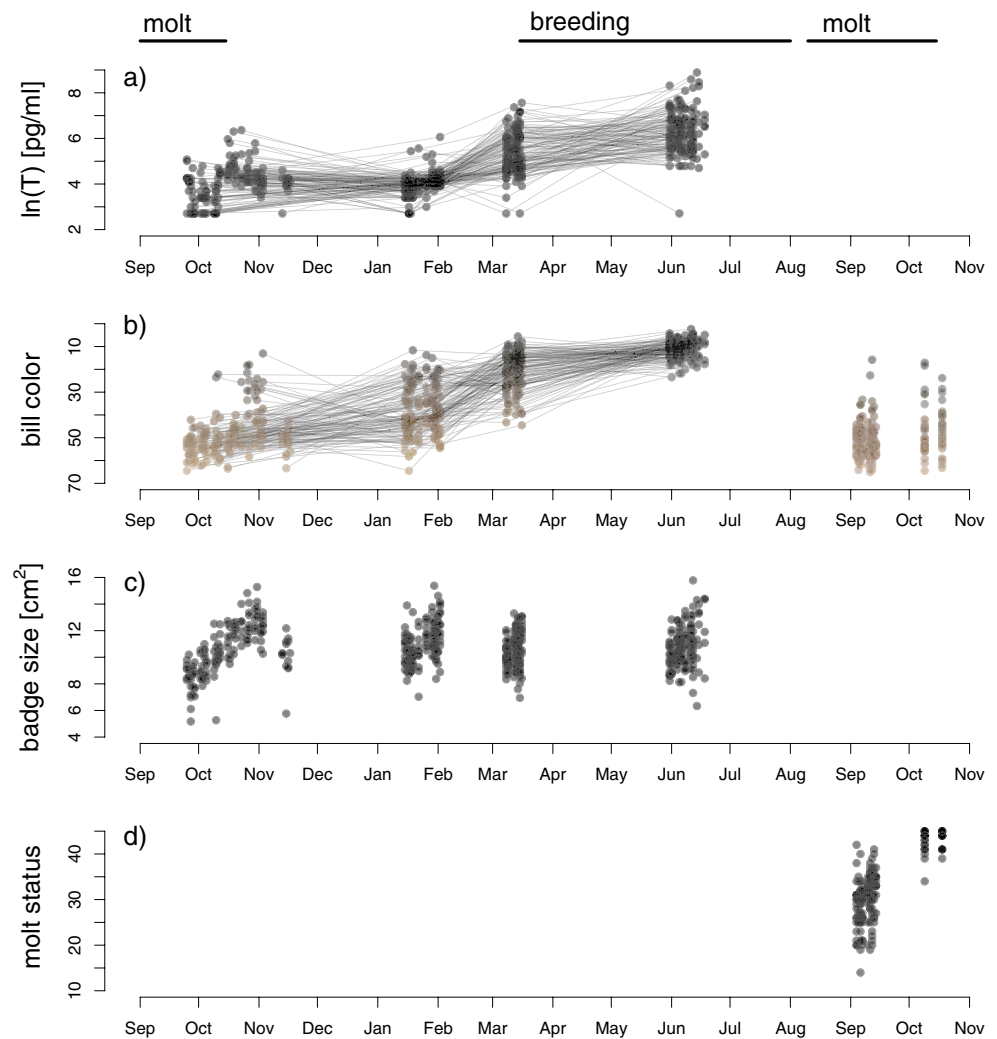
Badge size

Badge size was not related to plasma testosterone levels in any of the four periods ($r<0.09$, $t<1.10$, $p>0.27$; Figs. 2 and 3), even after accounting for tarsus length (variance explained by the model $R^2<0.06$, $t<0.64$, $p>0.53$). This did not change when using residuals of T in relation to date and day time. However, badge size was positively related to tarsus length ($r=0.24$, $t_{133}=2.90$, $p=0.004$) and to body mass in March ($r=0.17$, $t_{133}=2.03$, $p=0.04$) and June ($r=0.18$, $t_{133}=2.15$, $p=0.03$), but not in fall and January ($t<1.67$, $p>0.09$). The residuals of body mass on tarsus length were not related to our badge scores for any season ($t<0.86$, $p>0.39$). Thus, although there were weak positive correlations between badge size and overall size, there was no significant relationship between badge size and testosterone levels.

