An enhancer of \textit{Agouti} contributes to parallel evolution of cryptically colored beach mice

T. Brock Wooldridge\textsuperscript{a,b,c}, Andreas F. Kautt\textsuperscript{a,b,c}, Jean-Marc Lassance\textsuperscript{a,b,c}, Sade McFadden\textsuperscript{a,b,c}, Vera S. Domingues\textsuperscript{a,b,c,1}, Ricardo Mallarino\textsuperscript{a}, and Hopi E. Hoekstra\textsuperscript{a,b,c}

Contributed by Hopi E. Hoekstra; received February 16, 2022; accepted May 19, 2022; reviewed by Enrico Coen and Dolph Schluter

Identifying the genetic basis of repeatedly evolved traits provides a way to reconstruct their evolutionary history and ultimately investigate the predictability of evolution. Here, we focus on the oldfield mouse (\textit{Peromyscus polionotus}), which occurs in the southeastern United States, where it exhibits considerable color variation. Dorsal coats range from dark brown in mainland mice to near white in mice inhabiting sandy beaches; this light pelage has evolved independently on Florida’s Gulf and Atlantic coasts as camouflage from predators. To facilitate genomic analyses, we first generated a chromosome-level genome assembly of \textit{Peromyscus polionotus subgriseus}. Next, in a uniquely variable mainland population (\textit{Peromyscus polionotus allifrons}), we scored 23 pigment traits and performed targeted resequencing in 168 mice. We find that pigment variation is strongly associated with an \textasciitilde{}2-kb region \textasciitilde{}5 kb upstream of the \textit{Agouti signaling protein} coding region. Using a reporter-gene assay, we demonstrate that this regulatory region contains an enhancer that drives expression in the dermis of mouse embryos during the establishment of pigment prepatterns. Moreover, extended tracts of homozygosity in this \textit{Agouti} region indicate that the light allele experienced recent and strong positive selection. Notably, this same light allele appears fixed in both Gulf and Atlantic coast beach mice, despite these populations being separated by \textasciitilde{}1,000 km. Together, our results suggest that this identified \textit{Agouti} enhancer allele has been maintained in mainland populations as standing genetic variation and from there, has spread to and been selected in two independent beach mouse lineages, thereby facilitating their rapid and parallel evolution.

adaptation | camouflage | convergence | deer mice | pigmentation

To gain a complete picture of adaptation, we strive to understand both the molecular mechanisms and the evolutionary processes underlying trait evolution. On one hand, identifying the molecular basis of phenotypic adaptation can provide an opportunity to learn how traits vary—in particular, how specific changes in DNA can affect protein function or expression during development to produce the trait of interest. On the other hand, the evolutionary history of a specific allele can provide insights into when and why traits evolve. Importantly, an allele may be influenced by a combination of neutral and selective forces, which together explain its current distribution and frequency. Thus, the identification of a causal gene or better, a small gene region or mutation can serve as a handle with which to probe both the proximate (how) and ultimate (when/why) drivers of trait variation.

Cases of repeated evolution provide a particularly appealing context for understanding the drivers of adaptation. For example, one can ask the following question. Did similar phenotypes arise via the same or different molecular changes? While there are empirical examples of selection from new mutations (1–3), it has been suggested that rapid adaptation, in particular within species, may be fueled by selection on the same alleles from preexisting genetic variation (refs. 4–6; reviewed in ref. 7). Moreover, it has been argued that changes in \textit{cis}-regulatory elements may be the primary substrate of adaptation (8–10), although many examples of protein-coding changes (refs. 11–13; reviewed in ref. 14) or combinations of both regulatory and coding changes (15) have been identified. Nonetheless, when regulatory change has been implicated in repeated evolution, it is still rare that the causal regions, elements, or mutations have been identified (1, 16). This is in part due to the complexity of gene regulatory landscapes and the relative difficulty in testing the effects of a noncoding allele (17). By contrast, coding mutations are generally more amenable to identification and functional validation; therefore, when precise mutations have been shown to drive repeated evolution across species, they most commonly correspond to coding mutations (18). Thus, it remains difficult to determine the extent to which similar or different mutations contribute to repeated phenotypic evolution and where in the genome they occur.

Significance

Oldfield mice have independently colonized the white-sand beaches of Florida’s Gulf and Atlantic coasts, where they have evolved light fur that camouflages them from visually hunting predators. We find that fur color is strongly associated with DNA variation in a small regulatory region of the \textit{Agouti signaling protein}, which contains an enhancer that drives expression in mouse skin. This regulatory allele is found in all light-colored beach mice on both coasts, despite being separated by \textasciitilde{}1,000 km. Based on patterns of DNA variation within and between populations, our results suggest that this \textit{Agouti} allele arose in the mainland and then, has spread to and been selected in two independent beach mouse lineages, thereby facilitating their rapid and parallel evolution.

