Baseline haptoglobin concentrations are repeatable and predictive of certain aspects of a subsequent experimentally-induced inflammatory response

Kevin D. Matson *, Nicholas P.C. Horrocks, Maaike A. Versteegh, B. Irene Tieleman

Animal Ecology Group, Center for Ecological and Evolutionary Studies, University of Groningen, P.O. Box 11103, 9700 CC, Groningen, The Netherlands

A R T I C L E   I N F O

Article history:
Received 12 December 2011
Received in revised form 16 January 2012
Accepted 16 January 2012
Available online 31 January 2012

Keywords:
Acute phase proteins
Birds
Circannual patterns
Columbiformes
Ecological immunology
Inflammation
Lipopolysaccharide
Repeatability

A B S T R A C T

Ecologists sometimes assume immunological indices reflect fundamental attributes of individuals—an important assumption if an index is to be interpreted in an evolutionary context since among-individual variation drives natural selection. Yet the extent to which individuals vary over different timescales is poorly understood. Haptoglobin, an acute phase protein, is an interesting parameter for studying variability as it is easily quantified and concentrations vary widely due to the molecule’s role in inflammation, infection and trauma. We quantified haptoglobin in pigeon plasma samples collected over fourteen months and calculated repeatability to evaluate if haptoglobin concentration is a distinctive trait of individuals. We also explored the capacity of baseline haptoglobin concentrations to predict an array of physiological changes associated with a subsequent experimentally-induced inflammatory response. Maximum repeatability, which occurred over a short mid-winter interval, equaled 0.57. Baseline haptoglobin concentrations predicted response haptoglobin concentrations better than any other endotoxin-induced change. Overall, we identified several strengths and limitations of baseline [Hp] quantification. Acknowledging these qualities should lead to more refined conclusions in studies of the ecology and evolution of immune function.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Interest in incorporating aspects of immunology into studies of ecology and evolution continues to grow. Indices of immune function used by ecological immunologists are often assumed to reflect fundamental attributes of individuals (Hill, 2011). Such an assumption is important if the measures are interpreted in an evolutionary context since variation among individuals is a driver of natural selection (Hõrak et al., 2002; Tieleman et al., 2010, but see Pitala et al., 2007). The immune system is, however, dynamic. For many immune indices, the extent to which individuals vary over different timescales is poorly understood. Understanding immunological lability requires further investigation over timescales both short (i.e., hourly, daily, or pre- and post-challenge sampling; e.g., Millet et al., 2007; Horrocks et al., 2011) and long (i.e., monthly or seasonally sampling over one or more years; e.g., Hõrak et al., 2002; Buehler et al., 2008, Hegemann unpublished data).

Haptoglobin (Hp), in particular, and acute phase proteins, more generally, are especially interesting parameters for studying immunological lability. First and foremost, concentrations within individuals vary through time and mirror changes in health status and physiological condition (Hõrak et al., 2002, 2003). Under normal conditions (e.g., in the absence of acute pathologies), Hp circulates in the blood at low, but taxon-specific, concentrations. Concentrations can increase rapidly and manifoldly in response to acute infection, inflammation, or trauma (Delers et al., 1988; Millet et al., 2007). An elevated plasma haptoglobin concentration ([Hp]) often signifies the onset of a non-specific immune response, but changes might also reflect the role of Hp as an antioxidant (Gutteridge, 1987; Dobryszyczka, 1997; Quaye, 2008).

Hp and its functional equivalents have been identified, either genetically or functionally, in a wide range of taxa (Delers et al., 1988; Matson, 2006; Matson et al., 2006a; Wicher and Fries, 2006; Millet et al., 2007; Kakuschke et al., 2010). Mean [Hp] varies among bird species (Delers et al., 1988; Matson, 2006; Matson et al., 2006a; Millet et al., 2007), but the highest concentrations reported in birds are generally lower than the highest concentrations reported in mammals. The use of a simple, commercially-available, functional assay means that plasma [Hp] can be easily quantified in the wide variety of avian species that is often encountered with comparative ecological approaches (Matson, 2006; Matson et al., 2006a; Millet et al., 2007). Nevertheless, a firm grip on how to interpret [Hp] in the context of these comparative studies has remained elusive (e.g., Matson, 2006).

The connections between health status, physiological condition, and immune function embody one source of variability in many indices of immune function, including Hp. While such connections are central to some questions in ecological immunology, these very same connections can complicate interpretations of the causes and consequences of any observed differences (Hill, 2011). For example,
the quantification of [Hp] in the context of comparative ecological studies often permits only equivocal and unsatisfying conclusions about costs and trade-offs (Matson, 2006; Buehler et al., 2009b), even with the molecule’s unambiguous association with energetically-expensive systemic inflammation (Buehler et al., 2009a; van de Crommenacker et al., 2010). This lack of clarity results from gaps in the basic ecological understanding of acute phase proteins, such as the extent to which constitutive [Hp] predicts the magnitude of an acute phase response. Field studies where individuals are observed only at a single point in time (i.e., the point of capture and sampling) further complicate interpretation, and in these cases the degree to which plasma [Hp] is an attribute of individuals cannot be determined. More information might be known at the level of population (e.g., life history stage, habitat quality) or species (e.g., susceptibility to particular disease, geographic range) in field-studies, but these ecological correlates alone are unlikely to reveal the causes of differences in average [Hp] among groups (Matson, 2006; Buehler et al., 2009b).

To improve upon the current ecological understanding of avian Hp, we examined the variability in “baseline” or constitutive (i.e., non-experimentally-induced) [Hp] in the contexts of repeatability and predictive capacity. Repeatability, which is defined as the portion of total variation that is attributable to among-individual differences (Lessells and Boag, 1987), signals the degree to which baseline [Hp] is a distinctive trait of individuals. We defined predictive capacity as the strength of a relationship between a baseline physiological measurement (in our case, [Hp]) and a subsequent physiological change that results from an immunological challenge. We used this concept of predictive capacity to help identify potential consequences of among individual differences.