Dynamics in bill color and its relation to plasma T levels

Similar to testosterone, bill color changed dramatically throughout the course of the year (Fig. 2b). We found a highly significant relation between bill color and date (lme: $t_{423}=-42.31$; $p<0.0001$; random effect—bird ID). From September until March, individuals varied substantially in bill color, ranging from pale horn-colored to almost black. Nevertheless, the distribution of bill color shifted from a majority of individuals with pale horn-colored bills in September to a majority of individuals with dark bills in March. During the breeding season, between-individual variation in bill coloration was very low, i.e., all birds had

Fig. 2 Plasma testosterone levels (a, natural logarithm), bill color (b, brightness, actual bill color at sampling time presented), badge size (c, for each season, the mean of March and June badge is plotted), and molt status (d, BTO score) of house sparrows in the course of a year. We changed the transparency of points and lines to improve the visibility of overlapping points and lines. The gray lines in a and b connect measurements of the same individuals at different seasons. Parallel lines would be expected if individuals' relative phenotypes are consistent between seasons



dark bills (Bartlett Test of homogeneity of variances comparing breeding season with the other three seasons: Bartlett's $K_1^2 = 239.85$, $p < 0.001$). Within individuals, bill color was correlated between all seasons except the breeding season (Table 2). Bill color was also significantly repeatable between January measurements of two consecutive years ($r = 0.518 \pm 0.065$ (SE), $p < 0.001$, $n = 2 \times 127$). Bill

color was not related to body condition when accounting for date: Neither body mass nor tarsus length nor the residuals of body mass on tarsus length were correlated with bill color (lmes for overall analyses, lms for single seasons; all $p > 0.18$).

We found a significant overall relationship between bill color and plasma T levels (lme: $t_{421} = -24.23$, $p < 0.001$, random effect—bird ID; Fig. 2a, b). Bill color and T were also correlated in all periods separately, except in June when variation in bill color was very low (Table 3; Fig. 4). To exclude all date effects, we further used residuals of linear mixed effect models of bill color in relation to date and of T levels in relation to date: in this analysis, bill color and T were still highly correlated for all seasons combined ($t_{574} = -6.94$, $p < 0.0001$).

Bill color and molt status

BTO molt score was highly significantly related to date (lme: $t_{57} = 39.23$, $p < 0.001$, random effect—bird ID). After accounting for date, the score was not related to body mass,

Table 1 Correlations of plasma testosterone levels of house sparrows between seasons within individuals. We used Pearson correlation tests of the residuals of the natural logarithm of testosterone in relation to date and day time for each of the two seasons compared

Periods compared	<i>r</i>	95%CI	Statistic	<i>P</i> -value
Fall–January	0.24	0.08, 0.39	$t_{137} = 2.88$	0.005
Fall–March	0.07	-0.10, 0.23	$t_{136} = 0.78$	0.440
Fall–June	0.02	-0.15, 0.19	$t_{134} = 0.26$	0.799
January–March	0.19	0.03, 0.35	$t_{136} = 2.29$	0.023
January–June	-0.05	-0.21, 0.12	$t_{134} = -0.54$	0.589
March–June	0.10	-0.07, 0.27	$t_{133} = 1.20$	0.232

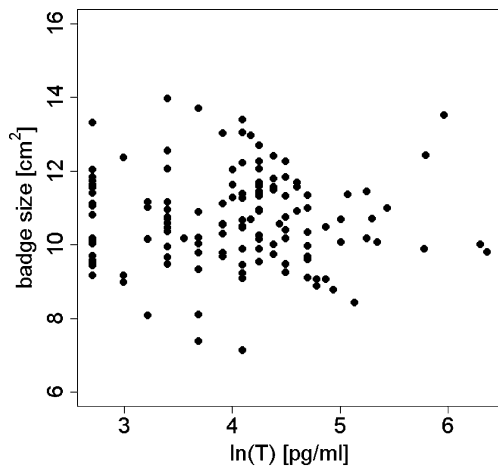


Fig. 3 Badge size in relation to plasma testosterone levels in house sparrows during the fall (Sept.–Oct.). We defined badge size as the mean of the March and June measurements and used the natural logarithm of plasma testosterone values ($p=0.9$, $r=0.01$)

tarsus length, or the residuals of mass on tarsus length (lmes for overall analyses, lms for September and October separately; all $p>0.07$).

Bill color was independent of molt status in October and for September and October combined (lme for overall analysis, lm for October separately; all $p>0.14$). However, when including date in the overall model and when looking at the September data alone, we found a highly significant negative relationship between bill color and BTO molt score (lme for overall analysis after accounting for date: $t_{56}=3.03$, $p=0.004$, random effect—bird ID; lm for September: $r=0.31$, $t_{142}=3.84$, $p<0.001$; Fig. 5). This means that birds got paler bills during the course of molt, but only as long as they were molting (hence no relation in October).

