Natural selection on morphology varies among years and by sex in Magellanic Penguins (Spheniscus magellanicus)

Authors: Koehn, Laura E., Hard, Jeffrey J., Akst, Elaine P., and Boersma, P. Dee

Source: The Auk, 133(4) : 783-805

Published By: American Ornithological Society

URL: https://doi.org/10.1642/AUK-16-50.1

The Auk
Ornithological Advances

AmericanOrnithology.org

Volume 133, 2016, pp. 783–805
DOI: 10.1642/AUK-16-50.1

RESEARCH ARTICLE

Natural selection on morphology varies among years and by sex in Magellanic Penguins (*Spheniscus magellanicus*)

Laura E. Koehn,[1,2]* Jeffrey J. Hard,[3] Elaine P. Akst,[4,5] and P. Dee Boersma[2,4]

[1] School of Aquatic and Fishery Sciences, University of Washington, Seattle, Washington, USA
[2] Department of Biology, University of Washington, Seattle, Washington, USA
[3] Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, NOAA, Seattle, Washington, USA
[4] The Wildlife Conservation Society, Bronx, New York, USA
[5] Molecular Genetics Laboratory, Smithsonian Conservation Biology Institute, National Zoological Park, Washington, DC, USA
* Corresponding author: laura.koehn216@gmail.com

Submitted March 4, 2016; Accepted July 11, 2016; Published September 21, 2016

ABSTRACT

The evolution of morphology in a population reflects several factors, including the influence of environmental variability on natural selection. We estimated natural selection on, and heritability of, 4 individual morphological traits (bill length, bill depth, flipper length, and foot length) and 2 multivariate morphological traits in adult Magellanic Penguins (*Spheniscus magellanicus*) at Punta Tombo, Argentina, from 1983 to 2010. We estimated heritability of morphology with parent–offspring regression and animal models, conditioning on sex because the species is sexually dimorphic. For the analysis of selection on each trait, we estimated both linear and quadratic selection gradients, based on the number of fledglings produced, for breeding males and females in each year. Estimates from animal models indicated that all 6 traits were heritable; in parent–offspring regressions, corresponding heritabilities were significantly higher in sons than in daughters in 100% of tests. Over 28 yr, we detected no selection in 21 yr for males and in 21 yr for females. For the years in which we did detect selection, the direction and intensity of selection on traits varied, being especially variable for females. We detected selection on primarily multivariate body size but also on male bill sizes and female bill and foot lengths. Selection on male flipper and foot lengths and on female bill depth was detectable only in relation to selection on body size. When there was selection in males, selection on body and bill sizes was mainly toward larger sizes and occurred in 4 of 6 yr with high chick starvation. The absence of detectable selection on morphology in most years suggests that it is not tightly linked to fitness and that the dynamic environment where Magellanic Penguins live helps maintain morphological variation. The temporal variability in selection likely fosters stability of morphology through time, a pattern that might not be evident in short-term studies.

Keywords: animal model, heritability, Magellanic Penguins, morphology, natural selection, reproductive success

La selección natural de la morfología varía entre años y según sexo en *Spheniscus magellanicus*

RESUMEN

La evolución de la morfología en una población refleja varios factores, incluyendo la influencia de la variabilidad ambiental en la selección natural. Estimamos la selección natural y la heredabilidad de cuatro rasgos morfológicos individuales: largo del pico, profundidad del pico, largo de la aleta y largo del pie, y dos rasgos morfológicos multivariados, en adultos de *Spheniscus magellanicus* en Punta Tombo, Argentina desde 1983 hasta 2010. Estimamos la heredabilidad de la morfología con una regresión entre progenitores y su descendencia y con modelos animales, segregando por sexo ya que estos pingüinos presentan dimorfismo sexual. Para el análisis de selección de cada rasgo, estimamos gradientes de selección lineales y cuadráticos, basados en el número de volantones producidos por machos y hembras reproductivos en cada año. Las estimaciones de los modelos animales indicaron que los seis rasgos eran heredables; en las regresiones progenitores-descendencia, las heredabilidades correspondientes fueron significativamente más altas en los hijos que en las hijas en el 100% de las evaluaciones. A lo largo de 28 años, no pudimos detectar selección en 21 años para los machos y en 21 años para las hembras. Para los años en que sí detectamos selección, la dirección y la intensidad de la selección en los rasgos varió, y fue particularmente variable para las hembras. Detectamos selección principalmente en el parámetro multivariado de tamaño corporal, pero también en el tamaño del pico de los machos y en el largo del pico y del pie de las hembras. La selección en el largo de la aleta y del pie del macho y en la profundidad del pico de la hembra fue detectable solamente en relación a la selección del tamaño corporal. Cuando se detectó selección en los machos, la selección en el tamaño corporal y del pico fue principalmente hacia tamaños más grandes y se presentó en cuatro de los seis años con baja hambruna de los polluelos. La ausencia de selección detectable en la morfología en la mayoría de los años sugiere que no está fuertemente vinculada a la adecuación sexual y que el ambiente dinámico donde vive *S. magellanicus* ayuda a mantener la variación morfológica. La variabilidad temporal en la selección probablemente fomenta la estabilidad de la morfología a través del tiempo, un patrón que podría no ser evidente en estudios de corto plazo.

Palabras clave: éxito reproductivo, heredabilidad, modelo animal, morfología, Pingüino Magallánico, selección natural

INTRODUCTION

Morphology can play an important role in the fitness of an individual. Survival and reproductive success, 2 measures of individual fitness, are linked to morphology in various ways, depending on a species' life-history characteristics. In some species, larger morphological traits enhance survival (Boag and Grant 1981, Brown and Brown 1999); whereas in other species, intermediate (Fox 1975) or smaller (Brown and Brown 2013) size results in higher survival. Morphological size can also enhance reproductive success by making an individual more attractive to the opposite sex (Pryke and Andersson 2002) or by aiding in competition for a mate (Pratt and Anderson 1982, Clutton-Brock et al. 1988).

Morphological traits that cause variation in fitness have the potential to evolve over time if those traits are sufficiently heritable and responsive to selection (Endler 1986). Heritability of and selection on morphological traits is well documented for a variety of avian populations, and these traits often have high heritability estimates (Boag and van Noordwijk 1987). A review of avian heritabilities by Merilä and Sheldon (2001) showed large, significant heritability estimates for morphological traits ranging from 0.4 to 0.6 (40–60% of phenotypic variation due to genetic, rather than environmental, sources of variation). Directional, stabilizing, and disruptive selection on morphology have been documented in several avian species (Grant and Grant 1993, Brown and Brown 1998, Hendry et al. 2009). In the Medium Ground-Finch (*Geospiza fortis*; Grant and Grant 1993), one of Darwin's finches in the Galápagos Islands, individuals with smaller bills had a higher chance of survival than those with larger bills after an El Niño, when smaller, soft seeds were the most common food. However, selection in the opposite direction occurred as well, when conditions were drier, favoring finches with larger bills that could eat harder seeds (Grant and Grant 2002). Also, the selective advantage of shorter wings, for avoiding being hit and killed by vehicles, decreased the average wing length in Cliff Swallows (*Petrochelidon pyrrhonota*; Brown and Brown 2013).

Sexual dimorphism of morphological traits can result from sexual selection or other forms of natural selection (Darwin 1874, Lande 1980). Originally, sexual dimorphism in birds was thought to be linked to polygamy (see review in Andersson 1994), but monogamous species with biparental care can be sexually dimorphic as well (see Owens and Hartley 1998). An alternative explanation for sexual dimorphism is that ecological factors drive natural selection and result in males and females using resources differently (Andersson 1994). Magellanic Penguins (*Spheniscus magellanicus*) are monogamous and sexually dimorphic; adult males are generally larger, with thicker, longer bills and longer feet and flippers than females (Boersma et al. 2013). Previous research has suggested that foraging habitats of male and female Magellanic Penguins may not be entirely overlapping, since larger individuals (males) dive deeper and for longer than smaller individuals (Walker and Boersma 2003).

For Magellanic Penguins at Punta Tombo, Argentina, body size and bill, flipper, and foot morphology likely play key roles in a penguin's ability to capture prey, migrate, secure a nest site, and attract a mate. Bill, flipper, and foot size may each influence Magellanic Penguin foraging success—and, therefore, survival and reproductive success—independent of body size. Flippers are used for propelling penguins through the water (Davis and Renner 2003) and allow penguins to catch prey and avoid predators, so flipper size may be most important for overall fitness. Additionally, Magellanic Penguins migrate long distances (Stokes et al. 2014) and are traveling farther to find prey than they did a decade ago (Boersma and Rebstock 2009), so even slightly longer flippers may benefit individuals. Feet are used as rudders and are important for foraging because they provide maneuverability (Davis and Renner 2003). Penguins use their bills to capture and handle prey (Wilson and Duffy 1986, Wilson and Wilson 1990), and bill size in birds is often positively correlated with prey size (Holmes and Pitelka 1968). Therefore, bill size should play a role in penguin foraging success and survival. Additionally, overall body size should be important to fitness. Although parental care is similar between sexes (Boersma et al. 1990), larger male penguins are more likely to win fights at the beginning of the reproductive season, acquire mates, and hold good-quality nests (Renison et al. 2002). Larger penguins also have a more favorable ratio of surface area to size and more storage capacity, though larger size is metabolically costly. Magellanic Penguin foraging depth and length of time submerged are related to body size as well (Walker and Boersma 2003).

We sought to determine how natural selection on these morphological traits (bill length, bill depth, flipper length, and foot length) and body size varied over time, estimate the inheritance of these traits, quantify whether selection varied between males and females, and determine whether trends in selection were occurring over time. We estimated heritabilities on traits and tested for natural selection in Magellanic Penguins at Punta Tombo using a 28 yr dataset of morphological size and reproductive success (Boersma et al. 1990, 2013). We hypothesized that selection would vary between the sexes because (1) Magellanic Penguins are sexually dimorphic, (2) competitive pressures on land are dissimilar between the sexes, (3) the sex ratio of breeders is skewed toward males (Boersma et al. 2013), and (4) foraging may vary by sex (Walker and Boersma 2003). We expected all traits (including body size) to be heritable and subject to selection because each of the 4 traits independently as well as overall body size are important for the survival and reproductive success of both sexes. We hypothesized that selection on each trait and on body size would be temporally variable because environmental conditions and reproductive success vary among years (Boersma 2008, Boersma and Rebstock 2009, 2014), and a long-term study on natural selection in Darwin's finches showed temporally variable selection (Grant and Grant 2002). The present study, to our knowledge, is the first long-term study of natural selection on morphology in a seabird and is one of only a few long-term selection studies on morphology in birds (see Grant and Grant 2002).

METHODS

Study Site and Trait Description

We followed individually marked Magellanic Penguins and their nests at Punta Tombo, Argentina (44.02°S, 65.11°W), starting in 1982 (Boersma et al. 1990, Boersma 2008). We marked individual penguins with either a stainless-steel flipper band or a web tag (Boersma and Rebstock 2010). In study nests, we used web tags to mark chicks when foot length was >9 cm, and we banded the flipper of a chick prior to fledging if the chick weighed $>1{,}800$ g and was seen after January 10. In some areas, all breeding adults were flipper banded; whereas in other areas, we web-tagged mates of known-age, banded penguins. We measured the lengths of 3 morphological traits (bill, flipper, and foot) and the depth of the bill (following Boersma 1974) when we banded an adult penguin and, subsequently, every 5 to 10 yr for breeding adult penguins in our study nests. Offspring were banded and measured as chicks and measured again as adults when we found them breeding in the colony. We used only adult measurements to estimate heritability (see below). We are reasonably confident of the parentage in this population because extrapair copulations are rare (not seen in 148 pairs), and although mate switching occurred in 9% of 148 pairs, females laid eggs well after the mate switch occurred, so it is likely the new partner was the father (Hood 1996).

Testing Assumptions in Estimating Heritability

We initially tested 2 important assumptions usually made when conducting both parent–offspring regression tests and animal models to estimate heritability. First, we assumed that homologous traits were measured in both the parents and offspring (Falconer and Mackay 1996). If penguins continue to grow after they are sexually mature and parents and offspring are not the same age when measured, then we would not be measuring homologous traits. Morphological traits should remain constant over time if adult size does not change with age after reproductive maturity is reached (see Starck and Ricklefs 1998). We tested whether this assumption was true by regressing the change in size between 2 measurements of an individual against the number of years elapsed between measurements, for each of the 4 traits and for each sex separately, in Magellanic Penguins that we measured more than once after they reached sexual maturity (for further methods and explanation, see Appendix). Second, if selection acts on offspring before they are measured as adults (i.e. differential survival of offspring based on a trait), heritability estimates may be biased (Grant 1983). To test this assumption, we compared the fledging sizes of offspring that returned as adults to the fledging sizes of their siblings that did not return as adults (for further methods and results, see Appendix). Therefore, offspring from our heritability analysis needed to have a sibling that survived to fledging to be included in this analysis ($n = 34$).

Heritabilities and Phenotypic and Genetic Correlations

Overview. To estimate heritabilities and genetic correlations for the 4 morphological traits, we used 2 methods: (1) conventional parent–offspring regression (see Lynch and Walsh 1998), a technique widely used in a variety of species, from the Horned Beetle (*Onthophagus sagittarius*; Watson and Simmons 2010) to salmonids (various species of salmon, trout, and chars; Carlson and Seamons 2008) to the Barn Swallow (*Hirundo rustica*; Saino et al. 2013); and (2) univariate and bivariate animal models (see Kruuk 2004, Wilson et al. 2010). When using individual trait values to estimate heritability, estimates from parent–offspring regression are typically similar to values estimated with univariate animal models (Åkesson et al. 2007). Although heritability estimates from animal models are sometimes lower than those estimated from parent–offspring regression, this is likely due to upward bias in the regression method, owing to shared environmental effects or fixed effects not accounted for by parent–offspring regression (Kruuk and Hadfield 2007). Animal-model analyses minimize this bias at the expense of increased complexity by incorporating information from all available relatives in each generation across the pedigree.

