The genetics of morphological and behavioral island traits in deer mice

Felix Baier & Hopi E. Hoekstra

Howard Hughes Medical Institute, Department of Molecular & Cellular Biology, Department of Organismic & Evolutionary Biology, Museum of Comparative Zoology, Harvard University, Biological Laboratories, 16 Divinity Avenue, Cambridge, MA 02138, email: hoekstra@oeb.harvard.edu

ABSTRACT

Animals on islands often exhibit dramatic differences in morphology and behavior compared to mainland individuals, a phenomenon known as the "island syndrome". These differences, such as changes in body size and aggression, are thought to be adaptations to island environments, where there are high resource levels, low predation, limited dispersal, and thus high population densities. However, the extent to which island traits have a genetic basis or instead represent plastic responses to environmental extremes is often unknown. Here, we revisit a classic case of island syndrome in deer mice (*Peromyscus maniculatus*) from British Columbia. Previous field studies suggested that Saturna Island mice evolved several island traits, including higher body weight and reduced aggression relative to mainland populations. Using historical collections, we show that Saturna Island mice and those from neighboring islands are approximately 35% (~5g) heavier than mainland mice. We then collected mice from two focal populations: Saturna Island and a nearby mainland population. First, using molecular data, we find that these populations are genetically distinct, having diverged approximately 10 thousand years ago. Second, we established laboratory colonies and find that Saturna Island mice are heavier both because they are longer and have disproportionately more lean mass. These trait differences are maintained in second-generation captive-born mice raised in a common environment, implying a strong heritable component. In addition, island-mainland hybrids are heavier when born to island mothers than to mainland mothers, revealing a maternal genetic effect on body weight. Next, using behavioral testing in the lab, we also find that wild-caught island mice are less aggressive than mainland mice. However, lab-raised mice born to these founders do not differ in aggression, regardless of whether they are tested in conditions that induce low or high aggression, suggesting the large behavioral difference observed between wild-caught island and mainland individuals is likely a plastic response. Together, our results reveal that these mice respond differently to environmental conditions on islands – evolving both heritable changes in a morphological trait and also expressing a plastic phenotypic response in a behavioral trait.

INTRODUCTION

Islands offer some of the most tractable examples of evolutionary adaptation and diversification (LOSOS AND RICKLEFS 2009). Because of their geographic isolation, they form self-contained environments that facilitate the interpretation of evolutionary patterns and processes. For example, many island populations of animals share consistent differences in demography, body size, reproductive rate, anti-predator behavior, and territorial aggression, a phenomenon known as the 'island syndrome' (FOSTER 1964; VAN VALEN 1973; LOMOLINO 1985; STAMPS AND BUECHNER 1985; ADLER AND LEVINS 1994). The repeated appearance of similar traits in similar ecological conditions across taxa suggests that they evolved as adaptations during the establishment of island populations (BLONDEL 2000; NOVOSOLOV *et al.* 2013). Thus, one implicit assumption is that these island syndrome traits have a genetic basis, resulting from directional selection and/or founder effects during island colonization (MAYR 1954; ADLER AND LEVINS 1994; BERRY 1996). However, island traits could also be driven, in part or fully, by phenotypic plasticity, but the role of plasticity in island evolution has received limited attention.

Plastic phenotypic changes are often considered primarily during the initial stages of the colonization of novel habitat. For example, it has been suggested that "environmental variation [on islands] may cause 'purely phenotypic' changes within a single generation [short-term changes] and genetically determined changes over evolutionary time [long-term changes]" (ADLER AND LEVINS 1994). Phenotypic plasticity can transiently reduce the intensity of selection and create a time lag between the initial exposure to novel environments and the necessity for genetic adjustment (SCHLICHTING AND PIGLIUCCI 1998), thus potentially enabling the colonization of novel environments like islands. It is thought that selection would then act on heritable differences in these traits over evolutionary time (MACARTHUR AND WILSON 1967). Eventually, through a process of genetic assimilation, a trait would become genetically and developmentally canalized, and environmental plasticity would be lost (WADDINGTON 1953). The few studies that have demonstrated phenotypic plasticity in island traits and genetic change over evolutionary time have focused primarily on morphological traits (LOSOS *et al.* 2000; AUBRET AND SHINE 2009; CORL *et al.* 2018). However, behavioral traits are thought to be

influenced by environmental factors more than morphology (WEST-EBERHARD 2003). Phenotypic plasticity in behavior could thus be critical to the colonization of novel island environments, and may even drive selection on morphological traits. Nonetheless, it remains unclear if these plastic behaviors ultimately undergo genetic assimilation themselves (WCISLO 1989; WEST-EBERHARD 1989; PRICE *et al.* 2003).

A key challenge in detecting heritable components underlying island syndrome traits has been that most reports are based on experiments in the field, or especially for behavioral traits, wild-caught individuals subsequently tested in the lab (e.g. HALPIN AND SULLIVAN 1978; HALPIN 1981; GRAY AND HURST 1998; ORROCK 2010). Detection of heritable differences in the wild, for example through the quantitative-genetic analysis of trait variance (SCHLICHTING AND PIGLIUCCI 1998), can be challenging. By contrast, a genetic component can be measured if wild-derived animals are born and raised in a common environment. Using this approach, one can assess the extent to which morphological and behavioral island syndrome traits associated with islands have a heritable genetic basis.

Here, we revisit a classic case of island syndrome in deer mice (*Peromyscus maniculatus*) from western Canada. Field studies in the 1970s have shown that the ecological conditions differ for deer mice on Saturna Island off the coast of British Columbia – namely they have a higher population density, lower dispersal rate, and smaller home range size than deer mice from a nearby mainland population (SULLIVAN 1977). In addition, wild-caught island deer mice have a larger body size (REDFIELD 1976; SULLIVAN 1977) and displayed fewer aggressive and defensive interactions than mainland deer mice (HALPIN AND SULLIVAN 1978; HALPIN 1981). Here, we investigate the extent of genetic versus environmental components underlying morphological and behavioral island syndrome traits by measuring body size and aggressive behavior both in wild-caught island and mainland deer mice as well as in their captive-born offspring.

RESULTS

Island mice are heavier than mainland mice

Using data from available museum specimens, we first compared 504 weight records of adult island (N = 328) and mainland (N = 176) deer mice from 49 locations in the Strait of Georgia and coastal British Columbia. With few exceptions, island populations had a greater median weight than neighboring mainland populations (Fig. 1A). The overall median weight of island specimens was 21g compared to a median weight of 16g in the mainland specimens (Kruskal-Wallis test, $P < 0.0001$), an increase of 31.25% (Fig. 1B). Although there was overlap in the distributions, there was only one record of a mouse more than 23g from a mainland population, whereas 30.48% of all island mice were heavier than 23g.

We next focused on two populations, one island (Saturna Island) and one on the mainland (Malcolm Knapp Research Forest; Fig. 1A). We collected wild mice from these two populations (N = 21 and 23, respectively) and found that these populations largely recapitulated the general pattern of body weight observed in the Strait of Georgia. The median weight of adult Saturna Island mice was 20g (± 1 SD 16.8-23.2g) compared to a median weight of 14.5g (± 1 SD 12.7-16.3g) in Malcolm Knapp Research Forest mice (Kruskal-Wallis test, $P < 0.0001$), an increase of 37.93% (Fig. 1B).

Island and mainland mice are genetically distinct populations

Using genetic data, we tested the genetic relatedness of our focal island and mainland populations and estimated their divergence time. We used a genotype-by-sequencing approach (ddRADseq; PETERSON *et al.* 2012) to obtain 80,248 genome-wide variants in a sample of deer mice from Saturna Island (N = 25), neighboring Pender Island (N = 8), and the Malcolm Knapp Research Forest (N = 21; Fig. 1A). Using these variants, we conducted a genetic principal component analysis (PCA) and found that these populations are genetically separable (Fig. 1C). In addition, none of the individuals examined showed evidence of recent admixture, suggesting that gene flow between island and mainland populations, as well as between neighboring islands, is limited (Fig. 1D). Lastly, we applied Bayesian divergence time estimations to a subset of high coverage genetic variants (N = 7,092) and specimens (N = 14) to obtain a time-calibrated phylogenetic tree of these populations, using *Peromyscus keeni* as an outgroup (Fig. 1E).

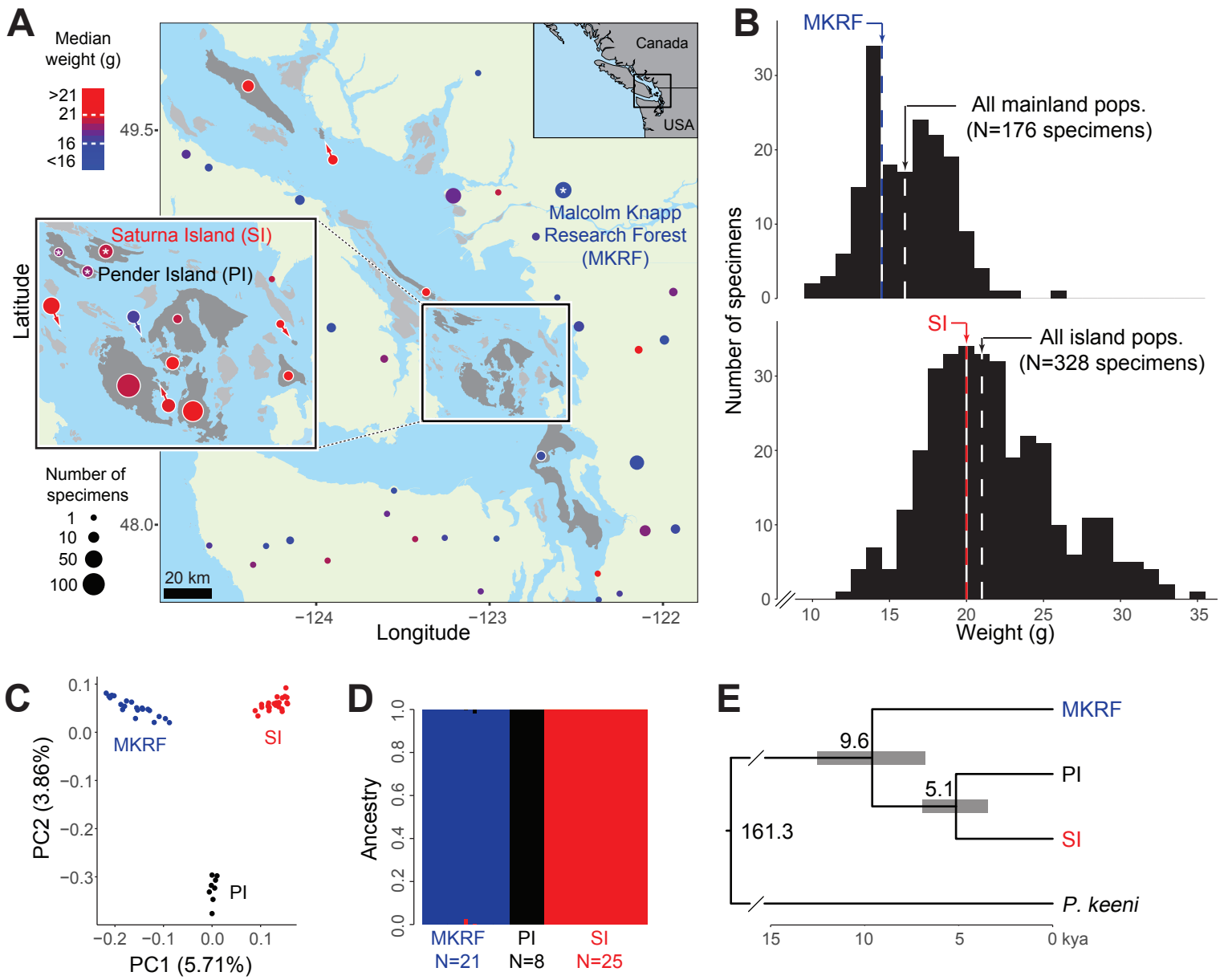


Figure 1. Body weight of island and mainland populations in the study area.
(**A**) Map of coastal British Columbia, Canada, showing the study area. The color of collection points indicates the median weight of *Peromyscus maniculatus* populations, and the diameter of points indicates sample size. Collection points from the same island and mainland records within 8 km of each other were combined into a single point. Only adult males were included. Sampled islands are shaded dark gray, non-sampled islands light gray, and mainland areas light green. Live deer mice were collected from Saturna Island (SI) and the Malcolm Knapp Research Forest (MKRF) on the mainland to start laboratory colonies (starred). (**B**) Histograms of body weight of mainland (top) and island (bottom panel) specimens. The median weights are indicated by dashed lines. (**C**) Genetic principal component analysis of 54 deer mice from Saturna Island, Pender Island and the Malcolm Knapp Research Forest, based on 80,248 genome-wide variants. The percent of variance explained by the first and second principal component is provided in parentheses. (**D**) Admixture analysis of the dataset from (C) with a cluster number of K=3. (**E**) Bayesian divergence time estimates of island and mainland *P. maniculatus* populations and the outgroup *P. keeni* based on 7,092 variants from 14 specimens. Gray bars represent 95% highest posterior density intervals. The scale bar is in thousands of years ago (kya).