We measured baseline [Hp] in 13 plasma samples collected repeatedly from 16 homing pigeons over the course of 14 months. We followed this baseline sampling scheme with an experiment designed to assess the physiological effects of an endotoxin-induced acute phase response. First, we investigated repeatability within intervals that comprised multiple consecutive samples but that differed in length (between two and thirteen approximately-monthly measurements) and time of year. We asked two questions: 1) to what extent is [Hp] a repeatable and distinctive trait of individuals? and 2) to what extent is this repeatability influenced by the time of year and the number of repeated measurements in a particular series of samples? We also tested the effects of sex and sample month on repeatability in specific sample series. Second, we explored the correlations between the within-individual means of baseline [Hp] and the within-individual changes in physiological parameters following an experimental immunological challenge. We asked the following question: within individuals, to what extent is baseline [Hp] predictive of the effects of an endotoxin-induced inflammatory response? For this purpose, we quantified changes in an array of physiological and immunological indices that are frequently used by ecologists: body mass, cloacal temperature, metabolic rate, oxidative status, blood glucose concentration, hemolysis, hemagglutination, bacteria killing abilities, and [Hp]. As with repeatability, we investigated if the predictive capacity of baseline [Hp] is influenced by the time of year and the number of repeated measurements used to calculate the predictor.

2. Materials and methods

2.1. Birds

We studied 16 homing pigeons (eight ♂, eight ♀; Columba livia domestica). Birds lived year-round in single sex groups of four in small outdoor avairies (roughly 2 m x 2 m x 2 m high). Birds accessed food (a mix of seed and pellet diets, grit, and vitamins) and water ad libitum. Hatched between late November and late December in 2005, all birds were about one year and five months old at the initiation of sampling and about two years and eight months old at the time of the endotoxin challenge (August 2008). For additional information about the study system, see van de Crommenacker et al. (2010).

2.2. Blood sampling, endotoxin challenge, and metabolic measurement

We used heparinized syringes to collect blood samples (~1 mL, <0.5% of body mass) from all birds by brachial venipuncture. We centrifuged the whole blood and collected and froze (~$-20°C$) the plasma fraction for future analyses. Approximately two to three hours elapsed between collection and freezing; blood samples were kept cool (~4°C) during this period.

We collected 13 baseline plasma samples from each bird at approximately-monthly intervals between April 2007 and June 2008 (mean interval = 36 days ± 11 days (s.d.)). On each baseline sample day, we collected blood from all birds over the course of the afternoon (between ~13:00 and ~18:00 h). Among cages, we followed a rotating sampling order with different starting cages on different baseline sample days; within cages, we caught and sampled birds in random order.

In August 2008 we collected pre- and post-endotoxin-challenge samples from eight birds (four ♂, four ♀). Pre- and post-endotoxin-challenge samples were always separated by a 48-hour interval. Both blood samples were collected from birds between 11:00 and 11:30, which was after 22 h of fasting and 18 h in a metabolic chamber at thermoneutral conditions. We administered the endotoxin challenge between 17:00 and 17:30, which was immediately before the birds’ second entry into the metabolic chamber and 18 h before the collection of post-challenge samples. The endotoxin challenge was an intraperitoneal injection of 2 mL per kg body mass of a 1.25 mg mL$^{-1}$ solution of lipopolysaccharide (LPS, L7261; Sigma, MO, USA) dissolved in phosphate buffered saline.

During the two 18 h periods each bird spent in a metabolic chamber, we used standard flow-through respirometry to measure oxygen consumption, which we divided by average nightly mass (mL h$^{-1}$ g$^{-1}$ body mass) (Gessaman, 1987). Because the respirometer was set up to measure two birds per night, the experimental protocol was staggered in time over a two-week period. For complete details about the metabolic set-up and equipment, see van de Crommenacker et al. (2010). This experiment was approved by the Animal Experimentation Committee of the University of Groningen (DEC license number 5095).

2.3. Haptoglobin quantification: baseline and endotoxin-challenge experiment

We quantified [Hp] (mg mL$^{-1}$) in all baseline blood samples and in pre- and post-challenge blood samples. We used a commercially-available functional assay (TP801; Tri-Delta Diagnostics, NJ, USA), which colorimetrically quantifies the heme-binding capacity of plasma. We followed the ‘manual method’ instructions provided by the kit manufacturer with a few minor modifications. Because preliminary measurements revealed low [Hp] in pigeons, we used twice the standard amount of plasma per well (15 μL instead of 7.5 μL) and adjusted the calculated concentrations accordingly. We measured absorbances at two wavelengths (450 and 630 nm) prior to the addition of the final reagent that initiated the color-change reaction. The pre-scan at the normal assay wavelength of 630 nm allowed us to correct for differences in plasma color and cloudiness by subtracting pre-scan absorbance values from final absorbance values. The 450 nm pre-scan enabled us to statistically analyze and correct for differences in plasma sample redness, an indication of hemolysis, which can affect the assay (see appendix).

Based on an among-plate standard that was run in duplicate in each of the five assay plates used in this study, within-plate variation averaged 3.7%, and between-plate variation equaled 3.5%. Based on 32 pigeon plasma samples that were run in duplicate in a single plate, within-plate variation averaged 5.0%.
2.4. Other measurements: endotoxin-challenge experiment only

2.4.1. Body mass and cloacal temperature

Using a digital balance (PB3002; Mettler-Toledo, Greifensee, Switzerland), we measured the body mass (g) of each bird before and after the metabolic measurement sessions that spanned the control and endotoxin-challenge nights. We calculated the percent of body mass lost during each 18 h session. Using a thermocouple thermometer (450-ATT; Omega Engineering, CT, USA), we measured the cloacal temperature (°C) of each bird immediately upon removal from a metabolic chamber.

2.4.2. Oxidative status

We measured total antioxidant capacity (TAC, in mmol L⁻¹ of HClO neutralized) using the OXY-Adsorbent test and reactive oxygen metabolites (ROMs, in mmol L⁻¹ of H₂O₂ equivalents) using the d-ROMs test (both kits: Diacon, Grosseto, Italy). For more details about these assays, see van de Crommenacker et al. (2010).

2.4.3. Blood glucose

We measured glucose concentrations (mmol L⁻¹) in fresh blood samples using a CardioChek PA Analyzer (1708), a small handheld diagnostic device, and PTS Panels Glucose Test Strips (1713; both by Polymer Technology Systems, IN, USA). All measurements fell well within the measuring range of the device (1.11–33.3 mmol L⁻¹), as outlined in the manufacturer’s instructions.