Because data for multiple generations are not readily available for most long-lived species like Magellanic Penguins, and our sample sizes are small (especially for individual offspring sex), we used both approaches (animal model and parent–offspring regression) to estimate trait heritabilities using all offspring. We first used parent–offspring regression with all offspring as an initial exploration of inheritance in this population. Because of the sexual dimorphism in this species, we also used parent–offspring regression to estimate heritability for each offspring sex. Finally, we used univariate and bivariate animal models with all offspring to estimate heritabilities and genetic correlations, because these models incorporated information from all relatives in the data, conditioned the estimates on sex as a fixed effect, and accounted for correlations among traits. For all analyses (heritabilities, selection analysis, and principal component analysis), all morphological data were centered and scaled using a Box-Cox transformation prior to analysis.

Parent–offspring regression. To estimate narrow-sense heritability (h^2; ratio of additive genetic variance to total phenotypic variance) of the 4 structural traits for individual offspring sexes, we compared the size of parents to the adult size of their offspring using parent–offspring regression. Only rarely do Magellanic Penguins breed before or at 4 yr of age (Rafferty et al. 2005). We used only parents whose chicks fledged in any year from 1983 to 2005, to allow offspring time to return to the colony and be resighted. We regressed the offspring's adult trait size on the average of its parents' sizes (hereafter "mid-parent"), where h^2 is equal to the slope of the linear regression (Falconer and Mackay 1996). When a parent or adult offspring had multiple measurements for a trait (maximum of 11 measurements for one offspring), we averaged them. We then averaged father and mother size to find the mid-parent size. We first used linear regression between all offspring trait sizes and all mid-parent trait sizes separately to estimate heritability. We then used linear regression between male offspring (hereafter "sons") size and mid-parent size for the 4 structural traits separately, and did the same for female offspring (hereafter "daughters") to estimate sex-specific heritabilities. Sons and daughters were considered separately because of the sexual dimorphism in this species (Boersma et al. 2013). For parents, we determined sex on the basis of behavior, copulation, cloaca size, time of arrival, and mating history. We determined the sex of offspring we found breeding in the same way as parents for ~45% of sons (45 of 99) and ~44% of daughters (18 of 41). For all other offspring, sex was determined by bill depth, because Magellanic Penguins can be sexed by bill depth with 92% accuracy (Boersma and Davies 1987). Any adult penguin with a bill depth <2.25 cm was classified as female, and any adult with a deeper bill was sexed as a male (Boersma and Davies 1987).

For each parent–offspring regression test, we weighted by family size (number of offspring from the same parents) and averaged offspring measurements within families to account for parents that had more than one offspring that returned to the colony as adults. Including each offspring separately from the same parents can be problematic for parent–offspring regressions because of pseudoreplication, but not for animal-model methods because of the way individual breeding value is estimated as a random effect (see Wilson et al. 2010). To count as a family for parent–offspring regressions with mid-parent size, offspring had to have both parents in common. For parent–offspring regressions of a single offspring sex, siblings of the opposite sex were not included in averages. For all regressions, both offspring measurements and parent measurements (mid-parent or single parent) were Box-Cox transformed before analysis.

We also estimated the heritability of multivariate morphometric traits from a principal component analysis (PCA) because of the correlations (see below) among the 4 morphological traits. A PCA was performed using measurements of all penguins, including both parents and offspring, and of both sexes (because there is not a large difference in the covariances among traits between sexes; see below). Traits were Box-Cox transformed and we ran the PCA on all 4 traits using the "prcomp" function in the "stats" package in R (R Development Core Team 2014). The first principal component (PC1) explained 75% of the total variance and is interpreted as overall body size because loadings of all individual traits on this axis are high and in the same direction. This first component was the only one to explain a significant amount of variance, based on a scree plot and broken-stick analysis. The second principal component (PC2) explained 10% of the variance and is interpreted as a body-shape contrast between bill size and flipper or foot size (or skeletal structure) because of high loadings for all traits, but in opposite directions (positive vs. negative) for bill length or depth vs. flipper and foot lengths. In other words, a penguin has a high score on PC2 if its bill size is larger or smaller than expected given the size of its flipper and foot.

Estimating heritabilities with parent–offspring regression across generations involved many tests. It is customary to adjust P (α) values when performing multiple comparisons from the same data (Bland and Altman 1995). For these analyses, we therefore used a false discovery rate (FDR) method that controls the proportion of falsely rejected hypotheses to account for multiple testing and to adjust the critical α for parent–offspring heritability tests (Benjamini and Yekutieli 2001, Narum 2006). This method gave an adjusted α of 0.017 ($k = 18$ tests, original $\alpha = 0.05$). For selection analyses (see below), we used a separate technique to address multiple testing because type II error (β) is more problematic for the selection tests than the type

I error (α) addressed by FDR techniques for the heritability tests.

We calculated phenotypic correlations and genetic correlations among traits to determine whether heritability and selection may be acting on multiple traits at once if the traits are highly correlated (either phenotypically or genetically). Phenotypic correlations were estimated for both males and females from Pearson correlations estimated in R with the function "rcorr" in the "Hmisc" package (R Development Core Team 2014). We estimated corresponding genetic correlations from the covariance and variance components specified by a bivariate animal model (see below).

Animal models. We estimated heritabilities for the population as a whole (to compare to parent–offspring regressions) and estimated genetic correlations among the 4 morphological traits using animal models (Kruuk 2004). An animal model is a generalized linear mixed model that uses the relationships among individuals in a pedigreed population to estimate the variance of breeding values (additive genetic variance) for individual traits. The univariate model we used to calculate heritability of a single trait was

$$y_i = \mu + a_i + s_i + e_i \qquad \text{(Equation 1)}$$

where μ is the population mean for a trait, y_i is each individual's phenotype for trait i, a_i is the random effect of the individual's breeding value (estimated from its relatives' phenotypes as a deviation from the population mean), s_i is the fixed effect of the individual's sex, and e_i is the residual error. We began with a univariate model for each trait and estimated its phenotypic variance (V_P) from $V_\mathrm{P} = V_\mathrm{G} + V_\mathrm{R}$, where V_G is the genetic variance and V_R is the residual variance. For each trait, we calculated the heritability (h^2), which is equal to the ratio of V_A to V_P. Additionally, using a bivariate form of the model, we calculated the genetic correlation (r_G) between each pair of morphological traits from their genetic variance (V_x or V_y) and covariance (Cov_{xy}):

$$r_\mathrm{G} = \mathrm{Cov}_{xy}/\sqrt{V_x V_y} \qquad \text{(Equation 2)}$$

where x and y are the 2 traits. Because the traits were phenotypically correlated, we also estimated heritabilities for each using a bivariate animal model.

In the animal-model framework, we conditioned our estimates of heritability and genetic correlation on sex because of the evident sexual dimorphism in Magellanic Penguins; each model included sex as a fixed effect. Univariate models incorporating an effect of sex showed better fits to the data, as evidenced by smaller Deviance Information Criterion (DIC) values (a Bayesian analogue of Akaike's Information Criterion). We did not control for hatch year when calculating estimates because hatch year was not known for all penguins, particularly parents, in the pedigree.

We fit the animal models to the data with a Bayesian Markov chain Monte Carlo (MCMC) method as executed in the R package "MCMCglmm" (Hadfield 2010, R Development Core Team 2014). For each random effect (e.g., individual breeding value), we specified uninformative or weakly informative priors to minimize any influence of the prior on the posterior estimates. Most priors that we used for random effects for univariate models were based on an inverse Wishart distribution or the chi-square distribution with parameter expansion. Vague priors help to ensure unbiased estimates of variance components and that posterior distributions primarily reflect information from the data. For bivariate models, a chi-square-based prior from de Villemereuil et al. (2013) led to convergence for most trait pairs. Convergence of estimates was not achieved for most trait pairs that included bill depth and for some that included flipper length (for any of the priors we tested), so univariate heritability estimates are also reported.

By default, MCMCglmm uses a broad normal distribution for fixed effect priors. For each analysis, we compared alternative models using the DIC of each and selected the model with the smallest DIC. We also compared DIC values for models with and without the pedigree information as a random effect to assess significance of heritability estimates (significant effect if the DIC of the model with the pedigree is a better fit; lower DIC) and examined 95% highest posterior density (HPD) intervals.

We computed heritability and genetic correlation estimates from the animal and residual variances generated by the model analyses, using single Markov chains 500,000–1,500,000 iterations long, with 50,000 burn-in iterations, and thinning rates of 1 in 5,000 to achieve effective sample sizes of 5,000–10,000 for each chain. We inspected Brooks-Gelman-Rubin and Heidelberg-Welch chain diagnostics (Brooks and Gelman 1998) to ensure that both lag autocorrelation and chain convergence were sufficient for each analysis.

Selection Analysis

If a trait is heritable and selection is acting on it, there is potential for adaptive evolution to occur. We tested whether there was selection on any of the 4 individual traits (bill length, bill depth, flipper length, and foot length) and overall body size for female and male adults, using annual reproductive success as a measure of fitness. We defined reproductive success for the selection analysis as number of chicks fledged (0, 1, or 2) per adult in a given year. We calculated annual reproductive success because we know the reproductive success for each penguin we sight in a given year. We could not reliably estimate lifetime reproductive success because we did not sight every bird each year in the colony and because many individuals were still alive and reproducing. For this analysis, we used only penguins (1) that were known to be breeding in a study nest in a given year, (2) that were sighted, and (3) for which we knew whether they had produced fledglings or not.

We tested for selection on each trait (bill length and depth, flipper and foot lengths) for each sex in each year between 1983 and 2010, as well as for selection on multivariate traits from a PCA. We tested for selection on body size because in years with selection on multiple traits, it could be that selection is acting on overall body size and not on the traits individually, because of correlations among the traits. We ran a separate PCA on all penguins in the selection analysis (different from the heritability analysis sample), which again included all 4 univariate traits. Again, PC1 was interpreted as overall body size and PC2 was interpreted as contrast between bill sizes vs. flipper and foot sizes. PC1 explained 73% of the variance and PC2 explained 12.5%, but PC1 was the only component to explain a significant amount of variance, based on a scree plot and broken-stick analysis. For the selection analysis, trait measurements were Box-Cox transformed for each year before analysis.

We calculated relative fitness for each male and female by dividing the individual's number of chicks fledged by the mean number of chicks fledged for that year. We estimated both linear and quadratic selection gradients in R using linear models:

$$w = \alpha + \beta x \qquad \text{(Equation 3)}$$

$$w = \alpha + \beta x + \gamma(x^2) \qquad \text{(Equation 4)}$$

where w is relative fitness (reproductive success); x is one of the morphological traits or PC scores; α is the y-intercept of the fitness function; β is the linear selection coefficient; and γ is the quadratic coefficient (in Equation 4). Significant linear coefficients signify directional selection, whereas significant quadratic coefficients signify stabilizing (negative coefficient) or disruptive (positive coefficient) selection; the quadratic equation also gives another estimate of the linear coefficient. We estimated linear and quadratic coefficients for all 4 traits and the 2 multivariate traits for both males and females in each year.

We calculated the effect size and power for each significant selection coefficient to determine the strength of the selective effect and the probability of detecting selection when there is a significant difference in size between fitness groups (and, thus, to minimize type II error as well as control type I error; see Sullivan and Feinn 2012). For each test, the effect size, based on the standardized selection coefficients (expressed in phenotypic standard deviations), is the magnitude of the difference at varying levels of reproductive success. An effect size $\leq$0.2 is generally considered a small effect size, one between 0.5 and 0.8 a moderate effect size, and one $>$0.8 an adequately large effect size (Cohen 1988). The associated power for each test gives the probability of detecting a significant difference when one actually exists, a more appropriate statistic than an adjusted α that corrects only for type I error. We considered any selection coefficient with power $>$0.8 to be significant, while coefficients with less power were not considered significant even with $P < 0.05$.

Finally, given that successful foraging may be related to Magellanic Penguin morphological size and because variability in reproductive success is most likely due to food availability (see Boersma et al. 1990, Boersma and Stokes 1995), we qualitatively compared the proportion of starved chicks in each year (out of the total number of chicks that died or fledged, or among chicks with known fate) to the yearly selection analysis. We used the proportion of chicks that starved as a proxy for food availability and foraging success (Boersma and Stokes 1995). Data on both the number of chicks starved in each year and the total number of chicks came from Boersma and Rebstock (2014), who showed that ~40% of chick deaths each year are due to starvation.

RESULTS

Testing Assumptions of Heritability Tests

Magellanic Penguins' bills, feet, and flippers remained similar in size after reaching sexual maturity. Regression slopes of change in trait size (bill, flipper, and foot) against years between measurements and individual were not significantly different from zero for males ($n = 25$, $P =$ 0.58–0.98) or for females ($n = 25$, $P = 0.11$–0.78). Consequently, our heritability estimates are not biased by growth variation among adults.