Using this approach, we estimated island mice diverged from mainland mice approximately 9.6 (6.8-12.5, 95% highest posterior density interval) thousand years ago (kya), while Pender Island and Saturna Island mice diverged approximately 5.1 (3.5-6.9) kya. These data are consistent with a postglacial isolation of deer mice on islands in the Strait of Georgia (MATHEWS *et al.* 1970) and little to no ongoing gene flow between island and mainland or among island populations.

Differences in body weight appear heritable and are driven by both differences in body length and proportional lean mass

To test whether differences in weight between Saturna Island and mainland mice observed in the field were retained across generations in a common environment, we established laboratory colonies of these two populations and then compared a set of wild-caught (N = 21) and captive-born (N = 25) mice. We used captive-born mice that were themselves born to captive-born parents (as two generations in captivity should minimize environmental, maternal and other non-heritable variation), and focused our analyses on male mice because the difference in weight was greater in males than in females (Suppl. Fig. 1). We found that the magnitude of difference in median weight between island and mainland mice observed in the field was retained in laboratory mice (island mice are 37.93% bigger in the field and 34.02% bigger in the lab; Fig. 2A); however, the overall weight of the mice was higher in the laboratory (captive-born mice were on average 34.99% heavier than wild-caught mice; Fig. 2A), likely due to an enriched and *ad libitum* diet.

We next asked how island mice attain their larger body weight and hypothesized that they might be heavier both because they are larger (i.e. longer) and/or because they are heavier relative to their body length. To distinguish among these possibilities, we first X-rayed mice to measure a set of 10 skeletal traits: length of the sacrum, skull, zygomatic bone, humerus, femur, ulna, tibia, and metatarsal calcaneus as well as the width of the skull and zygomatic bone. We found that median sacrum length, a proxy for body length, was approximately 14.39% larger in island than mainland mice for both field and lab individuals, but did not differ between field and lab populations (Fig. 2B), suggesting that

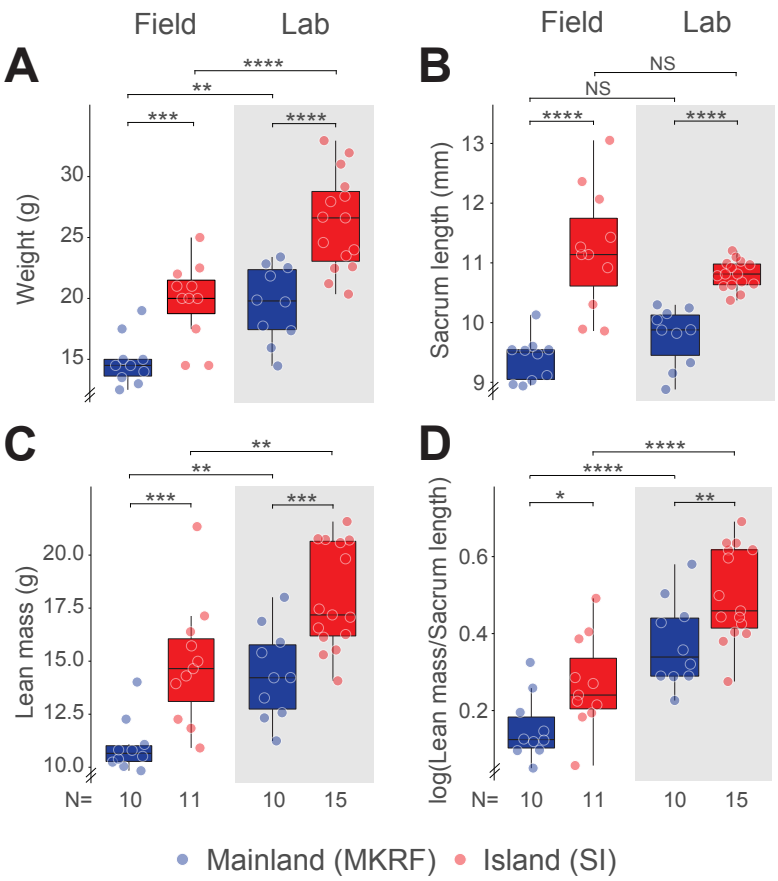


Figure 2. The genetic basis of body size traits in island mice.
(**A**) Body weight, (**B**) sacrum length, (**C**) lean mass, and (**D**) size-corrected lean mass in male wild- and captive-born island and mainland mice. Statistical significance evaluated by linear fixed effects models (see Methods for details). For all figures, box plots indicate median, interquartile range (IQR), and at most ± 1.5 IQR from hinges. NS=not significant, * P < 0.05, ** P < 0.01, *** P < 0.001, **** P < 0.0001.

island mice are longer than mainland mice but that environmental conditions (i.e. field versus lab) have little effect on body length. This pattern was largely recapitulated in other skeletal traits: size-corrected proxies for head length/width and several limb bones showed that island mice have disproportionately smaller heads and shorter front and hind legs in both field and lab populations, and little difference between field and lab conditions (Suppl. Fig. 2).

We next quantified body composition in field and laboratory mice using EchoMRI analyses. We found that island mice had more lean mass (Fig. 2C), but not fat mass (Suppl. Fig. 3), than mainland mice. Moreover, this increase in lean mass was not simply due to differences in body length (Fig. 2D). In addition, lean mass and relative lean mass was greater in captive-born than in wild-caught mice from the same location (Fig. 2C, D), suggesting that these traits are influenced by the environment. Collectively, these data show that island mice are heavier both because they are larger (i.e. longer) and because they have disproportionately more lean mass than mainland mice. In addition, because these trait differences are maintained both in a common environment and across generations, they likely have a strong genetic basis.

Island mice are born heavier with growth pulses at birth and weaning

To determine when these weight differences emerge across ontogeny, we weighed litters within 24 hours of birth, and for a subset continued weighing them every 4 days until weaning. We noted that island mice have slightly smaller litter sizes than mainland mice (Suppl. Fig. 4; Kruskal-Wallis test, $P = 0.034$); as litter size is often inversely correlated with birth weight, we thus included litter size as an explanatory variable in subsequent analyses. Because individual pups are difficult to track over time, we focused on comparisons between litters. We first found that island litters are heavier than mainland litters already at birth (Fig. 3A; linear fixed effects model, $t = -6.259$, $P = <0.0001$). This difference is not simply due to island mice ingesting more milk, because island litters were already heavier before their first milk meal (Suppl. Fig. 5; repeated measures linear mixed effects model; $t = 6.901$, $P = 6.96e-5$ [pre vs. post-feeding]; $t = 3.602$, $P = 0.004$ [Strain]). Furthermore, growth at birth in island mice is greater than in

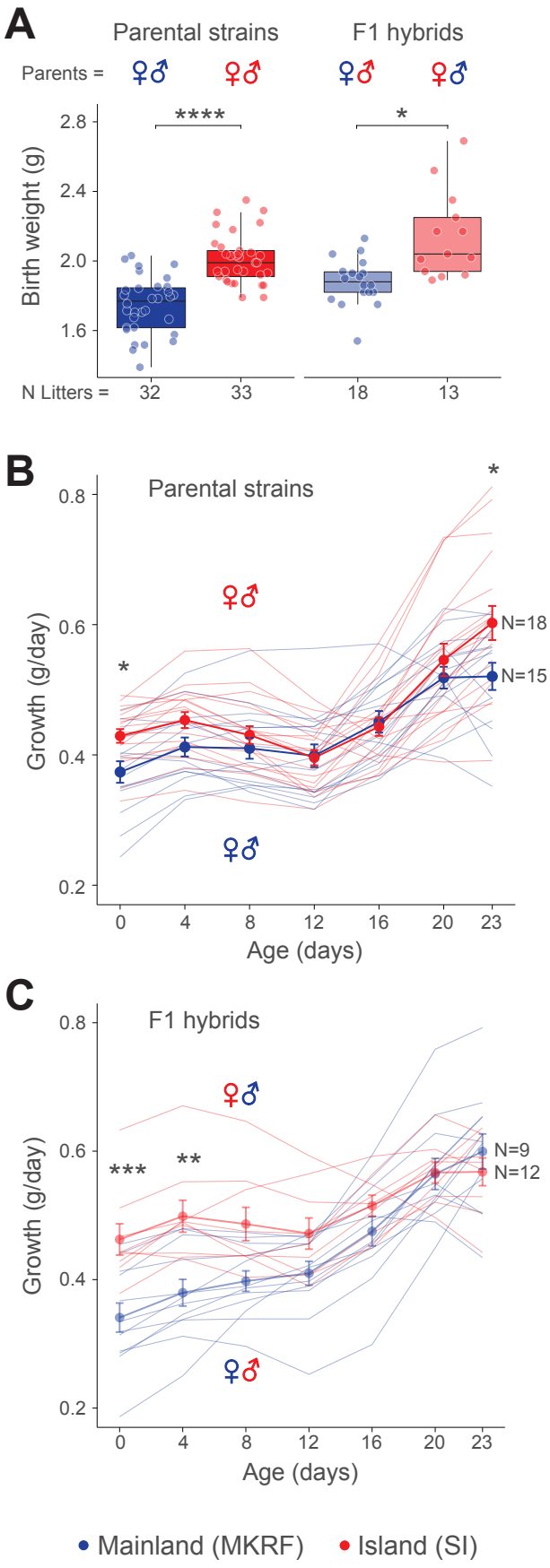


Figure 3. A maternal effect contributes to birth weight in island mice.
(**A**) Birth weight in island and mainland litters (left panel) and F1 hybrids by maternal strain (right panel). Points represent the mean weight of litter. Statistical significance evaluated by linear fixed effects model (see Methods for details). Growth curves of (**B**) island and mainland litters and (**C**) F1 hybrid litters by maternal strain. Points and error bars represent mean and SEM of growth of litters. Sample sizes are provided. * P < 0.05, ** P < 0.01, *** P < 0.001, **** P < 0.0001.

mainland mice (Fig. 3B; linear fixed effects model, t = -2.852, $P = 0.033$), but this significant difference disappears by postnatal day 4 (P4), and only reappears around weaning at P23 (Fig. 3B; linear fixed effects model, t = -2.748, $P = 0.039$). Notably, in both strains the post-natal growth peak at P4 slows down by P12, and growth only resumes and exceeds earlier levels subsequent to P16. Together, these results suggest that growth pulses around birth, and with the onset of sexual maturity at weaning, underlie the differences in weight between island and mainland mice.

A maternal effect contributes to birth weight and early growth in island mice

We next asked if these differences in weight are influenced by a maternal genetic effect by comparing first-generation (F1) hybrid litters born to mothers from either island or mainland populations. We first reciprocally paired litter mates from representative island and mainland parents. As before, we weighed litters at birth and then a subset every 4 days until weaning. We found that F1 litters born to island mothers are on average 14.11% heavier at birth than F1 litters born to mainland mothers (Fig. 3A; linear fixed effects model, t = -2.386, $P = 0.0241$), even when we controlled for family effects (wild cluster bootstrapped generalized linear model, 95% CI = 0.0022, 0.1749, $P = 0.029$). Furthermore, growth in F1 hybrid litters born to island mothers was significantly greater at birth (Fig. 3C; linear fixed effects model, t = -4.114, $P = 0.00048$) and at P4 (Fig. 3C; linear fixed effects model, t = -3.535, $P = 0.0034$), but we only saw a trend at P4 when we controlled for family effects (wild cluster bootstrapped generalized linear model, 95% CI = -0.1076, 0.6122, $P = 0.069$). Growth was not different on days 8-20 or at weaning (Fig. 3C; linear fixed effects model, t = 1.085, $P = 0.8400$). Finally, in F1 adults in which sex could be determined, we found a larger difference between offspring with island versus mainland mothers in male compared to female hybrids (Suppl. Fig. 6). Together, these results suggest that the growth difference in island versus mainland mice around birth, but not at the onset of sexual maturity at weaning, is likely governed by a maternal effect.

Parental care does not affect early growth differences

We next tested if differences in parental care contribute to the differences in growth observed between island and mainland mice. To test this hypothesis, we conducted a cross-fostering experiment by swapping island and mainland litters (i.e. island pups raised by mainland parents and vice versa) within 48 hours of birth. We then weighed the pups every 4 days until weaning (P23). We found little evidence that cross-fostering affected growth in either island or mainland mice (Suppl. Fig. 7A, B). For example, mainland litters raised by island parents grew the same (days 0 to 20) or slower (day 23) compared to non-fostered mainland litters (Suppl. Fig. 7A). These data suggest that the increases in growth in island litters and F1 hybrids born to island mothers are not due to higher parental care by island parents.

Wild-caught island mice are less aggressive than wild-caught mainland mice

We next compared levels of territorial aggression in both wild-caught and captive-born mice from island and mainland populations. We first developed a behavioral assay in the laboratory to test aggression levels in male mice based on the resident-intruder paradigm (Fig. 4A). Briefly, males were co-housed with females to initiate territoriality. Females were removed immediately before the trial from the cage. A conspecific, socially-housed male intruder was then introduced into a closed-off section of the cage. After habituation, the resident and intruder were allowed to interact, and behavior was video-recorded for 15 minutes. We conducted a second trial against a different intruder one day later. We then quantified a suite of social behaviors (EISENBERG 1962; EISENBERG 1963) as state (duration) events from the videos, and averaged resident behaviors across trials.