Discussion

As expected and in accordance with many other studies (reviewed in Anderson 2006), plasma testosterone levels of

Table 2 Correlations of bill color of house sparrows between seasons within individuals. We used Pearson correlation tests of residuals of bill color (brightness) in relation to date for each of the two seasons compared

Periods compared	<i>r</i>	95%CI	Statistic	<i>P</i> -value
Fall–January	0.49	0.35, 0.61	$t_{134}=6.44$	<0.001
Fall–March	0.33	0.18, 0.47	$t_{137}=4.14$	<0.001
Fall–June	0.05	-0.12, 0.22	$t_{134}=0.63$	0.532
January–March	0.60	0.48, 0.69	$t_{134}=8.59$	<0.001
January–June	0.02	-0.15, 0.19	$t_{131}=0.21$	0.838
March–June	-0.02	-0.18, 0.15	$t_{134}=-0.18$	0.856

Table 3 Correlations of bill color and plasma testosterone levels in house sparrows during different seasons. We used Pearson correlation tests of bill color (brightness) in relation to residuals of linear models of the natural logarithm of testosterone levels in relation to day time

Period	<i>r</i>	95%CI	Statistic	<i>P</i> -value
Fall	-0.25	-0.40, -0.09	$t_{137}=-3.06$	0.003
January	-0.24	-0.39, -0.08	$t_{134}=-2.88$	0.005
March	-0.25	-0.40, -0.09	$t_{136}=-3.02$	0.003
June	-0.09	-0.25, 0.07	$t_{148}=-1.13$	0.262

captive male house sparrows changed dramatically in the course of the year. We found the lowest values during the prebasic molt and the highest values during the breeding season, whereby there was only a slight increase between values in March, the beginning of breeding season, and June, the peak breeding season. The breeding season is the time with the most aggressive interactions between males

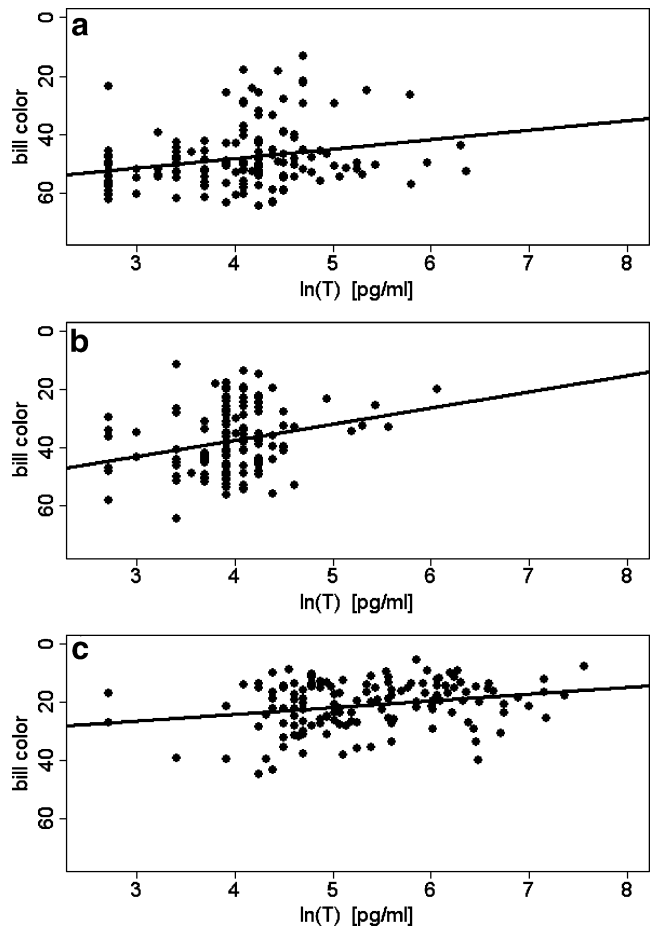


Fig. 4 Bill color in relation to plasma testosterone levels in house sparrows (a fall, b January, c March data). We measured bill color as brightness and used the natural logarithm of plasma testosterone values (a $p=0.003$, $r=-0.25$; b $p=0.005$, $r=-0.24$; c $p=0.003$, $r=-0.25$)