Offspring that returned to the colony as adults ($n = 34$: 17 males and 17 females) had longer flippers and feet at fledging than their siblings that fledged and did not return, meaning that there is differential survival of offspring based on size, at least when all years are combined. Therefore, heritability estimates for flipper length and foot length may be biased upward. Bills were similar in size between the groups ($P = 0.19$ for bill length; $P = 0.38$ for bill depth). Flippers were 0.31 cm longer (± SE) in chicks that returned ($\bar{x} = 15.01 \pm 0.08$ cm) compared to chicks that did not return ($\bar{x} = 14.70 \pm 0.09$ cm, $t_{33} = -2.5$, $P = 0.02$). Similarly, feet were 0.27 cm longer in chicks that returned ($\bar{x} = 11.97 \pm 0.08$ cm) than in their siblings that were not seen as adults ($\bar{x} = 11.70 \pm 0.08$ cm, $t_{33} = -2.5$, $P = 0.02$). A second-hatched chick was as likely to have the longest flippers and feet as a first-hatched chick, based on a chi-square test of independence (21 first chicks, 13 second

TABLE 1. Heritability estimates (narrow-sense h^2 ± SE) for bill length and depth and flipper and foot lengths in Magellanic Penguins, categorized by the relationship between parents and their offspring. (A) Mid-parent[a] bill length and depth are significantly heritable by all offspring ($n = 145$); foot length is significantly heritable before α adjustment for multiple testing. (B) Mid-parent bill and foot sizes are significantly heritable by sons ($n = 108$), but mid-parent sizes are not significantly heritable by daughters ($n = 46$). PC1 and PC2 are the first and second principal components from a principal component analysis. *P* values are compared to an adjusted α of 0.017 (calculated from a false-discovery-rate method): †$P < 0.05$ (not significant for adjusted α), *$P \leq 0.017$, **$P < 0.001$.

(A)

Trait	All offspring and mid-parent
Bill length	0.307 ± 0.076**
Bill depth	0.346 ± 0.079**
Flipper length	0.123 ± 0.084
Foot length	0.213 ± 0.081*
PC1	0.568 ± 0.178*
PC2	0.245 ± 0.101*

(B)

Trait	Sons and mid-parent	Daughters and mid-parent
Bill length	0.399 ± 0.088**	0.285 ± 0.131†
Bill depth	0.264 ± 0.092*	0.188 ± 0.150
Flipper length	0.201 ± 0.098†	–0.093 ± 0.146
Foot length	0.330 ± 0.093**	0.098 ± 0.147
PC1	0.570 ± 0.138**	0.111 ± 0.281
PC2	0.345 ± 0.118*	–0.020 ± 0.189

[a] Regression of the offspring's adult trait size on the average of its parents' sizes.

chicks, $\chi^2_1 = 1.88$, $P = 0.17$), so larger size at fledging was not due to hatch order (a first-hatched chick was as likely to survive as a second-hatched chick).

Heritabilities and Phenotypic and Genetic Correlations

Parent–offspring regression. When considering male and female offspring together, heritabilities estimated by parent–offspring regression using mid-parent size were significant for bill length and depth, foot length, multivariate body size (PC1), and multivariate bill vs. skeletal size contrast (PC2) (at $\alpha = 0.017$) but not for flipper length ($P = 0.15$; Table 1A and Appendix Figure 3). From regressions of offspring on mid-parent involving progeny of one sex, we found that heritability estimates of morphology were higher for sons than for daughters (Table 1B, Appendix Figure 4; 100% of tests). For each parent and offspring sex separately, parent morphology (except for flipper length) was significantly heritable in sons but not in daughters. Bill length was significantly heritable in daughters from parents before but not after FDR adjustment for multiple testing ($P = 0.035$, $\alpha = 0.017$). Resemblance of daughters to their parents was weak for all traits. Offspring flipper length was not significantly heritable in either sex (Table 1).

Animal model. In general, heritabilities estimated with the univariate animal model for the whole population and those estimated with parent–offspring regressions for all offspring (and for sons separately) were similar (Table 2A), although regression estimates were generally larger as expected (estimates from the univariate and bivariate animal models: 0.17–0.41; from parent–offspring regressions: 0.21–0.67). Additionally, the univariate animal model revealed significant heritability estimates for flipper length (though lower than bill- and foot-size heritability estimates). Under a univariate animal model for each trait with sex as a fixed effect, we detected significant heritability estimates for bill length, bill depth, foot length, body size, and bill size vs. skeletal size (flipper or foot length). The heritability estimate for flipper length was significant (the 95% HPD interval did not overlap zero, and the model including pedigree information had a smaller DIC), but the value was low, indicating low genetic variance for flipper length, as also shown by parent–offspring regression. For all traits, models with pedigree information produced lower DICs, indicating that all traits are significantly heritable.

For bivariate animal models, we also estimated significant heritabilities for bill, flipper, and foot sizes (Table 2B). Estimates from most bivariate models that included bill depth would not converge for bill depth based on the plots of MCMC chains. However, heritability estimates for bill depth converged in the bivariate model with flipper length (though the estimate for flipper length did not converge), so we only report the heritability estimate for bill depth from the one bivariate model. Similarly, the heritability estimate for flipper length converged only for the bivariate model that included flipper length and foot length.

Phenotypic correlations among all 4 traits were significantly positive for both males ($n = 239$) and females ($n = 181$) (Table 3A). Using the bivariate animal model with sex as a fixed effect, genetic correlations between bill and foot lengths and between flipper and foot lengths were significantly positive (Table 3B), but they could not be assessed for 4 of 6 trait pairs because of nonconvergence based on plots of MCMC chains. There appears to be a weak family effect on genetic correlations, but we could not estimate correlations precisely because of limited power, owing to the relatively small number of families. Therefore, we could not determine whether the majority of traits are genetically correlated.

Selection Analysis

We detected interannual variation in natural selection, in both intensity and form, in both males and females

TABLE 2. Heritability estimates (h^2) for bill length and depth and for flipper and foot lengths in Magellanic Penguins, using (A) univariate and (B) bivariate animal models. Traits are significantly heritable, based on the 95% highest posterior density (HPD) intervals and comparison of Deviance Information Criterion (DIC) values from models with and without pedigree information, when conditioned on sex. "NA" indicates traits for which convergence of estimates was not achieved.

(A)

Trait	h^2 (95% HPD)	Fixed effect of sex (95% HPD)	Random effect of individual breeding value (95% HPD)	DIC (DIC without pedigree)
Bill length	0.40 (0.22–0.59)	1.42 (1.30–1.56)	0.19 (0.10–0.30)	858.92 (941.37)
Bill depth	0.26 (0.02–0.43)	1.69 (1.60–1.80)	0.063 (0.006–0.12)	651.21 (677.81)
Flipper length	0.17 (0.002–0.34)	1.28 (1.13–1.43)	0.08 (0.001–0.21)	1,014.91 (1028.87)
Foot length	0.33 (0.14–0.53)	1.36 (1.23–1.51)	0.18 (0.07–0.30)	930.79 (982.37)
PC1 (body size)	0.30 (0.13–0.52)	–2.89 (−3.07 to −2.70) [a]	0.30 (0.12–0.50)	1,168.50 (1,218.46)
PC2 (bill vs. skeletal)	0.21 ($8.0e^{-6}$ to 0.41)	0.24 (0.12–0.36)	0.08 ($2.9e^{-6}$ to 0.17)	812.20 (836.63)

(B)

Bivariate model	h^2 (95% HPD)	DIC (DIC without pedigree)
Bill length and bill depth	Bill length = 0.41 (0.23–0.58) Bill depth = NA	1,442.98 (1,587.70)
Bill length and flipper length	Bill length = 0.40 (0.22–0.59) Flipper length = NA	1,817.13 (1,914.01)
Bill length and foot length	Bill length = 0.40 (0.22–0.59) Foot length = 0.33 (0.08–0.53)	1,734.73 (1,868.27)
Bill depth and flipper length	Bill depth = 0.24 ($1.1e^{-6}$ to 0.42) Flipper length = NA	1,649.17 (1,693.15)
Bill depth and foot length	Bill depth = NA Foot length = 0.34 (0.10–0.53)	1,546.26 (1,628.71)
Flipper length and foot length	Flipper length = 0.20 ($2.2e^{-6}$ to 0.34) Foot length = 0.35 (0.11–0.54)	1,858.11 (1,920.38)

[a] PC1 loadings are negative, creating negative fixed effects.

TABLE 3. Genetic and phenotypic correlation matrices for 4 morphometric traits in Magellanic Penguins, estimated from bivariate animal models. (A) Phenotypic correlation matrices show that the sizes of bill, flipper, and foot are significantly positively correlated with each other in males and females (*r*; *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$). (B) Genetic correlation values (95% highest posterior density [HPD] intervals) reflect estimates for the traits conditioned on the fixed effect of sex. With Box-Cox transformed data, bivariate models including bill depth, and the model with bill length and flipper length, would not converge; therefore, there is not sufficient power to estimate genetic correlations between bill depth and other traits, and these trait pairs are marked "NA."

(A)

	Males (*n* = 239)				Females (*n* = 181)			
Trait	Bill length	Bill depth	Flipper length	Foot length	Bill length	Bill depth	Flipper length	Foot length
Bill length	–	0.29***	0.32***	0.31***	–	0.25***	0.38***	0.40***
Bill depth		–	0.15*	0.28***		–	0.25***	0.28***
Flipper Length			–	0.36***			–	0.51***
Foot length				–				–

(B)

Trait	Bill length	Bill depth	Flipper length	Foot length
Bill length	–	NA	NA	0.47 (0.03–0.81)
Bill depth	–	–	NA	NA
Flipper length	–	–	–	0.97 (0.50->0.99)
Foot length	–	–	–	–

for 3 of the 4 individual traits and for the 2 multivariate (PCA) traits (Tables 4 and 5; Appendix Tables 6 and 7). In most years (21 of 28 yr for males; 21 of 28 yr for females) we did not detect selection. For years in which we did detect selection, which traits were selected upon, in which direction, and the magnitude varied by year and sex. Also, which years had detectable selection varied by sex, with only 3 of 28 yr having selection for both males and females. Additionally, in many years in which we detected selection on individual traits, we also saw selection on multivariate body size, making it likely that the detected selection on individual traits is primarily a reflection of selection on body size. Effect sizes and power analysis showed that we had adequate power to detect a standardized selection coefficient with a magnitude ≥ 0.105. We did not have sufficient power to detect weaker selection than that, and although we report smaller coefficients because we found a significant P value (at $\alpha = 0.05$; see Appendix Tables 6 and 7), we do not discuss these results further because of the small effect size and low power of these tests.

For males, we detected significant selection with an effect size >0.12 and sufficient power (>0.8) in 7 of 28 yr for bill length and bill depth, independently of selection on multivariate body size, and for flipper length and foot length but only in relation to selection on body size (Table 4; see also Appendix Table 6). Nonetheless, we did not detect selection in most years (21 of 28 yr). Among the significant tests, the range in magnitude of significant selection coefficients was 0.125–0.450. Significant directional selection on bill size (length and depth) with adequate power occurred in 6 yr (1987, 1991, 1994, 1997, 2000, and 2002), with larger bill sizes (length and depth) corresponding to higher reproductive success (Figure 1 and Table 4). However, in 3 of those years, we also detected selection on body size (PC1), indicating that there was selection on overall body size in 1997, 2000, and 2002, but not on individual traits. We found significant selection on foot length with high enough power for detection only in 1997 and 2000, and on flipper length only in 2000, but we also found evidence of significant selection on body size (PC1) in those years. We detected selection on the bill vs. skeletal contrast trait (PC2) alone in 2004, indicating that natural selection favored smaller bill size and larger flippers and feet that year. Bill length and depth exhibited significant selection independently of body size (unlike flipper and foot lengths) in 3 of 28 yr, with selection favoring larger sizes in those years.

We detected significant selection with sufficient power in males for 4 of 6 yr (67%) with the highest chick starvation (chick starvation $>50\%$; order of highest to lowest: 2000, 1987, 1984, 1997, 1990, and 2002; Table 4). In those 4 yr (1987, 1997, 2000, and 2002) we detected significant selection on bill size (length, depth, or both) and body size, all toward larger sizes. The only years in which we found significant selection on foot length or flipper length (as part of selection on body size) were the 2 yr with the highest and third-highest chick starvation (2000 and 1997). We found significant selection with adequate power in only 3 other years, and there was no selection in 4 of the 5 yr with the lowest chick starvation (2008, 2004, 2007, 1985, and 1999). Finally, body size (PC1), bill vs. skeletal size contrast (PC2), and bill length and foot length had large directional selection coefficients toward larger sizes in 1984 (another year with high chick starvation), although these coefficients were not significant because of high variance (Figure 1).

When power was adequate, we also found evidence of significant selection in females in 7 of 28 yr, which included selection on bill length and foot length (independent of selection on body size) and selection on bill depth in relation to selection on body size (Table 5; see also Appendix Table 7). However, selection was variable in direction and intensity and we found no significant selection in 21 of 28 yr. The significant selection coefficients ranged from 0.105 to 0.843 in magnitude. Similar to the results in males, we mainly found significant selection on body size and bill sizes (bill length and depth, with or without selection on body size). However, unlike in males, selection varied in direction depending on the year, with smaller sizes favored in the 1980s and larger sizes favored in the 1990s and 2000s (with some disruptive selection in 2000 on body size and bill length; Figure 2 and Table 5). There was sufficient power to detect selection on foot length in 1987 and 1991 (in the absence of selection on body size), with directional selection toward shorter feet in 1987 but toward longer feet in 1991. We found no year with significant selection on flipper length; however, the linear coefficient for flipper length in 2009 had only a marginally significant test P value ($P = 0.063$, $\alpha = 0.05$) despite the high power of the test (see Appendix for further explanation). There were additional trait–year combinations with high (>0.12) linear selection coefficients (see Figure 2), but these coefficients were not significant, owing to high variance.

Selection on traits in females corresponded less with food availability (as assessed by chick starvation) than did selection on traits in males. We detected significant selection on bill sizes and/or body size in females (but in opposing directions) only in the 2 yr with highest chick starvation (1987 and 2000); and in 1 yr with high chick starvation (1987), we detected significant selection for shorter feet.