We first tested the male mice that were collected in the wild and paired to establish laboratory colonies (N=7 island, N=6 mainland). At the time of behavioral testing, these individuals had been co-housed for ~1.5-2 years with a female in the lab. We performed a principal component analysis (PCA) to reduce the dimensionality of the behavioral dataset (Fig. 4B). The first two principal components (PCs) explained a cumulative proportion of 47.92% of the variance; 95% confidence intervals for the two strains only

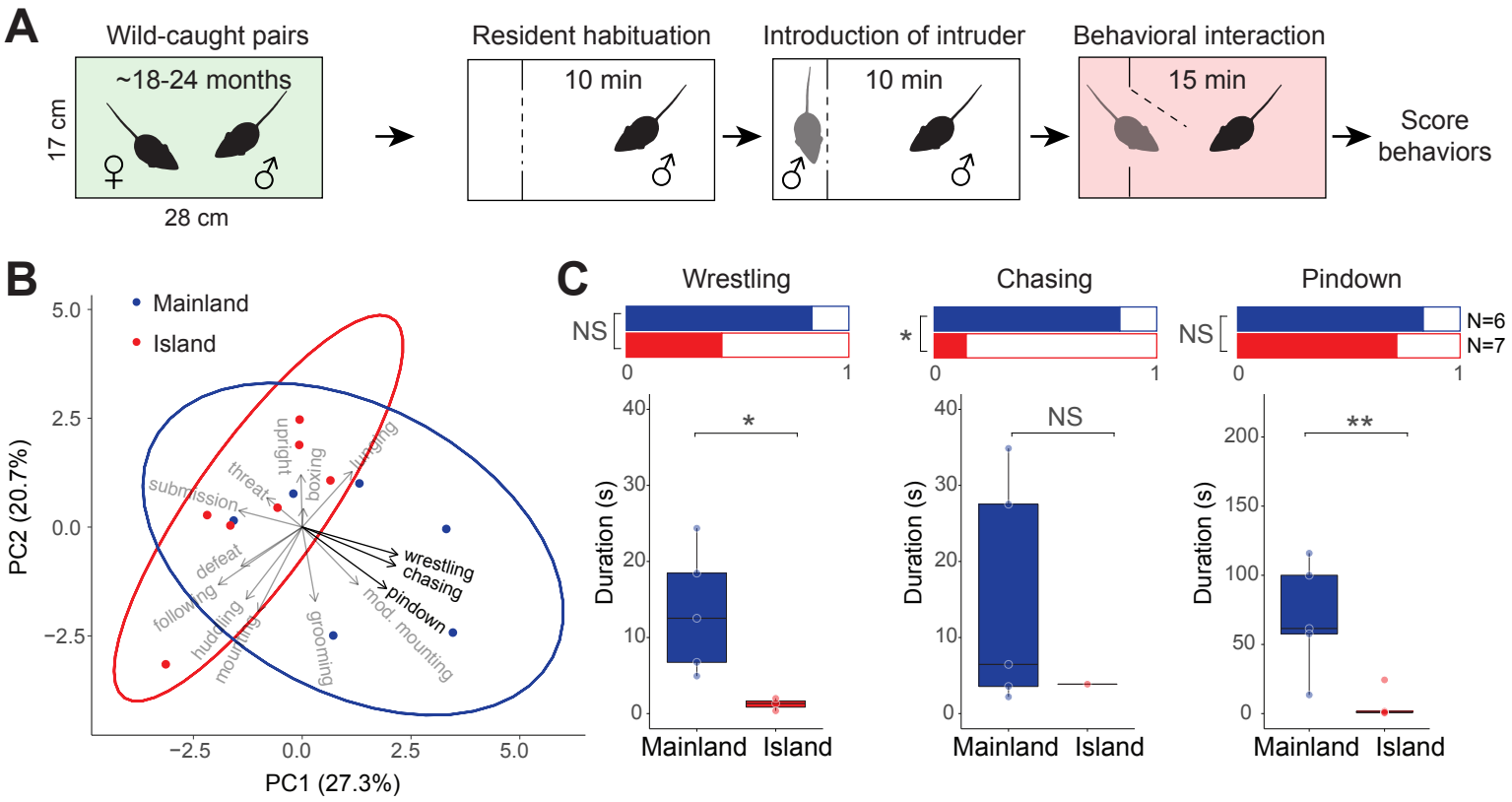


Figure 4. Territorial aggression in wild-caught island and mainland mice.
(**A**) Schematic of the behavioral assay. For each resident, we ran the assay twice separated by a day using two different conspecific intruders. We scored aggressive, defensive, and cohesive behaviors and averaged scores across trials. (**B**) Principal component analysis (PCA) of the duration of wild-caught resident behaviors. Ellipses represent 95% confidence intervals. Arrow directions indicate how behaviors contribute to the two PCs, and arrow lengths are proportional to the strength of contribution. Behaviors highlighted in bold were selected for additional analysis. (**C**) Aggressive behavior in wild-caught founder animals. Bar graphs (top) show the proportions of animals engaging in wrestling, chasing, and pindown behaviors. Boxplots (bottom) show duration of each behavior for those individuals that engaged in that behavior. Statistical significance of proportions and durations evaluated by generalized linear models and linear fixed effects models, respectively (see Methods for details). Sample sizes are provided. NS=not significant, * $P < 0.05$, ** $P < 0.01$.

partially overlapped, and 67% of mainland animals were outside the confidence interval for island mice. The behaviors that contributed most to the first PC were wrestling duration (12.94%), chasing duration (12.65%), and pindown duration (11.42%).

Using this dataset, we tested for differences in these behaviors between wild-caught island and mainland mice. We used a two-stage hurdle model approach to separately test for differences in the proportion of animals displaying a behavior as well as the number and duration of bouts once a behavior was initiated. We found no significant difference in the proportion of wild-caught island and mainland mice showing wrestling behavior (83.3% vs. 42.86%, logistic regression model, z = -1.421, $P = 0.155$), but mainland animals that wrestled did so significantly longer than island animals (Fig. 4C, Suppl. Fig. 8A; linear fixed effects model, t = -2.512, $P = 0.0458$). For chasing, significantly more mainland animals chased than island animals (83.3% vs. 14.29%, logistic regression model, z = -2.211, $P = 0.027$), but we found no difference in duration once animals chased (Fig. 4C, Suppl. Fig. 8A; linear fixed effects model, t = -0.666, $P = 0.5418$). Similar to wrestling behavior, we found no significant difference between populations in the proportion of animals exhibiting pindown behavior (83.3% vs. 71.43%, logistic regression model, z = -0.503, $P = 0.615$), but mainland animals that exhibited pindown behavior did so significantly longer than island animals (Fig. 4C, Suppl. Fig. 8A; linear fixed effects model, t = -3.460, $P = 0.0086$). These differences in aggression could not be explained by differences in the time that animals had spent in captivity, the time since a litter was last sired, or the weight difference between resident and intruder (Suppl. Fig. 9). Together, these results suggest that wild-caught island mice show reduced territorial aggression compared to wild-caught mainland mice, consistent with previous studies on wild mice from these populations (HALPIN AND SULLIVAN 1978; HALPIN 1981).

Captive-born mice tested in standard conditions are not aggressive

We next compared aggression in captive-born island and mainland mice to determine if the behavioral differences observed in wild mice were retained in a controlled laboratory environment and over generations. We initially paired residents with a virgin female for one week prior to the behavioral testing (Fig. 5A), a well-established

procedure to induce territoriality and aggression in laboratory mice (BLANCHARD AND BLANCHARD 1988). However, both captive-born island and mainland mice displayed low levels of aggression in this assay (Fig. 5B, Suppl. Fig. 8B). We found neither a significant difference in the proportion of island versus mainland mice that displayed wrestling (21.73% vs. 24.39%, logistic regression model, z = 0.240, $P = 0.81$), chasing (30.43% vs. 17.07%, logistic regression model, z = -1.227, $P = 0.22$), or pindown (47.83% vs. 53.66%, logistic regression model, z = 0.448, $P = 0.654$), nor did we find a significant difference in the duration of wrestling (linear fixed effects model, t = -1.938, $P = 0.0746$), chasing (linear fixed effects model, t = -1.432, $P = 0.1777$), or pindown (linear fixed effects model, t = 0.268, $P = 0.7905$) once animals engaged in these behaviors. These results are in stark contrast to differences observed in wild-caught mice; both island and mainland lab-reared animals behaved similarly to the less aggressive wild-caught island mice.

Reproductively active captive-born mice do not differ in aggression levels

While one week of pairing with a female typically induces sexual activity and territoriality in male laboratory mice, it may not be sufficient for deer mice. To explore this possibility, we tested for a subset of residents when they sired their first litter. We found that the majority of males sired litters approximately three weeks after they were paired (Suppl. Fig. 10A), suggesting that few mice had become sexually active by the time of testing after only a single week. Furthermore, we identified mice among this subset that had not sired a litter by the time of first testing, and re-tested them after they had sired and been co-housed with a litter until weaning. As expected, residents wrestled significantly more after they had reproduced (Suppl. Fig. 10B; repeated measures linear mixed effects model; t = 2.578, $P = 0.0327$ [Trial]; t = 0.337, $P = 0.7462$ [Strain]), suggesting that reproductive experience is crucial to aggression in these deer mouse populations.

As aggression in male deer mice peaks at the time of and just following copulation with a receptive female (DE CATANZARO 1981; DEWSBURY 1981; DEWSBURY 1984), we next aimed to test for island-mainland differences in aggression at this time. We used

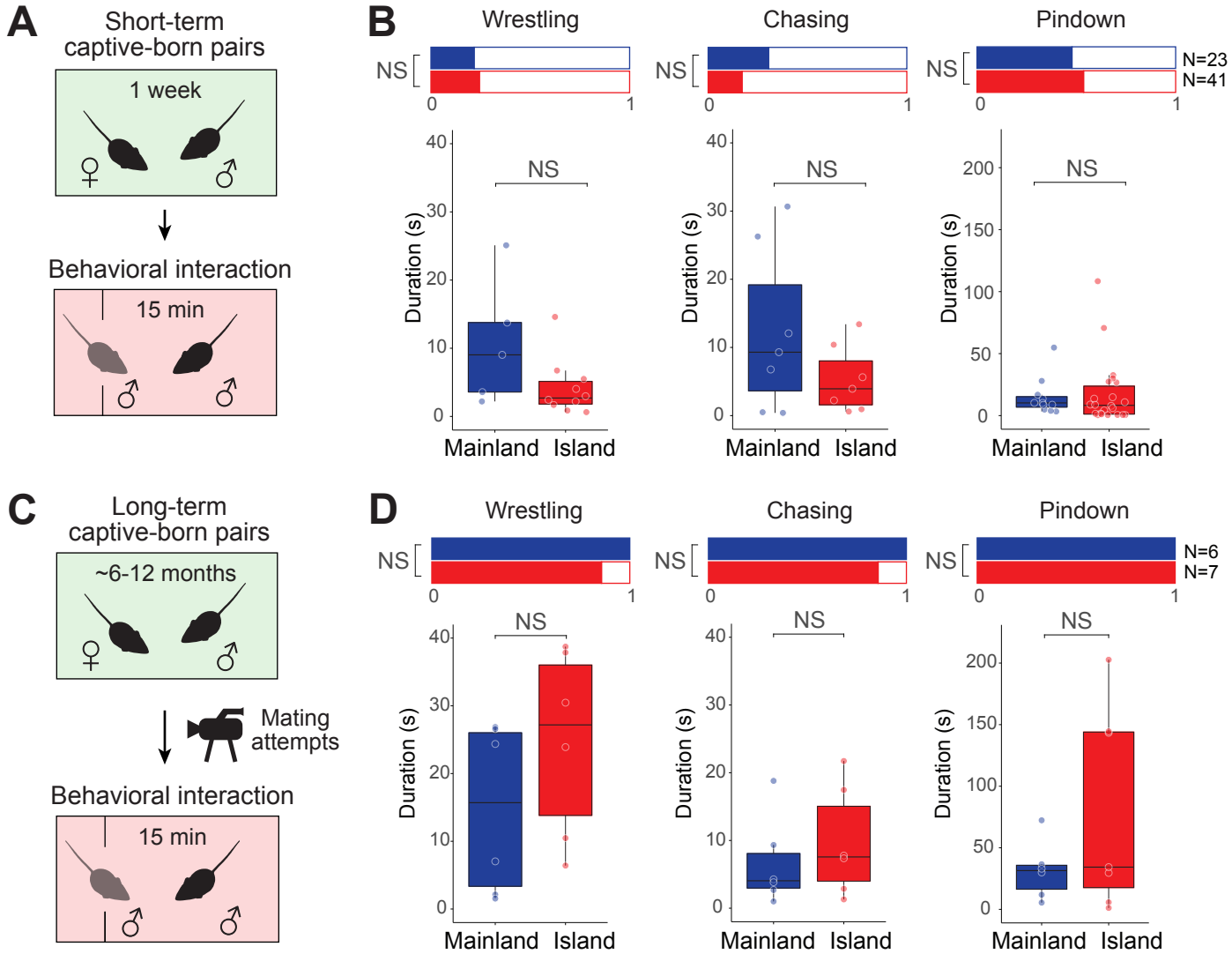


Figure 5. Territorial aggression in captive-born island and mainland mice.
(**A**) Schematic of the behavioral assay, a standard paradigm to induce aggression in laboratory mice, using captive-born males paired with a female for one week. For each resident, we ran the assay twice separated by a day using two different conspecific intruders. (**B**) Aggressive behavior in short-term pairings. Bar graphs (top) show the proportions of animals engaging in wrestling, chasing, and pindown behaviors. Boxplots (bottom) show duration of each behavior for those individuals that engaged in that behavior. Statistical significance of proportions and durations evaluated by generalized linear models and linear fixed effects models, respectively (see Methods for details). (**C**) Schematic of the behavioral assay using captive-born males paired with a female for 6-12 months, and then testing males in the resident-intruder paradigm approximately one hour after the onset of mating attempts and again one day later. (**D**) Aggressive behaviors in long-term pairings. Bar graphs (top) show the proportions of animals engaging in wrestling, chasing, and pindown behaviors. Boxplots (bottom) show duration of each behavior for those individuals that engaged in that behavior. Statistical significance of proportions and durations evaluated by generalized linear models and linear fixed effects models, respectively (see Methods for details). Sample sizes are provided. NS=not significant.

first-generation captive-born breeding animals that had been paired for 6-12 months and took advantage of the fact that females undergo post-partum estrus (Fig. 5C). We identified pregnant females in our colony and continuously video-monitored their behavior in their home cage. We found that male mice will begin to attempt copulation about 12 hours after the birth of a litter for about 1-2 hours (Suppl. Fig. 11). Therefore, we tested male mice in our behavioral assay approximately one hour after the onset of copulation and again one day later.