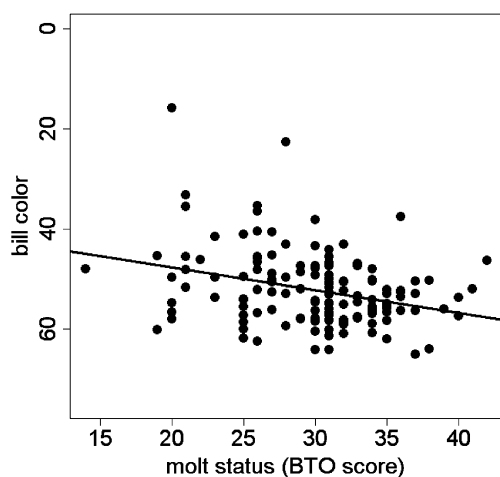


Fig. 5 Bill color in relation to molt status in house sparrows (September data). We measured bill color as brightness and molt status according to Ginn and Melville (2000, p. 28, Fig. 7b) ($p < 0.001$, $r = 0.31$)

establishing and defending territories, nest sites, and mates (e.g., Wingfield et al. 1990; Hau 2007). In contrast, molt is the time when aggression is lowest as molting birds are more vulnerable and their flight abilities are reduced (Swaddle and Witter 1997; Anderson et al. 2004). As house sparrows show flocking behavior all year round and tend to establish dominance ranks, inter-individual aggression occurs most of the year (Anderson 2006) as reflected in increases of plasma testosterone levels after the termination of molt, but long before breeding commences.

In contrast to our first prediction, within-individual testosterone levels were only very weakly correlated between periods. Thus, males with relatively high (or low) levels in one period did not have relatively high (or low) levels in another period. This result suggests that caution is required when interpreting point-sample testosterone measurements. Our result contrasts strongly with Buchanan et al. (2003) who did find some consistency in individual testosterone levels between breeding and post-breeding season, albeit with a much lower sample size ($n = 19$) and medium effect ($R^2 = 19.3$, $p = 0.027$).

We did not find any relationship between badge size and testosterone levels in our population of house sparrows during any of the studied periods, thus failing to support our second prediction. This is surprising as the result strongly contrasts with previous studies (Evans et al. 2000; Buchanan et al. 2001; Gonzalez et al. 2001). However, our analysis is based on a much larger sample size, and four separate measurements of T per individual. Since we did find very strong correlations between testosterone levels and another ornament (bill color, see below), it seems unlikely that the lack of a correlation with badge size was a type II error. In our study population, badge size therefore does not seem to predict testosterone-related dominance

and aggression (Ketterson and Nolan 1992). We therefore suggest that if badge size does reflect dominance in house sparrows, then it is more likely that it indicates dominance related to both body size (this study; Buchanan et al. 2001) and age (see Nakagawa et al. 2007; Morrison et al. 2008) rather than dominance related to testosterone per se. For a better understanding of the potential signaling function of badge size in our population, the determination of dominance ranks and of the degree of inter-individual aggression would be necessary. Note that testosterone could still be an important factor for the badge ornament despite the above results: For example, badge size could be related to true baseline testosterone levels or to true maximum levels (released during agonistic interactions), but these are hard to measure. Alternatively, testosterone could be negatively related to the length of the white tips of the badge feathers which may hide color signals during times when the signaling function is less needed as suggested by Gonzalez et al. (2001). At any rate, the relationship between badge size and circulating testosterone in our population of house sparrows is, at best, weak.

The lack of support for our first two predictions has important implications for the discussion of signal honesty in general (see also Kempenaers et al. 2008). The Immunocompetence Handicap Hypothesis (Folstad and Karter 1992) and the Oxidation Handicap Hypothesis (Alonso-Alvarez et al. 2007) explain the honesty of signals of quality via the testosterone-dependent immunosuppression and depressed resistance to oxidative stress, respectively (e.g., Evans et al. 2000; Gonzalez et al. 2001). However, both hypotheses assume that testosterone levels are correlated between seasons, so that a signal is honest not only at the time it is produced (i.e., molt for the badge) but also when it is used (i.e., all year round for the badge). Our observed inconsistency of testosterone levels between different seasons violates this basic assumption. Moreover, our results suggest only a weak (at best) correlation of testosterone levels with badge size, and therefore a relationship to physiological effects of testosterone such as immunosuppression is an unlikely explanation for the badge signal's honesty. Overall, we conclude that the honesty of the badge cannot be ensured via testosterone-related physiological effects but rather via different costs, such as social costs.