2.4.4. Hemolysis and hemagglutination

We followed the procedure outlined by Matson et al. (2005) to assess hemolysis and hemagglutination. Red blood cells (RBC) from rabbits (RBA050, HemoStat Laboratories, CA, USA) were incubated in serially diluted pigeon plasma samples. Agglutination and lysis were recorded as titers (−log₂ of the last plasma dilution that shows each reaction). Agglutination and lysis were scored from assay plate images recorded 20 min after incubation, and lysis was scored from images recorded 90 min and 24 h after incubation. Blind to sample and plate identity, KDM scored randomized images from each time point a minimum of two times. If these first two scores were <1 titer apart, then we used the mean value in further analyses. If these first two scores were ≥1 titer apart (≥5% of samples), then the sample was scored a third time, and we used the median value in further analyses.

Lysis was recorded 24 h after incubation because this reaction was absent (22 of 40 samples) or very low (<1 titer, an additional 9 of 40 samples) at 90 min post-incubation. Lysis increased in all individuals between the two recordings; overall, this increase was highly significant (t(39) = 18.5, P < 0.0001). To avoid methodological redundancy, we constructed (using principal component analysis) a single hemolysis variable that combined the two responses, which were highly correlated (Spearman Rank Correlation, ρ = 0.8, P < 0.0001).

2.4.5. Bacteria killing abilities

We followed the basic procedures outlined in Tieleman et al. (2005) and Matson et al. (2006b). We reconstituted lyophilized pellets of two microbial strains, *Escherichia coli* (ATCC 8739) and *Staphylococcus aureus* (ATCC 6538), according to the manufacturer’s instructions (0483E7 and 0485E7; MicroBioLogics, MN, USA). We diluted blood (1:10 for *E. coli*; 1:2 for *S. aureus*) and plasma (1:5 for *E. coli*) in CO₂-independent medium (18045–054, Invitrogen, Breda, The Netherlands). We added 20 μL of a microbial suspension to tubes containing 200 μL of diluted blood or plasma. We incubated these mixtures at 41 °C for 10 min for *E. coli* and 90 min for *S. aureus*. Then, in duplicate, 75 μL was spread onto agar plates. We also directly plated 75 μL of control mixtures (200 μL of medium, 20 μL of microbial suspension). Plates were incubated overnight, and colonies were counted the next day. We calculated the following index: killing ability = 1 − (count_{control-plate}/count_{test-plate}). We constructed a single *E. coli* killing ability variable that combined the correlated responses of plasma and of whole blood.

2.5. Statistical analyses

2.5.1. Variation and repeatability

Using R version 2.6.2 (R Development Core Team, 2008) and linear mixed-effects (lme) models that included individual as a random factor, we investigated the repeatability of baseline [Hp]. We obtained the requisite among- and within-individual variances from the lme models and used the following equation to calculate repeatability: \( R = \frac{\text{variance among}}{\text{variance within} + \text{variance among}} \) (Lessells and Boag, 1987). Statistical significance of repeatability was evaluated with likelihood-ratio tests and \( \chi^2 \)-statistics after contrasting models with and without individual as a random factor.

We calculated repeatability using all series that comprised multiple consecutive samples but that differed in number (between two and thirteen) and time of year (e.g., two samples in summer, seven samples spanning fall and winter, etc.). We first used a model with no fixed effects to analyze raw and Box-Cox transformed data. The Box-Cox transformation (\( \lambda = 0.06 \)) substantially improved normality by eliminating a strong right skew (before = 5.32, after = −0.02) and reducing kurtosis (before = 38.62, after = 4.92). We used Akaike’s information criterion (AIC) to rank and compare models that included combinations of sex, sample month, and different plasma redness indices (see appendix). The best model, which had AIC, of 3.35 less than the second best model, included sex, sample month, the interaction between sex and sample month, and an index that accounted for within-individual variation in plasma redness (450nm, see appendix). Using sample series of different lengths, we then calculated and compared repeatabilities using the best model (i.e., the full model) described above, the sex-only model, the sample-month-only model, and the null model.

2.5.2. Predictive capacity

Prior to investigating the predictive capacity of baseline [Hp], we tested the effect of endotoxin challenge on each physiological parameter. If there was clearly no effect of challenge (\( P > 0.1 \)) on a parameter, then we did not investigate that parameter any further. Only TAC fell into this category (Wilcoxon Signed Ranks Test, Z = 1.3, P = 0.21). All other parameters increased significantly (Z > 1.8, P < 0.05) or marginally (Z > 1.7, P < 0.08) following endotoxin challenge.

When investigating the predictive capacity for physiological responses other than Hp, we used intra-individual differences (post-challenge minus pre-challenge) in parameters as the dependent variables. For the Hp response, we used as dependent variables the post-challenge [Hp] that were either uncorrected or corrected for plasma redness (\( 450\text{nm} \), see appendix). Using sample series of different lengths, we then calculated and compared repeatabilities using the best model (i.e., the full model) described above, the sex-only model, the sample-month-only model, and the null model. We calculated repeatability using all series that comprised multiple consecutive samples. We correlated all single-sample and average values of baseline [Hp] with each response variable (i.e., a total of 91 correlations per response). We evaluated these correlations using non-parametric (Kendall’s τ rank correlation coefficient) and parametric (Pearson’s correlation coefficient) statistics, and both approaches resulted in qualitatively similar conclusions. We report Kendall’s τ because it is 1) free of normality assumptions (that are difficult to meaningfully test with small sample sizes), 2) more statistically conservative, and 3) simple to interpret. Given the problems associated with multiple comparisons and our high number of correlations, we did not concern ourselves with statistical significance of individual correlations. Instead, we used descriptive statistics to summarize the 91 correlations between baseline [Hp] and each response, and we calculated the percentage of these correlations for which the absolute value of Kendall’s τ > 0.5, which is the point where concordant pairs of data points outnumber discordant
ones. We also examined the effects of the time of year and the number of repeated measurements used to calculate the mean baseline [Hp] on the strength of these correlations.

3. Results

3.1. Variation and repeatability

3.1.1. The complete data set

Plasma [Hp] varied among individuals (Fig. 1A) and sample months (Fig. 1B). Repeatability of untransformed [Hp], when calculated using the 16 individuals over the entire sample series and a null model, equaled 0.02 ($\chi^2_{12} = 0.4$, $P = 0.55$, $n = 13$ measures per individual). Repeatability of Box-Cox transformed [Hp] using the same samples and null model equaled 0.17 ($\chi^2_{12} = 18.7$, $P < 0.0001$). With these transformed data and the full model, the interaction between sex and sample month ($\chi^2_{12} = 35.0$, $P = 0.0005$) and the plasma redness index ($\chi^2_{12} = 13.4$, $P = 0.0002$) were significant, and repeatability increased to 0.21 ($\chi^2_{12} = 24.2$, $P < 0.0001$).

