Predicting the bending properties of long bones: Insights from an experimental mouse model

Sarah J. Peacock1 | Brittney R. Coats2 | J. Kyle Kirkland3 | Courtney A. Tanner4 | Theodore Garland, Jr.5 | Kevin M. Middleton1

1Department of Pathology and Anatomical Sciences, University of Missouri, Columbia, Missouri
2ITW David Speer Academy, Chicago, Illinois
3Department of Kinesiology, Michigan State University, East Lansing, Michigan
4Blivet Solutions, Inc., Chicago, Illinois
5Department of Evolution, Ecology, and Organismal Biology, University of California, Riverside, California

Correspondence
Sarah J. Peacock, Department of Pathology & Anatomical Sciences, University of Missouri, Medical Sciences M263, 1 Hospital Dr., Columbia, MO 65212, USA. Email: sjpd58@mail.missouri.edu

Funding Information
National Institutes of Health National Institute of Arthritis and Musculoskeletal Diseases, Grant/Award Number: 5G11HD052368-03; University of Missouri Department of Pathology and Anatomical Sciences; University of Missouri Bond Life Sciences Fellowship Program; University of Missouri Gus T. Ridgel Fellowship; California State University, San Bernardino; Associated Students Incorporated of California State University, San Bernardino; University of Missouri School of Medicine; National Science Foundation, Grant/Award Number: IOS-1121273

Abstract

Objectives: Analyses of bone cross-sectional geometry are frequently used by anthropologists and paleontologists to infer the loading histories of past populations. To address some underlying assumptions, we investigated the relative roles of genetics and exercise on bone cross-sectional geometry and bending mechanics in three mouse strains: high bone density (C3H/He), low bone density (C57BL/6), and a high-runner strain homozygous for the Myh4Minimsc allele (MM).

Methods and Materials: Weanlings of each strain were divided into exercise (wheel) or control (sedentary) treatment groups for a 7-week experimental period. Morphometrics of the femoral mid-diaphysis and mechanical testing were used to assess both theoretical and ex vivo bending mechanics.

Results: Across all measured morphological and bending traits, we found relatively small effects of exercise treatment compared to larger and more frequent interstrain differences. In the exercised group, total distance run over the experimental period was not a predictor of any morphological or bending traits. Cross-sectional geometry did not accurately predict bone response to loading.

Discussion: Results from this experimental model do not support hypothesized associations among extreme exercise, cross-sectional geometry, and bending mechanics. Our results suggest that analysis of cross-sectional phenotypes alone is insufficient to predict loading response, and questions the common assumption that cross-sectional geometry differences are indicative of differential loading history.

KEYWORDS
biomechanics, loading, mouse model, skeletal plasticity

INTRODUCTION

Bone response to loading has been a frequent research focus due to both its clinical and biological relevance (Burr, Martin, Schaffler, & Radin, 1985; Duncan and Turner, 1995; Lanyon, 1992; Lieberman, 2003; Meakin, Udeh, Galea, Lanyon, & Price, 2015; Sugiyama et al., 2012). Because intense and/or chronic skeletal loading can induce an anabolic response (Hamrick, Skedros, Pennington, & McNeil, 2006; Moss, 1997; Rubin, 1984; Rubin and Lanyon, 1985), analyses of bone morphology and cross-sectional geometry have been used as proxies for inferring a bone’s functional response to loading in a wide range of vertebrates (Kelly, Czech, Wight, Blank, & Garland, 2006; Middleton, Kelly, & Garland, 2008a; Middleton et al., 2008b; Simons, Hieronymus, & O’connor, 2011), including humans and other primates (Demes, Jungers, & Selpien, 1991; Demes, Jungers, & Walker, 2000; Holt, 2003; Maggiano et al., 2008; Polk et al., 2000; Ruff, Larsen, & Hayes, 1984; Ruff, 2009; Shaw and Stock, 2013; Stock and Pfeiffer, 2001; Trinkaus and Ruff, 1999). However, the extent to which cross-sectional phenotypes can be used to reconstruct past behaviors has been questioned by other studies (Demes et al., 1998; Lieberman, Polk, & Demes, 2004;
were chosen at weaning (~21 days) from eight dams (two per dam) that had undergone 12 generations of inbreeding with continued selection for high levels of voluntary wheel running (see Supporting Information 1 for additional information about inbreeding protocol and Swallow, Carter, and Garland (1998a) for details about the selection experiment prior to founding of the MM inbred line at generation 51). After 12 generations of brother-sister mating, we estimate that at least 93% of alleles are homozygous, following the equations adapted from Green (1981) for residual heterozygosity at the ith generation under brother–sister mating:

$$\text{Heterozygosity} = 1 - \frac{F_i}{2^{n-i}}$$

where $F_i$ is the $i$th observation in the Fibonacci sequence ($F_i = F_{i-1} + F_{i-2}$ given that $F_1 = 1$ and $F_2 = 2$). 93% homozygosity is likely an underestimate because after 51 generations of selection, the source HR line was likely already homozygous at many loci.