In clear contrast to badge size, we found that bill color was strongly correlated with testosterone levels in three of the four studied periods (no correlation in June, when variation in bill color was reduced). Furthermore, these relationships were independent of condition and repeatable in similar environmental conditions in consecutive years. These findings are in accordance with our third prediction and with other studies that have found that the degree of blackening in the bill coincided with an increase in

testosterone levels (Keck 1933, 1934; Witschi 1936; Pfeiffer et al. 1944; Haase 1975; Donham et al. 1982). This suggests that, at any time outside the breeding season, bill color is a very good predictor of relative testosterone levels. A complete change of bill color takes about three and a half weeks in male house sparrows (Anderson 2006). Therefore, bill color probably reflects average baseline testosterone levels (i.e., a running average) over a short period of time (a few weeks at most). Our results strongly suggest that any ornamental signaling function in male house sparrows that is directly related to testosterone-dependent information will be much more likely found with bill coloration rather than with the badge. More specifically, we suggest that bill color might serve as a signal of “strategy” (see Dale 2006), rather than as a signal of quality per se. Although signals of strategy are distinct from signals of overall quality (see Dale 2006), strategy signals can still indirectly reveal relative quality provided only good-quality males pursue more costly strategies (as might be associated with increased aggression associated with higher T levels or higher T levels maintained over a longer period of time).

We thus hypothesize that, between molt and breeding, some males in the population keep their testosterone levels relatively high, indicate this with darker bills, and are more dominant because of testosterone-related aggressiveness, while other males maintain low testosterone levels, indicate this with paler bills, and are thus rather subordinate while avoiding costs of high testosterone levels. The former is similar to what some studies have described as ‘autumn sexuality’ which may be related to higher breeding success (reviewed in Hegner and Wingfield 1986) and to any inherent qualities associated with different durations of elevated testosterone levels (see, e.g., Kempenaers et al. 2008). Such a potential signal for behavioral strategies could be used either in the context of mediating competitive interactions in feeding flocks (with potentially variable flock sizes and group members, as suggested for melanin-based ornaments in New and Old world sparrows (Tibbetts and Safran 2009)) and/or alternatively in the context of establishing early breeding territories/sites and acquiring a mate early (as suggested for red grouse (Mougeot et al. 2005)). Indeed, McGraw (2004) noted that the melanin-based black cap of male American goldfinches (*Carduelis tristis*) was particularly more variable in the non-breeding season and suggested that such “remnant” winter ornaments might be expected to evolve in animals that live in large non-breeding groups (e.g., status signaling systems) or in those where mates begin associating before breeding onset. In both scenarios, a signal announcing the degree of aggressiveness a male is willing to engage in could help to resolve encounters without true fights as is often suggested for badge size (Møller 1987). Because bill color is a much more dynamic ornament and because badge size

does not signal testosterone levels in our population, we suggest that bill color signals testosterone-related aggressiveness in the non-breeding season in house sparrows. That is, we suggest that, in house sparrows, bill color is the more likely “badge-of-status” than is the badge. This is in accordance with another study that suggested bill darkening in fall to be a response to social competition (Hegner and Wingfield 1986).

We found a negative relation between bill color and molt in September—the peak period of molt. This reflects the general relationship between molt and decreased testosterone (Schleussner 1990; Nolan et al. 1992) when the majority of males had pale bills. The hypothesis that male bill color is a signal of strategy predicts that all molting males signal low aggression perhaps because they are more vulnerable at this time. Molt can therefore be considered the mirror image to breeding as in both seasons all males are pursuing the same strategy at the same time. Nevertheless, bill color variation during molt is larger than during breeding. This may reflect (1) that the timing and speed of molt are more flexible in response to external and internal factors (Hahn et al. 1992) and (2) that individuals do not need to synchronize to successfully complete molt (in contrast to breeding).

In summary, we found that individual plasma testosterone levels were highly variable and not repeatable. We furthermore found that badge size and bill color in male house sparrows likely signal different kinds of information. Bill color is a strongly T-related trait, whereas badge size is not. We hypothesize that bill color indicates different aggression-related strategies during the non-breeding season. More detailed studies on the function of bill color in the context of social interactions are needed for a better understanding of this testosterone-related trait in male house sparrows.

Acknowledgements Many thanks to Ariane Mutzel and Andrea Wittenzellner for their help with taking blood samples and to Barbara Helm for help with the molt scoring. We are grateful to Mihai Valcu and Holger Schielzeth for statistical advice and programming, to Kevin McGraw and three anonymous reviewers for comments on the manuscript, and to Magnus Rudorfer, Edith Bodendorfer, Annemarie Grötsch, and Frances Weigel for animal care. We thank the Behavioural Ecology and Evolutionary Genetics Group at the MPIO for discussion and comments. Funding for this research was provided by the Max Planck Society. The study was performed in agreement with the current laws of Upper Bavaria, Germany (under license: permit nr. 55.1-8642.3-3-2006 of the “Regierung Oberbayern”, with several extensions).

Conflict of interest The authors declare that they have no conflict of interest.