3.1.2. Shorter sample series: effects of time of year and number of measurements

We calculated repeatabilities for all possible sub-series ($n = 77$) that comprised between two and twelve consecutive samples (Fig. 2). Among all sub-series, repeatability of untransformed [Hp] ranged from 0.00 to 0.53 (mean = 0.09, s.d. = 0.12). Of these 77 repeatability values, 16 were significant ($\chi^2_{1} > 4.1$, $P < 0.041$). After Box-Cox transformation, repeatability ranged from 0.00 to 0.57 (mean = 0.19, s.d. = 0.14, Fig. 2), and 41 of the 77 repeatability values were significant ($\chi^2_{1} > 4.3$, $P < 0.037$), with 14 of these being highly significant ($\chi^2_{1} > 15.0$, $P < 0.0001$). The maximum repeatability values calculated using raw and transformed data arose from the same two consecutive samples: December 2007 (sample 8) and January 2008 (sample 9).

3.1.3. Other sources of variation: effects of sex and sample month

We explored the effects of sex and sample month on repeatability of baseline [Hp] (Table 1). Here we focused on a “long series” that comprised ten consecutive but non-overlapping samples over one year (July 2007 to June 2008, samples 4 to 13) and a “short series” of the two consecutive samples that showed the highest repeatability (December 2007 to January 2008, samples 8 and 9). Box–Cox transformation increased repeatability when calculated with null models over both of these series (untransformed long, $R = 0.07$; untransformed short, $R = 0.53$; compared with transformed results shown in Table 1, model A); thus, we used the transformed [Hp] data in further analyses. In these analyses of the short series (Table 1, models B, C, and D), repeatability consistently remained high (0.50 to 0.57) and significant (all $\chi^2_{1} \geq 4.6$, $P < 0.032$), and neither sex nor sample month were significant (all $\chi^2_{1} \leq 2.9$, $P > 0.088$).

In the analyses of the long series, sex and sample month accounted for significant variation in [Hp] and impacted repeatability in terms of magnitude ($0.13 \leq R \leq 0.28$) but not in terms of significance (all $\chi^2_{1} \geq 8.1$,

Fig. 1. Variability among individuals (A) and among samples and between sexes (B) in baseline haptoglobin concentration ([Hp]). In (A) box plots of Box-Cox transformed data ($\lambda = 0.06$) are plotted in accordance with the y-axis on the left. A second y-axis on the right, which is still scaled to the transformation, provides untransformed values in original units (mg mL$^{-1}$). In (B) the means of raw data are plotted with error bars indicating the standard error.

Fig. 2. Summary of repeatability values of baseline haptoglobin concentration ([Hp]). Repeatability, R, equaled (variance among / (variance within + variance among)). Variances came from linear mixed effects models. Box-Cox transformed data ($\lambda = 0.06$) and the null model (i.e., A in Table 1) were used to calculate repeatability for all possible series that comprised multiple (between two and thirteen) consecutive samples. The x-axis identifies the first sample in a series (by month and year); the y-axis identifies the last sample in a series (by month and year). The outermost diagonal provides the repeatability values for each two-sample series; the lower-leftmost bubble provides the repeatability over the complete sample series. Magnitude of repeatability is indicated by the diameter of the bubbles.
Effects of additional explanatory variables on repeatability of haptoglobin concentration ([Hp]) over two series. Repeatability, R, equalled \( \text{variance}_{\text{among}} / (\text{variance}_{\text{within}} + \text{variance}_{\text{among}}) \). Variances came from linear mixed effects models and Box-Cox transformed data (\( \lambda = 0.06 \)). Significance of repeatability was determined by contrasting models with and without individual as a random factor, using likelihood-ratio tests and \( \chi^2 \)-statistics. Likelihood-ratio tests and \( \chi^2 \)-statistics were also used to test the significance of the explanatory variables by contrasting models with and without each variable.

**Table 1**

<table>
<thead>
<tr>
<th>Model</th>
<th>Explanatory Var.</th>
<th>( \chi^2 )</th>
<th>R</th>
<th>( \chi^2 )</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Null</td>
<td>1 17.9 -0.0001 0.22</td>
<td>1 6.2 0.0130 0.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>Sex</td>
<td>1 6.3 0.0124</td>
<td>1 2.9 0.0887</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Repeatability</td>
<td>1 8.1 0.0044 0.13</td>
<td>1 4.6 0.0316 0.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>Month</td>
<td>9 36.5 -0.0001</td>
<td>1 0.1 0.8155</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Repeatability</td>
<td>1 26.9 -0.0001 0.28</td>
<td>1 6.2 0.0128 0.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>Plasma redness*</td>
<td>9 31.5 0.0002</td>
<td>1 0.0 0.9241</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Month*Sex</td>
<td>9 31.5 0.0002</td>
<td>1 0.0 0.9241</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Repeatability</td>
<td>1 22.2 -0.0001 0.25</td>
<td>1 4.7 0.0308 0.50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**3.2. Predictive capacity**

3.2.1. Capacity of baseline haptoglobin to predict response haptoglobin

Baseline [Hp] exhibited some capacity to predict response [Hp] following endotoxin challenge (Fig. 3). Of the complete set of correlations (n = 91), 41 exceeded the threshold of |\( \tau | = 0.5, and the overall median value of \( \tau \) was 0.50 (Table 2). Correcting for redness in the endotoxin-response plasma samples slightly improved the predictive capacity: 43 exceeded the threshold of |\( \tau | = 0.5, but the overall median was unchanged (Table 2). The predictive capacities of baseline [Hp] from different single months (the outermost diagonal in Fig. 3) revealed a peak in \( \tau \) over several consecutive autumn samples (August, September, and November 2007); this autumn-centered pattern was relatively robust to increases in sample series length (i.e., the addition of one or more months to one or both ends of the series). In general, the predictive capacity of baseline [Hp] for response [Hp] increased as more consecutive samples (i.e., longer sample series) were used to calculate the predictor (Fig. 3). The only negative relationships (albeit weak ones) were observed between baseline and response [Hp] in summer 2007 (April, May, June, and/or July).