The remaining two inbred strains C3H/HeNHsd (hereafter "C3H/He"‡) and C57BL/6NHsd (hereafter "C57BL/6") mice were purchased from a commercial vendor (Harlan Laboratories) at weaning. These strains were chosen because they are standard models for the study of skeletal loading responses. They represent the ends of a continuum of bone mineralization among inbred strains (Beamer, Donahue, Rosen, & Baylink, 1996) and provide context for interpretation of the bending properties of MM bones.

Rodent chow (Teklad Rodent Diet 8604; Harlan) and water were supplied ad libitum. After a seven-week experimental period starting at 21 days, mice were euthanized via CO2 inhalation. All procedures were approved by the California State University, San Bernardino Institutional Animal Care and Use Committee.

### 2.2 Voluntary wheel-running behavior

Half of each group of mice was randomly assigned to either a wheel access group or a nonwheel control (sedentary) group ($n = 8$ of each strain in each group). Mice without wheels were housed in groups of four in standard cages according to strain, while mice given wheels were individually housed in cages that included a 0.73 m circumference running wheel (Nalgene Cages, Bend, OR; see Morgan, Garland, & Carter, 2003 for additional details). During the course of the experiment, five C3H/He mice died for unknown reasons, leaving a final group size of 11 ($n = 7$ nonwheel control; $n = 4$ wheel access). One MM histological section was damaged and could not be analyzed ($n = 8$ nonwheel control; $n = 7$ wheel access). All C57BL/6 mice were analyzed ($n = 16$).

Although single housing induces some stress and may alter baseline activity, it is necessary in order to measure an individual mouse’s activity. Single housing of sedentary animals was not possible due to space constraints. Mice were free to use the exercise wheel as much or as little as they chose at all times. The number of wheel revolutions was computed using a magnet attached to the wheel which triggered a reed switch connected to a digital I/O interface (USB-120A; Eagle Appliances Ltd., Cape Town, South Africa). Time stamps for each revolution were recorded to the nearest millisecond, using custom-written

---

**2 | METHODS**

### 2.1 Experimental design

We compared bone phenotypes in three strains of mice: (1) high-runner (HR) mice homozygous for the Myh4<sup>MM</sup> allele (hereafter "MM"‡), (2) a high bone density phenotype, C3H/HeNHsd, and (3) a low bone density phenotype, C57BL/6NHsd. MM mice display an approximately 50% reduction in the mass of the triceps surae muscle complex resulting from the near complete absence of type IIb muscle fibers (Guderley, Houle-Leroy, Diffee, Camp, & Garland, 2006; Guderley, Joanisse, Mokas, Bilodeau, & Garland, 2008; Bilodeau, Guderley, Joanisse, & Garland, 2009; Houle-Leroy, Guderley, Swallow, & Garland, 2003; Talmadge, Acosta, & Garland, 2014). Male MM mice
software. Revolutions were summed for each daily 24-h period starting at 1200 (noon) and ending at 1159 using a custom R package (http://github.com/kmiddleton/binning) and converted to meters for analysis (Figure 1). Cages were changed weekly, which resulted in slightly shorter running durations on those days; however, this gap occurred during the light phase when mice ran minimally.

### 2.3 Gross morphometrics and histomorphometry

Following euthanasia, both hind limbs of each mouse were dissected. In preparation for whole-bone morphometrics, left femora were manually defleshed, and traits were measured with digital calipers to 0.01 mm: femur length (superior articular surface of the femoral head to the farthest extent of the distal condyles), anteroposterior and mediolateral diameters at the mid-femoral diaphysis, proximal width (greater trochanter to the medial femoral head), distal width (across the distal condyles), and femoral head proximodistal height and anteroposterior depth. Left femora were subsequently frozen in saline for use in mechanical testing (see below).

Right femora were dried and defleshed using dermestid beetles. Once cleaned of residual muscle tissues, the bones were individually embedded in low viscosity epoxy (Epo-Thin; Buehler, Ltd., Lake Bluff, IL) under a vacuum and allowed to cure overnight. Approximately 1-mm-thick sections were cut from the mid-diaphysis of each femur below the third trochanter using a low-speed saw (IsoMet; Buehler, Ltd., Lake Bluff, IL).

The cross-sections were mounted on slides and imaged using light microscopy (Figure 2a). Each image was captured digitally and then prepared for analysis of cortical cross-sectional geometry (GNU Image Manipulation Program; www.gimp.org). Image color levels were adjusted to reduce background color and to improve bone contour visibility (Figure 2b). The images were then manually segmented to remove medullary contents (Figure 2c), and a threshold was applied to yield a black and white image of the cortex (Figure 2d). Repeatability of cross-sectional area measurements using this method was 0.99, estimated as the ratio of within group to total variance from a linear mixed model (Nakagawa and Schielzeth, 2010). Thus, despite slight variation in choice of threshold value and manual segmentation, this method yields highly repeatable results. Cross-sectional images were analyzed for area, second moments of area, and section modulus following Lieberman et al. (2004), using BoneJ (Doube et al., 2010).