TABLE 4. Standardized selection coefficients (with SD in parentheses) and direction of selection by year for male Magellanic Penguins where selection was significant (at $\alpha < 0.05$). Selection coefficients with sufficient power ($1 - \beta > 0.8$) are in bold. Traits are specified as "linear" or "quadratic," depending on which model gave the significant linear coefficient; and with a superscript 2 if the quadratic term in the quadratic model was significant. If there was a significant linear term in both the linear and quadratic models, the results for both models are shown. "NA" indicates years without significant selection coefficients. "PC1" is a multivariate trait from a principal component analysis (PCA) that represents body size based on high positive loadings for bill sizes and foot and flipper length. "PC2" is another multivariate trait, where bill sizes had high negative loadings and flipper and foot lengths had high positive loadings, representing a contrast between bill sizes and skeletal size. The proportion (%) of chicks starved (out of all chicks) in a given year is shown for comparison with years that had significant selection. There is significant selection on bill sizes individually (length and depth) and on body size. The only selection on flipper length and foot length occurred in years with selection on body size, pointing to selection on overall body size, not individual traits.

Year	*n*	Trait	Coefficient	Selection	Chicks starved
1983	483	NA	NA	NA	0.273
1984	785	NA	NA	NA	0.658
1985	732	Bill depth^2	0.076 (0.032)	Disruptive	0.207
1986	781	NA	NA	NA	0.467
1987	747	**Bill length (linear)**	**0.359 (0.121)**	**Longer**	0.697
		Bill length (quadratic)	**0.359 (0.121)**	**Longer**	
1988	613	NA	NA	NA	0.280
1989	626	NA	NA	NA	0.389
1990	511	NA	NA	NA	0.553
1991	369	**Bill depth (linear)**	**0.429 (0.171)**	**Deeper**	0.339
		Bill depth (quadratic)	**0.425 (0.171)**	**Deeper**	
1992	417	Bill depth^2	−0.092 (0.034)	Stabilizing	0.323
1993	697	NA	NA	NA	0.289
1994	626	**Bill length (quadratic)**	**0.127 (0.059)**	**Longer**	0.286
		Bill length^2	0.068 (0.029)	Disruptive	
1995	396	NA	NA	NA	0.393
1996	440	NA	NA	NA	0.286
1997	437	**Bill length (linear)**	**0.194 (0.071)**	**Longer**	0.658
		Bill length (quadratic)	**0.162 (0.077)**	**Longer**	
		Foot length (linear)	**0.176 (0.072)**	**Longer**	
		Foot length (quadratic)	**0.171 (0.073)**	**Longer**	
		PC1 (linear)	**0.135 (0.059)**	**Larger**	
1998	448	NA	NA	NA	0.471
1999	474	NA	NA	NA	0.209
2000	364	**Bill depth (linear)**	**0.276 (0.128)**	**Deeper**	0.862
		Flipper length (quadratic)	**0.450 (0.134)**	**Longer**	
		Flipper length^2	**0.165 (0.050)**	**Disruptive**	
		Foot length^2	**0.165 (0.069)**	**Disruptive**	
		PC1 (linear)	**0.218 (0.098)**	**Larger**	
		PC1 (quadratic)	**0.208 (0.098)**	**Larger**	
		PC1^2	0.083 (0.032)	Disruptive	
2001	330	NA	NA	NA	0.459
2002	347	**Bill length (linear)**	**0.265 (0.107)**	**Longer**	0.538
		Bill length (quadratic)	**0.261 (0.109)**	**Longer**	
		Bill depth (linear)	**0.213 (0.107)**	**Deeper**	
		Bill depth^2	**−0.158 (0.076)**	**Stabilizing**	
		PC1 (linear)	**0.230 (0.092)**	**Larger**	
2003	398	Bill length (quadratic)	0.118 (0.051)	Longer	0.394
		Bill length^2	0.062 (0.030)	Disruptive	
2004	408	**PC2 (quadratic)**	**0.181 (0.072)**	**Smaller bill, larger flipper/foot**	0.168
		PC2^2	−0.113 (0.052)	Stabilizing	
2005	474	Bill length (linear)	0.092 (0.042)	Longer	0.438
		Bill length (quadratic)	0.076 (0.043)	Longer	
		Bill length^2	−0.036 (0.023)	Stabilizing	
2006	396	NA	NA	NA	0.257
2007	592	NA	NA	NA	0.185
2008	577	NA	NA	NA	0.117
2009	628	NA	NA	NA	0.409
2010	602	NA	NA	NA	0.382

TABLE 5. Standardized selection coefficients (with SD in parentheses) and direction of selection by year for female Magellanic Penguins where selection was significant (at $\alpha < 0.05$). Coefficients with sufficient power ($1 - \beta > 0.8$) are in bold. Traits are specified as "linear" or "quadratic," depending on which model gave the significant linear coefficient; and with a superscript 2 if the quadratic term in the quadratic model was significant. If there was a significant linear term in both the linear and quadratic models, the results for both models are shown. "NA" indicates years without significant selection coefficients. "PC1" is a multivariate trait from a principal component analysis that represents body size based on high positive loadings for bill sizes and foot and flipper length. "PC2" is another multivariate trait, where bill sizes had high negative loadings and flipper and foot lengths had high positive loadings, representing a contrast between bill sizes and skeletal size. The proportion (%) of chicks starved (out of all chicks) in a given year is shown for comparison with years that had significant selection. There was significant selection on bill length and foot length individually, and on body size. The only selection on bill depth and flipper length occurred in years with selection on body size, pointing to selection on overall body size, not individual traits.

Year	*n*	Trait	Coefficient	Selection	Chicks starved
1983	452	**Bill depth (linear)**	**−0.155 (0.046)**	**Shallower**	0.273
		Bill depth (quadratic)	**−0.157 (0.045)**	**Shallower**	
		Bill depth^2	0.069 (0.031)	Disruptive	
		Flipper length (linear)	−0.103 (0.046)	Shorter	
		Flipper length (quadratic)	−0.103 (0.046)	Shorter	
		PC1 (linear)	**−0.175 (0.056)**	**Smaller**	
		PC1^2	0.044 (0.043)	Disruptive	
1984	783	NA	NA	NA	0.658
1985	718	**Bill depth (linear)**	**−0.125 (0.045)**	**Shallower**	0.207
		Bill depth (quadratic)	**−0.125 (0.045)**	**Shallower**	
		PC1 (linear)	**−0.105 (0.053)**	**Smaller**	
1986	775	NA	NA	NA	0.467
1987	734	**Bill length (linear)**	**−0.243 (0.120)**	**Shorter**	0.697
		Foot length (linear)	**−0.239 (0.120)**	**Shorter**	
1988	612	NA	NA	NA	0.280
1989	622	NA	NA	NA	0.389
1990	509	NA	NA	NA	0.553
1991	362	**Foot length (linear)**	**0.434 (0.179)**	**Longer**	0.339
		Foot length (quadratic)	**0.444 (0.180)**	**Longer**	
1992	356	NA	NA	NA	0.323
1993	559	**Bill length (linear)**	**0.148 (0.055)**	**Longer**	0.289
		Bill length (quadratic)	**0.167 (0.058)**	**Longer**	
		Bill depth (linear)	**0.131 (0.055)**	**Deeper**	
		Bill depth (quadratic)	**0.118 (0.057)**	**Deeper**	
		PC1 (linear)	**0.114 (0.055)**	**Larger**	
		PC2^2	**−0.130 (0.052)**	**Stabilizing**	
1994	457	NA	NA	NA	0.286
1995	384	Bill length^2	−0.092 (0.038)	Stabilizing	0.393
1996	369	Flipper length (quadratic)	0.080 (0.046)	Longer	0.286
		Flipper length^2	−0.051 (0.026)	Stabilizing	
1997	334	NA	NA	NA	0.658
1998	381	NA	NA	NA	0.471
1999	320	NA	NA	NA	0.209
2000	252	**Bill length (quadratic)**	**0.566 (0.175)**	**Longer**	0.862
		Bill length^2	**0.211 (0.070)**	**Disruptive**	
		Bill depth (linear)	**0.359 (0.148)**	**Deeper**	
		Bill depth (quadratic)	**0.391 (0.163)**	**Deeper**	
		PC1 (quadratic)	**0.843 (0.338)**	**Larger**	
		PC1^2	**0.145 (0.071)**	**Disruptive**	
		PC2 (linear)	**−0.502 (0.218)**	**Larger bill, smaller flipper/foot**	
2001	171	NA	NA	NA	0.459
2002	312	Bill length^2	−0.136 (0.053)	Stabilizing	0.538
2003	388	**PC2 (linear)**	**0.228 (0.084)**	**Smaller bill, large flipper/foot**	0.394
		PC2 (quadratic)	**0.229 (0.090)**	**Smaller bill, large flipper/foot**	
2004	406	NA	NA	NA	0.168
2005	468	NA	NA	NA	0.438
2006	458	NA	NA	NA	0.257
2007	569	Bill length	0.080 (0.038)	Longer	0.185
2008	587	NA	NA	NA	0.117
2009	619	NA	NA	NA	0.409
2010	558	NA	NA	NA	0.382

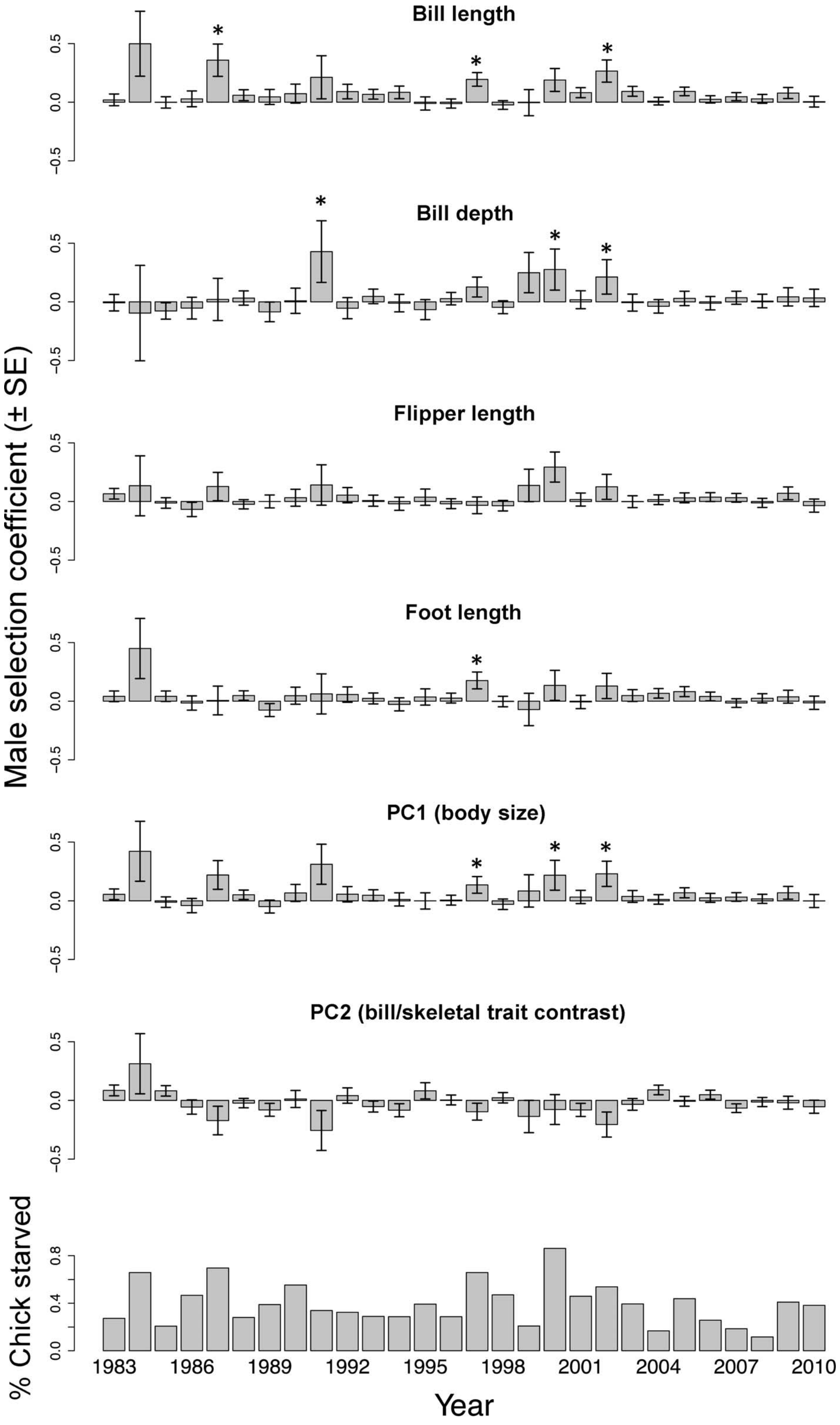


FIGURE 1. Male selection coefficients across 28 yr for 4 morphological traits and 2 multivariate traits that reflect body size and bill vs. skeletal size from a principal component analysis (PC1 and PC2 = first and second principal components), and the proportion (%) of chicks starved by year. Bars with asterisk denote significant coefficients ($P < 0.05$, effect size > 0.12, and power > 0.8). Coefficients are only for the linear model (not the quadratic model); therefore, linear coefficients in the quadratic model are not represented in this graph. For sample sizes, see Table 4.