3.2.2. Capacity of baseline haptoglobin to predict other physiological responses

None of the physiological responses correlated with baseline [Hp] to the same extent as response [Hp] (Table 2). Four response variables were predicted to a limited extent by baseline [Hp]: S. aureus killing ability, cloacal temperature, hemolysis, and ROMs. With these four variables, median values for \( r \) ranged from 0.36 (ROMs) to 0.43 (S. aureus killing ability), and the number of correlations for which |\( \tau | > 0.5 \) ranged from nine (ROMs) to 34 (S. aureus killing ability) (Table 2). One response variable, E. coli killing ability, exhibited an inverse relationship with baseline [Hp]. With E. coli killing ability, the median value for \( \tau \) was −0.33, and with six correlations |\( \tau | < 0.5 \) were shown. Median |\( \tau | \) informs about the overall strength of the correlation, median \( \tau \) informs about the overall direction of the correlation, maximum and minimum \( \tau \) tell about the overall variability.

---

**Fig. 3**

Summary of the capacity of baseline haptoglobin concentration ([Hp]) to predict response haptoglobin concentration. All predictive capacities (i.e., correlations) are based on the same response data, but each ellipse represents a different predictor dataset. Predictor datasets comprised either a single baseline sample or multiple consecutive baseline samples. In the case of the latter, mean baseline haptoglobin concentration values were calculated and used as predictors. The x-axis identifies the first sample in a series (by month and year); the y-axis identifies the last sample in a series (by month and year). All series are inclusive of all samples between the first and last. For example, the Apr-07 to Jul-07 series (i.e., the leftmost ellipse containing a \( \chi^2 \)) also includes May-07 and June-07 values in the baseline haptoglobin concentration mean. The outermost diagonal provides the predictive capacities for each single baseline sample; the lower-leftmost ellipse provides the predictive capacity of the mean baseline haptoglobin concentration value calculated from the complete series of thirteen samples. All haptoglobin concentration data are untransformed and not corrected for sample redness. Strength of correlation (Kendall’s \( \tau \)) is indicated by the narrowness and shading (black, |\( \tau | > 0.5\); gray, |\( \tau | = 0.5\); white, |\( \tau | < 0.5\)) of the ellipse. Direction of correlation is indicated by the direction of the ellipse (negative (−) and zero (0) correlations values are also noted).

---

**Table 2**

Summary of the correlations between baseline haptoglobin concentration ([Hp]) and response variables. Kendall’s \( \tau \) was used to evaluate 91 correlations per response variable. The number and percentage of correlations stronger than 0.5 are shown. Median |\( \tau | \) informs about the overall strength of the correlation, median \( \tau \) informs about the overall direction of the correlation, maximum and minimum \( \tau \) tell about the overall variability.

<table>
<thead>
<tr>
<th>Response var.</th>
<th>No.</th>
<th>No.</th>
<th>%</th>
<th>Median</th>
<th>Median</th>
<th>Max</th>
<th>Min</th>
<th>( \tau )</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Hp], redness corrected</td>
<td>43 0 47 0.50 0.50</td>
<td>0.93</td>
<td>0.50</td>
<td>0.93</td>
<td>0.50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>[Hp]</td>
<td>41 0 45 0.50 0.50</td>
<td>0.79</td>
<td>0.50</td>
<td>0.79</td>
<td>0.50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S. aureus killing ability</td>
<td>30 0 37 0.43 0.43</td>
<td>0.81</td>
<td>0.24</td>
<td>0.81</td>
<td>0.24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cloacal temperature</td>
<td>15 0 16 0.42 0.42</td>
<td>0.87</td>
<td>0.04</td>
<td>0.87</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hemolysis</td>
<td>14 0 15 0.40 0.40</td>
<td>0.62</td>
<td>0.25</td>
<td>0.62</td>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ROMs</td>
<td>11 0 10 0.36 0.36</td>
<td>0.79</td>
<td>0.14</td>
<td>0.79</td>
<td>0.14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. coli killing ability</td>
<td>1 5 7 0.33</td>
<td>−0.33</td>
<td>0.62</td>
<td>−0.33</td>
<td>0.62</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oxygen consumption</td>
<td>3 0 3 0.14 0.07</td>
<td>0.57</td>
<td>0.43</td>
<td>0.57</td>
<td>0.43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hematocrit correction</td>
<td>0 2 2 0.22</td>
<td>−0.15</td>
<td>0.37</td>
<td>−0.15</td>
<td>0.37</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glucose</td>
<td>0 1 1 0.21</td>
<td>−0.21</td>
<td>0.36</td>
<td>−0.21</td>
<td>0.36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mass loss</td>
<td>0 0 0 0.07 0.00</td>
<td>0.36</td>
<td>0.43</td>
<td>0.36</td>
<td>0.43</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In order to explore similarities and differences in the effects of sample series length and time of year, all 91 values of \( \tau \) were plotted (as in Fig. 3) for each of the five responses outlined above (not shown). Visual examination of these plots revealed that predictive capacity generally increased when more consecutive samples (i.e., longer sample series) were used for calculating average baseline [Hp]. Another similarity occurred during summer 2007 (April, May, June, and/or July) when all five responses showed correlations with baseline [Hp] that were the reverse of all other series (i.e., negative for S. aureus killing ability, cloacal temperature, hemolysis, and ROMs; positive E. coli killing ability). Other similarities that were clear and consistent among responses were more difficult to detect. For example, evaluating and comparing among responses the predictive capacities of single-month samples (i.e., equivalent to outermost diagonal in Fig. 3) indicated that each response was predicted best by a different single-month sample.

Four response variables were predicted particularly poorly by baseline [Hp]: oxygen consumption, hemagglutination, blood glucose, and mass loss. With these four variables, median values for \( \tau \) ranged from -0.21 (glucose) to 0.07 (oxygen consumption), and there were only three or fewer correlations per parameter for which \(|\tau|\) exceeded 0.50 (Table 2).