As expected, aggression levels were very high in reproductively active mice (Fig. 5D, Suppl. Fig. 8C). Nearly all residents of both populations exhibited wrestling, chasing, and pindown behaviors. Surprisingly, however, there was no significant difference in the proportion of island and mainland mice exhibiting wrestling (100% vs. 85.71%, logistic regression model, z = -0.003, $P = 0.998$), chasing (100% vs. 85.71%, logistic regression model, z = -0.003, $P = 0.998$), or pindown behavior (100% vs. 100%, logistic regression model, z = 0, $P = 1$), nor was there a difference in the duration of wrestling (linear fixed effects model, t = 1.309, $P = 0.2199$), chasing (linear fixed effects model, t = 0.725, $P = 0.4851$), or pindowns (linear fixed effects model, t = 1.409, $P = 0.186$). Thus, although we were able to induce higher aggression levels in captive-born mice, we still found no behavioral differences between captive-born island and mainland mice, implying that the aggression differences observed in wild-caught mice are driven, at least in large part, by environmental effects.

DISCUSSION

Here, we investigated the genetic basis of morphological and behavioral traits associated with the island syndrome in deer mice from British Columbia. Previous work in the 1970s reported that wild-caught Saturna Island deer mice had higher body weight (REDFIELD 1976; SULLIVAN 1977) and reduced aggression levels compared to mainland mice (HALPIN AND SULLIVAN 1978; HALPIN 1981). We recapitulated these differences in both body weight and aggression in our populations of wild-caught deer mice. First, we found that body weight is indeed higher in Saturna Island (as well as most other islands in the region) than mainland mice, and further showed that this difference is driven by both

the increased body length and disproportionately larger lean mass of Saturna Island mice. We also found striking differences in aggression: island mice showed reduced aggression in several behaviors compared to mainland mice. Thus, Saturna Island mice indeed conform to the predicted island syndrome phenotype. However, because measurements were taken using wild-caught individuals, it was unclear if these morphological and behavioral differences had a genetic basis or instead were plastic responses to environmental extremes (HALPIN AND SULLIVAN 1978; HALPIN 1981).

We showed that island mice diverged from their mainland ancestors approximately 10 thousand years ago and subsequently remained largely isolated, thus being subject to prime conditions for the evolution of heritable adaptive traits. Using established laboratory colonies, we found, indeed, that body size differences persisted in captive-born mice raised in common conditions, were thus no longer environmentally sensitive, and instead appeared largely heritable. Although examples of density-dependent plasticity in body weight, including maternal effects, are known from natural populations (e.g. DANTZER *et al.* 2013), our results are consistent with other studies that have demonstrated the genetic basis of gigantism and skeletal evolution in island rodents (ROTH AND KLEIN 1986; GRAY *et al.* 2015; PARMENTER *et al.* 2016). Our data on island-mainland hybrid growth showed that the larger body size in island mice was also partly driven by a maternal genetic effect. The maternal effect was not recapitulated by cross-fostering, suggesting that it cannot be explained simply by post-natal nursing differences. A study in island and mainland deer mice from California showed that four day old mainland embryos that were transferred into island foster mothers had a significantly increased neonatal weight (ROTH AND KLEIN 1986), arguing for increased maternal uterine investment in island deer mice. We cannot exclude the possibility of maternal genomic imprinting (e.g. CHEVERUD *et al.* 2008) without an embryo transfer experiment. Regardless, it is clear that both offspring and maternal genotypes driving larger body size have evolved in island deer mice.

Unlike body size differences, however, our behavioral experiments demonstrated that aggression levels in captive-born island and mainland deer mice remained indistinguishable across a spectrum of low to high overall aggression intensity, suggesting that additional environmental factors are necessary to express the aggression

differences reported previously (HALPIN AND SULLIVAN 1978; HALPIN 1981) and replicated in our study of wild-caught mice. Similar to our findings, a comparative study of both captive and wild insular macropodid marsupials suggested that loss of some anti-predator behaviors may be heritable, while loss of others is more phenotypically plastic (BLUMSTEIN AND DANIEL 2005). By contrast, fixed differences between island-mainland deer mice populations were found in other behaviors, for example, climbing behavior (PULSIFER AND HERMAN 1989).

The evolution of these island syndrome traits can be framed in the context of the phenotypic plasticity (e.g. PRICE *et al.* 2003; WEST-EBERHARD 2003; CRISPO 2007; LEVIS AND PFENNIG 2016). The larger body size in island mice was retained in captivity and thus may be the result of genetic assimilation, a process of genetic and developmental canalization of an adaptive phenotype in a new environment until all plasticity is lost. By contrast, territorial aggression in island rodents could be considered a behavioral polyphenism if aggression levels are indeed driven entirely by environmental parameters such as population density. Alternatively, the underlying process could be considered a genetic accommodation if island mice were better genetically equipped than mainland populations to adjust territorial aggression to environmental conditions on islands. For example, anecdotal observations from our colonies suggest that island mice are overall less skittish than mainland mice, raising the possibility that a more benign temperament could facilitate the downregulation of aggression.

Consistent with our observations in *Peromyscus maniculatus*, high population density, increased body size and reduced aggression have been documented in two other rodents, *Microtus breweri* (TAMARIN 1978; REICH *et al.* 1982) and *Myodes glareolus skomerensis* (JOHNSON 1975). Several non-exclusive hypotheses have been proposed to explain the evolution of larger body size and reduced territorial aggression in island rodents like these (GLIWICZ 1980; STAMPS AND BUECHNER 1985; ADLER AND LEVINS 1994). Central to these is the observation that population density in island rodents is typically increased, likely because of the absence of emigration, fewer competing species and thus more available resources, and fewer predators on islands. The rate of social interactions, and thereby also the cost of territorial defense, likely increase with density. In addition, reproductive output in rodents is often adjusted to population density. Rodents in high

density island populations could thus face pressure to reduce direct conflict in favor of indirect competition (territorial defense hypothesis), and to reallocate resources to produce fewer, but more competitive offspring (reallocation hypothesis). Both paradigms would likely select for offspring and parental genotypes driving larger body size.

Our results are consistent with the notion that variation in morphological traits is more heritable and less sensitive to environmental conditions than behavioral traits, in part perhaps because they are less closely connected with fitness (STIRLING *et al.* 2002). We speculate that plasticity in territorial behavior may have been critical to sustaining populations in the face of strong selection on islands, and that this behavioral plasticity thus contributed to the selective environment in which parental and offspring genotypes driving larger offspring body size became advantageous on islands (WCISLO 1989; WEST-EBERHARD 1989). It is not obvious why reduced territorial aggression did not become genetically assimilated over evolutionary time, perhaps 10 thousand years (or ~25 thousand generations) is not enough time. However, it remains to be seen if other examples of behavioral island syndrome traits show a similar extent of plasticity, and if this plasticity is specific to certain behaviors, such as population density-dependent traits.

While the island syndrome has been reported in a wide range of organisms, few studies have tested if these traits are genetically encoded. More research is needed into the drivers, sequence, and mechanisms of trait evolution during the colonization of islands and subsequent establishment of populations to fully understand one of the most iconic and widespread patterns of repeated evolution in nature.

ACKNOWLEDGEMENTS

We thank J. Chupasko for help planning field work and permits, N. Bedford, E. Hager, and K. Turner for assistance in the field and J. Kenagy for help with field logistics. We also thank H. Wellington for generating the X-ray images and taking skeletal trait measurements; A. Shultz for advice on the population-genetic analysis; and A. Nelson for advice on the behavioral assay design. K. Pritchett-Corning and the Office of Animal Resources at Harvard University were instrumental in setting up breeding colonies from wild-caught mice. Data science specialists S. Worthington and S. Goshev at the Institute

for Quantitative Social Science, Harvard University, provided statistical support. Fieldwork was conducted with permits from the Gulf Island National Park Reserve (GINP-2014-17276), the Pender Island Parks and Recreation Commission, the Malcolm Knapp Research Forest at the University of British Columbia, and the Ministry of Forests, Lands and Natural Resource Operations (NASU14-154228, NASU14-154230). Fieldwork was supported by a Putnam Expedition Grant of the Harvard Museum of Comparative Zoology. FB was supported by a HHMI International Student Research Fellowship, a Joan Brockman Williamson Graduate Research Fellowship, and a Herchel Smith Graduate Fellowship. HEH is an Investigator of the Howard Hughes Medical Institute.

METHODS

Body weight of museum specimens. We retrieved records (species, sex, age, body weight, body length and tail length) from museum specimens of *Peromyscus maniculatus* collected in Washington and British Columbia using the Arctos and Vertnet databases. To minimize confounding effects of reproductive status and age, we excluded females and specimens of unknown sex, juveniles, subadults, and embryos. We then removed specimens with a body length ≤ 70 mm and/or a tail length ≤ 60 mm, which represents minimum values in a sample of adult *P. maniculatus* we collected in British Columbia (see below). To remove specimens with outdated taxonomy that no longer are classified as *P. maniculatus*, notably *P. keeni* (HOGAN *et al.* 1993), we filtered out records labelled as *P. keeni* or any of its synonyms or subspecies. We also excluded specimens with both a tail/body ratio of ≥ 1.1 and a tail length ≥ 95mm, which are morphological criteria to identify *P. keeni* (ZHENG *et al.* 2003). Our final sample comprised 504 mice, including 53 from our own collection (see 'Fieldwork').

Fieldwork. We sampled deer mice (N = 120) in the Gulf Island National Park Reserve on Saturna Island and Pender Island (*Peromyscus maniculatus saturatus*) and in the Malcolm Knapp Research Forest in Maple Ridge (*P. m. austerus* and *P. keeni*) in October 2014. Of these, we collected 101 individuals and accessioned them into the Mammal Department of the Museum of Comparative Zoology at Harvard University as voucher specimens (accession numbers: XXX-XXX).

Species assignment. Two species of deer mice co-occur at our mainland sampling location, *Peromyscus maniculatus* and *P. keeni.* To verify the species identity of our samples, we sequenced a partial fragment (381bp) of the *cytochrome b* gene from all field-collected individuals using previously reported primers and PCR conditions (ZHENG *et al.* 2003). *P. keeni* have distinct *cytb* haplotypes from *P. maniculatus* (Suppl. Fig. 12), and these individuals were excluded from subsequent analyses.

Admixture analysis and divergence dating. We extracted genomic DNA from 65 wild-caught mice (Saturna Island, N=28; Pender Island, N=9; mainland, N=28) using an AutoGenPrep 965 (AutoGen). The quality of extractions was verified with Quant-iT (Thermo Fisher Scientific). We used a genotype-by-sequencing approach (ddRADseq; (PETERSON *et al.* 2012) to obtain genome-wide markers for analyses of genetic admixture and divergence time dating. In brief, we digested genomic DNA samples with NlaIII and MluCI enzymes (New England Biolabs) and ligated fragments to biotinylated barcoded adapters. We size-selected for 216-276bp fragments using a PippinPrep (Sage Science), cleaned fragments with Streptavidin-coupled M-280 Dynabeads (Thermo Fisher Scientific), and PCR amplified fragments using 12 uniquely indexed PCR primers with Phusion DNA polymerase (New England Biolabs). After a bead cleanup and evaluation of the library quality on a 2200 TapeStation (Agilent Technologies), we generated 125bp paired-end reads with a standard v4 run on an Illumina HiSeq 2500 sequencer.