References

Abramoff MD, Magelhaes PJ, Ram SJ (2004) Image processing with ImageJ. *Biophoton Int* 11:36–42

- Adkins-Regan E (2005) Hormones and animal social behavior. Princeton University Press, Princeton
- Alonso-Alvarez C, Bertrand S, Faivre B, Chastel O, Sorci G (2007) Testosterone and oxidative stress: the oxidation handicap hypothesis. *Proc R Soc Biol Sci Ser B* 274:819–825
- Anderson TR (2006) Biology of the ubiquitous house sparrow. Oxford University Press, New York
- Anderson KE, Davis GS, Jenkins PK, Carroll AS (2004) Effects of bird age, density, and molt on behavioral profiles of two commercial layer strains in cages. *Poult Sci* 83:15–23
- Blas J, Pérez-Rodríguez L, Bortolotti GR, Vinuela J, Marchant TA (2006) Testosterone increases bioavailability of carotenoids: insights into the honesty of sexual signaling. *Proc Natl Acad Sci U S A* 103:18633–18637
- Bókony V, Garamszegi LZ, Hirschenhauser K, Liker A (2008) Testosterone and melanin-based black plumage coloration: a comparative study. *Behav Ecol Sociobiol* 62:1229–1238
- Buchanan KL, Evans MR, Goldsmith AR, Bryant DM, Rowe LV (2001) Testosterone influences basal metabolic rate in male house sparrows: a new cost of dominance signalling? *Proc R Soc Biol Sci Ser B* 268:1337–1344
- Buchanan KL, Evans MR, Goldsmith AR (2003) Testosterone, dominance signalling and immunosuppression in the house sparrow, *Passer domesticus*. *Behav Ecol Sociobiol* 55:50–59
- Calisi RM, Hews DK (2007) Steroid correlates of multiple color traits in the spiny lizard, *Sceloporus pyrocephalus*. *J Comp Physiol B* 177:641–654
- Candolin U (2003) The use of multiple cues in mate choice. *Biol Rev (Camb)* 78:575–595
- Clough D, Heistermann M, Kappeler PM (2009) Individual facial coloration in male *Eulemur fulvus rufus*: a condition-dependent ornament? *Int J Primatol* 30:859–875
- Dale J (2006) Intraspecific variation in coloration. In: Hill GE, McGraw KJ (eds) Bird coloration vol 2. Harvard University Press, Cambridge, pp 36–86
- Dijkstra PD, Hekman R, Schulz RW, Groothuis TGG (2007) Social stimulation, nuptial colouration, androgens and immunocompetence in a sexual dimorphic cichlid fish. *Behav Ecol Sociobiol* 61:599–609
- Donham RS, Wingfield JC, Mattocks PW, Farmer DS (1982) Changes in testicular and plasma androgens with photoperiodically induced increase in plasma LH in the house sparrow. *Gen Comp Endocrinol* 48:342–347
- Evans MR, Goldsmith AR, Norris SRA (2000) The effects of testosterone on antibody production and plumage coloration in male house sparrows (*Passer domesticus*). *Behav Ecol Sociobiol* 47:156–163
- Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Pearson, Harlow
- Folstad I, Karter AJ (1992) Parasites, bright males, and the immunocompetence handicap. *Am Nat* 139:603–622
- Ginn HB, Melville DS (2000) Moults in birds, vol 19. British Trust for Ornithology, Norwich
- Gonzalez G, Sorci G, Smith LC, de Lope F (2001) Testosterone and sexual signalling in male house sparrows (*Passer domesticus*). *Behav Ecol Sociobiol* 50:557–562
- Grafen A (1990) Biological signals as handicaps. *J Theor Biol* 144:517–546
- Haase E (1975) Effects of testosterone propionate on secondary sexual characters and testes of house sparrows, *Passer domesticus*. *Gen Comp Endocrinol* 26:248–252
- Hahn TP, Swingle J, Wingfield JC, Ramenofsky M (1992) Adjustments of the prebasic Molt schedule in birds. *Ornis Scand* 23:314–321
- Hau M (2007) Regulation of male traits by testosterone: implications for the evolution of vertebrate life histories. *BioEssays* 29:133–144