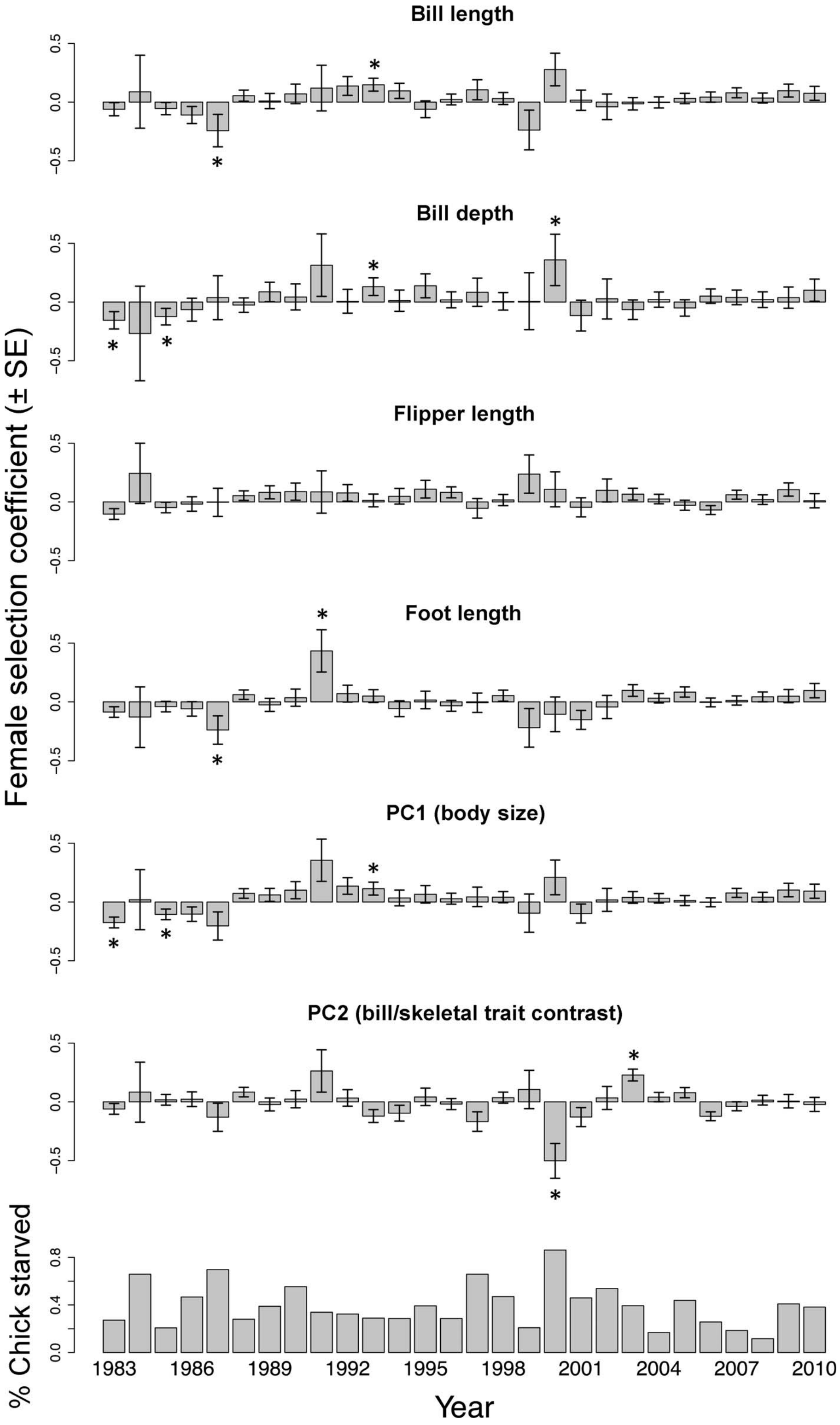


FIGURE 2. Female selection coefficients across 28 yr for 4 morphological traits and 2 multivariate traits that reflect body size and bill vs. skeletal size from a principal component analysis (PC1 and PC2 = first and second principal components), and the proportion (%) of chicks starved, by year. Bars with asterisk denote significant coefficients ($P < 0.05$, effect size > 0.10, and power > 0.8). Coefficients are only for the linear model (not the quadratic model); therefore, linear coefficients in the quadratic model are not represented in this graph. For sample sizes, see Table 5.

DISCUSSION

In Magellanic Penguins studied at Punta Tombo, Argentina, we found that several prominent morphological traits, including overall body size, are heritable. We detected natural selection on most traits, but selection was undetectable in most of the 28 yr we examined, and when selection was significant, it varied temporally in direction as well as magnitude. We found that selection appeared to act primarily on overall body size, although we also found selection on some individual traits, specifically bill length and depth in males and bill and foot lengths in females. The selection that we detected varied by sex and was more variable for females. For males, selection acted most commonly on bill and body size, and it was predominantly directional toward longer, deeper bills and larger overall body size. Selection on female traits was also primarily on body and bill size, but the direction varied annually and was not detected in the same years as it was for males (only 3 of 28 yr showed detectable selection in both sexes).

The unpredictability in direction of selection (particularly in females) and the absence of detectable selection in many years are patterns similar to those found in a long-term study of selection on morphometric traits in Darwin's finches (Grant and Grant 2002), where selection varied in direction from year to year and was absent in some years. The large selection coefficients in years with detectable selection, and their relationship with individual variation in fledgling production, implies that these traits, especially overall body size, directly influence fitness (if annual fledging success is highly correlated with lifetime reproductive success). This is, however, true only in some years, given that selection on traits in both sexes was undetectable in most years we examined. Moreover, the influence of size (smaller or larger) on fitness is variable for females. Our results show that there may be a connection between selection in males and prey availability that would help explain the direct influence of these traits on fitness, though one or multiple other unmeasured factors likely act to conserve variation in years with no detectable selection, potentially keeping the mean size of traits stable over time.

All the morphological traits we measured, including body size, are significantly heritable in this population. The observed variability in these traits and corresponding estimates of heritability and genetic correlation indicate that ample phenotypic and genetic variation is available for natural selection to facilitate an adaptive response in morphology. Bouzat et al. (2009) showed high genetic diversity in Magellanic Penguins using nuclear and mitochondrial DNA markers. Although we estimated significant heritabilities for bill size, foot length, and body size with parent–offspring regression, the explanatory power of these relationships was low (low r^2; see Appendix), which suggests high phenotypic variance, likely resulting from the high variability in direction and magnitude of selection among years. Both parent–offspring regression and animal-model analyses indicated significant heritability estimates for all traits (except for flipper length in parent–offspring regressions). However, unlike the animal models, parent–offspring regressions with all offspring are not conditioned on the fixed effect of sex (and they generally consider sexual dimorphism by performing separate analyses for each sex), which leads to some disparity in estimates between these methods. It is likely that flipper length is significantly heritable, given that the animal model is more comprehensive in its inclusion of relatives.

When males and females were analyzed separately, traits were not significantly heritable in daughters. Although point estimates were often as large for daughters as they were for sons, the phenotypic variance was high as well, leading to nonsignificant heritability estimates. The large variance around female size could be caused by the ample phenotypic variation from opposing selection forces we detected in females (directional selection for both smaller and larger traits, depending on the year) but not in males. The lack of selection observed in females, and no evidence for selection on flipper length independent of selection on body size, may help maintain phenotypic variation. Forero et al. (2001) speculated that female Magellanic Penguins either had high heritability for flipper length or that females with larger flippers had higher survival, leading to low sexual dimorphism for flipper length. Since heritability of flipper length is lower in females than in males and the estimate is very small (depending on the test used), the low sexual dimorphism of flipper length appears to result from higher survival of females with longer flippers. We show that fledglings with longer flippers are more likely to return to the colony, an observation that also lends support to differential survival being the likely explanation.

The differential survival of offspring with longer flippers and feet could lead to inflated heritability estimates. It is unlikely that the offspring that did not return survived to maturity, given that only 0.3% of individuals banded at Punta Tombo from 1982 to 2005 were ever sighted elsewhere (Boersma 2008). The inflation of heritability estimates could mean that even less phenotypic variation is explained by genetic variation, which seems probable given the variability in selection from year to year, likely reflecting variability in the environment. We note that our proxy for fitness for the selection analyses was number of offspring fledged in a given year, not number of fledglings that returned. Therefore, total reproductive fitness for a parent could be lower if the parent fledged multiple chicks but none returned (based on the different survival of fledglings), thereby diluting evolutionary impacts. At the same time, there is still selection on traits in parents occurring within the breeding season, before chicks fledge—selection that also has evolutionary consequences.

Two other important considerations can influence the reliability of heritability estimates. The first of these is selection (differential survival of distinct phenotypes) occurring before offspring reach adult size (Grant 1983). In the Pied Flycatcher (*Ficedula hypoleuca*), when using offspring adult sizes compared to fledgling sizes, heritability estimates were lower because stabilizing selection occurred after the offspring fledged and before they reached adult size (Potti and Merino 1994). We found evidence of selection before maturity in Magellanic Penguins, in that offspring that fledged but did not return to the colony had significantly shorter flippers and feet at fledging than their siblings that did return to the colony as adults. Therefore, the heritability estimates for flipper and foot sizes are likely inflated and should be considered upper bounds. The second consideration of heritability bias is environmental covariance, or common environments between parents and offspring that may inflate resemblance (Merilä and Sheldon 2001), and many studies in avian heritability have discussed potential biases from environment correlations between parents and offspring (see Barbraud 2000, Keller et al. 2001). Some studies have addressed this through cross-fostering experiments, allowing offspring to be raised by other parents to eliminate shared environments (Wiggins 1989, Gustafsson and Merilä 1994). However, studies on heritability of body size in avian species using foster-parent experiments show little or no evidence of inflation of estimates due to environment correlation, suggesting weak environmental covariance (Dhondt 1982, Wiggins 1989, Gustafsson and Merilä 1994), so shared environments may not have appreciably biased our heritability estimates.

The selection that we detected was more consistent across years for males than for females, though still undetectable in most years for both sexes. Selection tended to favor larger males and longer and deeper male bills, particularly in years of high chick starvation. When looking at the proportion of chicks starved in a year (a proxy for prey availability), there were some consistencies in the years in which selection was significant for males, implying that larger size and larger bills may contribute to the foraging success and reproductive success of males. Additionally, we detected no selection on any traits for males in most years (21 of 28 yr) with the lowest chick starvation, which suggests that when ample food is available, size may not be an important determinant of reproductive success. Predation, climatic events, nesting density, fights, and foraging distance (see Stokes and Boersma 1998, 2000, Boersma and Rebstock 2009, 2014) can also affect reproductive success in Magellanic Penguins. Similar environmental variables influence selection of traits in other birds, including weather (Brown and Brown 1998), food size (Boag and Grant 1981), and human interference (Brown and Brown 2013). Additionally, there is some intrasexual competition in male Magellanic Penguins (Renison et al. 2002) that may impose selection on morphology, similar to results in Snow Petrels (*Pagodroma nivea*; Barbraud 2000). Any weak selection that fluctuates temporally—or selection due to survival (not reproductive success) that acts in opposing directions—could help maintain genetic and phenotypic variation in morphology. Even though average male body and bill sizes could increase over time (owing to occasional instances of strong directional selection favoring larger sizes), the variation among years in selection, weak undetectable selection, and lack of selection in most years likely fosters stability of morphology through time.

We found less consistency in direction of selection on traits in females than in males and a weaker relationship with chick starvation—though, again, selection was rarely detected. This suggests that multiple environmental factors are likely interacting in a given year and that the selective forces exerted on the traits are varying annually, potentially even changing in direction. This is in contrast to Red Knots (*Calidris canutus canutus*), in which climate change has led to consistently strong selective pressure against small bills (see van Gils et al. 2016). Additionally, the significant selection in females in the 2 yr with highest chick starvation (1987 and 2000) was in opposing directions (for bill size or body size), and there was also selection in medium and low starvation years. Body mass or condition (Sæther et al. 1997), foraging experience (Limmer and Becker 2009), and behavioral qualities such as prey choice (Golet et al. 2000) and foraging distance (Boersma and Rebstock 2009) may all influence females' foraging success more than their morphology. Females skip breeding seasons in resource-poor years more often than males (Boersma and Rebstock 2010), so potentially only females in good body condition are breeding. Other parental qualities such as body condition (as opposed to body size), as well as breeding experience and age, can affect reproductive success in seabirds (Wooller et al. 1990, Chastel et al. 1995). In Magellanic Penguins, higher parent quality (which included multiple factors) led to larger eggs and, thus, larger chicks and higher fledging success (Reid and Boersma 1990). Therefore, parent quality related to experience, laying date, and other factors may be more tightly linked than body size or other morphological traits to overall reproductive success in Magellanic Penguins.

For years with selection, the selection was rarely similar for males and females in a given year, either in magnitude or direction (but see 2000), which may contribute to the sexual dimorphism in this species. For both sexes, most years (21 of 28 for both males and females) had no detectable selection but there was more evidence of selection on bills than on flippers or feet, and bills are more sexually dimorphic (see Boersma et al. 2013). There was more directional selection toward smaller body and bill sizes in females (3 of 28 yr) than in males (0 of 28 yr), particularly in the 1980s. However, sexual dimorphism is also likely driven by female mate selection and sexual selection. Punta Tombo has a biased adult breeding sex ratio, about 1.5:1 (male:female; Boersma et al. 2013), and larger males win fights for higher-quality nests (Renison et al. 2002). Nonetheless, at least some portion of the sexual dimorphism in this species is likely due to variation in natural selection between the sexes.