Author affiliations: \textsuperscript{a}Department of Organismic & Evolutionary Biology, Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138; \textsuperscript{b}Department of Molecular & Cellular Biology, Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138; \textsuperscript{c}HHMI, Harvard University, Cambridge, MA 02138; and \textsuperscript{1}Department of Molecular Biology, Princeton University, Princeton, NJ 08544

Author contributions: T.B.W., A.F.K., V.S.D., and H.E.H. designed research; T.B.W., A.F.K., J.-M.L., S.M., V.S.D., R.M., and H.E.H. performed research; T.B.W., A.F.K., J.-M.L., V.S.D., and R.M. analyzed data; and T.B.W., A.F.K., and H.E.H. wrote the paper.

Reviewers: E.C., John Innes Centre; and D.S., The University of British Columbia.

The authors declare no competing interest.

Copyright © 2022 The Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution License 4.0 (CC BY).

1Present address: Nature Ecology & Evolution, London N1 9Ww, United Kingdom.

To whom correspondence may be addressed. Email: hoekstra@eob.harvard.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2202862119/-/DCSupplemental.

Published July 1, 2022.
Variation in pigmentation has long served as a model for the study of adaptation. At the molecular level, the genes and pathways involved in vertebrate pigmentation have been well characterized (19). At the phenotypic level, color can vary dramatically in the wild, be measured straightforwardly, and have clear links to fitness (20). One classic example of repeated color evolution involves beach mice in the southeastern United States. Beach mouse subspecies on Florida’s Gulf and Atlantic coasts have independently evolved light coloration from a dark-colored mainland ancestor (21). Previous work identified three genomic regions involved in differences between Gulf coast beach mice and mainland mice (22), in which two pigment genes have thus far been implicated: the Melanocortin-1 receptor [Mc1r (19)] and the Agouti signaling protein [ASIP (23)]. The interaction between Mc1r and Agouti mediates the switch from dark (eumelanin) to light pigment (pheomelanin) production in mammals (24–26).

In Gulf coast beach mice, a single missense mutation in Mc1r reduces the receptor’s signaling ability, thereby contributing to light pigmentation in some (but not all) populations (19), but there are no differences in Mc1r expression level (22). By contrast, differences in Agouti transcript abundance are associated with pigment variation, with higher expression in Gulf coast beach mice than mainland mice, but there are no differences in the Agouti coding sequence (22). Therefore, while changes in genes at multiple levels of the pigment pathway have been implicated in the evolution of camouflaging coloration in Gulf coast beach mice (i.e., Mc1r and Agouti), we have an incomplete understanding of the regulatory mechanism driving differences in Agouti expression.

The genes (and mutations) contributing to the light coats of the Atlantic coast beach mice have remained even more elusive. For example, the Mc1r amino acid change found in Gulf coast mice is absent from Atlantic coast mice (19), and no new Mc1r mutations are associated with color variation or have a measurable effect on Mc1r function (21). Moreover, there are no differences in the Agouti coding region between mainland and Atlantic coast beach mice (22), and changes in Agouti expression have not been measured. Without knowledge of the genes or mutations underlying light pigmentation in the Atlantic coast beach mice, an outstanding question remains: Is the remarkable similarity in coloration between Gulf and Atlantic coast beach mice due to the same or distinct pigmentation alleles?

Here, we return to the classic case of adaptation in Gulf and Atlantic coast beach mice, first described over a century ago (27, 28), and capitalize on naturally occurring color variation differences in both subspecies in the phylogeny mirror their geographic distribution, a pattern that is supported by a genetic analysis of genome-wide single nucleotide polymorphisms (SNPs) derived from putatively neutral regions in a targeted sequence-capture dataset, we generated a highly supported phylogeny confirming the independent origin of beach mice on the Gulf and Atlantic coasts from an ancestral mainland form (Fig. 1C), consistent with previous studies (21, 29–31). The Gulf coast beach mice form a paraphyletic group with adjacent mainland populations, all of which share a common ancestor between 3.5 and 7.2 thousand years ago (yka). Similarly, the Atlantic coast beach mice share a common ancestor with their closest mainland counterparts 2.9 to 6.4 yka, suggesting that both Gulf and Atlantic beach lineages originated at approximately the same time. In general, we find that the relationships of subspecies in the phylogeny mirror their geographic distribution, a pattern that is supported by a genetic analysis of genome-wide single nucleotide polymorphisms (SNPs) derived from putatively neutral regions in a targeted sequence-capture dataset (Fig. 1D). The evolutionary history of both Gulf and Atlantic beach mice as well as several mainland populations provides a demographic context in which to understand the evolution of crypsis.