### 2.4 Mechanical testing

Mechanical testing was performed on freshly thawed bones that had been previously frozen in 0.9% saline (see above). The authors acknowledge that the material properties of bone have been shown to change with hydration level. However, all bones were treated similarly and can be assumed to be equally moist. As our interest is not in the specific values of the mechanical properties tested, but rather in the differences between strains, this is sufficient for our needs. To test bending behavior, left femora were placed into a custom-built 3-point bending apparatus (following Turner and Burr, 2001) with a span of 6 mm and a support radius of 0.75 mm to avoid shear (Turner and Burr, 2001). Following the methods of Kodama et al. (2000), femora were loaded via a materials testing machine (Instron 5942; Instron, Inc., Norwood, MA) anteroposteriorly at a rate of 10 mm/min at the mid-femoral diaphysis until fracture, as determined by the point where the load rate dropped by 40% (Figure 3).

Raw load–displacement data were converted into flexure stress ($\sigma$) and flexure strain ($\varepsilon$) following Turner (1989) and Turner and Burr (2001), using the equations:

\[
\sigma = \frac{P \cdot s \cdot r}{4 \cdot T}
\]

\[
\varepsilon = \frac{6 \cdot D \cdot d}{5 \cdot s^2}
\]

where $P$ is the load, $s$ is the span length (6 mm), $r$ is the radius of the bone in the direction of loading, $I$ is the second moment of area about...
the axis of loading, \( D \) is the vertical displacement under load, and \( d \) is the diameter in the direction of loading. \( I \) was determined via histomorphometric analysis of the contralateral limb using BoneJ (see above). Although it would have been ideal to use the same leg for both tests, the logistical constraints of undecalcified bone histology precluded this. Previous studies in HR mice have shown significantly lower limb skeleton asymmetry than in randomly bred controls (Garland and Freeman, 2005), so we assume that minimal bias was introduced by using cross sectional data from the contralateral limb.

The yield point of a stress–strain curve is the point which separates the elastic and plastic loading regions (Gere and Timoshenko, 2001). This point was identified algorithmically with piecewise regression in R (R Core Team, 2017), using the package “segmented” (Muggeo, 2003). We used the slope of the elastic portion of the stress–strain curve (the portion of the curve to the left of the yield point) as an estimate of Young’s modulus, a measure of the intrinsic material stiffness. From the stress–strain curve, we estimated ultimate strength (maximum value of the stress–strain curve), ultimate strain

---

**FIGURE 2** Comparative femoral cross-sections. (a) An example of the raw femoral cross-section image data prior to thresholding. (b) The same cross-section image after color adjustment to reduce excess background color. (c) An outline has been added to the cross-section to indicate the areas of bone that will be kept as a silhouette; everything else is removed before analysis. (d) The final silhouette after background removal and conversion to black and white. (e) Comparative cross-sectional morphology of MM, C57BL/6, and C3H/He femora from mice of approximately the same body size (about 27 g). The C3H/He femur has significantly greater cortical thickness than the other two strains, which is not unexpected given that C3H/He mice are the high bone mass phenotype mouse model. In comparison, the C57BL/6 femur appears much more gracile, while the MM femur is somewhat intermediate to C3H/He and C57BL/6 morphology. Scale bar equals 0.5 mm.
(strain at ultimate strength), failure strength (stress at failure), failure strain (strain at failure), and modulus of toughness (area under the stress–strain curve) both to ultimate strength and to failure (Figure 3).

2.5 | Statistical analysis

We used a Bayesian inference framework for all statistical analyses. Bayesian analysis has gained traction in recent years with the development of tools to facilitate model definition and has been suggested as applicable for questions in physical anthropology (Konigsberg and Frankenberg, 2013). Benefits of a Bayesian approach for our analysis include (1) allowing comparisons of posterior distributions of differences between strains or exercise vs. sedentary treatment without reliance on specific post-hoc tests or intermediate ANOVA or ANCOVA models (Kruschke, 2015; McElreath, 2015), (2) obviating the need for post-hoc multiple comparisons adjustments (Gelman and Tuerlinckx, 2000; Gelman, Hill, & Yajima, 2012), and (3) facilitating the interpretation of parameter estimates relative to a value of interest (e.g., difference = 0), using a 95% highest density interval (HDI), which includes the most credible values for the parameter estimate. In this case, a 95% HDI, which excludes 0, constitutes credible evidence that parameter value is different from 0 (Kruschke, 2015).