The selection coefficients we estimated for morphological traits in Magellanic Penguins are similar to values estimated for other natural populations and bird species, and the patterns we observed in selection are similar to those seen in another long-term study of selection on bird morphology. In a 30 yr study of Darwin's finches, one species, the Medium Ground-Finch, showed changes in direction of selection on bill and body size between years, and in many years no selection was detectable (Grant and Grant 2002). Additionally, in a second species, the Common Cactus-Finch (*G. scandens*), Grant and Grant (2002) saw significant selection only toward larger bills, with weaker, nonsignificant selection toward smaller bills, similar to our results for males. The significant selection coefficients we found (0.105–0.566) are similar in magnitude to the significant values for the 2 populations of Darwin's finches (rarely >0.5 and rarely significant when <0.1), and the high variability in selection coefficients we observed among years was akin to that observed by Grant and Grant (2002). Also, selection coefficients for traits in other natural populations are similar to those that we found for both males and females (Kingsolver et al. 2001). The heritabilities we estimated for traits in Magellanic Penguins are within the ranges of heritability estimates documented for morphological traits in other avian species (see Merilä and Sheldon [2001] and Jensen et al. [2003], in which values ranged from 0.4 to 0.6 and from 0.28 to 1.06, respectively). Evidently, these patterns of genetic and phenotypic variation in, and natural selection on, penguin morphology are common among birds, which suggests that birds with similar life histories and habitats experience a broadly similar range of environmental conditions to which they must successfully adapt if they are to persist.

Conclusion

From a 28 yr study of Magellanic Penguins at Punta Tombo, Argentina, we provide evidence for natural selection on body size, bill length and depth, and foot length and show that these traits are heritable. However, we also find that selection is undetectable in most years, selection varies between the sexes, and selection is temporally variable in direction as well as magnitude. In the few years when we did detect selection, it consistently favored larger body size and bill sizes in males and was more variable for females. Morphology in this species appears unlikely to be changing appreciably as a result of temporal variability in the relationship between size and reproductive success. The variability in selection between the sexes, among years, and over 3 decades is coupled with appreciable genetic and phenotypic variation in morphology in this population (consistent with Bouzat et al. 2009), and the dynamics of natural selection on these penguins are likely to be primary catalysts for maintaining this variation in the face of environmental variability. The temporal variability in selection also constrains consistent morphological change over time, a pattern that would be evident only in a long-term study. Like those of Grant and Grant (2002), our results highlight the importance of long-term studies in identifying patterns and trends in the environment that can influence patterns of phenotypic and genetic responses in morphology and life history through natural selection.

ACKNOWLEDGMENTS

Data were collected under a joint agreement between the Wildlife Conservation Society (WCS) and the Office of Tourism, Province of Chubut, Argentina, since 1982. We thank the Province of Chubut and the La Regina family for logistical support and access to the penguin colony, and the many students and field volunteers who collected data over the years. P. Garcia Borboroglu, W. Conway, G. Harris, and P. Harris facilitated the research and helped with permits. Research funding for our dataset includes WCS; Pew Fellows Program in Marine Conservation; ExxonMobil Foundation; Disney Worldwide Conservation Fund; National Geographic Society; the Chase, Cunningham, CGMK, Offield, Peach, Thorne, Tortuga, and Kellogg foundations; the Wadsworth Endowed Chair in Conservation Science at the University of Washington; and Friends of the Penguins. We thank Trimble for donating 6 tablet computers and Canon for contributing 2 pairs of binoculars for our research. E. P. Akst did the penguin chick sexing at the Center for Conservation and Evolutionary Genetics Lab, facilitated by R. Fleischer and supported by a National Science Foundation (NSF) Graduate Fellowship, the Smithsonian Institution, and an NSF Research Training Grant to the University of Maryland and the Smithsonian Institution and WCS. Toby Bradshaw, Tim Essington, Olivia Kane, Ginger Rebstock, Gordon Orians, Wayne Arendt, and anonymous reviewers improved the manuscript.

Ethics statement: The work was done under an animal care permit (protocol no. 2213-02).

Funding statement: None of the funders had any influence on the content of the submitted or published manuscript. None of the funders require approval of the final manuscript to be published.

Author contributions: L.E.K. and P.D.B. conceived and designed the project. P.D.B. collected data. L.E.K. and J.J.H. analyzed the data. L.E.K., P.D.B., and J.J.H. wrote or

substantially edited the paper. E.P.A. edited the paper and performed experiments and analyzed data for a subsection of the paper.

LITERATURE CITED

Åkesson, M., S. Bensch, and D. Hasselquist (2007). Genetic and phenotypic associations in morphological traits: A long term study of Great Reed Warblers *Acrocephalus arundinaceus*. Journal of Avian Biology 38:58–72.

Andersson, M. B. (1994). Sexual Selection. Princeton University Press, Princeton, NJ, USA.

Barbraud, C. (2000). Natural selection on body size traits in a long-lived bird, the Snow Petrel *Pagodroma nivea*. Journal of Evolutionary Biology 13:81–88.

Benjamini, Y., and D. Yekutieli (2001). The control of the false discovery rate in multiple testing under dependency. Annals of Statistics 29:1165–1188.

Bland, J. M., and D. G. Altman (1995). Multiple significance tests: The Bonferroni method. British Medical Journal 310:170.

Boag, P. T., and P. R. Grant (1981). Intense natural selection in a population of Darwin's finches (*Geospizinae*) in the Galápagos. Science 214:82–85.

Boag, P. T., and A. J. van Noordwijk (1987). Quantitative genetics in wild bird populations. In Avian Genetics: A Population and Ecological Approach (F. Cooke and P. A. Buckley, Editors). Academic Press, London, UK. pp. 45–78.

Boersma, P. D. (1974). The Galapagos Penguin: A study of adaptations for life in an unpredictable environment. Ph.D. dissertation, The Ohio State University, Columbus, OH, USA.

Boersma, P. D. (2008). Penguins as marine sentinels. BioScience 58:597–607.

Boersma, P. D., and E. M. Davies (1987). Sexing monomorphic birds by vent measurements. The Auk 104:779–783.

Boersma, P. D., E. Frere, O. Kane, L. M. Pozzi, K. A.Putz, Raya Rey, G. A. Rebstock, A. Simeone, J. Smith, A. Van Buran, P. Yorio, and P. Garcia Borboroglu (2013). Magellanic Penguin (*Spheniscus magellanicus*). In Penguins: Natural History and Conservation (P. Garcia Borboroglu and P. D. Boersma, Editors). University of Washington Press, Seattle, WA, USA. pp. 232–263.

Boersma, P. D., and G. A. Rebstock (2009). Foraging distance affects reproductive success in Magellanic Penguins. Marine Ecology Progress Series 375:263–275.

Boersma, P. D., and G. A. Rebstock (2010). Effects of double bands on Magellanic Penguins. Journal of Field Ornithology 81:195–205.

Boersma, P. D., and G. A. Rebstock (2014). Climate change increases reproductive failure in Magellanic Penguins. PLOS One 9:e85602. doi:10.1371/journal.pone.0085602

Boersma, P. D., and D. L. Stokes (1995). Mortality patterns, hatching asynchrony, and size asymmetry in Magellanic Penguin (*Spheniscus magellanicus*) chicks. In The Penguins: Ecology and Management (P. Dann, I. Norman, and P. Reilley, Editors). Surrey Beatty, Chipping Norton, Australia. pp. 3–25.

Boersma, P. D., D. L. Stokes, and P. M. Yorio (1990). Reproductive variability and historical change of Magellanic Penguins (*Spheniscus magellanicus*) at Punta Tombo, Argentina. In Penguin Biology (L. S. Davis and J. T. Darby, Editors). Academic Press, San Diego, CA, USA. pp. 15–43.

Bouzat, J. L., B. G. Walker, and P. D. Boersma (2009). Regional genetic structure in the Magellanic Penguin (*Spheniscus magellanicus*) suggests metapopulation dynamics. The Auk 126:326–334.

Brooks, S. P., and A. Gelman (1998). General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 7:434–455.

Brown, C. R., and M. B. Brown (1998). Intense natural selection on body size and wing and tail asymmetry in Cliff Swallows during severe weather. Evolution 52:1461–1475.

Brown, C. R., and M. B. Brown (1999). Natural selection on tail and bill morphology in Barn Swallows *Hirundo rustica* during severe weather. Ibis 141:652–659.

Brown, C. R., and M. B. Brown (2013). Where has all the road kill gone? Current Biology 23:R233–R234.

Carlson, S. M., and T. R. Seamons (2008). A review of quantitative genetic components of fitness in salmonids: Implications for adaptation to future change. Evolutionary Applications 1: 222–238.

Chastel, O., H. Weimerskirch, and P. Jouventin (1995). Influence of body condition on reproductive decision and reproductive success in the Blue Petrel. The Auk 112:964–972.

Clutton-Brock, T. H., S. D. Albon, and F. E. Guinness (1988). Reproductive Success in Male and Female Red Deer. University of Chicago Press, Chicago, IL, USA.

Cohen, J. (1988). The effect size. In Statistical Power Analysis for the Behavioral Sciences, second edition. Lawrence Erlbaum Associates, Hillsdale, NJ, USA.

Darwin, C. (1874). The Descent of Man, and Selection in Relation to Sex, second edition. John Murray, London, UK.

Davis, L. S., and M. Renner (2003). Penguins. Yale University Press, New Haven, CT, USA.

de Villemereuil, P., O. Gimenez, and B. Doligez (2013). Comparing parent–offspring regression with frequentist and Bayesian animal models to estimate heritability in wild populations: A simulation study for Gaussian and binary traits. Methods in Ecology and Evolution 4:260–275

Dhondt, A. A. (1982). Heritability of Blue Tit tarsus length from normal and cross-fostered broods. Evolution 36:418–419.

Endler, J. A. (1986). Natural Selection in the Wild. Princeton University Press, Princeton, NJ, USA.

Falconer, D. S., and T. F. C. Mackay (1996). Introduction to Quantitative Genetics, fourth edition. Longmans Green, Harlow, UK.

Forero, M. G., J. L. Tella, J. A. Donázar, G. Blanco, M. Bertellotti, and O. Ceballos (2001). Phenotypic assortative mating and within-pair sexual dimorphism and its influence on breeding success and offspring quality in Magellanic Penguins. Canadian Journal of Zoology 79:1414–1422.

Fox, S. F. (1975). Natural selection on morphological phenotypes of the lizard *Uta stansburiana*. Evolution 29:95–107.

Golet, G. H., K. J. Kuletz, D. D. Roby, and D. B. Irons (2000). Adult prey choice affects chick growth and reproductive success in Pigeon Guillemots. The Auk 117:82–91.

Grant, B. R., and P. R. Grant (1993). Evolution of Darwin's finches caused by a rare climatic event. Proceedings of the Royal Society of London, Series B 251:111–117.

Grant, P. R. (1983). Inheritance of size and shape in a population of Darwin's finches, *Geospiza conirostris*. Proceedings of the Royal Society of London, Series B 220:219–236.

Grant, P. R., and B. R. Grant (2002). Unpredictable evolution in a 30-year study of Darwin's finches. Science 296:707–711.

Griffiths, R., M. C. Double, K. Orr, and R. J. Dawson (1998). A DNA test to sex most birds. Molecular Ecology 7:1071–1075.

Gustafsson, L., and J. Merilä (1994). Foster parent experiment reveals no genotype–environment correlation in the external morphology of *Ficedula albicollis*, the Collared Flycatcher. Heredity 73:124–129.

Hadfield, J. D. (2010). MCMC Methods for multi-response generalized linear mixed models: The MCMCglmm R package. Journal of Statistical Software 33:1–22.

Hendry, A. P., S. K. Huber, L. F. De León, A. Herrel, and J. Podos (2009). Disruptive selection in a bimodal population of Darwin's finches. Proceedings of the Royal Society of London, Series B 276:753–759.

Holmes, R. T., and F. A. Pitelka (1968). Food overlap among coexisting sandpipers on northern Alaskan tundra. Systematic Zoology 17:305–318.

Hood, L. C. (1996) Adrenocortical response to stress in incubating Magellanic Penguins (*Spheniscus magellanicus*) and mate switching in Magellanic Penguins at Punta Tombo, Argentina. M.S. thesis, University of Washington, Seattle, WA, USA.

Jensen, H., B.-E. Sæther, T. H. Ringsby, J. Tufto, S. C. Griffith, and H. Ellegren (2003). Sexual variation in heritability and genetic correlations of morphological traits in House Sparrow (*Passer domesticus*). Journal of Evolutionary Biology 16:1296–1307.

Keller, L. F., P. R. Grant, B. R. Grant, and K. Petren (2001). Heritability of morphological traits in Darwin's finches: Misidentified paternity and maternal effects. Heredity 87: 325–336.

Kingsolver, J. G., H. E. Hoekstra, J. M. Hoekstra, D. Berrigan, S. N. Vignieri, C. E. Hill, A. Hoang, P. Gilbert, and P. Beerli (2001). The strength of phenotypic selection in natural populations. The American Naturalist 157:245–261.

Kruuk, L. E. B. (2004). Estimating genetic parameters in natural populations using the 'animal model.' Philosophical Transactions of the Royal Society, Series B 359:873–890.

Kruuk, L. E. B., and J. D. Hadfield (2007). How to separate genetic and environmental causes of similarity between relatives. Journal of Evolutionary Biology 20:1890–1903.

Lande, R. (1980). Sexual dimorphism, sexual selection, and adaptation in polygenic characters. Evolution 34:292–305.

Limmer, B., and P. H. Becker (2009). Improvement in chick provisioning with parental experience in a seabird. Animal Behaviour 77:1095–1101.

Lynch, M., and B. Walsh (1998). Genetics and Analysis of Quantitative Traits, vol. 1. Sinauer Associates, Sunderland, MA, USA.

Merilä, J., and B. C. Sheldon (2001). Avian quantitative genetics. Current Ornithology 16:179–255.

Narum, S. R. (2006). Beyond Bonferroni: Less conservative analyses for conservation genetics. Conservation Genetics 7: 783–787.

Owens, I. P., and I. R. Hartley (1998). Sexual dimorphism in birds: Why are there so many different forms of dimorphism? Proceedings of the Royal Society of London, Series B 265: 397–407.