### 4. Discussion

#### 4.1. Strengths and weaknesses of measuring baseline [Hp]

Haptoglobin is an easily measured index of immune function that can be incorporated into experimental or comparative studies of individuals, populations, or species. Our study identified additional merits that bolster the overall usefulness of assaying Hp. First, the maximum repeatability suggests that this immunological index is a distinctive trait of individuals. Second, baseline [Hp] values have the capacity to predict endotoxin-induced changes in concentration of this acute phase protein. Our analysis, however, also identified two important limitations associated with baseline [Hp] quantification and interpretation. First, the high repeatability, which we found in winter and over the short-term, decreased markedly when we calculated it in other times of year and over the intermediate- and long-term. Second, the predictive capacity of baseline [Hp] values did not extend to other endotoxin-induced physiological changes, such as inflammation-related increases in energy metabolism. Taken together, these newly-illustrated strengths and weaknesses of baseline [Hp] data should lead to more refined conclusions from studies of the ecology and evolution of immune function that employ this measure.

#### 4.2. Repeatability

Of all [Hp] repeatability calculations, the maximum value (\( R = 0.57 \)) was calculated between two consecutive samples that were collected 29 days apart. The second highest repeatability value (0.52) was calculated among three consecutive samples: the two samples that exhibited the maximum repeatability and the preceding sample that was collected 48 days prior. The inter-sample time span of the maximum repeatability in the current study is intermediate in relation to the inter-measurement time spans of other recent studies investigating repeatability of avian physiological parameters, which can span days, months, seasons, or years (Bech et al., 1999; Hõrak et al., 2002; Norte et al., 2008; Ochs and Dawson, 2008; Versteegh et al., 2008; Broggi et al., 2009; Hatch and Smith, 2010; Tieleman et al., 2010; Berzins et al., 2011; Bushuev et al., 2011; Horrocks et al., 2011). These studies report, for at least some parameters, significant repeatabilities that are similar to the maximum value reported in the current study. We conclude that baseline [Hp] can be seen as a distinctive trait of individuals; therefore meaningful inter-individual correlations with other distinctive (i.e., repeatable) traits should be possible, at least during this particular window of time.

In general, intra-individual repeatability of avian plasma protein concentrations is less-well studied than the repeatability of other physiological parameters (e.g., metabolic rate). Repeatability of protein concentrations is often significant over the short-term, but the measured proteins are sometimes rather broadly categorized (Ots et al., 1998; Hõrak et al., 2002; Norte et al., 2008). We are aware of only two instances in which repeatabilities of concentrations of a single (functionally-defined) acute phase proteins have been previously reported for birds. One study shows that ovotransferrin is highly significantly repeatable (\( R = 0.32 \)) when measured using four of the same samples as in the current study (November 2007, December 2007, January 2008, and February 2008; Horrocks et al., 2011). The other study reports non-significant repeatabilities for haptoglobin both within and between bouts of incubation by a free-living seabird (Cepphus grilis, Berzins et al., 2011).

As series increased in length and included more consecutive samples, repeatability of baseline [Hp] generally declined, a phenomenon similar to previous reports (Hõrak et al., 2002; Norte et al., 2008; but see also Rannig et al., 2005). In the 10 sample “long” series, accounting for sample month explained a significant amount of variation in baseline [Hp] values but improved repeatability only to a limited degree. Thus, the diminished repeatability exhibited by long series could not be brought up to the levels exhibited by some shorter series simply by reducing seasonal variation. Consequently, intra-individual variation that does not adhere tightly to synchronized monthly or seasonal cycles is likely to blame for the diminished repeatability that we observed. This type of variation might be attributed to inflammation events that are either stochastic in nature or at least not highly synchronized among all birds. Even among captive animals housed in close confines, highly synchronized (e.g., ±6–12 h) acute phase responses and attendant spikes in [Hp] are not probable given the time required for most disease transmissions and the short duration of avian acute phase responses (Nomoto, 1996; Adler et al., 2001).

The time of year also affected repeatability. When we considered the repeatability values based on two consecutive samples, we found that the majority of these values were low (median = 0.20), and four were very low (<0.1); May–June, June–July, July–August 2007 and March–April 2008). The three repeatabilities that exceeded 0.40 (January–February 2008 (0.41), August–September 2007 (0.43), and December 2007–January 2008 (0.57)) included samples collected between late summer and mid-winter. This autumn-centered pattern is reinforced by the repeatabilities of other short and mid-length series (three to nine consecutive samples). Overall, two winter samples (December 2007 and January 2008) factored prominently in many high repeatability values. Of the 30 repeatabilities greater than 0.25, 26 included one or both of these months, which is more than expected by chance alone (i.e., 87% versus the 65% of all repeatability values that include one or both of these months, \( P = 0.033 \), Fisher’s exact test).

The high repeatability associated with the relatively low [Hp] in winter suggests that this is a quiescent period in terms of inflammatory responses in pigeons. Winter measurements have been used in other studies of repeatability (Versteegh et al., 2008; Tieleman et al., 2010). In these studies, the winter was deemed to be a quiescent period based on an understanding of the study subject’s biology, and repeatabilities were not compared among different times of the year or among sample series of different lengths. Quiescent periods characterized by limited inflammatory responses could result from a reduction in the magnitude or frequency of inflammatory challenges, a dampening in the immunological responsiveness to a challenge, or both. Previous studies show immunological responsiveness is maintained year round. For example, birds can mount energetically expensive immune responses in winter (Svensson et al., 1998; Ots et al., 2001; Hegemann unpublished data), and the magnitude of the response to a standard
challenge is constant across all stages of the annual cycle, including winter in *Alauda arvensis* (Hegemann unpublished data). And, while cold exposure limits antibody responses in *Cyanistes (Parus) caeruleus* (Svensson et al., 1998), cold exposure has little effect on innate immune function in *Calidris canutus* (Buehler et al., 2008). Overall, this apparent constancy of innate immunological responsiveness hints that periods of decreased (or increased) inflammation may instead be caused by seasonal dynamics in the challenges that trigger inflammation (Altizer et al., 2006). In the case of multicellular parasites and arthropod vectors, the low temperatures associated with temperate winters likely limit disease transmission (Altizer et al., 2006). Indeed, a study of a population of songbirds (*Cyanistes caeruleus*) in the United Kingdom reports a complete absence of avian malaria (*Plasmodium* spp.) infections during December and January (Cosgrove et al., 2008).