We used three types of linear models: (1) Prediction of single traits (total distance run, body mass, femoral length) by strain, by wheel access, or by the full factorial model (one- or two-way ANOVA-like models), (2) prediction of traits by strain and wheel access with femoral length as a continuous covariate (an ANCOVA-like model), and (3) prediction of traits by total cumulative distance for animals in the wheel treatment group only (results presented in Supporting Information, 2). The authors recognize that body mass is a more commonly-used covariate than femoral length within anthropological research; however, as body mass is more highly influenced by activity levels, and considering that variation in activity is an important component of our experimental design, we decided that using femur length was more appropriate in the context of this study. In the final set of models, we were interested in the slope parameter predicting traits from total cumulative distance. These models were run for each mouse strain separately, to avoid confounding strain with total distance (Figure 1). We neither calculated nor report p values, but in their stead include summaries of posterior distributions of models: priors conditioned on the observed data. Reporting parameter estimates rather than significance levels for specific null hypothesis tests allows us to avoid multiple comparison issues (Gelman et al., 2012; Kruschke, 2015).

In a Bayesian framework, each model generates posterior distributions of parameter estimates from observations (data) conditioned on priors using Markov chain Monte-Carlo (MCMC), where the priors are the initial probabilities for each possible value of each parameter. Because each MCMC sample contains simultaneous estimates for all parameters in a model, differences between estimates (e.g., effect of wheel access) can be calculated directly from posterior distributions (Figure 4). We summarized the differences between posterior distributions for comparisons of interest using the median and 95% highest density interval (HDI), which excludes 0, constitutes credible evidence that parameter value is different from 0 (Kruschke, 2015).

We used three types of linear models: (1) Prediction of single traits (total distance run, body mass, femoral length) by strain, by wheel access, or by the full factorial model (one- or two-way ANOVA-like models), (2) prediction of traits by strain and wheel access with femoral length as a continuous covariate (an ANCOVA-like model), and (3) prediction of traits by total cumulative distance for animals in the wheel treatment group only (results presented in Supporting Information, 2). The authors recognize that body mass is a more commonly-used covariate than femoral length within anthropological research; however, as body mass is more highly influenced by activity levels, and considering that variation in activity is an important component of our experimental design, we decided that using femur length was more appropriate in the context of this study. In the final set of models, we were interested in the slope parameter predicting traits from total cumulative distance. These models were run for each mouse strain separately, to avoid confounding strain with total distance (Figure 1). We neither calculated nor report p values, but in their stead include summaries of posterior distributions of models: priors conditioned on the observed data. Reporting parameter estimates rather than significance levels for specific null hypothesis tests allows us to avoid multiple comparison issues (Gelman et al., 2012; Kruschke, 2015).

In a Bayesian framework, each model generates posterior distributions of parameter estimates from observations (data) conditioned on priors using Markov chain Monte-Carlo (MCMC), where the priors are the initial probabilities for each possible value of each parameter. Because each MCMC sample contains simultaneous estimates for all parameters in a model, differences between estimates (e.g., effect of wheel access) can be calculated directly from posterior distributions (Figure 4). We summarized the differences between posterior distributions for comparisons of interest using the median and 95% highest density interval (HDI), which excludes 0, constitutes credible evidence that parameter value is different from 0 (Kruschke, 2015).

We used three types of linear models: (1) Prediction of single traits (total distance run, body mass, femoral length) by strain, by wheel access, or by the full factorial model (one- or two-way ANOVA-like models), (2) prediction of traits by strain and wheel access with femoral length as a continuous covariate (an ANCOVA-like model), and (3) prediction of traits by total cumulative distance for animals in the wheel treatment group only (results presented in Supporting Information, 2). The authors recognize that body mass is a more commonly-used covariate than femoral length within anthropological research; however, as body mass is more highly influenced by activity levels, and considering that variation in activity is an important component of our experimental design, we decided that using femur length was more appropriate in the context of this study. In the final set of models, we were interested in the slope parameter predicting traits from total cumulative distance. These models were run for each mouse strain separately, to avoid confounding strain with total distance (Figure 1). We neither calculated nor report p values, but in their stead include summaries of posterior distributions of models: priors conditioned on the observed data. Reporting parameter estimates rather than significance levels for specific null hypothesis tests allows us to avoid multiple comparison issues (Gelman et al., 2012; Kruschke, 2015).

In a Bayesian framework, each model generates posterior distributions of parameter estimates from observations (data) conditioned on priors using Markov chain Monte-Carlo (MCMC), where the priors are the initial probabilities for each possible value of each parameter. Because each MCMC sample contains simultaneous estimates for all parameters in a model, differences between estimates (e.g., effect of wheel access) can be calculated directly from posterior distributions (Figure 4). We summarized the differences between posterior distributions for comparisons of interest using the median and 95% highest density interval (HDI), which excludes 0, constitutes credible evidence that parameter value is different from 0 (Kruschke, 2015).
FIGURE 5  Summarized linear models of gross and histomorphometry. Results of Bayesian linear models with femoral length as a continuous covariate are summarized by the median (solid circle) and 95% HDI (bars). (a) Slope estimates for the covariate femoral length are compared to the null hypothesized isometric slope (open circle). In only one trait (distal femoral width) did the 95% HDI not include the null slope value. (b) Summaries of posterior distributions of differences of wheel: no wheel treatment groups for each mouse strain show that very few traits show an HDI that deviates from 0, indicating very small differences between treatment groups. The range of the x axis is equal between Figures 5b and 6 to facilitate comparisons.