Potti, J., and S. Merino (1994). Heritability estimates and maternal effects on tarsus length in Pied Flycatchers, *Ficedula hypoleuca*. Oecologia 100:331–338.

Pratt, D. M., and V. H. Anderson (1982). Population, distribution, and behaviour of giraffe in the Arusha National Park, Tanzania. Journal of Natural History 16:481–489.

Pryke, S. R., and S. Andersson (2002). A generalized female bias for long tails in a short-tailed widowbird. Proceedings of the Royal Society of London, Series B 269:2141–2146.

Rafferty, N. E., P. D. Boersma, and G. A. Rebstock (2005). Intraclutch egg-size variation in Magellanic Penguins. The Condor 107:921–926.

R Development Core Team (2014). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Reid, W. V., and P. D. Boersma (1990). Parental quality and selection on egg size in the Magellanic Penguin. Evolution 44: 1780–1786.

Renison, D., D. Boersma, and M. B. Martella (2002). Winning and losing: Causes for variability in outcome of fights in male Magellanic Penguins (*Spheniscus magellanicus*). Behavioral Ecology 13:462–466.

Sæther, B.-E., S.-H. Lorentsen, T. Tveraa, R. Andersen, and H. C. Pedersen (1997). Size-dependent variation in reproductive success of a long-lived seabird, the Antarctic Petrel (*Thalassoica antarctica*). The Auk 114:333–340.

Saino, N., M. Romano, D. Rubolini, C. Teplitsky, R. Ambrosini, M. Caprioli, L. Canova, and K. Wakamatsu (2013). Sexual dimorphism in melanin pigmentation, feather coloration and its heritability in the Barn Swallow (*Hirundo rustica*). PLOS One 8:e58024. doi:10.1371/journal.pone.0058024

Starck, J. M., and R. E. Ricklefs (1998). Avian Growth and Development: Evolution within the Altricial–Precocial Spectrum. Oxford University Press, Oxford, UK.

Stokes, D. L., and P. D. Boersma (1998). Nest-site characteristics and reproductive success in Magellanic Penguins (*Spheniscus magellanicus*). The Auk 115:34–49.

Stokes, D. L., and P. D. Boersma (2000). Nesting density and reproductive success in a colonial seabird, the Magellanic Penguin. Ecology 81:2878–2891.

Stokes, D. L., P. D. Boersma, J. L. de Casenave, and P. Garcia-Borboroglu (2014). Conservation of migratory Magellanic Penguins requires marine zoning. Biological Conservation 170:151–161.

Sullivan, G. M., and R. Feinn (2012). Using effect size–or why the *P* value is not enough. Journal of Graduate Medical Education 4:279–282.

van Gils, J. A., S. Lisovski, T. Lok, W. Meissner, A. Ożarowska, J. de Fouw, E. Rakhimberdiev, M. Y. Soloviev, T. Piersma, and M. Klaassen (2016). Body shrinkage due to Arctic warming reduces Red Knot fitness in tropical wintering range. Science 352:819–821.

Walker, B. G., and P. D. Boersma (2003). Diving behavior of Magellanic Penguins (*Spheniscus magellanicus*) at Punta Tombo, Argentina. Canadian Journal of Zoology 81:1471–1483.

Watson, N. L., and L. W. Simmons (2010). Male and female secondary sexual traits show different patterns of quantitative genetic and environmental variation in the horned beetle *Onthophagus sagittarius*. Journal of Evolutionary Biology 23:2397–2402.

Wiggins, D. A. (1989). Heritability of body size in cross-fostered Tree Swallow broods. Evolution 43:1808–1811.

Wilson, A. J., D. Reale, M. N. Clements, M. M. Morrissey, E. Postma, C. A. Walling, L. E. Kruuk, and D. H. Nussey (2010). An ecologist's guide to the animal model. Journal of Animal Ecology 79:13–26.

Wilson, R. P., and D. C. Duffy (1986). Prey seizing in African Penguins *Spheniscus demersus*. Ardea 74:211–214.

Wilson, R. P., and M.-P. T. Wilson (1990). Foraging ecology of breeding *Spheniscus* penguins. In Penguin Biology (L. S. Davis and J. T. Darby, Editors). Academic Press, San Diego, CA, USA. pp. 181–203.

Wooller, R. D., J. S. Bradley, I. J. Skira, and D. L. Serventy (1990). Reproductive success of Short-tailed Shearwaters *Puffinus tenuirostris* in relation to their age and breeding experience. Journal of Animal Ecology 59:161–170.

APPENDIX

METHODS

Testing Assumptions of Heritability Tests

To test the assumption that Magellanic Penguins will not continue to grow once they reach adult size, we regressed change in size between 2 measurements against number of years between those 2 measurements for each of the 4 traits (bill length, bill depth, flipper length, foot length) and for each sex separately for penguins that had more than one measurement after they reached 15 mo of age and had adult plumage (sexual maturity). Using a random-number generator, we randomly chose 25 breeding females and 25 breeding males with 2 or more measurements (after 15 mo of age) and regressed trait size against number of years between measurements, using a general linear mixed-effects model as specified by the "lme4" package in R (RDCT 2014), with years between measurements as a fixed effect and band number (identification) as a random effect. To test whether penguins grew after becoming adults, we compared the full mixed-effects model with a null model with zero slope, using a likelihood ratio test. If growth is not continuous and size does not change, the model with an estimated slope will not be significantly different from the model with a slope of zero. If there were more than 2 measurements for an individual for 1 yr, we used the average size of each trait for that year. The time between measurements ranged from 1 to 17 yr for both females and males. The number of measurements for each male ranged from 2 to 10, and for females it ranged from 2 to 7.

To test whether selection is occurring on penguin offspring before they reach adult size and before heritability is calculated, we compared fledging sizes of offspring that returned as adults to the fledging sizes of their siblings that did not return as adults. We used paired *t*-tests to see whether selection had acted differentially on the offspring. We used 34 offspring (17 males and 17 females) that returned as adults and 34 siblings that did not return (hatched in the same nest as one of the 34 offspring that returned) for each trait. We assumed that most siblings that did not return had died, because few chicks banded at Punta Tombo are found breeding in other colonies (Boersma 2008). We also assumed a sex ratio at fledging of 1:1. We tested our assumption by examining the hatchling sex ratios at Punta Tombo, using 61 blood samples from pairs of sibling chicks and polymerase chain reaction (PCR) amplification of CHD-W and CHD-Z genes with primers from Griffiths et al. (1998). We found a nonsignificant female bias of 1.6 to 1 for chicks at hatching ($n = 33$, $\chi^2 = 0.16$, $\mathrm{df} = 1$, $P > 0.05$), so we assumed a sex ratio at fledging of ~1:1 (if there is selection, it is acting on both males and females).

RESULTS

Selection Analysis

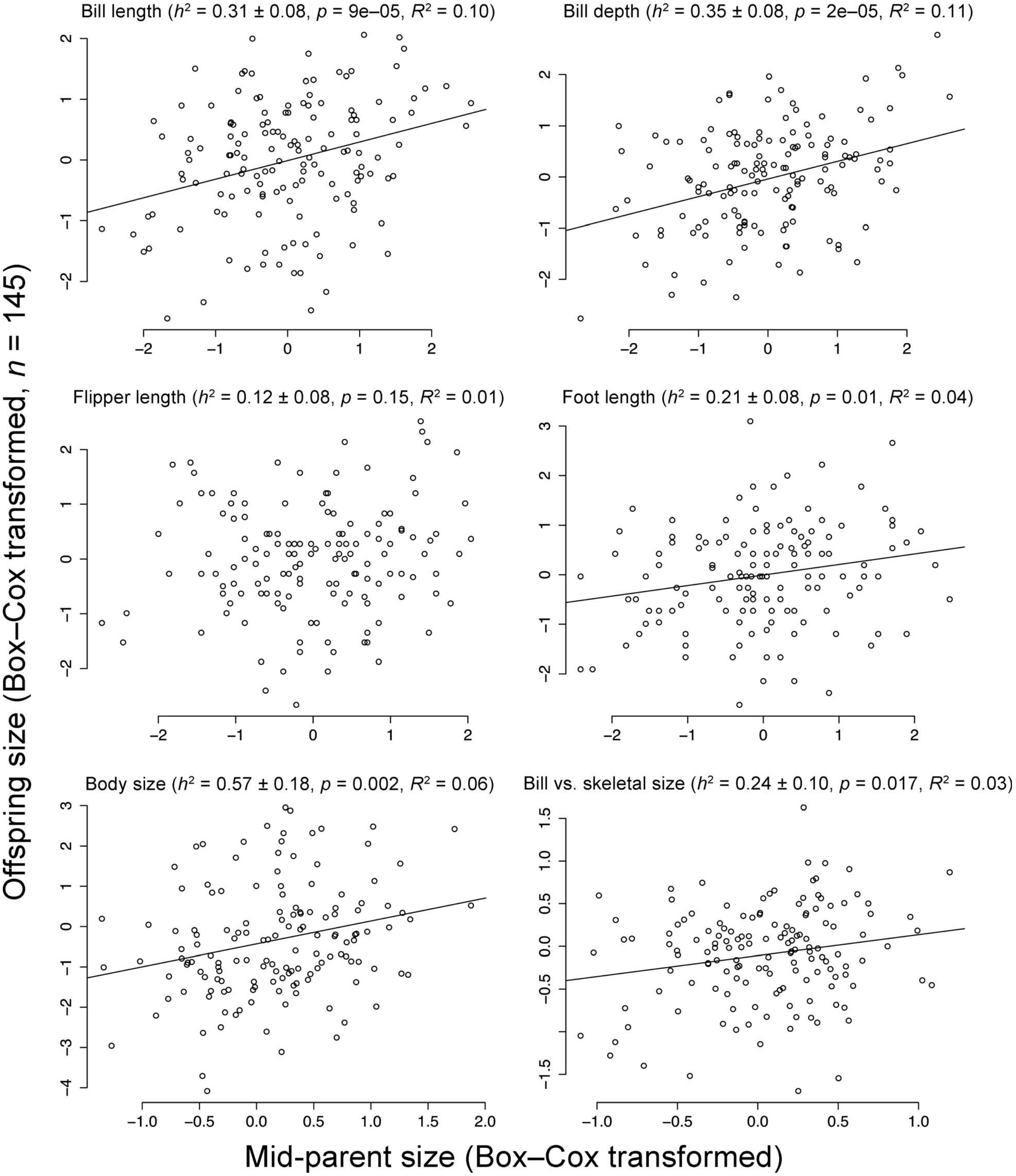
A few tests revealed natural selection on specific traits with high power (>0.8) but "nonsignificant" test P values ($P > 0.05$). Particularly, selection tests for female bill length in 1992, flipper length in 2009, and male bill length in 2000 had high power for the linear term in both linear and quadratic models, but test P values >0.05 ($P = 0.052$, 0.063, and 0.139). The result for male bill length in 2000 is especially surprising given the relatively large P value, but it may be due to the strong correlation (genetically and phenotypically) between bill length and depth and the significant selection on bill depth in that year. For females, the high power but nonsignificant P values suggest that there was likely weak directional selection on bill length in 1992 and weak directional selection on flipper length in 2009, but the relationship between size and reproductive success may not be strictly linear, leading to the larger P values.

APPENDIX TABLE 6. Test results, by year, for male Magellanic Penguins: standardized selection coefficients; standard deviations of coefficients (SD); test *P* values; power of each test; and direction of selection for bill length, bill depth, flipper length, foot length, and two multivariate body-size traits from a principal component analysis (PC1 = overall body size; PC2 = bill–skeletal contrast). Selection coefficients in bold are tests with sufficient power (>0.8). Traits with significant selection are specified as "linear" or "quadratic," depending on which model gave the significant linear coefficient, and with a superscript 2 if the quadratic term in the quadratic model was significant. If there was a significant linear term in both the linear and quadratic models, the results for both models are shown unless the Akaike's Information Criterion value of one was lower by >2. "NA" indicates years without significant selection coefficients. The percentage of chicks starved (out of all chicks) in a given year is shown for comparison with years that had significant selection.