**Results**

**Assembly of a High-Quality Chromosome-Level Genome for *P. polionotus*.** We first generated whole-genome sequencing (WGS) data and assembled a de novo high-quality reference genome for the oldfield mouse, *Peromyscus polionotus subgriseus* (BioProject no. PRJNA494229). The final genome was 2.645 Gb in length with an N50 scaffold length of 13 Mb. We could anchor 97% of the de novo assembled bases into 23 autosomes and the X chromosome using high-density genetic linkage maps for *Peromyscus*. Our results indicate that the assembly contains 95.4 and 94.8% of single-copy core mammalian and euchromatic genes, respectively. Our annotation strategy, which combined comparative in silico and evidence-based approaches, identified 18,502 protein-coding genes having orthologs in the *Mus* genome, 536 paralogs of *Mmu* genes, and 1,912 additional genes showing homology with known proteins from curated databases. This high-quality genome enables evolutionary analyses of genome-wide variation across populations of this species.

**Recent and Independent Evolution of Beach Mice on the Gulf and Atlantic Coasts.** To better estimate the timing and pattern of divergence in the beach and mainland subspecies (Fig. 1A), we sampled six beach and five mainland populations, all together representing 9 of the 14 extant *P. polionotus* subspecies (Fig. 1B) as well as the closely related sister species, *Peromyscus maniculatus subterracea*. Using 1,000 randomly distributed genome-wide single nucleotide polymorphisms (SNPs) derived from putatively neutral regions in a targeted sequence-capture dataset, we generated a highly supported phylogeny confirming the independent origin of beach mice on the Gulf and Atlantic coasts from an ancestral mainland form (Fig. 1C), consistent with previous studies (21, 29–31). The Gulf coast beach mice form a paraphyletic group with adjacent mainland populations, all of which share a common ancestor between 3.5 and 7.2 thousand years ago (yka). Similarly, the Atlantic coast beach mice share a common ancestor with their closest mainland counterparts 2.9 to 6.4 yka, suggesting that both Gulf and Atlantic beach lineages originated at approximately the same time. In general, we find that the relationships of subspecies in the phylogeny mirror their geographic distribution, a pattern that is supported by a genetic analysis of genome-wide single nucleotide polymorphisms (SNPs) derived from putatively neutral regions in a targeted sequence-capture dataset (Fig. 1D).

The evolutionary history of both Gulf and Atlantic beach mice as well as several mainland populations provides a demographic context in which to understand the evolution of crypsis.

**Phenotypic Variation in a Single Mainland Population (*Peromyscus polionotus albifrons*).** We sampled one mainland population neighboring beach habitat, *P. p. albifrons*, that exhibited a wide range of coat colors—from light and sparsely pigmented coats similar to those of beach mice to the dark and extensively pigmented coats typical of mainland mice. To characterize and quantify this variation, we measured 23 coat-color traits in 168 skin specimens of *P. p. albifrons* (Fig. 2A). All traits were related to either the distribution of pigmentation (e.g., tail-stripe length) or intensity of pigment (e.g., dorsal hue, brightness) and are known to vary among beach mouse populations (21). To establish reference points with which to compare the *albifrons* population, we scored the same 23 traits in representative mice from Gulf coast, Atlantic coast, and mainland subspecies (SI Appendix, Table S1).

We found that many pigment traits are highly correlated in the *albifrons* population (Fig. 2B). A principal components analysis
cies similar in appearance (and pPC1 score) to the mainland subspecies observed in the population represent the darkest mice, which are passed as many traits as pPC1 or show the same strength of association with population structure, not surprising given that the extent of pigmentation on the cheek, rump, whisker, and ankle—heavily load on phenotypic PC1 (pPC1) and that a species’s pPC1 score is a strong predictor of overall lightness or darkness (Fig. 2B and SI Appendix, Fig. S1). Remaining traits also form distinct clusters, but none of these additional clusters encompass as many traits as pPC1 or show the same strength of association with overall coloration (Fig. 2B). The highest pPC1 values observed in the population represent the darkest mice, which are similar in appearance (and pPC1 score) to the mainland subspecies P. p. albifrons (Fig. 2C). Additionally, while the lightest albifrons individuals are still darker than the geographically proximate beach subspecies P. p. leucoviridis—the palest form of the Gulf beach mice—many individuals with intermediate pPC1 scores are comparable with a typical Atlantic beach mouse (e.g., P. polionotus niveiventris) (Fig. 2C). Despite this range in coloration that encompasses both beach and mainland phenotypes, none of these pigment traits show a significant association with population structure, not surprising given that the P. p. albifrons population has little detectable genetic structure (SI Appendix, Fig. S2).