- Hegner RE, Wingfield JC (1986) Gonadal development during autumn and winter in house sparrows. *Condor* 88:269–278
- Hill GE (2002) A red bird in a brown bag: the function and evolution of colorful plumage in the House Finch. Oxford University Press, New York
- Hill GE, McGraw KJ (2006) Bird coloration: function and evolution, vol 2. Harvard University Press, Cambridge
- Humphrey PS, Parkes KC (1959) An approach to the study of molts and plumages. *Auk* 76:1–31
- Huyghe K, Husak JF, Herrel A, Tadić Z, Moore IT, Van Damme R, Vanhooydonck B (2009) Relationships between hormones, physiological performance and immunocompetence in a color-polymorphic lizard species, *Podarcis melisellensis*. *Horm Behav* 55:488–494
- Jawor JM, Breitwisch R (2003) Melanin ornaments, honesty, and sexual selection. *Auk* 120:249–265
- Jensen H, Steinsland I, Ringsby TH, Sæther BE (2008) Evolutionary dynamics of a sexual ornament in the house sparrow (*Passer domesticus*): the role of indirect selection within and between sexes. *Evolution* 62:1275–1293
- Johnstone RA, Norris K (1993) Badges of status and the cost of aggression. *Behav Ecol Sociobiol* 32:127–134
- Keck WN (1933) Control of the bill color of the male English sparrow by injection of male hormone. *Proc Soc Exp Biol Med* 30:1140–1141
- Keck WN (1934) The control of the secondary sex characters in the English sparrow, *Passer domesticus* (Linnaeus). *J Exp Zool* 67:315–347
- Kempnaers B, Peters A, Foerster K (2008) Sources of individual variation in plasma testosterone levels. *Philos Trans R Soc Lond B Biol Sci* 363:1711–1723
- Ketterson ED, Nolan V (1992) Hormones and life histories—an integrative approach. *Am Nat* 140:S33–S62
- Kurtz J, Kalbe M, Langefors Å, Mayer I, Milinski M, Hasselquist D (2007) An experimental test of the immunocompetence handicap hypothesis in a teleost fish: 11-ketotestosterone suppresses innate immunity in three-spined sticklebacks. *Am Nat* 170:509–519
- Lessells CM, Boag PT (1987) Unrepeatable repeatabilities—a common mistake. *Auk* 104:116–121
- Lewis RJ (2009) Chest staining variation as a signal of testosterone levels in male Verreaux's sifaka. *Physiol Behav* 96:586–592
- Liker A, Barta Z (2001) Male badge size predicts dominance against females in house sparrows. *Condor* 103:151–157
- Maynard Smith JFRS, Harper DGC (1988) The evolution of aggression—can selection generate variability? *Philos Trans R Soc Lond B Biol Sci* 319:557–570
- McGraw KJ (2004) Winter plumage coloration in male American goldfinches: do reduced ornaments serve signaling functions in the non-breeding season? *Ethology* 110:707–715
- McGraw KJ, Ardia DR (2007) Do carotenoids buffer testosterone-induced immunosuppression? An experimental test in a colourful songbird. *Biol Lett* 3:375–378
- McGraw KJ, Dale J, Mackillop EA (2003) Social environment during molt and the expression of melanin-based plumage pigmentation in male house sparrows (*Passer domesticus*). *Behav Ecol Sociobiol* 53:116–122
- Møller AP (1987) Variation in badge size in male house sparrows *Passer domesticus*—evidence for status signaling. *Anim Behav* 35:1637–1644
- Møller AP (1990) Sexual-behavior is related to badge size in the house sparrow *Passer domesticus*. *Behav Ecol Sociobiol* 27:23–29
- Møller AP, Erritzøe J (1992) Acquisition of breeding coloration depends on badge size in male house sparrows *Passer domesticus*. *Behav Ecol Sociobiol* 31:271–277
- Montgomerie R (2006) Analyzing colors. In: Hill GE, McGraw KJ (eds) Bird coloration, vol 1. Harvard University Press, Cambridge, pp 90–147