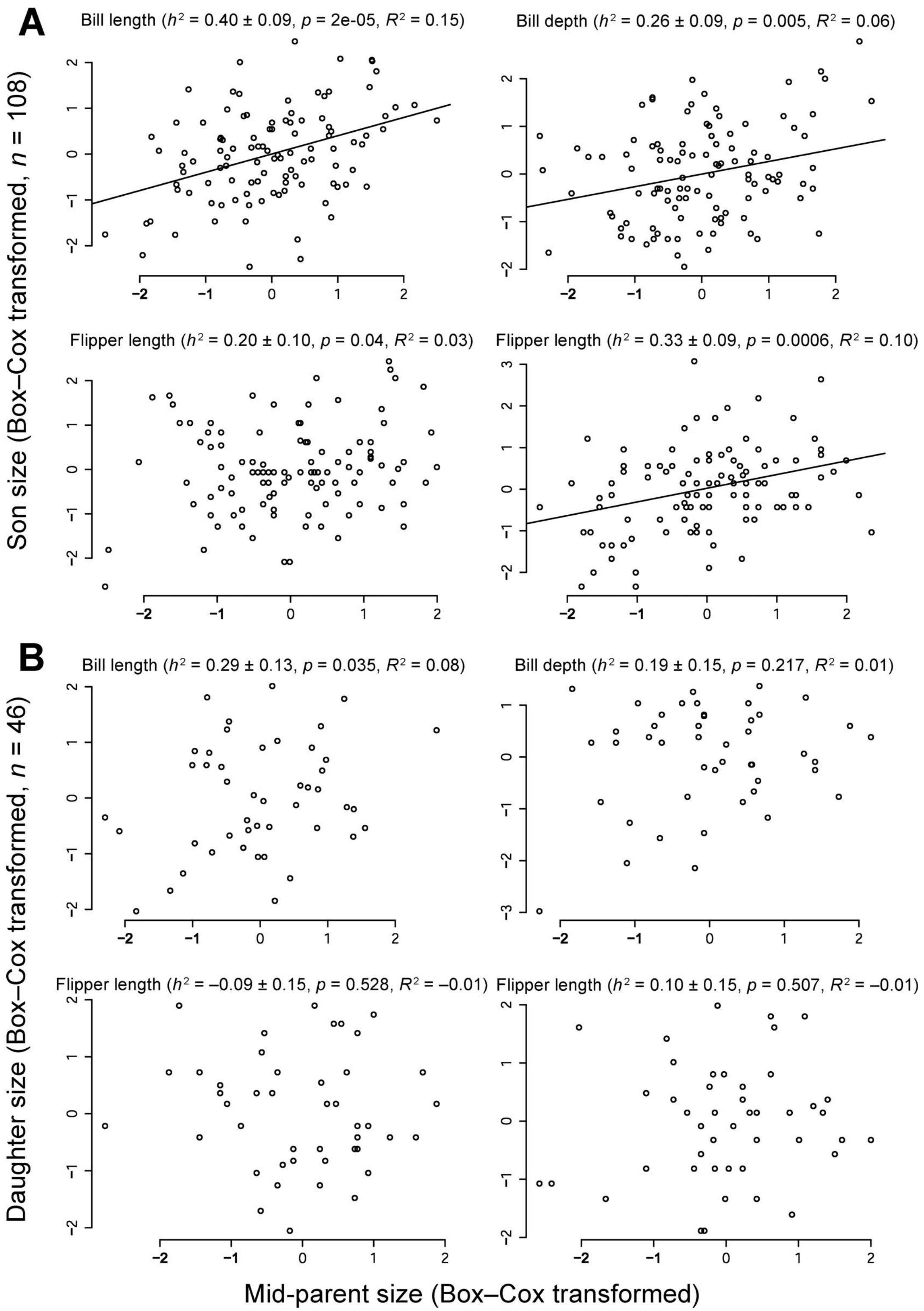
Year	*n*	Trait	Coefficient	SD	Test *P* value	Power of test	Direction of selection	Chicks starved (%)
1983	483	NA	NA	NA	NA	NA	NA	0.273
1984	785	NA	NA	NA	NA	NA	NA	0.658
1985	732	Bill depth^2	0.076	0.032	0.013	0.6589	Disruptive	0.207
1986	781	NA	NA	NA	NA	NA	NA	0.467
1987	747	**Bill length (linear)**	**0.359**	**0.121**	**0.003**	**>0.9999**	**Longer**	0.697
		Bill length (quadratic)	**0.359**	**0.121**	**0.008**	**>0.9999**	**Longer**	
1988	613	NA	NA	NA	NA	NA	NA	0.280
1989	626	NA	NA	NA	NA	NA	NA	0.389
1990	511	NA	NA	NA	NA	NA	NA	0.553
1991	369	**Bill depth (linear)**	**0.429**	**0.171**	**0.013**	**>0.9999**	**Deeper**	0.339
		Bill depth (quadratic)	**0.425**	**0.171**	**0.022**	**>0.9999**	**Deeper**	
1992	417	Bill depth^2	−0.092	0.034	0.017	0.5913	Stabilizing	0.323
1993	697	NA	NA	NA	NA	NA	NA	0.289
1994	626	**Bill length (quadratic)**	**0.127**	**0.059**	**0.023**	**0.9369**	**Longer**	0.286
		Bill length^2	0.068	0.029	0.023	0.5218	Disruptive	
1995	396	NA	NA	NA	NA	NA	NA	0.393
1996	440	NA	NA	NA	NA	NA	NA	0.286
1997	437	**Bill length (linear)**	**0.194**	**0.071**	**0.007**	**0.9919**	**Longer**	0.658
		Bill length (quadratic)	**0.162**	**0.077**	**0.013**	**0.9588**	**Longer**	
		Foot length (linear)	**0.176**	**0.072**	**0.014**	**0.9788**	**Longer**	
		Foot length (quadratic)	**0.171**	**0.073**	**0.047**	**0.9728**	**Longer**	
		PC1 (linear)	**0.135**	**0.059**	**0.022**	**0.8796**	**Bigger**	
1998	448	NA	NA	NA	NA	NA	NA	0.471
1999	474	NA	NA	NA	NA	NA	NA	0.209
2000	364	**Bill depth (linear)**	**0.276**	**0.128**	**0.032**	**0.9998**	**Deeper**	0.862
		Flipper length (quadratic)	**0.450**	**0.134**	**<0.001**	**>0.9999**	**Longer**	
		Flipper length^2	**0.165**	**0.050**	**<0.001**	**0.9328**	**Disruptive**	
		Foot length^2	**0.165**	**0.069**	**0.035**	**0.9328**	**Disruptive**	
		PC1 (linear)	**0.218**	**0.098**	**0.028**	**0.9939**	**Bigger**	
		PC1 (quadratic)	**0.208**	**0.098**	**0.003**	**0.9897**	**Bigger**	
		PC1^2	0.083	0.032	0.003	0.4743	Disruptive	
2001	330	NA	NA	NA	NA	NA	NA	0.459
2002	347	**Bill length (linear)**	**0.265**	**0.107**	**0.013**	**0.9995**	**Longer**	0.538
		Bill length (quadratic)	**0.261**	**0.109**	**0.046**	**0.9993**	**Longer**	
		Bill depth (linear)	**0.213**	**0.107**	**0.048**	**0.9897**	**Deeper**	
		Bill depth^2	**−0.158**	**0.076**	**0.017**	**0.9019**	**Stabilizing**	
		PC1 (linear)	**0.230**	**0.092**	**0.016**	**0.9957**	**Bigger**	
2003	398	Bill length (quadratic)	0.118	0.051	0.021	0.7600	Longer	0.394
		Bill length^2	0.062	0.030	0.021	0.3409	Disruptive	
2004	408	**PC2 (quadratic)**	**0.181**	**0.072**	**0.031**	**0.9775**	**Smaller bill, larger flip/foot**	0.168
		PC2^2	−0.113	0.052	0.031	0.7369	Stabilizing	
2005	474	Bill length (linear)	0.092	0.042	0.030	0.6388	Longer	0.438
		Bill length (quadratic)	0.076	0.043	0.027	0.5030	Longer	
		Bill length^2	−0.036	0.023	0.027	0.1943	Stabilizing	
2006	396	NA	NA	NA	NA	NA	NA	0.257
2007	592	NA	NA	NA	NA	NA	NA	0.185
2008	577	NA	NA	NA	NA	NA	NA	0.117
2009	628	NA	NA	NA	NA	NA	NA	0.409
2010	602	NA	NA	NA	NA	NA	NA	0.382

APPENDIX TABLE 7. Test results, by year, for female Magellanic Penguins: standardized selection coefficients; standard deviations of coefficients (SD); test *P* values; power of each test; and direction of selection for bill length, bill depth, flipper length, foot length, and two multivariate body-size traits from a principal component analysis (PC1 = overall body size; PC2 = bill–skeletal contrast). Selection coefficients in bold are tests with sufficient power (>0.8). Traits with significant selection are specified as "linear" or "quadratic," depending on which model gave the significant linear coefficient, and with a superscript 2 if the quadratic term in the quadratic model was significant. If there was a significant linear term in both the linear and quadratic models, the results for both models are shown unless the Akaike's Information Criterion value of one was lower by >2. "NA" indicates years without significant selection coefficients. The percentage of chicks starved (out of all chicks) in a given year is shown for comparison with years that had significant selection.

Year	*n*	Trait	Coefficient	SD	Test *P* value	Power of test	Direction of selection	Chicks starved (%)
1983	452	**Bill depth (linear)**	**−0.155**	**0.046**	**<0.001**	**0.9501**	**Shallower**	0.273
		Bill depth (quadratic)	**−0.157**	**0.045**	**<0.001**	**0.9543**	**Shallower**	
		Bill depth^2	0.069	0.031	<0.001	0.4285	Disruptive	
		Flipper length (linear)	−0.103	0.046	0.025	0.7060	Shorter	
		Flipper length (quadratic)	−0.103	0.046	0.026	0.7059	Shorter	
		PC1 (linear)	**−0.175**	**0.056**	**0.002**	**0.9808**	**Smaller**	
		PC1^2	0.044	0.043	0.005	0.2386	Disruptive	
1984	783	NA	NA	NA	NA	NA	NA	0.658
1985	718	**Bill depth (linear)**	**−0.125**	**0.045**	**0.006**	**0.9556**	**Shallower**	0.207
		Bill depth (quadratic)	**−0.125**	**0.045**	**0.014**	**0.9556**	**Shallower**	
		PC1 (linear)	**−0.105**	**0.053**	**0.046**	**0.8782**	**Smaller**	
1986	775	NA	NA	NA	NA	NA	NA	0.467
1987	734	**Bill length (linear)**	**−0.243**	**0.120**	**0.043**	**>0.999**	**Shorter**	0.697
		Foot length (linear)	**−0.239**	**0.120**	**0.046**	**>0.999**	**Shorter**	
1988	612	NA	NA	NA	NA	NA	NA	0.280
1989	622	NA	NA	NA	NA	NA	NA	0.389
1990	509	NA	NA	NA	NA	NA	NA	0.553
1991	362	**Foot length (linear)**	**0.434**	**0.179**	**0.016**	**>0.999**	**Longer**	0.339
		Foot length (quadratic)	**0.444**	**0.180**	**0.045**	**>0.999**	**Longer**	
1992	356	NA	NA	NA	NA	NA	NA	0.323
1993	559	**Bill length (linear)**	**0.148**	**0.055**	**0.007**	**0.9678**	**Longer**	0.289
		Bill length (quadratic)	**0.167**	**0.058**	**0.017**	**0.9892**	**Longer**	
		Bill depth (linear)	**0.131**	**0.055**	**0.017**	**0.9263**	**Deeper**	
		Bill depth (quadratic)	**0.118**	**0.057**	**0.042**	**0.8732**	**Deeper**	
		PC1 (linear)	**0.114**	**0.055**	**0.038**	**0.8525**	**Bigger**	
		PC2^2	**−0.130**	**0.052**	**0.012**	**0.9229**	**Stabilizing**	
1994	457	NA	NA	NA	NA	NA	NA	0.286
1995	384	Bill length^2	−0.092	0.038	0.040	0.5615	Stabilizing	0.393
1996	369	Flipper length (quadratic)	0.080	0.046	0.033	0.4558	Longer	0.286
		Flipper length^2	−0.051	0.026	0.033	0.2524	Stabilizing	
1997	334	NA	NA	NA	NA	NA	NA	0.658
1998	381	NA	NA	NA	NA	NA	NA	0.471
1999	320	NA	NA	NA	NA	NA	NA	0.209
2000	252	**Bill length (quadratic)**	**0.566**	**0.175**	**0.002**	**>0.999**	**Longer**	0.862
		Bill length^2	**0.211**	**0.070**	**0.002**	**0.9550**	**Disruptive**	
		Bill depth (linear)	**0.359**	**0.148**	**0.016**	**>0.999**	**Deeper**	
		Bill depth (quadratic)	**0.391**	**0.163**	**0.049**	**>0.999**	**Deeper**	
		PC1 (quadratic)	**0.843**	**0.338**	**0.040**	**>0.999**	**Bigger**	
		PC1^2	**0.145**	**0.071**	**0.040**	**0.7424**	**Disruptive**	
		PC2 (linear)	**−0.502**	**0.218**	**0.022**	**>0.999**	**Bigger bill, smaller body**	
2001	171	NA	NA	NA	NA	NA	NA	0.459
2002	312	Bill length^2	−0.136	0.053	0.037	0.7740	Stabilizing	0.538
2003	388	**PC2 (linear)**	**0.228**	**0.084**	**0.007**	**>0.999**	**Bigger body, smaller bill**	0.394
		PC2 (quadratic)	**0.229**	**0.090**	**0.026**	**>0.999**	**Bigger body, smaller bill**	
2004	406	NA	NA	NA	NA	NA	NA	0.168
2005	468	NA	NA	NA	NA	NA	NA	0.438
2006	458	NA	NA	NA	NA	NA	NA	0.257
2007	569	Bill length	0.080	0.038	0.037	0.6030	Longer	0.185
2008	587	NA	NA	NA	NA	NA	NA	0.117
2009	619	NA	NA	NA	NA	NA	NA	0.409
2010	558	NA	NA	NA	NA	NA	NA	0.382



APPENDIX FIGURE 3. Scatter plots of relationships between parent size ("mid-parent"; see text) and offspring size and heritability estimates (narrow-sense $h^2 \pm$ SE) for bill, flipper, foot, and body sizes in Magellanic Penguins. Solid lines indicate significant ($P <$ 0.0183) linear relationships. Mid-parent bill sizes, foot length, and body size (based on a principal components analysis) are significantly heritable in offspring, based on the adjusted α ($\alpha = 0.0183$ from a false-discovery-rate adjustment). All traits for offspring and parents were Box-Cox transformed for normality before running analyses.



APPENDIX FIGURE 4. Scatter plots of relationships between parent size ("mid-parent"; see text) and offspring size (sons or daughters) and heritability estimates (narrow-sense $h^2 \pm$ SE) for bill, flipper, and foot sizes in Magellanic Penguins. Solid lines indicate highly significant ($P < 0.0150$) linear relationships. All traits for mid-parent, sons, and daughters were Box-Cox transformed for normality before analyses. **(A)** Mid-parent bill and foot sizes are significantly heritable in sons, based on the adjusted α ($\alpha = 0.0150$ from a false-discovery-rate adjustment). **(B)** Mid-parent sizes are not significantly heritable in daughters.