Association between Pigmentation and a Noncoding Region of Agouti. Capitalizing on the extensive color variation observed within the panmictic P. p. albifrons population, we performed single-variant association mapping using the sequence-capture data from this population. These data include 6,547 putatively neutral biallelic SNPs from across the genome as well as the genomic regions encompassing two pigmentation genes, Agouti and Mc1r (190 and 150 kb in length, respectively, including all exons and known regulatory regions). In our scan, we detected a single region associated with pPC1 that exceeded the genome-wide significance threshold (P < 1.23 × 10−5 corrected for the number of effective tests) (Fig. 3A) in the Agouti locus. A closer investigation of this region revealed three SNPs significantly associated with pigment variation, spanning 1,756 bp, in strong linkage disequilibrium (mean r2 = 0.85). A single SNP on chromosome 4 (chr4) at position chr4: 9,845,301 showed a markedly stronger association with pPC1 than the other two (Fig. 3B). This SNP is located between two untranslated exons (exons 1D and 1E), is 120 bp upstream of a cluster of Short Interspersed Nuclear Elements (SINE) in reverse orientation relative to the transcription of Agouti, and is 5,641 bp upstream of the first coding exon (exon 2). Genotype–phenotype regressions show an additive effect of this locus, which explains 36% of the variance in pPC1, as well as a substantial degree of additive variation in pPC1-correlated traits, such as dorsal brightness (19%) or tail-stripe length (7.2%) (Fig. 3C). Together, these data point to a small noncoding region of Agouti containing a mutation(s) having a major effect on variation in overall pigmentation in P. p. albifrons.

The Candidate Agouti Region Is Capable of Regulatory Activity. To determine if this ~2-kb Agouti region associated with pPC1 is capable of regulatory activity, we first determined
whether the region overlaps with known regulatory elements (Fig. 4A). In the homologous region and ±10 kb surrounding sequence in Mus, we observe few known regulatory elements, none of which are associated with dermal tissues (SI Appendix, Table S2). Moreover, it does not overlap with any previously identified regions associated with pigment variation in other Peromyscus species (e.g., ref. 32) (SI Appendix, Fig. S3). As sequence conservation can be indicative of conserved molecular function, we next examined sequence similarity across 27 rodents in the 5-kb upstream and downstream of the top-associated SNP (which also includes the two linked variants and SINE elements). Surprisingly, conservation within rodents was minimal, with only a subset of the species—the superfamily Muroidea—showing greater than 50% sequence similarity for the majority of the region (Fig. 4B and SI Appendix, Fig. S4). These data suggest that if this region has regulatory function, it is likely to have evolved recently.

To assess whether the candidate region of Agouti contains functional enhancers, we cloned a 2.6-kb sequence that spans 0.5 kb upstream and 2.1 kb downstream of the most strongly associated variant (i.e., chr4: 9,845,301) and includes the two additional associated variants (chr4: 9,845,152, chr4: 9,846,908) as well as a small downstream region conserved in rodents (Fig. 4B and SI Appendix, Fig. S5). We then inserted this sequence upstream of a minimal promoter and lacZ reporter gene (Fig. 4C). Given the currently limited transgenic techniques available for Peromyscus, the resulting construct was injected into embryos of Mus (strain FVB/NJ), and embryos were collected at embryonic stage (E) 14.5, a time point when Agouti expression plays a key role in the establishment of pigment prepatterns in both Mus and Peromyscus (23). Of the 14 embryos with independent genomic integrations of the lacZ construct (verified by PCR), we observed consistent lacZ expression in the skin of eight embryos, although expression was spatially variable across embryos (Fig. 4D and SI Appendix, Fig. S5). Histological analysis showed that lacZ was localized to the dermis, corresponding to the known site of endogenous Agouti expression during embryonic development (Fig. 4E). Together, the results of these experiments suggest that this previously undescribed ~2-kb noncoding region contains a cis-regulatory element (or possibly multiple elements constituting a cis-regulatory module) capable of driving Agouti dermal expression during embryonic development.