FIGURE 6  Summarized linear models of gross and histomorphometry. In pairwise comparisons of differences of HDIs for (a) sedentary and (b) wheel access groups, each line summarizes the difference between one pair of strains in one treatment group (e.g., MM: C3H/He). HDIs that do not cross 0 are interpreted as having strong credibility that the difference deviates from 0. For example, AP diameter is larger in MM than in either C3H/He or C57BL/6 in both sedentary and wheel access treatments, but the latter two are not different from each other. The range of the x axis is equal between Figures 5b and 6 to facilitate comparisons.
density interval (HDI). When the parameter of interest (e.g., difference $\neq 0$) falls outside the HDI, we interpret this situation as plausible evidence that the parameter estimate does not include the value in question. For example, Figure 4 shows the posterior densities for body mass estimates in wheel access and sedentary MM mice and the difference between the posterior estimates. The 95% HDI for the difference does not encompass 0, which we interpret as credible evidence for lower mass in the wheel access treatment group. The most credible difference is $\Delta = 4.2$ g.

With the exception of body mass, which was estimated on the original scale, all morphological variables were log$_{10}$-transformed and centered before analyses. Centering (subtracting the grand mean from all values) improves model fitting while preserving estimation of slopes for covariates on the original scale. Cumulative distance run was standardized ($Z$) prior to analysis, because centering did not adequately account for the very large variance differences in cumulative distances between strains. For all parameters, we specified broad but weakly regularizing priors that temper the effects of extreme values: normal distributions centered on 0 for most parameters, normal distributions centered on null hypothesized isometric slopes for continuously varying covariates, and Cauchy distributions with a shape parameter of 2 for standard deviations. Models were initially run with four replicate chains for 2,000 MCMC iterations with 50% for warmup to assess suitability of prior specifications and check for adequate sampling, judged from trace plots, and ratio of between chain to within chain variance (Check : $R \approx 1$; Gelman and Rubin, 1992; Brooks and Gelman, 1998). To fully explore posterior distributions, final models were run as a single chain for 200,000 iterations, also with 50% warmup, yielding 100,000 samples from the posterior distribution for all parameters. HDI were calculated from these distributions.

All statistical analyses were performed in R version 3.3.2 (R Core Team, 2016). Bayesian inference was carried out using the rstan package for R (version 2.14.1; https://cran.r-project.org/web/packages/rstan/index.html), an interface to Stan, a Hamiltonian Monte Carlo sampler (Carpenter et al., 2015; Hoffman and Gelman, 2014). We used the rethinking R package (version 1.59; https://github.com/mcelreath/rethinking), which includes convenience functions for building, sampling, and summarizing models (McElreath, 2015). All code for models and analysis is included as Supporting Information.

3 | RESULTS

3.1 | Body mass

Across all strains, mice with access to a running wheel were about 4 g lighter than the sedentary controls: MM $= -3.9$ g (95% HDI of Wheel vs Sedentary: $-6.3$ to $-1.5$ g; Figure 4), C3H/He $= -4.0$ g (95% HDI: $-6.9$ to $-1.2$ g), C57BL/6 $= -4.5$ g (95% HDI: $-6.7$ to $-2.2$ g). This pattern of reduced body mass with access to running wheels has been previously observed in HR mice (e.g., Koteja, Swallow, Carter, & Garland, 1999; Swallow, Garland, Carter, Zhan, & Sieck, 1998b; Swallow, Koteja, Carter, & Garland, 2001; and subsequent); however, here we find a similar pattern and magnitude in the two fully inbred mouse strains as well. Descriptive statistics (means and standard errors of the mean) for all traits analyzed are included in Supporting Information, Table 1.
3.2 | Gross morphometrics and histomorphometry

Bayesian linear models for the gross morphometric (with femoral length as a covariate) and histomorphometric analyses are summarized in Figures 5 and 6 and Supporting Information, Table 1. Overall, we find that 95% HDIs of covariate estimates (i.e., slopes of trait vs femoral length) encompass the null hypothesized isometric slope value, with the exception of distal femoral width, which scales with negative allometry (Figure 5a). Across all measures of limb bone robusticity (diameters, widths, areas, and second moments of area), we find relatively few effects of activity within the three strains of mice (Figure 5b). Indeed, the only HDIs that did not include the null hypothesized difference of 0 were AP femoral diameter and femoral head depth in C3H/He mice. In stark contrast, we find widespread differences in morphometric traits between strains (Figure 6). Because the linear model included a strain by activity interaction term, differences were calculated pairwise between the strains separately for each activity group. Pairwise comparisons among both sedentary and exercise treatment groups show differences (HDIs that do not include 0) between all combinations of strains (Figure 6a,b). In general, we find that C3H/He mice have mediolaterally narrower femora than C57BL/6 or MM but are larger in cross-sectional area than C57BL/6 (but not MM). Lower cross-sectional areas in C3H/He mice lead to lower overall second moments of area than in the other two strains (Figure 6).