We used a custom-made pipeline to demultiplex reads and combine them by sample into one R1 and R2 file. We then used BWA 0.7.15 (LI AND DURBIN 2009) to map reads to an in-house *de novo* assembly (v2.1) of the *P. maniculatus bairdii* (strain BW) reference genome. After calling variants for each sample separately with GATK 3.5 (MCKENNA *et al.* 2010) using "HaplotypeCaller" (including options -minPruning 1, -minDanglingBranchLength 1, and -het 0.005), we produced one vcf file containing all samples with GATK's "GenotypeGVCFs" (including option het -0.005). After quality filtering (QualByDepth (QD) < 2; FisherStrand (FS) > 60 for SNPs, > 200 for indels; RMSMappingQuality (MQ) < 40; MappingQualityRankSumTest (MQRankSum) < 12.5; ReadPosRankSumTest (ReadPosRankSumTest) < -8 for SNPs, < -20 for indels; StrandOddsRatio (SOR) > 3 for SNPs, > 10 for indels), we used vcftools (DANECEK *et al.* 2011) to explore depth and missingness. We then removed samples of *P. keeni*, and filtered sites by minimum and maximum depth across individuals (-min-meanDP 1 and -max-meanDP 264).

To prepare the input file for the principal component analysis (PCA) and admixture analysis, we next filtered by minimum and maximum depth within individuals (-minDP 4 and -maxDP 25) and by missingness across individuals (-max-missing 0.6), and then removed samples with <25% of these sites. To account for the genotype uncertainty

inherent in low-coverage data, we generated genotype probabilities using ANGSD 0.911 (KORNELIUSSEN *et al.* 2014) with the -doPost 1 and -doGeno 32 options. This dataset contained 80,248 variants across 54 *P. maniculatus* samples. For the PCA, we used the ngsCovar function in ngsTools 0.615 (FUMAGALLI *et al.* 2014) to compute a covariance matrix. For the admixture analysis, we generated a beagle file with the -doGlf function in ANGSD, and used the ngsAdmix function in ngsTools to calculate admixture proportions. We plotted PCs and admixture proportions following this tutorial: https://github.com/mfumagalli/ngsTools/blob/master/TUTORIAL.md.

To prepare the input file for the divergence dating analysis, we selected a subset of high coverage samples (3 Saturna Island *P. maniculatus*, 5 Pender Island *P. maniculatus*, 3 mainland *P. maniculatus*, 3 *P. keeni*), and filtered by minimum and maximum depth within individuals (-minDP 5 and -maxDP 25) and missingness across individuals (-max-missing 0.8). This resulted in a dataset of 7,092 variants across 14 samples. We then converted the vcf file to nexus format with custom Python code, and used SNAPP 1.3.0 (BRYANT *et al.* 2012) to create a phylogenetic tree with divergence time estimates. Briefly, we loaded the nexus file into a SNAPP template in BEAUti 2.4.8 (BOUCKAERT *et al.* 2014), assigned species IDs to samples, calculated mutation rates, logged every 500 trees, and used a chain length of 10^6. We ran 10 separate iterations of the resulting XML file in BEAST 2.4.8 (BOUCKAERT *et al.* 2014), and combined log and tree files across iterations in LogCombiner with a by-iteration burn-in of 10%. We checked for convergence and high ESS values in Tracer 1.6 (RAMBAUT *et al.* 2018), and used TreeAnnotator to calculate a maximum clade credibility tree with median heights. We produced the final tree using FigTree 1.4.2.

Establishment of laboratory colonies and animal husbandry. After quarantine of the wild-caught mice at Charles River Laboratories, we established colonies at Harvard University with 18 individuals from Saturna Island and 11 individuals from the Malcolm Knapp Research Forest. We housed animals on Bed-o'Cobs 1/4" bedding (The Andersons, Maumee, Ohio) in ventilated standard rodent cages (Allentown Inc., Allentown, NJ) on a 16h light: 8h dark cycle at 23°C. Animals were provided with a red translucent polycarbonate hut, Enviro-Dri nesting material, and a cotton nestlet. Animals

were given *ad libitum* access to irradiated LabDiet Prolab Isopro RMH 3000 5P74 (LabDiet) and water. We weaned litters at 23 days of age and kept animals in groups of up to five individuals of the same sex and strain. Subsequent experiments were conducted using either wild-caught mice or first-, second- and third-generation captive-born mice as noted. All experiments were approved by the Harvard University Faculty of Arts and Sciences Institutional Animal Care and Use Committee (protocol #14-08-211).

Weight measurements. Mice in the field were weighed using a Pesola spring scale to the nearest gram. Mice in the lab were weighed with a Jennings TB500 digital precision scale to the nearest 0.01g. To weigh newborn litters before their first milk meal, we isolated and monitored highly pregnant females through remote surveillance of the cage with multiple cameras. We thus were able to intercept females within 20-60 minutes of birth, and verified that pups had not yet had a meal by checking for the presence of milk in the stomach (Suppl. Fig. 5B).

EchoMRI. EchoMRI experiments were conducted in the Metabolic Core of Brigham and Women's Hospital Boston. Frozen mice were thawed to room temperature and measured individually in a calibrated EchoMRI 3-in-1 machine.

X-ray measurements. To measure skeletal traits, we X-ray imaged mice on a digital X-ray system in the Digital Imaging Facility of the Museum of Comparative Zoology. We used ImageJ to measure 10 skeletal elements: length of the sacrum, skull, zygomatic bone, humerus, femur, ulna, tibia, and metatarsal calcaneus as well as the width of the skull and zygomatic bone.

F1 hybrids. We reciprocally paired one female and one male litter mate each from five island and mainland parents. Of these 10 pairs, four island female x mainland male and three mainland female x island male pairs successfully reproduced.

Cross-fostering. Within 48 hours of birth, we exchanged pups between island and mainland parents. We fostered the same number of pups from each litter because we did

not want to dilute nursing effects of island mothers, which have slightly smaller litter sizes (Suppl. Fig. 4), on offspring body weight.

Behavioral experiments. We measured territorial aggression with the resident-intruder paradigm. To induce territoriality, residents were first paired with a female. For one set of experiments (i.e. Fig. 5A-B), 9 to 13 week-old residents were isolated with a female for one week. For a second set of experiments (i.e. Fig. 4 and Fig. 5C-D), residents were breeding males and had been paired with a female for several months. Intruders were raised in social groups with at least two other mice, and were 9-13 weeks old. Resident and intruder had different parents and had not previously met. We conducted all behavioral experiments during the dark phase under red light. About 3-4 hours before the trial, the female was removed from the resident's cage, resident and intruder were weighed, and the intruder was marked for easy identification with permanent marker on the tail.

At the start of the experiment, we placed the resident's cage into black blinders, removed the lid and the red hut, and placed a custom-built divider with a gate and a transparent lid onto the cage. The divider set apart a small portion of the cage for the intruder, but was perforated to enable contact. The resident was habituated to the divider for 10 mins. Then we introduced the intruder into the area separated by the divider, and both mice were habituated for another 10 mins. Next, we opened the gate and the intruder was allowed to enter the resident's area. We video-recorded behavior from above the cage at 15 fps; trials lasted for 15 mins. One day later, we tested the resident a second time against a different intruder.

To test the aggressive behavior of males exposed to a receptive female, we used breeding males from our colony. Once a female was pregnant, we set up continuous video-monitoring of the cage, and once a litter was found, noted the approximate time of birth. Females typically reached post-partum estrus approximately 6-14 hours after the birth of the litter, as evidenced by the onset of male mating behavior (e.g. by vigorous pursuit [DEWSBURY 1979]). Males were tested approximately one hour after the onset of mating attempts and after several mating attempts had occurred.

Behavioral analysis. We annotated videos using the software Observer (Noldus). We scored aggressive (lunging, pindown, upright threats, boxing, wrestling, and chasing), defensive (submission, defensive threat, and defeat), and cohesive behaviors (following, mounting, grooming, and huddling) as start-stop events performed by both the resident and the intruder.

Statistical analysis. We tested for differences in weight, body composition, and skeletal traits with linear fixed effects models with a strain (island/mainland) by origin (field/lab) interaction. We tested for differences in birth weight of parental strains (or F1 hybrids) using linear fixed effects models with the natural logarithm of weight as the dependent variable and the natural logarithm of litter size and strain (or maternal strain, respectively) as explanatory variables. We tested for differences in growth rate between parental strains (or F1 hybrids) using linear fixed effects models with the natural logarithm of growth rate as the dependent variable and the natural logarithm of litter size and an interaction of strain (or maternal strain, respectively) and day as explanatory variables, and adjusted p-values for multiple testing with the Holm method. Growth rates were obtained by taking the first derivative of cubic smoothing splines fit to the weight data using smoothPspline {pspline}. Our sample size for the F1 hybrids was moderate, therefore we ran additional tests that controlled for family effects. We did not include parent ID (as a proxy for family effects) into the model to avoid model estimation problems introduced by the perfect collinearity of parent ID and strain. Instead, we accounted for family effects by adjusting the standard errors of the model using wild cluster bootstrapping as implemented in cluster.wild.glm {clusterSEs}. To compare litter weight before and after the first milk meal, we used a repeated measures linear mixed effects model with the natural logarithm of weight as dependent variable, the natural logarithm of litter size and an interaction of feeding status (pre- vs. post-feeding) and strain as explanatory variables and litter ID as random effect. We tested for differences in weight by maternal strain in adult hybrid mice with a linear fixed effects model with the natural logarithm of weight as the dependent variable and strain as the explanatory variable. In the cross-fostering experiment, we tested for differences in the natural logarithm of growth rate of conspecific non-fostered and fostered litters with a linear

fixed effects model, with the natural logarithm of litter size and an interaction of state (fostered vs. non-fostered) and day as explanatory variables, and adjusted p-values for multiple testing with the Holm method.

To reduce the dimensionality in the behavior of the wild-caught resident animals, we performed a PCA on behaviors averaged across trials using prcomp {stats}, with variables scaled to unit variance and shifted to be zero centered. We next selected variables that contributed most to the first PC and focused on these behaviors for subsequent analyses. Specifically, we compared levels in these behaviors in wild-caught residents, captive-born residents one week after they were paired with a female, and breeding pairs after the onset of mating attempts. For this, we used a two-model hurdle approach to test for differences between strains. This included a binary model to test whether different proportions of individuals express a behavior, and count and duration models to test whether differences exist in the number of bouts and duration once animals express a given behavior. We initially evaluated the impact of trial number and weight difference between resident and intruder on aggression by comparing repeated measures linear mixed effects models with and without these factors using likelihood ratio tests. Because these factors did not improve models for most behaviors, we averaged data across trials (average number of bouts were rounded) and ran logistic regression models using glm {stats} with binomial (binary) and quasipoisson (count) error distributions and linear fixed effects models (duration). To compare wrestling duration before and after reproduction, we fit a repeated measures mixed effects linear model on the averaged dataset with reproductive status and strain as explanatory variables and resident ID as random effect. To evaluate other potential impacts on the summed duration of wrestling, chasing, and pindown in wild-caught mice, we fit a repeated measures mixed effects linear model with days since capture, days since last litter was sired, weight difference, and strain as explanatory variables and resident ID as random effect.

All linear models were fit with lm {stats} or lmer {lme4}. *P* values were calculated using lsmeans and contrast {lsmeans} or summary {lmerTest}.

LIST OF REFERENCES

Adler, G. H., and R. Levins, 1994 The island syndrome in rodent populations. The Quarterly Review of Biology 69**:** 473-490.

Aubret, F., and R. Shine, 2009 Genetic assimilation and the postcolonization erosion of phenotypic plasticity in island tiger snakes. Current Biology 19**:** 1932-1936.

Berry, R. J., 1996 Small mammal differentiation on islands. Philosophical Transactions: Biological Sciences 351**:** 753-764.

Blanchard, D. C., and R. J. Blanchard, 1988 Ethoexperimental approaches to the biology of emotion. Annual Review of Psychology 39**:** 43-68.

Blondel, J., 2000 Evolution and ecology of birds on islands: Trends and prospects. Vie et Milieu 50**:** 205-220.

Blumstein, D. T., and J. C. Daniel, 2005 The loss of anti-predator behaviour following isolation on islands. Proceedings of the Royal Society B: Biological Sciences 272**:** 1663-1668.

Bouckaert, R., J. Heled, D. Kühnert, T. Vaughan, C.-H. Wu *et al.*, 2014 BEAST 2: A software platform for Bayesian evolutionary analysis. PLoS Computational Biology 10**:** e1003537.

Bryant, D., R. Bouckaert, J. Felsenstein, N. A. Rosenberg and A. RoyChoudhury, 2012 Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Molecular Biology and Evolution 29**:** 1917-1932.

Cheverud, J. M., R. Hager, C. Roseman, G. Fawcett, B. Wang *et al.*, 2008 Genomic imprinting effects on adult body composition in mice. Proceedings of the National Academy of Sciences 105**:** 4253-4258.

Corl, A., K. Bi, C. Luke, A. S. Challa, A. J. Stern *et al.*, 2018 The genetic basis of adaptation following plastic changes in coloration in a novel environment. Current Biology 28**:** 2970-2977.