### 4.3. Predictive capacity

The predictive capacity of baseline [Hp] differed among physiological response variables, times of year, and sample series lengths (i.e., the number of consecutive repeated measures used to calculate a mean baseline value). Baseline [Hp] generally exhibited the capacity to predict post-challenge [Hp], but this capacity was highest from fall into winter. Furthermore, calculating mean baseline [Hp] using a greater number of consecutive repeats improved the predictive capacity for response [Hp]. No other response variable correlated with baseline [ Hp] as well as response [Hp], but cloacal temperature, hemolysis, ROMs, *S. aureus* killing ability, and *E. coli* killing ability exhibited some similarities to response [Hp] in terms of time of year and sample series length. Patterns of predicative capacity display temporal similarity with patterns of repeatability. That is, baseline [Hp] did the best job of predicting response [Hp] during roughly the same time of year that baseline [Hp] exhibited the highest repeatability. This similarity suggests that from fall to winter baseline [Hp] is both a good predictor of a future inflammatory response and a distinctive trait of individuals. Similarly, a study of free-living passerine (*Parus major*) chicks revealed that the blood cell parameters that are significant predictors of nestling survival to fledging also show the highest intra-brood repeatabilities (Nadolski et al., 2006). Yet in the current study, mean baseline [Hp] calculated over longer sample series generally served as better predictors, even though these longer sample series exhibited lower repeatabilities.

The distinction between those responses to endotoxin challenge that were predicted by baseline [Hp] and those that were not highlights an important interpretational limitation. Baseline [Hp] should probably not be viewed as a comprehensive forecaster of in vivo inflammation may instead be caused by seasonal dynamics in the challenges that trigger inflammation (Altizer et al., 2006). In the case of multicellular parasites and arthropod vectors, the low temperatures associated with temperate winters likely limit disease transmission (Altizer et al., 2006). Indeed, a study of a population of songbirds (*Cyanistes caeruleus*) in the United Kingdom reports a complete absence of avian malaria (*Plasmodium* spp.) infections during December and January (Cosgrove et al., 2008).

### 4.3. Predictive capacity

The predictive capacity of baseline [Hp] differed among physiological response variables, times of year, and sample series lengths (i.e., the number of consecutive repeated measures used to calculate a mean baseline value). Baseline [Hp] generally exhibited the capacity to predict post-challenge [Hp], but this capacity was highest from fall into winter. Furthermore, calculating mean baseline [Hp] using a greater number of consecutive repeats improved the predictive capacity for response [Hp]. No other response variable correlated with baseline [Hp] as well as response [Hp], but cloacal temperature, hemolysis, ROMs, *S. aureus* killing ability, and *E. coli* killing ability exhibited some similarities to response [Hp] in terms of time of year and sample series length. Patterns of predicative capacity display temporal similarity with patterns of repeatability. That is, baseline [Hp] did the best job of predicting response [Hp] during roughly the same time of year that baseline [Hp] exhibited the highest repeatability. This similarity suggests that from fall to winter baseline [Hp] is both a good predictor of a future inflammatory response and a distinctive trait of individuals. Similarly, a study of free-living passerine (*Parus major*) chicks revealed that the blood cell parameters that are significant predictors of nestling survival to fledging also show the highest intra-brood repeatabilities (Nadolski et al., 2006). Yet in the current study, mean baseline [Hp] calculated over longer sample series generally served as better predictors, even though these longer sample series exhibited lower repeatabilities.

The distinction between those responses to endotoxin challenge that were predicted by baseline [Hp] and those that were not highlights an important interpretational limitation. Baseline [Hp] should probably not be viewed as a comprehensive forecaster of inflammation or a marker of immunity (sensu Lee, 2006). Notably, ecologically relevant changes in metabolic rate were not predicted by baseline [Hp]. It is possible that the responses that were uncorrelated with baseline [Hp] were not themselves distinctive characteristics of individuals. For among individual correlations to be meaningful, both the predictor and the response should be repeatable. However, calculating within-individual repeatability of an in vivo inflammatory response is not straightforward since subsequent immunological responses to a particular endotoxin challenge are expected to differ in terms of magnitude and mechanism (e.g., differences between primary, secondary, and hyper-vaccination responses; but see Cucco et al., 2008). With only a single challenge and a single response window per bird, repeatabilities of the responses could not be calculated in the current study. In summary, the predicative capacity results suggest baseline [Hp] can be a valuable addition to comparative studies that do not allow experimental manipulations, but care must be taken to avoid over-interpreting this index as a proxy for all of the physiological changes associated with endotoxin-induced inflammation.

### Acknowledgments

We thank V. Goerlich, B. Riedstra, and the Behavioural Biology Group for supplying the pigeons. We thank T.C. Broesamle and Animal Ecology Group members for pigeon care assistance. A. A. Cohen shared R code, which we adapted slightly to produce Fig. 3. Two anonymous reviewers provided useful suggestions that improved our initial submission. K.D.M. was supported by a Veni fellowship (863.08.026) from the Netherlands Organization for Scientific Research. B.L.T. was also supported a Veni Fellowship (863.04.023) and by a Rosalind Franklin Fellowship. J. van de Crommenacker and J. Komdeur (supported by WOTRO grant W82-289) provided the oxidant-antioxidant data.

### Appendix. The effects of plasma redness on haptoglobin quantification and repeatability

#### 1. Introduction

We investigated the effects of plasma redness on haptoglobin (Hp) quantification and repeatability. Plasma from hemolysed blood is redder and has higher levels of free heme. Because the Hp assay is based on the heme-binding capacity of Hp, erythrocyte lysis and the release of heme into blood plasma can affect the results. This point is noted in the instructions that are provided by the kit manufacturer.

#### 2. Methods

As indicated in the main methods section of the associated article, we measured absorbances at 450 nm prior to the addition of the final reagent that initiated the reaction. These pre-scan values functioned as an index of plasma redness. We examined variation within and among individuals in this index (450dev, total), and we calculated the index’s repeatability to determine if plasma redness is a characteristic of individuals. We then partitioned the variation in 450dev into a) averages among all samples per individual (450avg) and b) within-individual deviations of each sample from that individual’s average (450dev) (van de Pol and Wright, 2009). The three plasma redness parameters served as covariates in further analyses aimed at explaining variation in haptoglobin concentration ([Hp]). First, we compared and ranked models in their abilities to explain variation in [Hp]. To do this we used Akaike's information criterion (AIC) and a model set that included all permutations of sex, sample month, and their interaction in the presence and absence of each plasma redness covariate (ten models each, including the null, for 450avg, 450dev, and 450total). Second, we calculated and compared repeatability values derived from models that included redness covariates. We investigated repeatability of [Hp] over a long (10 samples) and a short (2 samples) sub-series (as explained in the associated article) and with and without the variables of sex and sample month. We calculated repeatability using the same protocols described in the associated article (i.e., linear mixed-effects (lme) models that included individual as a random factor; R Development Core Team 2008).