The Light Agouti Allele Shows a Signature of Positive Selection. We next tested if there was evidence of natural selection acting on the light-associated allele at this regulatory element, which is found at 86% frequency in the P. p. albifrons population (SI Appendix, Table S3). In the region surrounding the top-associated SNP (chr4: 9,845,301), we found that haplotype homozygosity decays more quickly for the dark allele than for the light allele, a signal consistent with recent positive selection for light pigmentation (Fig. 5A). This signal of extended haplotype homozygosity (EHH) is statistically significant, with all 3 candidate SNPs identified in our association analysis (as well as 15 additional SNPs in this region) showing a significantly positive integrated haplotype score (IHS; P < 0.05) (Fig. 5B). We did not detect a signal of selection at these candidate SNPs in any other population, although low sample sizes and lack of polymorphisms limit our power. Together, these data support the hypothesis that natural selection, most likely favoring light coloration, has led to an increase in light Agouti allele frequency in the P. p. albifrons population.

The Light Agouti Allele Is Fixed in Both Gulf and Atlantic Beach Mouse Populations. Given the evidence for nonneutral evolution at this Agouti regulatory element, we next aimed to infer whether it exhibits a unique evolutionary history relative to the rest of the genome. Using Saguaro (33), we generated local phylogenies in variably sized genomic windows across the Agouti locus for the combined dataset of beach and mainland populations and then compared these phylogenies with the population tree constructed from genome-wide neutral loci. At Agouti, we find that much of the locus fits a topology (cactus 6) that coarsely mirrors the population-level phylogeny (Fig. 5D). In contrast, one unique topology (cactus 4) is
exclusively derived from two small regions that include the top-associated SNP and closely match the regulatory element identified in our mapping experiment (i.e., chr4: 9,841,443 to 9,842,079 and chr4: 9,844,852 to 9,847,023) (Fig. 5C). This unique topology clusters all individuals homozygous for Agouti light alleles into a single clade with short branch lengths, consistent with a recent origin for this allele, while all dark allele homozygous individuals fall into a clade with longer branch lengths similar to that observed in the population tree (Fig. 5D). Therefore, not only does this unsupervised approach identify the same region of Agouti as was localized in both the independent association and the selection analyses, but it also points to a single origin for the light Agouti allele. These results show that the light Agouti allele arose once and is now shared by all Gulf and Atlantic beach mice, even though these lineages are geographically distant and independently colonized their respective beach environments from mainland ancestors.

The light Agouti allele, however, occurs in several mainland populations in addition to the albifrons population and is likely maintained as standing genetic variation. Consistent with this hypothesis, in the albifrons population, we observe elevated levels of sequence divergence (e.g., D_{XY}) at the Agouti regulatory element between light and dark alleles (SI Appendix, Fig. S7), suggesting that the light Agouti allele may have been introduced to P. p. albifrons through introgression from another population. This scenario, in which an allele is maintained in several ancestral (mainland) populations and repeatedly selected for in multiple (beach) lineages, suggests that parallel genotypic evolution (sensu ref. 34) has been an important factor in the evolution of light-colored beach mice.

**Discussion**

The questions of how adaptation proceeds at the molecular level and how predictable the process is have long been of...
interest to evolutionary biologists. Debate around, for example, the locus of adaptation (i.e., coding vs. regulatory mutations), the source of adaptive mutations (i.e., de novo mutations vs. preexisting variation), and the repeatability of this process (i.e., do the same or different mutations lead to similar independently evolved traits) has been lively (7, 14, 18, 34–37). Here, we provide insight into these questions from a classic system, first described over a century ago (27): the recent adaptation of two independent lineages of beach mice to white-sand habitat through the evolution of camouflaging color. While the Agouti gene has been shown to contribute to the evolution of the Gulf coast beach mouse color pattern through changes affecting both its expression level (22) and spatial domain during embryogenesis (23), the molecular basis of these regulatory changes remained unclear. Furthermore, whether Agouti contributed to light pigmentation in the Atlantic coast beach mice remained unknown. Without this information, the question of if the genetic basis of light pigment in the Gulf and Atlantic coast beach mice was the same or different was unknowable. Here, we uncovered a regulatory element in the Agouti gene and provided evidence that an allele of this element is associated with lighter pigmentation, which has been selected repeatedly in both lineages of beach mice.

To identify mutation(s) that contribute to changes in pigmentation, we first identified a population (P. p. albifrons) that was phenotypically variable, ranging from light beach to dark mainland coat color. While most populations show little variation in pigmentation, this mainland population appears unique, likely because of its geographic proximity (~25 km) to the beach habitat, its patchwork of light sandy and dark loamy soil, and intermediate level of vegetative cover relative to the open beach and dense mainland habitats. By conducting genetic association mapping in this variable population, we were able to narrow in on a small—approximately 2-kb—noncoding region that is strongly associated with overall pigmentation. This region having a causal effect on pigmentation is bolstered by two additional results. First is its ability to drive expression in the dermis of Mus embryos at a developmental stage relevant to the establishment of pigmentation precursors. Second, patterns of DNA polymorphism show a strong signature of positive selection in this same small region. Interestingly, this region had not been previously identified as functionally important in Mus (26) or Peromyscus sp. (32, 38). Moreover, this region is not highly conserved (in rodents), suggesting it may have evolved regulatory function only recently. This regulatory element further supports the observation that Agouti regulation is highly modular (26, 32, 38), which could, in turn, explain why Agouti expression may be the target of repeated evolutionary tinkering across vertebrates [for example, in rabbits (5), dogs (39), buffalo (40), and birds (41, 42)—akin to other highly modular genes, such as Pax6 in vertebrates (43), Pitr1 in stickleback fish (1, 44), Ebony in Drosophila (45), and Asip2bl Agtp2 in cichlid fishes (46)].