3.3 | Mechanical testing

Three-point bending analyses revealed moderate differences in estimated biomechanical traits but a similar overall pattern as the morphometric analyses above (Figure 7 and Supporting Information, Table 1). Within mouse strains, only 2 of 21 comparisons deviated from the null difference of 0: ultimate strain in C57BL/6 and ultimate toughness in C3H/H3 (Figure 7a). These results contrast with pairwise comparisons among the strains for sedentary and wheel access treatments (Figure 7bc), wherein we find far-more HDIs that exclude 0 (16 of 42). These results demonstrate that estimated Young’s modulus is credibly higher in C3H/He mice than in either MM or C57BL/6, and that MM is also credibly higher than C57BL/6. Furthermore, C3H/He strain mice show consistent patterns indicative of their relatively stiff bones, including reduced ultimate strain and higher modulus of toughness relative to MM, and lower failure strain than C57BL/6 (Figure 7).

3.4 | Total distance

In the wheel access treatment only, analyses of gross and histomorphometric traits and biomechanical loading traits were repeated separately for each strain with standardized total distance run as a continuous covariate. It was necessary to analyze each strain separately because strain is confounded with total distance (see “Voluntary running behavior” in Supporting Information). In all morphometric and bending traits, HDIs of the linear regression slope included the null value of zero, indicating that there was minimal association between total cumulative distance and the measured traits (Supporting Information, Figure 1). Although some median slope estimates deviated substantially from 0 (Supporting Information, Figures 2 and 3), they were associated with very wide 95% HDIs (Supporting Information, Figure 1).

4 | DISCUSSION

4.1 | Cross-sectional morphology and the exercise response

The goals of this study were to examine the connections between locomotor behavior and observed cross-sectional geometry after several weeks of “training” opportunity and the relationship between inferred and directly measured bone loading performance. Our overall aim was to better define the limits of inference about bone mechanics and locomotor behavior from observed bone morphology.

Previous studies have shown significant differences in cross-sectional properties and loading responses between C3H/He and C57BL/6 mice (Akhter et al., 2000; Kodama et al., 2000; Sheng et al., 1999), and our results agree with these and are here extended to include MM mice. Our first hypothesis, that mice with different genetic backgrounds will exhibit differences in cross-sectional morphology and histomorphometrics, was widely supported (Figure 5). For example, C3H/He mice showed the highest values for cross-sectional area, which is consistent with their observed comparative cross-sectional morphology (Figures 2e and 6). C3H/He mice also exhibited the lowest values for both anteroposterior and mediolateral second moments of area, which is also predicted given their smaller femoral diameters than C57BL/6 or MM strains (Figure 6). MM mice had higher mediolateral second moment of area values than both C3H/He and C57BL/6 mice but were similar to C57BL/6 mice in their anteroposterior second moment of area values.

In general, we find much more prevalent differences in whole-bone measurement and cross-sectional properties comparing among strains than we do within strains with exercise treatment (see below). This result is relevant to our understanding of how to best interpret observed variation in bone phenotype among archaeological populations. If bone shape and response to loading are driven more by genetic background than by exercise, as has been demonstrated by research using both inbred (Middleton et al., 2008b; Wallace et al., 2010, 2012) and outbred (Wallace, Judex, & Demes, 2015) mouse strains and suggested by studies on human populations (Cowgill, 2010; Meiring, Avidon, Norris, & McVeigh, 2013; Weiss, 2003), it becomes very difficult to extrapolate population-level changes in activity based solely on observed bone morphology.

Although we recognize that there is greater genetic variation within human populations than among them (Lewontin, 1972; Rosenberg et al., 2002), we do expect archaeological samples representative of different human populations to have some degree of among-group genetic differences, which may involve the skeletal system. In this context, we view genetically distinct mouse strains as informative for considering the expected results for archeological human populations with different genetic backgrounds. Studies have found genetic differences between archaeological populations with different subsistence strategies (Bramanti et al., 2009; Hervella et al., 2012; Skoglund et al., 2012,
2014). As these populations differ in their genetic backgrounds and in their habitual behaviors, any observed changes in their skeletal morphology cannot be definitively ascribed to one cause over the other. Therefore, we suggest caution in interpreting cross-sectional differences in fossil assemblages as resulting from different loading environments when other underlying population-level differences and selection histories remain unknown (Wallace et al., 2010, 2015; Wescott, 2006).