#### 3. Results

#### 3.1. Repeatability of plasma redness

Plasma redness ([Hp]) was significantly repeatable when calculated using the 16 individuals over the entire sample series: 0.11 ± 0.06 (χ² 1 = 8.5, P = 0.0036, n = 13 measures per individual). Sample month, when added to the model, significantly accounted for variation in plasma redness (χ² 1 = 70.1, P < 0.0001) and increased repeatability to 0.17 ± 0.07 (χ² 1 = 18.1, P < 0.0001).

#### 3.2. Ability of plasma redness parameters to explain [Hp]

Comparisons among all models (ΔAICc, Appendix Table A.1) revealed that 450dev best explained variation in [Hp]. Models that included
Appendix Table A.1

A summary of the models compared using Akaike’s information criterion (AIC). Model statistics are listed separately for each plasma redness covariate and models are ranked within each covariate by ΔAICc. [No 450avg statistics are shown, AICc was always >385] Equivalent models that differ only in plasma redness covariate are in the same row. AICc can be used to compare models between covariates. Superscript letters following some models identify the corresponding models in Table 1 of the associated article and Appendix Table A.2. All models included individual as a random factor. The analyzed dataset comprised sixteen individuals over the entire series of thirteen samples (n = 208).

<table>
<thead>
<tr>
<th>Model</th>
<th>No. of parameters</th>
<th>RED = 450total</th>
<th>RED = 450dev</th>
<th>RED = 450avg</th>
<th>450total - 450dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>RED + Sex + Month + Sex:Month</td>
<td>D,F</td>
<td>29</td>
<td>−154.66</td>
<td>377.10</td>
<td>0.00</td>
</tr>
<tr>
<td>RED + Sex + Month</td>
<td>17</td>
<td>−172.32</td>
<td>381.85</td>
<td>4.75</td>
<td>0.080</td>
</tr>
<tr>
<td>RED + Month</td>
<td>16</td>
<td>−174.33</td>
<td>383.50</td>
<td>6.40</td>
<td>0.035</td>
</tr>
<tr>
<td>Sex + Month + Sex:Month</td>
<td>28</td>
<td>−159.71</td>
<td>384.50</td>
<td>7.40</td>
<td>0.024</td>
</tr>
<tr>
<td>Sex + Month</td>
<td>16</td>
<td>−178.72</td>
<td>392.30</td>
<td>15.20</td>
<td>0.000</td>
</tr>
<tr>
<td>Month</td>
<td>15</td>
<td>−181.72</td>
<td>395.95</td>
<td>18.84</td>
<td>0.000</td>
</tr>
<tr>
<td>RED + Sex</td>
<td>15</td>
<td>−196.64</td>
<td>403.57</td>
<td>26.47</td>
<td>0.000</td>
</tr>
<tr>
<td>RED</td>
<td>14</td>
<td>−198.67</td>
<td>405.54</td>
<td>28.44</td>
<td>0.000</td>
</tr>
<tr>
<td>Sex</td>
<td>14</td>
<td>−203.43</td>
<td>413.06</td>
<td>37.96</td>
<td>0.000</td>
</tr>
<tr>
<td>Null</td>
<td>A</td>
<td>3</td>
<td>−206.43</td>
<td>418.97</td>
<td>41.87</td>
</tr>
</tbody>
</table>

450total were inferior overall and inferior compared to equivalent models that included 450dev (2−ΔAICc<4, Burnham and Anderson, 1998). All models that included 450avg were considered implausible (ΔAICc>4) when compared to the best overall model and to equivalent models that included 450dev or 450total (ΔAICc > 385 for all 450avg models). These results were reflected in the parameter weights of plasma redness covariates: 450dev = 0.996, 450total = 0.978, and 450avg = 0.419. The ten models in the 450dev analysis ranked in precisely the same order as the ten in 450total analysis. Overall, the only plausible models were the full models (sex, sample month, and their interaction) that included either 450dev (AICc = 373.75) or 450total (AICc = 377.10).

3.3. Effects of plasma redness parameters on repeatability of [Hp]
Our additional investigations into repeatability revealed that the effects of plasma redness parameters were limited in magnitude and inconsistent between sub-series (Appendix Table A.2). With the long (one year) series, the inclusion of a redness parameter (450total or 450dev) resulted in small (<0.05) increases in repeatability of [Hp] compared to the null model, and these redness parameters were significant when present as the only explanatory variables in a model (models E and G, Appendix Table A.2) and insignificant when present as covariates in models that also included sex, sample month, and the interaction between sex and sample month (models D and F, Appendix Table A.2). With the short (two sample) series, the redness parameters either reduced or had no effect on repeatability, and these parameters were never statistically significant.

4. Discussion

Plasma sample redness (i.e., 450total) was significantly repeatable; thus, the parameter can be viewed as a distinctive trait of individuals, with some pigeons having redder plasma than others. Plasma samples also differed in redness among months, and correcting for sample month increased the repeatability of plasma redness. The precise causes of this variation in redness could include methodological inconsistencies (e.g., in blood collection technique) or a range of general (e.g., nutrition status or sensitivity to stress) or more-specific (e.g., erythrocyte fragility) physiological differences.

An individual’s deviation from its average redness among samples, 450dev, always explained variation in [Hp] better than 450avg (or 450total). Thus, it is within-individual variation, not among-individual variation, in plasma redness that potentially affects the quantification of [Hp]. This role of within-individual variation in plasma redness, particularly after accounting for variation among sample months (i.e., model D, Appendix Table A.2), points to causes of variation that are circumstance-specific (i.e., a particular bird on a particular sampling day).

Including 450dev or 450total in statistical analyses did little to improve repeatability of [Hp] regardless of the terms’ statistical significance. Nonetheless, plasma sample redness holds the potential to bias results and affect conclusions. For example, population differences in erythrocyte fragility and lysis could be misinterpreted as population differences in [Hp]. Thus, we encourage those using this haptoglobin assay to photometrically evaluate plasma redness and consider its potential effects on the biological questions at hand. Distinctions between within- and among-individual variation in redness are only possible in studies with repeated measures within individuals. When repeated measures are unavailable, 450total, the simplest plasma redness covariate, can still provide valuable insights.

References