This 2-kb Agouti regulatory element likely contains causal mutation(s) that affect pigmentation. In total, there are 11 fixed differences between the light and dark Agouti alleles found in beach and mainland mice. While precisely which variants are causal remains unclear, the three of these SNPs that are significantly associated with overall pigmentation (pPC1) in the polygenic albisfons population represent the strongest candidates (SI Appendix, Table S5). We also observe several complex insertions/deletions (INDELS) and repetitive elements in the same region, which may themselves affect Agouti expression and drive an association signal in linked SNPs. While many of these variants disrupt predicted transcription factor (TF) binding sites identified in Mus (SI Appendix, Table S4), such predictive approaches have poor specificity, and it is unclear if or when any of these sites may be active. Additionally, because this region is not well conserved even among rodents, results from gene-editing experiments in Mus, particularly ones that could result in subtle variation between mutants, may be challenging to interpret. However, future surveys of additional individuals in the admixed albisfons population may allow us to pinpoint the causal mutation(s). Similarly, the establishment of dermal cell lines that express Agouti in the correct trans environment—cell lines that are currently unavailable—could allow us to test the effects of specific mutations or combinations of mutations via cell-based assays.

Nonetheless, with a regulatory element identified, one can now more easily determine the source of this variation. In our survey of Agouti variation across both beach and mainland populations, we found that the light Agouti allele is fixed in both beach lineages (see below) and also segregating or even fixed in some dark mainland populations. One possibility is that the light Agouti allele first arose de novo in one beach lineage and was subsequently introduced through gene flow to the other beach lineage via the mainland population. However, an alternative scenario may be more plausible; the mutation first arose in mainland mice, was maintained as standing variation, and then, was introduced to independent beach lineages via gene flow. First, beach mice are known to have very small
populations (31, 47); thus, the opportunity for a new adaptive mutation to arise is low (29). Second, because the beach habitat is relatively young [8 to 10 kya (48)] and colonization of that habitat occurred relatively recently (29), there has been only limited time for a new mutation to arise (much less spread from the beach and across the mainland to the opposite coast via migration). Third, the estimated migration rate from the beach to the mainland is ~10-fold lower than the reverse (29); this asymmetric migration makes it more likely that substantial genetic variation is being contributed from the mainland rather than to the mainland. Thus, a likely scenario is one in which the light Agouti allele originated in a mainland population, was maintained as a standing genetic variation, and then, selected repeatedly in mice that independently colonized the Gulf and Atlantic beaches. This hypothesis is consistent with the results from Mc1r, in which the age estimate for the emergence of the causal Mc1r mutation predates the age of the beach habitat (29).

Interestingly, this scenario was predicted almost a century ago by Francis Sumner (28) based on reports of light-colored mice occurring on the mainland near isolated beach habitats (27). However, while the mainland P. p. albifrons population represents a likely reservoir of genetic variation, it is unlikely that the light Agouti allele originated in this specific population. High sequence divergence in the focal Agouti regulatory region between light and dark alleles suggests that the light Agouti allele was introduced to the albifrons population through introgression from another unsampled (likely mainland) lineage. Regardless of the precise geographic origin of the causal mutation(s), the natural history of the region and the demographic history of these populations together suggest that mainland mice are the likely source of the light Agouti allele, which was then used for the rapid and repeated evolution of light-colored beach mice.