Crispo, E., 2007 The Baldwin effect and genetic assimilation: Revisiting two mechanisms of evolutionary change mediated by phenotypic plasticity. Evolution 61**:** 2469-2479.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks *et al.*, 2011 The variant call format and VCFtools. Bioinformatics 27**:** 2156-2158.

Dantzer, B., A. E. M. Newman, R. Boonstra, R. Palme, S. Boutin *et al.*, 2013 Density triggers maternal hormones that increase adaptive offspring growth in a wild mammal. Science 340**:** 1215-1217.

de Catanzaro, D., 1981 Facilitation of intermale aggression in mice through exposure to receptive females. Journal of Comparative and Physiological Psychology 95**:** 638-645.

Dewsbury, D. A., 1979 Copulatory behavior of deer mice (*Peromyscus maniculatus*): I. Normative data, subspecific differences, and effects of cross-fostering. Journal of Comparative and Physiological Psychology 93**:** 151-160.

Dewsbury, D. A., 1981 Social dominance, copulatory behavior, and differential reproduction in deer mice (*Peromyscus maniculatus*). Journal of Comparative and Physiological Psychology 95**:** 880-895.

Dewsbury, D. A., 1984 Aggression, copulation, and differential reproduction of deer mice (*Peromyscus maniculatus*) in a semi-natural enclosure. Behaviour 91**:** 1-23.

Eisenberg, J. F., 1962 Studies on the behavior of *Peromyscus maniculatus gambelii* and *Peromyscus californicus parasiticus*. Behaviour 19**:** 177-207.

Eisenberg, J. F., 1963 The intraspecific social behavior of some cricetine rodents of the genus *Peromyscus*. American Midland Naturalist 69**:** 240-246.

Foster, J. B., 1964 Evolution of mammals on islands. Nature 202**:** 234-235.

Fumagalli, M., F. G. Vieira, T. Linderoth and R. Nielsen, 2014 ngsTools: methods for population genetics analyses from next-generation sequencing data. Bioinformatics 30**:** 1486-1487.

Gliwicz, J., 1980 Island populations of rodents: Their organization and functioning. Biological Reviews 55**:** 109-138.

Gray, M. M., M. D. Parmenter, C. A. Hogan, I. Ford, R. J. Cuthbert *et al.*, 2015 Genetics of rapid and extreme size evolution in island mice. Genetics 201**:** 213-228.

Gray, S. J., and J. L. Hurst, 1998 Competitive behaviour in an island population of house mice, *Mus domesticus*. Animal Behaviour 56**:** 1291-1299.

Halpin, Z., 1981 Adult-young interactions in island and mainland populations of the deermouse *Peromyscus maniculatus*. Oecologia 51**:** 419-425.

Halpin, Z. T., and T. P. Sullivan, 1978 Social interactions in island and mainland populations of the deer mouse, *Peromyscus maniculatus*. Journal of Mammalogy 59**:** 395-401.

Hogan, K. M., M. C. Hedin, H. S. Koh, S. K. Davis and I. F. Greenbaum, 1993 Systematic and taxonomic implications of karyotypic, electrophoretic, and mitochondrial-DNA variation in *Peromyscus* from the Pacific Northwest. Journal of Mammalogy 74**:** 819-831.

Johnson, R. P., 1975 Scent marking with urine in two races of the bank vole (*Clethrionomys glareolus*). Behaviour 55**:** 81-93.

Korneliussen, T. S., A. Albrechtsen and R. Nielsen, 2014 ANGSD: Analysis of next generation sequencing data. BMC Bioinformatics 15**:** 356.

Levis, N. A., and D. W. Pfennig, 2016 Evaluating 'plasticity-first' evolution in nature: Key criteria and empirical approaches. Trends in Ecology & Evolution 31**:** 563-574.

Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25**:** 1754-1760.

Lomolino, M. V., 1985 Body size of mammals on islands: The island rule reexamined. The American Naturalist 125**:** 310-316.

Losos, J. B., D. A. Creer, D. Glossip, R. Goellner, A. Hampton *et al.*, 2000 Evolutionary implications of phenotypic plasticity in the hindlimb of the lizard *Anolis sagrei*. Evolution 54**:** 301-305.

Losos, J. B., and R. E. Ricklefs, 2009 Adaptation and diversification on islands. Nature 457**:** 830-836.

MacArthur, R. H., and E. O. Wilson, 1967 *The Theory of Island Biogeography*. Princeton University Press, Princeton.

Mathews, W. H., J. G. Fyles and H. W. Nasmith, 1970 Postglacial crustal movements in southwestern British Columbia and adjacent Washington state. Canadian Journal of Earth Sciences 7**:** 690-702.

Mayr, E., 1954 Change of genetic environment and evolution, pp. 157-180 in *Evolution as a Process*, edited by J. Huxley, A. C. Hardy and E. B. Ford. Allen and Unwin, London.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010 The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research 20**:** 1297-1303.

Novosolov, M., P. Raia and S. Meiri, 2013 The island syndrome in lizards. Global Ecology and Biogeography 22**:** 184-191.

Orrock, J. L., 2010 When the ghost of predation has passed: Do rodents from islands with and without fox predators exhibit aversion to fox cues? Ethology 116**:** 338-345.

Parmenter, M. D., M. M. Gray, C. A. Hogan, I. N. Ford, K. W. Broman *et al.*, 2016 Genetics of skeletal evolution in unusually large mice from Gough Island. Genetics 204**:** 1559-1572.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher and H. E. Hoekstra, 2012 Double Digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS ONE 7**:** e37135.

Price, T. D., A. Qvarnström and D. E. Irwin, 2003 The role of phenotypic plasticity in driving genetic evolution. Proceedings of the Royal Society of London. Series B: Biological Sciences 270**:** 1433-1440.

Pulsifer, M. D., and T. B. Herman, 1989 Comparative arboreal behaviors of wild-caught and captive-born deer mice, *Peromyscus maniculatus* Wagner, from Isle Haute and mainland Nova Scotia. Canadian Journal of Zoology 67**:** 789-794.

Rambaut, A., A. J. Drummond, D. Xie, G. Baele and M. A. Suchard, 2018 Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic Biology 67**:** 901-904.

Redfield, J. A., 1976 Distribution, abundance, size, and genetic variation of *Peromyscus maniculatus* on the Gulf Islands of British Columbia. Canadian Journal of Zoology 54**:** 463-474.

Reich, L. M., K. M. Wood, B. E. Rothstein and R. H. Tamarin, 1982 Aggressive behaviour of male *Microtus breweri* and its demographic implications. Animal Behaviour 30**:** 117-122.

Roth, V. L., and M. S. Klein, 1986 Maternal effects on body size of large insular *Peromyscus maniculatus*: Evidence from embryo transfer experiments. Journal of Mammalogy 67**:** 37-45.

Schlichting, C. D., and M. Pigliucci, 1998 *Phenotypic Evolution: A Reaction Norm Perspective*. Sinauer Associates, Sunderland, Massachusetts.

Stamps, J. A., and M. Buechner, 1985 The territorial defense hypothesis and the ecology of insular vertebrates. The Quarterly Review of Biology 60**:** 155-181.

Stirling, D. G., D. Réale and D. A. Roff, 2002 Selection, structure and the heritability of behaviour. Journal of Evolutionary Biology 15**:** 277-289.

Sullivan, T. P., 1977 Demography and dispersal in island and mainland populations of the deer mouse, *Peromyscus maniculatus*. Ecology 58**:** 964-978.

Tamarin, R. H., 1978 Dispersal, population regulation, and K-selection in field mice. The American Naturalist 112**:** 545-555.

Van Valen, L. M., 1973 Pattern and the balance of nature. Evolutionary Theory 1**:** 31-49.

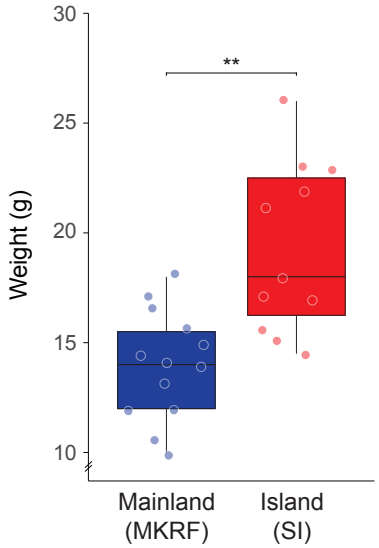
Waddington, C. H., 1953 Genetic assimilation of an acquired character. Evolution 7**:** 118-126.

Wcislo, W. T., 1989 Behavioral environments and evolutionary change. Annual Review of Ecology and Systematics 20**:** 137-169.

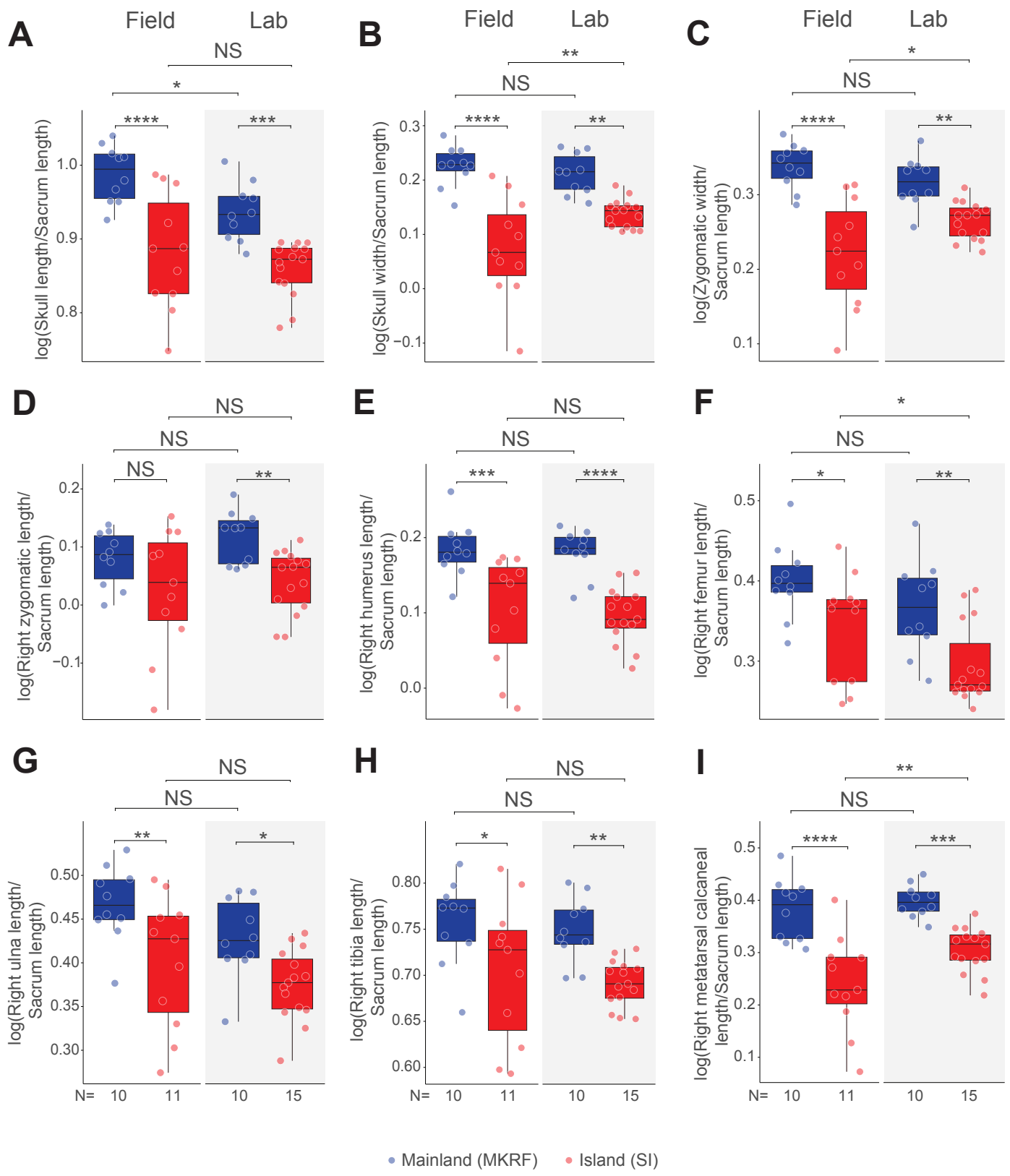
West-Eberhard, M. J., 1989 Phenotypic plasticity and the origins of diversity. Annual Review of Ecology and Systematics 20**:** 249-278.

West-Eberhard, M. J., 2003 *Developmental Plasticity and Evolution*. Oxford University Press, Oxford; New York.