The second hypothesis, that mice allowed wheel access will show a corresponding bone loading response that can be quantified histomorphometrically, was not widely supported. A potential explanation for this result may be that wheel running, although a common locomotor behavior in mice (Novak, Burghardt, & Levine, 2012; Swallow et al., 1998a) does not provide an adequate loading environment (external or internal) to produce the expected morphological responses (but see Kelly et al., 2006). If the muscle forces generated by voluntary running on a wheel are not enough to induce a plastic response, then no difference would be seen between the bones of mice with wheel access and those without. In the context of our study, it is possible that the increased activity associated with the mice given wheel access was not sufficient to reach threshold for bone remodeling (Frost, 2001).

Previous studies using this mouse model system have produced mixed results, with some studies showing widespread effects of activity (Kelly et al., 2006) while others have not (Middleton et al., 2008b, 2010). These three studies differed from the one described here in use of (1) a larger diameter wheel (1.1 m circumference), which may alter gait patterns, (2) noninbred animals of both sexes and from different generations in the artificial selection protocol, and (3) the use of a different covariate for statistical analysis (here we used femur length rather than body mass, which allowed us to avoid any analytical complications arising from the impact of exercise on body mass). The inbreeding process in the MM mice used here may have reduced the magnitude of phenotypic plasticity in response to habitual exercise; however, we doubt this as other studies have found an effect of exercise on bone phenotype using inbred strains (Akhter, Cullen, Pedersen, Kimmel, & Recker, 1998; Kelly et al., 2006; Kodama et al., 2000; Plochocki, Rivera, Zhang, & Ebba, 2008).

4.2 Relationship between cross-sectional geometry and mechanical loading behavior

Our third hypothesis, that predictions of bending mechanics based on cross-sectional geometry would reflect the results of mechanical testing, was unsupported. The bones were loaded anteroposteriorly (i.e., about the mediolateral axis), and mediolateral second moment of area is the variable used to predict loading response in that direction. Histomorphometric analysis indicated that, based on their mediolateral second moments of area, the femora of MM mice should be the most resistant to bending and C3H/He the least, with C57BL/6 mice intermediate (Figure 6). However, the results of mechanical testing indicate that C3H/He femora are the most resistant to bending, while MM and C57BL/6 femora are not different from one another (difference ≈ 0). We found that C3H/He mice have significantly stiffer bones that are able to absorb more energy to maximum strength than both MM and C57BL/6 mice (Figure 7 and Supporting Information, Table 1). This result is supported by the observation that C3H/He mice have relatively high bone mineralization (Beamer et al., 1996; Sheng et al., 1999). Our findings support other studies that indicate that mineralization is an important component of bone strength (Donnelly, Chen, Boskey, Baker, & van Der Meulen, 2010; Keller, 1994; Jepsen, Akkus, Majeska, & Nadeau, 2003; Jepsen, Silva, Vashisht, Guo, & van der Meulen, 2015; van Der Meulen, Jepsen, & Mikic, 2001).

Similarly, although the values of mediolateral second moments of area suggested that MM femora should be more resistant to bending than those of C57BL/6 mice, mechanical testing resulted in no difference in loading response between the two strains. Thus, while the cross-sectional geometry of MM mice differs grossly from that of C57BL/6 mice (Figure 2e), the bones of MM mice do not actually behave very differently when loaded. These results suggest that the changes in bone morphology associated with the MM genotype do not actually provide them with a biomechanical (i.e., locomotor fitness in terms of resistance to bending) advantage related to increased voluntary locomotion. Furthermore, lack of differences in bending resistance between C57BL/6 and MM mice suggest that differences in hind limb muscle mass and/or mass-specific contractile strength (Syme, Evashuk, Grintuch, Rezende, & Garland, 2005) may not necessarily be correlated with differences in bone biomechanical behavior in this system. The observed similarities between MM and C57BL/6 bones in response to mechanical testing despite differences in femoral cross-sectional shape suggest that MM and C57BL/6 mice may have similar mineralization profiles.

Overall, our results support the idea that estimations of bone loading response that are made based solely on cross-sectional geometry, without consideration of bone mineralization, will be incomplete (Jepsen et al., 2015; van Der Meulen et al., 2001). While the expected phenotypic variation within inbred mouse strains is more analogous to a group of identical siblings than to heterogeneous human populations, the importance of bone mineralization on loading behavior is likely to be true on a population level as well as an individual or family group level. The implication of these results is that differences in bone morphology among populations, whether driven by genetic background or plastic responses to muscle forces, may not indicate differences in bone functional properties or behavioral patterns (Cowgill, 2014; Cowgill, Warrener, Pontzer, & Ocobock, 2010; Garland et al., 2011; Maki, 2013; Osborne et al., 2012; Wallace et al., 2010; Wescott, 2006).

4.3 Plasticity versus genetic adaptation

Our results have important implications for the interpretation of loading mechanics in archaeological and fossilized remains. Observation of bone cross-sectional properties is a technique frequently used to infer past behavior and mechanical loading history (Bridges, 1991; Dobson and Trinkaus, 2002; Lieberman et al., 2004; Ruff and Hayes, 1983, 1984; Shackelford, Marshall, & Peters, 2013; Weiss, 2005). However, as demonstrated here and elsewhere (Akhter et al., 2000; Schriefer...
et al., 2005), the mechanical properties of bone are dependent on both their size and their mineral content. As such, histomorphometric analyses without accompanying assessment or knowledge of the mineral composition of the bone in question will only convey a partial sense of its expected response to loading. Because mineralization data are not available from fossilized remains, predictions of mechanical properties made from fossil bones should be made with caution.