This evolutionary scenario then raises the question of how the light Agouti allele is maintained in mainland populations, where it may be deleterious. Indeed, previous field experiments demonstrated that models of light mice experienced higher rates of predation than dark models in a dark soil mainland habitat (49). We raise two possible and nonmutually exclusive explanations. First, the light allele may persist in mainland mice because epistasis limits its phenotypic effect; many mainland mice that carry the light Agouti allele (even homozygotes) appear to have relatively dark coloration typical of a mainland mouse. In Peromyscus, Agouti is known to contain multiple mutations that affect pigmentation (32) and to interact with epistatic interactions between mutation(s) in the this Agouti regulatory element and other mutations in Agouti or elsewhere in the genome explain why, in some populations, the light Agouti allele has minimal effect on pigmentation (36), thus limiting its visibility to selection. Indeed, previous work in beach mice demonstrated a role for epistasis between Mc1r and Agouti (22). A second possibility is that the light allele may have a phenotypic effect, even if small, in some mainland populations that persists due to varying selection pressures. In mainland
populations, such as *P. p. albifrons*, soil coloration is not uniformly dark but rather, patchy, with sometimes large regions of surprisingly light beach-like substrate. Such mainland areas have light sandy soil due to the geological history of the southeastern United States, which has experienced successive episodes of glacial advance and retreat, depositing light sediments inland and forming sand-dune habitats that remain to this day (50). Thus, the light *Agouti* allele may, at least in some mainland populations, be beneficial, consistent with a signature of positive selection on the light allele in the *P. p. albifrons* population. However, the *P. p. albifrons* population also harbors dark *Agouti* alleles, possibly due to spatially variable selection in patchy habitats (e.g., ref. 51) or due to migration from surrounding dark mainland populations (52, 53). Additional sampling through the range of *P. polionotus*, including measurements of soil color, combined with whole-genome resequencing may shed further light on these two hypotheses to explain the prevalence of the light *Agouti* allele in mainland populations.

The distribution of the light allele—segregating in some mainland populations and fixed in both beach mouse populations—is consistent with a scenario in which the same allele was independently selected in the two beach lineages from standing genetic variation. We note, however, that without specific information about the causal mutation(s), we cannot rule out the formal possibility that independent mutations with similar phenotypic effects evolved within the same small regulatory region. However, given the monophyly of the light allele, this would necessitate the unlikely scenario in which independent evolution of causal mutation(s) would have had to occur on the same *Agouti* haplotype. Moreover, given that the Gulf and Atlantic clades are closely related and recently derived from similar mainland ancestors (22), some may argue that this was an ideal scenario for repeated selection on shared ancestral variation (4, 35). Sharing the same *Agouti* light mutation(s) would provide a simple mechanistic explanation for why the Gulf and Atlantic coast beach mice are so similar in pigmentation (19).

Together, these results suggest a scenario in which a cis-acting regulatory mutation(s) in *Agouti* likely evolved in the mainland and was independently selected in both the Gulf and Atlantic coast beach mice, contributing to their rapid, parallel evolution. This evolutionary history is in stark contrast with previous results for a second pigmentation gene *McIr* (19); a coding change (i.e., a single amino acid mutation) contributes to light coloration in beach mice in Gulf coast beach mice but not Atlantic coast beach mice (21). Thus, together, these two genes demonstrate how—even within a single species and associated with the same adaptive trait—evolution may take very different genetic paths to similar phenotypic ends.

**Materials and Methods**

**Specimen and Tissue Collection.** Over two expeditions, in the summer and winter of 2009, we collected 168 *P. p. albifrons* mice from a single population occupying a habitat with patches of both light-colored sand and dark loam–clay soil in Lafayette Creek Wildlife Management Area of Walton County, Florida, ~25 km inland from the Gulf of Florida (SI Appendix, Table S1). Mice were captured overnight using Sherman live traps. Following euthanasia, we sampled liver tissue from each individual and placed the tissues in 95% ethanol until cultured overnight using Sherman live traps. Following euthanasia, we sampled soil in Lafayette Creek Wildlife Management Area of Walton County, Florida, over two expeditions, in the summer and winter of 2009, we collected 168 *P. p. albifrons* mice from a single population occupying a habitat with patches of both light-colored sand and dark loam–clay soil in Lafayette Creek Wildlife Management Area of Walton County, Florida, ~25 km inland from the Gulf of Florida (SI Appendix, Table S1). Mice were captured overnight using Sherman live traps. Following euthanasia, we sampled liver tissue from each individual and placed the tissues in 95% ethanol until their scores were identical. For the full data set, each trait was scored by a single individual across all specimens to ensure consistency; two researchers scored traits, with each scoring half the traits in all individuals. Second, to measure pigment intensity, we used a FLAME UV/Vis spectrometer with a pulsed xenon light source, a 400-nm reflectance plate, and OceanView software (Ocean Optics) to measure five reflectance spectra from each of three body regions (dorsal stripe, flank, and ventrum). We used a custom R script to obtain brightness, hue, and saturation values in the visible spectrum (400 to 700 nm) with 1-nm bin width using a segment classification approach (54) with formulae as described for CLR v1.05 (55). For all traits, we took five measurements and then, calculated the median value for each body region for each individual. In total, we measured these 23 traits on 168 *P. p. albifrons* specimens as well as representative individuals from the Gulf (*P. leucopus*), n = 13), Atlantic (*P. p. niveiventris*, n = 15), and mainland (*P. polionotus*, n = 17) populations.