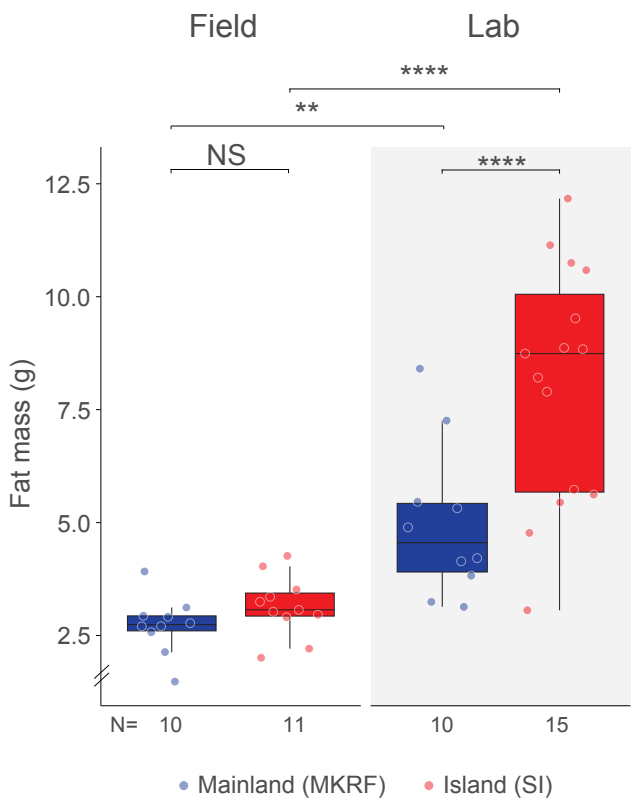
Zheng, X., B. S. Arbogast and G. J. Kenagy, 2003 Historical demography and genetic structure of sister species: deermice (*Peromyscus*) in the North American temperate rain forest. Molecular Ecology 12**:** 711-724.



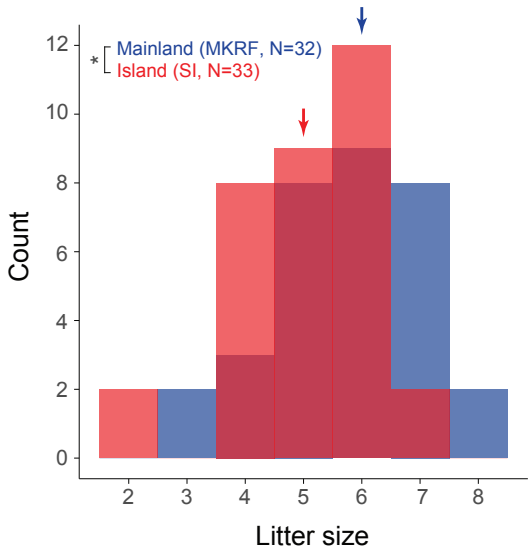
Suppl. Fig. 1. Body weight in adult female wild-caught island and mainland mice. Statistical significance evaluated by linear fixed effects model (see Methods for details). ** $P < 0.01$.



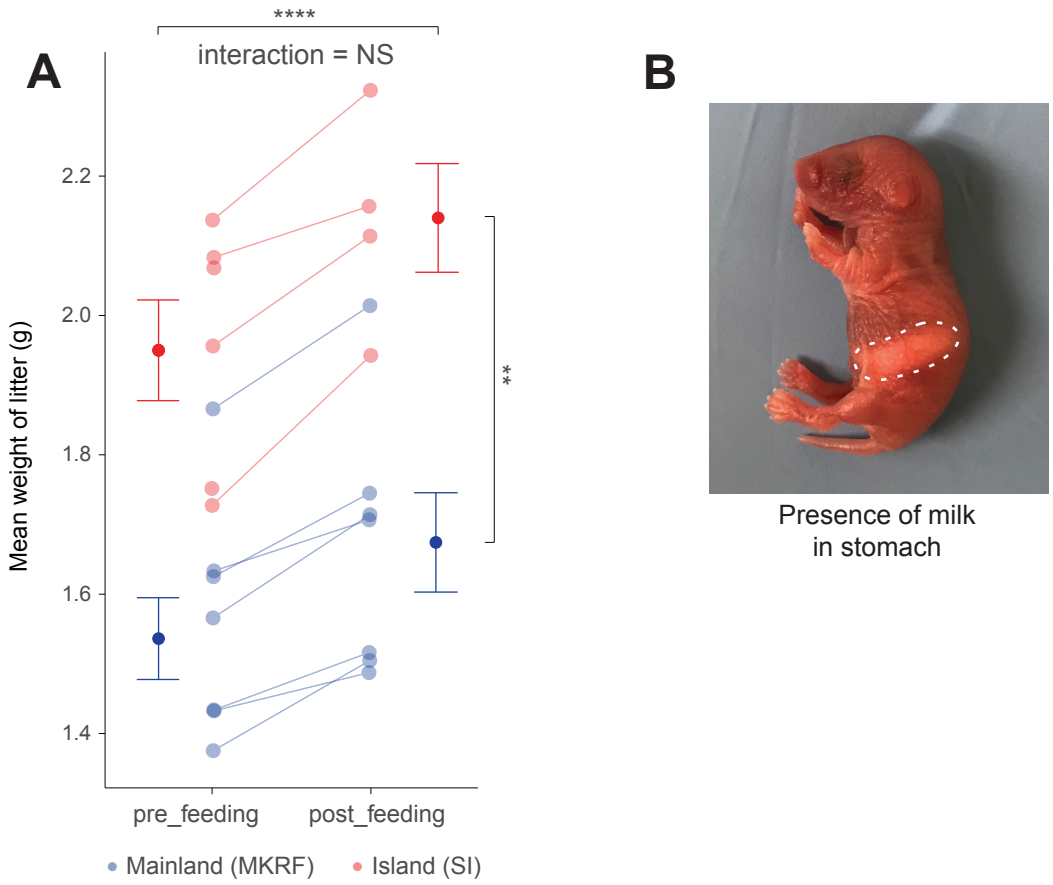
Suppl. Fig. 2. Size-corrected (**A**) skull length, (**B**) skull width, (**C**) zygomatic width, (**D**) right zygomatic length, (**E**) right humerus length, (**F**) right femur length, (**G**) right ulna length, (**H**) right tibia length, and (**I**) right metatarsal calcaneal length in male wild- and captive-born island and mainland mice. Statistical significance evaluated by linear fixed effects model (see Methods for details). NS=not significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.



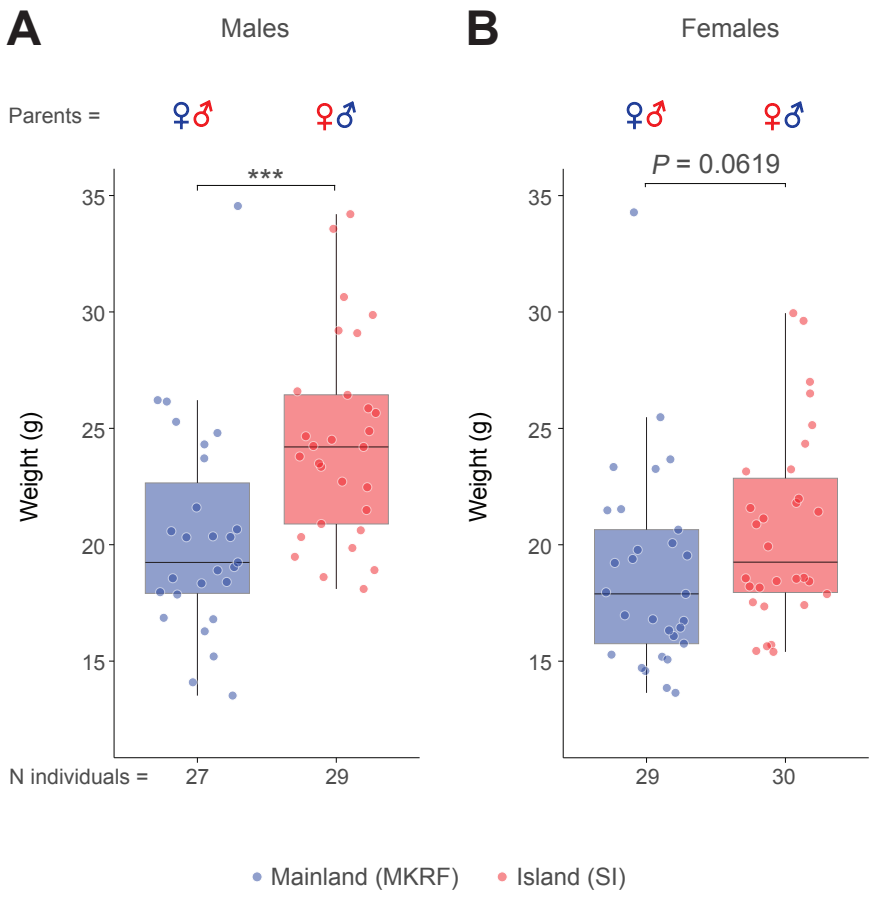
Suppl. Fig. 3. Fat mass in adult male wild- and captive-born island and mainland mice. Statistical significance evaluated by linear fixed effects model (see Methods for details). NS=not significant, ** $P < 0.01$, **** $P < 0.0001$.



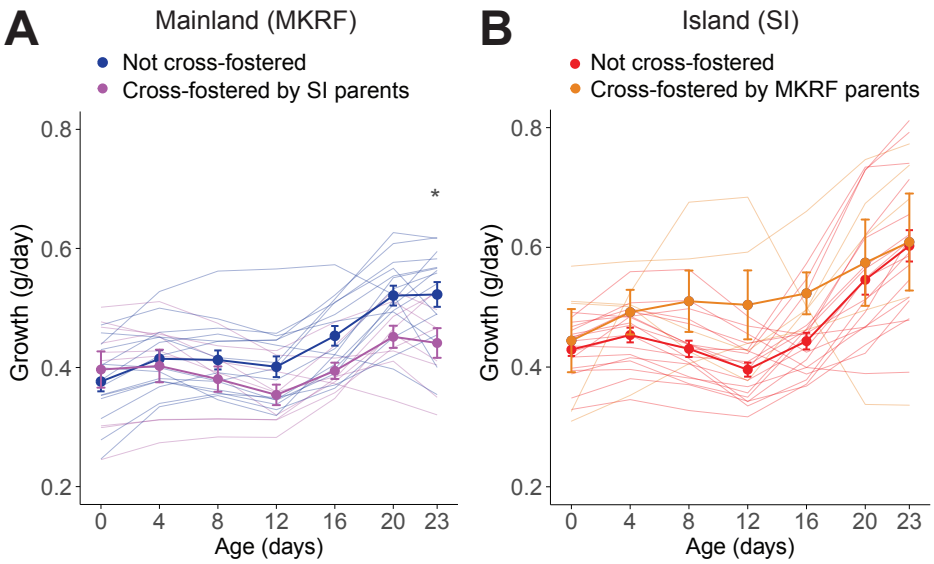
Suppl. Fig. 4. Litter size of island and mainland mice. Arrows indicate the median litter size for each strain. Statistical significance evaluated by Kruskal-Wallis test.



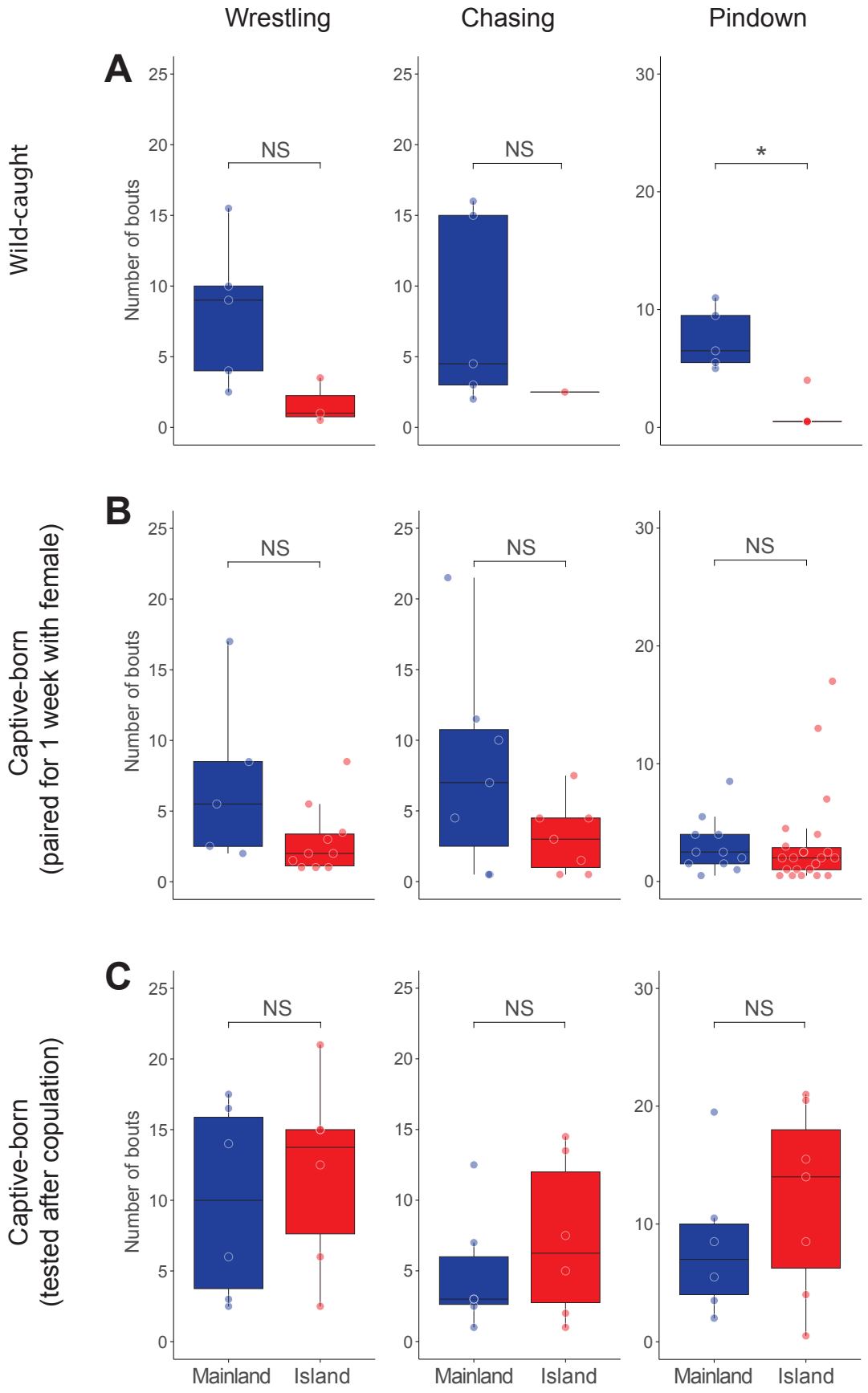
Suppl. Fig. 5. (**A**) Mean weight of litters before and after the first milk meal. Statistical significance evaluated by repeated measures linear mixed effects model (see Methods for details). NS=not significant, ** $P < 0.01$, **** $P < 0.0001$. (**B**) Example of pup with milk present in stomach (outlined).