Skeletal morphology, including how morphology changes in response to environmental influences, is often hypothesized to be adapted to a particular function or behavior (Carter and Orr, 1992; Evseev, Cardini, Morozova, & O’Higgins, 2014; Hylander, 1979; Marchi, 2008; Meloro, Elton, Louys, Bishop, & Ditchfield, 2013; Ruff, Holt, & Trinkaus, 2006; Shaw and Stock, 2009; Susman, 1979). However, because evolutionary adaptations and individual-level plastic responses can result in similar phenotypes, it can be difficult to determine how to best interpret observed morphological patterns. Reconstructions of behavior in fossil remains present a particular challenge because, at the time of analysis, the fossilized bones have been decoupled from the selective history and/or loading environments that formed them. Individual bone phenotypes of fossils are often interpreted as being purely indicative of plastic responses to loading regimes that correspond to habitual activity patterns. However, because little is known about the genetic backgrounds of most of these populations, it is difficult to definitively determine whether the observed morphology is derived from individual-level plastic responses or population-level genetic history. Our results have shown it is quite possible for genetically based differences within a species to dwarf the degree of environmental differences. While we are only able to characterize three genotypes here and thus cannot make more general statements about the relative importance of genetic variation and environmental variation on a population wide scale, we do consider it likely that differences observed between populations reflect some degree of genetically based differences rather than differences due to behavior alone.

In other words, an osteological correlate of increased activity could either indicate that the population themselves led an active lifestyle, or that their ancestors did (Cowgill, 2010; Lieberman, 1997; Wallace et al., 2010; Weiss, 2003). Similarly, because evolutionary adaptations are mediated by the genome, a particular phenotype may be apparent even in the absence of a specific environmental stimulus. For example, Wallace et al. (2010, 2012) showed that at least some of the limb bone morphology of mice artificially selected for increased voluntary activity (the high-runner lines, from which the MM strain used here were derived) reflects evolutionary adaptation based in genetics, rather than individual plastic responses to loading. An additional complication is that an individual’s capacity for plasticity is under genetic control (Garland and Kelly, 2006; Kelly, Panhuis, & Stoehr, 2012; West-Eberhard, 2003) and may also be sex-dependent. Thus, while the specific morphology itself may be the result of phenotypic plasticity, the range of possible plastic responses may be genetically constrained. These factors are important to consider when interpreting behavior from the fossil record, because they demonstrate that alterations in behavior do not necessarily correspond directly to alterations in bone morphology.

5 | CONCLUSION

Our results highlight two main concerns associated with inferring past behaviors from the cross-sectional morphology of fossilized bones. First, we found that observed differences in femoral morphology were more reflective of differences in genetic background than differences in activity. Second, we found that our predictions of bone behavior under loading, based solely on femoral cross-sectional geometry and without consideration of genetically based differences in bone material properties, were incorrect. Together these results suggest that analysis of skeletal morphology alone is not sufficient to predict loading behavior. Our intention here is not to suggest that skeletal geometry has no impact on how a bone behaves when loaded; we recognize the importance of bone shape and the role of plastic adaptation in adjusting skeletal morphology in response to strain. Rather, our goal is to draw attention to the limitations associated with reconstructing past behaviors from fossilized remains and to help underscore the importance of bone mineral content for understanding bone strength.

Tests of the underlying genetic or environmental determinants of bone morphology, such as those discussed previously, are generally not possible for fossils. Instead, comparisons between the morphology of fossil populations and populations with known behaviors, such as the studies by Stock and Pfeiffer (2001) and Shaw and Stock (2013), may provide a viable alternative. However, such analyses should be made with the understanding that they are incomplete. Further studies of bone loading response in controlled experimental settings using animals with known genetic backgrounds will be critical in furthering our understanding of phenotypic variation in fossil populations.

ACKNOWLEDGEMENTS

The authors wish to thank Cindy Chrisler for assistance with animal husbandry. They also acknowledge Monica Ahsan, Elizabeth King, and Karthik Panchanathan for fruitful discussion of Bayesian analysis. Kristina Aldridge, Scott Maddux, Ian Wallace, and an anonymous reviewer provided helpful comments on earlier drafts of this article.

ORCID

Sarah J. Peacock http://orcid.org/0000-0002-1064-9482
Theodore Garland, Jr. http://orcid.org/0000-0002-7916-3552
Kevin M. Middleton http://orcid.org/0000-0003-4704-1064

REFERENCES


**SUPPORTING INFORMATION**
Additional Supporting Information may be found online in the supporting information tab for this article.

---