- Morrison EB, Kinnard TB, Stewart IRK, Poston JP, Hatch MI, Westneat DF (2008) The links between plumage variation and nest site occupancy in male house sparrows. *Condor* 110:345–353
- Mougeot F, Irvine JR, Seivwright L, Redpath SM, Piernety S (2004) Testosterone, immunocompetence, and honest sexual signaling in male red grouse. *Behav Ecol* 15:930–937
- Mougeot F, Dawson A, Redpath SM, Leckie F (2005) Testosterone and autumn territorial behavior in male red grouse *Lagopus lagopus scoticus*. *Horm Behav* 47:576–584
- Mundinger PC (1972) Annual testicular cycle and bill color change in Eastern American goldfinch. *Auk* 89:403–419
- Murphy TG, Rosenthal MF, Montgomerie R, Tarvin KA (2009) Female American goldfinches use carotenoid-based bill coloration to signal status. *Behav Ecol* 20:1348–1355
- Nakagawa S, Ockendon N, Gillespie DOS, Hatchwell BJ, Burke T (2007) Assessing the function of house sparrows' bib size using a flexible meta-analysis method. *Behav Ecol* 18:831–840
- Nolan V, Ketterson ED, Ziegenfus C, Cullen DP, Chandler CR (1992) Testosterone and avian life histories—effects of experimentally elevated testosterone on prebasic molt and survival in male dark-eyed juncos. *Condor* 94:364–370
- Peters A, Denk AG, Delhey K, Kempenaers B (2004) Carotenoid-based bill colour as an indicator of immunocompetence and sperm performance in male mallards. *J Evol Biol* 17:1111–1120
- Peters A, Delhey K, Goymann W, Kempenaers B (2006) Age-dependent association between testosterone and crown UV coloration in male blue tits (*Parus caeruleus*). *Behav Ecol Sociobiol* 59:666–673
- Pfeiffer CA, Hooker CW, Kirschbaum A (1944) Deposition of pigment in the sparrow's bill in response to direct applications as a specific and quantitative test for androgen. *Endocrinology* 34:389–399
- R Development Core Team (2006) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>
- R Development Core Team (2008) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>
- Roberts ML, Buchanan KL, Evans MR (2004) Testing the immunocompetence handicap hypothesis: a review of the evidence. *Anim Behav* 68:227–239
- Roelants H, Schneider F, Goritz F, Streich J, Blottner S (2002) Seasonal changes of spermatogonial proliferation in roe deer, demonstrated by flow cytometric analysis of c-kit receptor, in relation to follicle-stimulating hormone, luteinizing hormone, and testosterone. *Biol Reprod* 66:305–312
- Safran RJ, McGraw KJ (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
- Salvador A, Veiga JP, Martin J, Lopez P, Abelenda M, Puerta M (1996) The cost of producing a sexual signal: testosterone increases the susceptibility of male lizards to ectoparasitic infestation. *Behav Ecol* 7:145–150
- Schleussner G (1990) Experimental induction of eccentric primary molt in adult male European starlings (*Sturnus vulgaris*). *J Ornithol* 131:151–155
- Setchell JM, Smith T, Wickings EJ, Knapp LA (2008) Social correlates of testosterone and ornamentation in male mandrills. *Horm Behav* 54:365–372
- Swaddle JP, Witter MS (1997) The effects of molt on the flight performance, body mass, and behavior of European starlings (*Sturnus vulgaris*): an experimental approach. *Can J Zool* 75:1135–1146
- Thompson CW, Moore MC (1991) Throat colour reliably signals status in male tree lizards, *Urosaurus ornatus*. *Anim Behav* 42:745–753
- Tibbetts EA, Dale J (2004) A socially enforced signal of quality in a paper wasp. *Nature* 432:218–222
- Tibbetts EA, Safran RJ (2009) Co-evolution of plumage characteristics and winter sociality in New and Old World sparrows. *J Evol Biol* 22:2376–2386
- Václav R (2006) Within-male melanin-based plumage and bill elaboration in male house sparrows. *Zool Sci* 23:1073–1078
- Vaclav R, Hoi H (2002) Different reproductive tactics in house sparrows signalled by badge size: is there a benefit to being average? *Ethology* 108:569–582
- Veiga JP (1993) Badge size, phenotypic quality, and reproductive success in the house sparrow—a study on honest advertisement. *Evolution* 47:1161–1170
- Verhulst S, Dieleman SJ, Parmentier HK (1999) A tradeoff between immunocompetence and sexual ornamentation in domestic fowl. *Proc Natl Acad Sci U S A* 96:4478–4481
- Wingfield JC (1984) Environmental and endocrine control of reproduction in the song sparrow, *Melospiza melodia*. II. Agonistic interactions as environmental information stimulating secretion of testosterone. *Gen Comp Endocrinol* 56:417–424
- Wingfield JC (1985) Short-term changes in plasma-levels of hormones during establishment and defense of a breeding territory in male song sparrows, *Melospiza melodia*. *Horm Behav* 19:174–187
- Wingfield JC, Hegner RE, Dufty AM, Ball GF (1990) The challenge hypothesis—theoretical implications for patterns of testosterone secretion, mating systems, and breeding strategies. *Am Nat* 136:829–846
- Wingfield JC, Lynn SE, Soma KK (2001) Avoiding the 'costs' of testosterone: ecological bases of hormone–behavior interactions. *Brain Behav Evol* 57:239–251
- Witschi E (1936) The bill of the sparrow as an indicator for the male sex hormone. I. Sensitivity. *Proc Soc Exp Biol Med* 33:484–486
- Witschi E, Woods RP (1936) The bill of the sparrow as an indicator for the male sex hormone. II. Structural basis. *J Exp Zool* 73:445–459