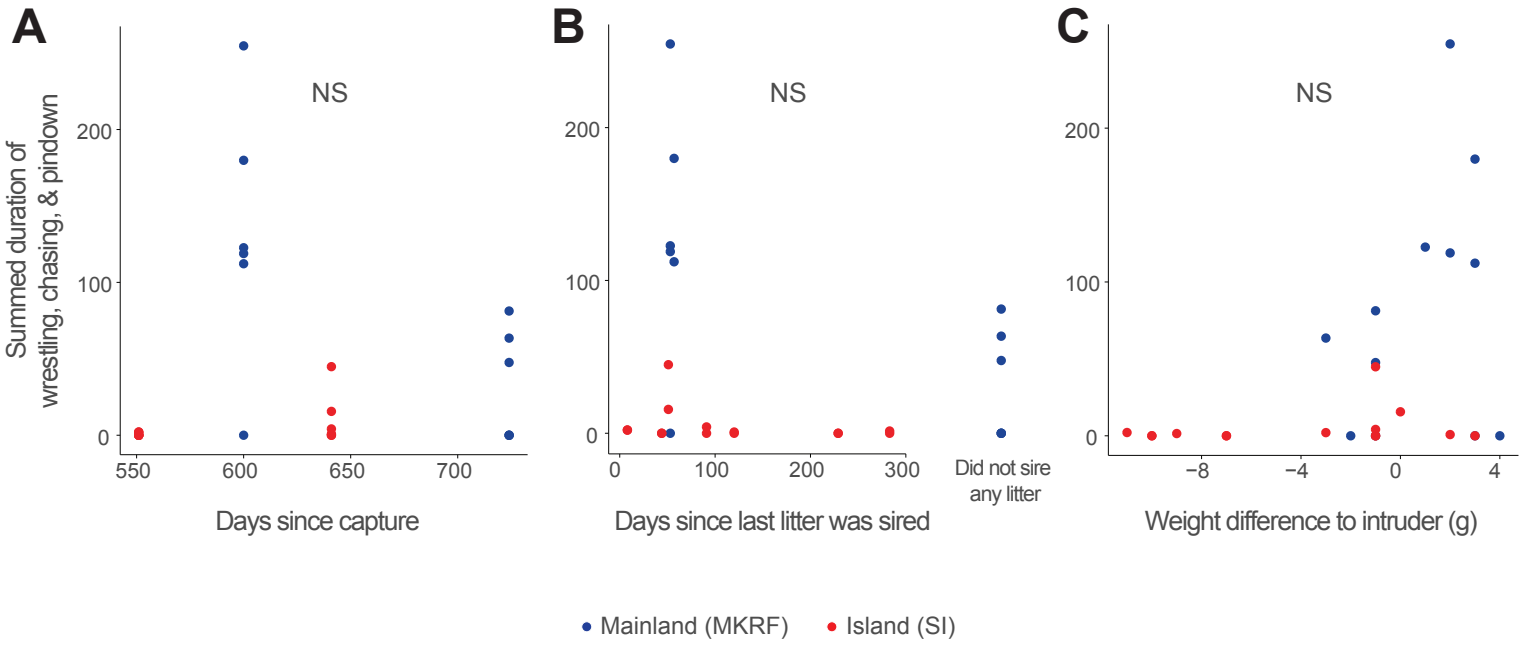
Suppl. Fig. 6. Weight of adult hybrid mice plotted by maternal strain identity. The natural logarithm of weights of (**A**) males and (**B**) females were separately compared by maternal strain with linear fixed effects models (see Methods for details). *** $P < 0.001$.



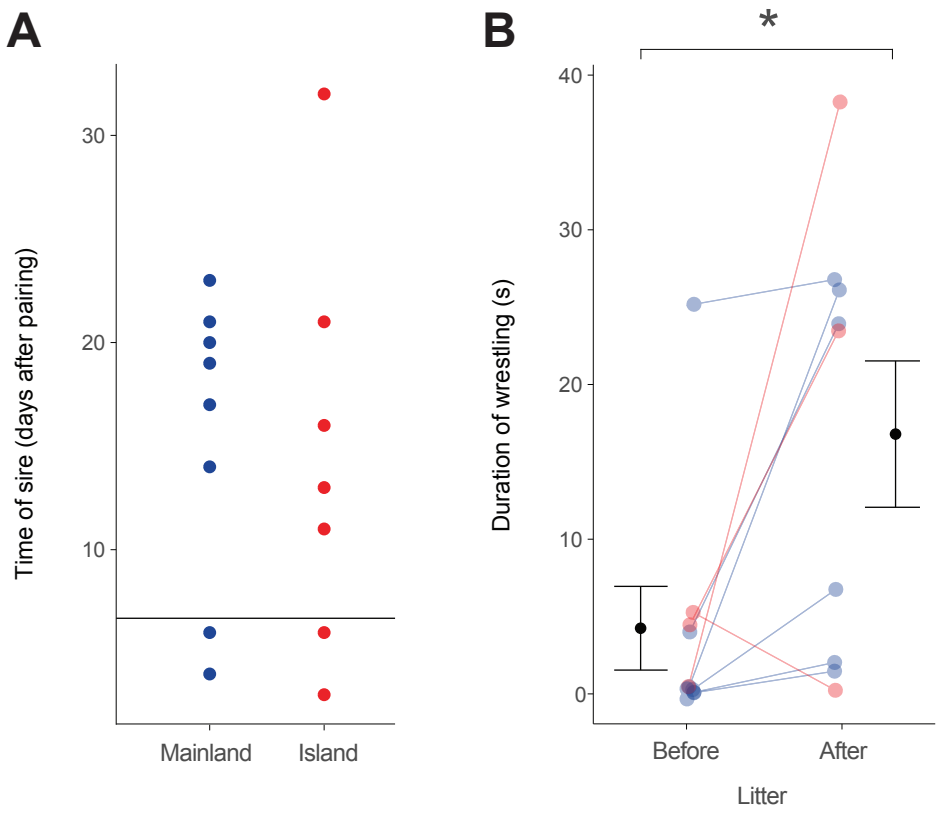
Suppl. Fig. 7. Growth rates of (**A**) mainland litters cross-fostered by island parents, compared to mainland litters that were not cross-fostered, (**B**) island litters cross-fostered by mainland parents, compared to island litters that were not cross-fostered. Statistical significance evaluated by linear fixed effects model (see Methods for details). * $P < 0.05$. Points and error bars represent mean and SEM of growth of litters.



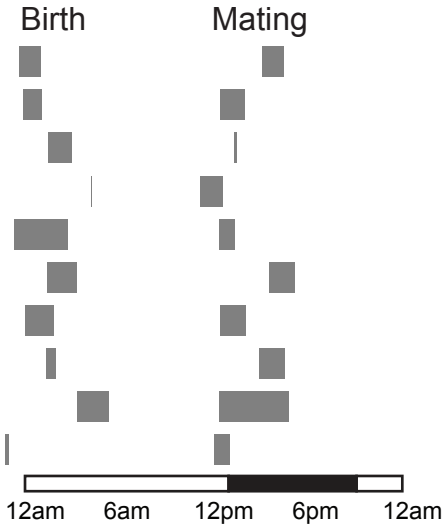
Suppl. Fig. 8. Count data of wrestling, chasing, and pindown behaviors in wild-caught founder animals (**A**), captive-born males paired with a female for one week before testing (**B**), and captive-born breeding males tested after verified copulation (**C**). Statistical significance evaluated by generalized linear models (see Methods for details). NS=not significant, * $P < 0.05$.



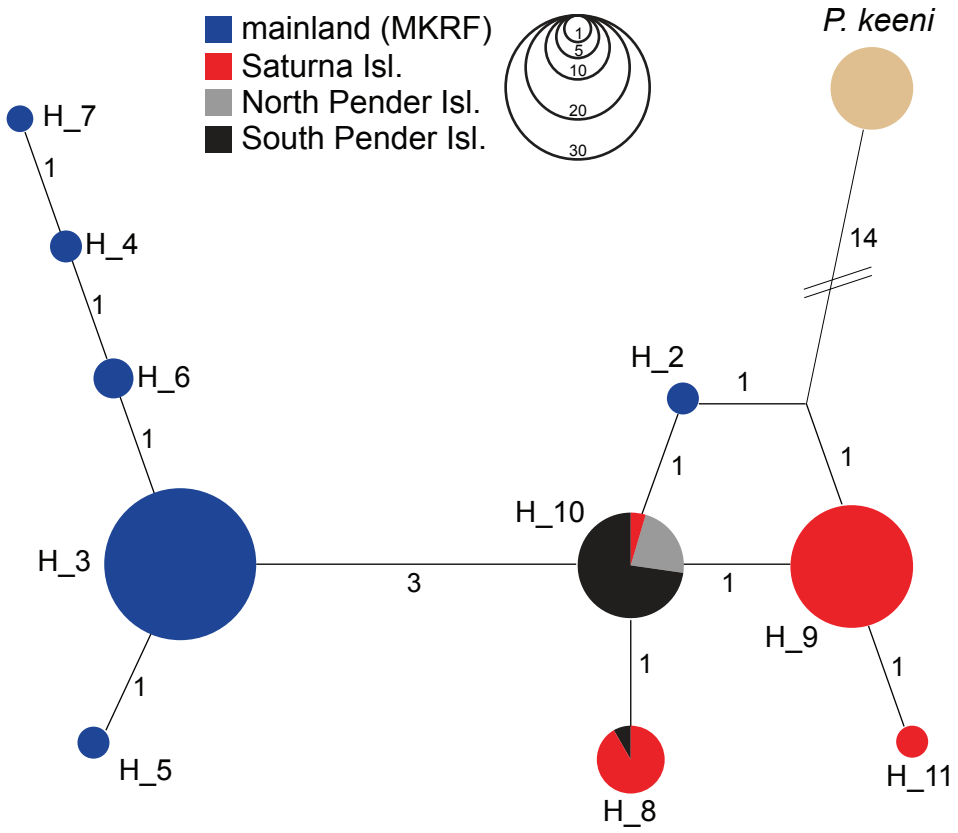
Suppl. Fig. 9. Effect of (**A**) time since capture at testing, (**B**) time since a litter was last sired, and (**C**) weight difference between resident and intruder on the aggressive behavior of wild-caught deer mice (N=7 island mice, N=6 mainland mice). Replicate trials are plotted separately. Statistical significance evaluated with a repeated measures mixed effects linear model (see Methods for details). NS=not significant.



Suppl. Fig. 10. Effect of reproductive experience on territorial aggression. (**A**) Time of mating after pairing in a subset of mice from the experiment in Fig. 5A-B. We backdated for females that went on to give birth when the litter was sired. The black line marks the time when males were tested in the resident-intruder assay (7 days after pairing); at this time, few males in this subset had sired litters. (**B**) We re-tested a subset of these males after they had sired a litter, and compared wrestling duration before and after siring a litter with a repeated measures linear mixed effects model (see Methods for details). Only mice were included that had not sired a litter by the time of first testing. Mean and SEM across strains is shown in black. * $P < 0.05$.



Suppl. Fig. 11. Schematic showing birth and mating events in breeding pairs. Each row corresponds to one pair. The gray boxes indicate the maximum duration of behaviors (it was not possible to observe females continuously, e.g. when they were inside the nest at the time of birth). The light cycle (16h:8h, light:dark) is shown at the bottom.



Suppl. Fig. 12. Haplotype network based on 381bp of the mitochondrial cytochrome b gene. Numbers on edges indicate the number of substitutions separating the adjacent haplotypes. The diameter of nodes is proportional to the number of specimens that carry the respective haplotype (as indicated in the schematic at the top of the graph).