**Trait Correlations and Phenotypic PCA.** To test for correlations among traits, we calculated pairwise trait correlations using the cor() method (pearson, user="complete.obs") and cor.mtest() functions in base R, correcting for the number of pairwise tests to determine statistical significance (Bonferroni method). To account for trait correlations and to reduce the dimensionality of our dataset, we performed a Principal Component Analysis (PCA) of all pigmentation traits using the FactoMineR v2.3 (56) and factoextra v1.0.6 (57) R libraries. More specifically, we first estimated the best number of dimensions for imputing missing data with estim_ncpPCA (method="KFold"), imputed missing data based on the estimated number with imputePCA, ncp = 5, and then, performed the PCA on the imputed dataset using the PCA() function. Pairwise trait correlations and the phenotypic PCA were based on *P. p. albifrons* individuals only. To compare overall pigmentation (largely captured by pPC1) among populations, pigment scores for individuals from other populations were projected onto the *albifrons* principal component space post hoc using the predict() R function.

**Genome Sequencing and Assembly.** DNA was extracted using standard laboratory procedures from the liver of one female (*P. polionotus subgriseus*, PO stock) obtained from our laboratory colony. By choosing a female individual, we have equal coverage for the autosomes and the X chromosome, but the Y chromosome is not part of the assembly. We prepared libraries with Illumina TruSeq DNA Sample Prep Kit v2 according to the manufacturer’s instructions and performed de novo sequencing on an Illumina HiSeq platform using a combination of short paired-end libraries and longer mate-pair libraries suitable for use with the ALLPATHS-LG genome assembler v. 58. All libraries were constructed and sequenced at the Broad Institute Sequencing Platform. In total, we generated 240.27 Gb of raw sequence data, representing a total physical coverage of 290× and a sequence coverage of 68×. We assembled these reads using ALLPATHS-LG.

We used ALLMAPS v.0.6.2 (59) in combination with five genetic maps based on interspecific crosses [Restriction site-Associated DNA sequencing (RAD-seq) based (60–63)] gene based (64)] to assemble the scaffolds into the pseudochromosomes. DNA sequences corresponding to 182 genes and RAD-seq markers used to build the genetic maps were aligned against the genome using BLAT v.36 × 2 (65). Markers that could not be unambiguously mapped to a single location in the genome were filtered out. A total of 58,922 markers were included in the dataset. During a first iteration, ALLMAPS revealed that a total of 66 scaffolds housed markers associated with more than one linkage group and were likely misassembled. These were subsequently split, and the position of the break
points was determined based on the ALLMAPS predictions and the location of discordantly mapped reads. In most cases, these corresponded to assembly gaps. After correcting for these assembly errors, ALLMAPS was run an additional time to generate the pseudochromosomes. Our final assembly includes 531 scaffolds, encompassing 2,575,648,500 bp (97.4% of the total assembled sequence), distributed in 23 autosomes and the X chromosome. The orientation of 461 scaffolds corresponding to 2,566,039,849 bp (97.0% of the total sequence) could be determined due to the presence of more than one marker. We assigned chromosome names based on previous reports from interspecific reciprocal whole-chromosome painting, which allowed us to assign linkage groups with known genes to Peromyscus chromosomes (64, 66). The chosen chromosome assignments reflect the standardized Peromyscus cytogenetic nomenclature (67).

**Genome Annotation.** We annotated repetitive elements using a combination of RepeatModeler (68) and RepeatMasker v. open-4.0.8 (69) using Peromyscus- and rodent-specific repeat libraries. To annotate protein-coding genes, we used a recently developed annotation strategy making use of multiple genome alignments and an existing high-quality annotation set called the Comparative Annotation Toolkit (CAT) v.4.414000 (70). While permitting the finding of newly discovered genes via ab initio gene modeling, this approach allows to identify orthology relations readily and with high accuracy. We first aligned the old-field mouse chromosome-level assembly to the assemblies of the laboratory mouse (Mus musculus; GRCm38), the rat (Rattus norvegicus; Rnor_6.0), the prairie vole (Microtus ochrogaster; Mioch1.0), and the prairie deer mouse (Peromyscus maniculatus bairdii; Pman2.1.3) using ProgressiveCactus v.0.0 (71, 72). We reasoned that, including more species that represent progressive levels of evolutionary divergence would improve the accuracy of the ancestral sequence reconstruction process that takes place during the preparation of the whole-genome alignment. Using CAT, we annotated the oldfield mouse genome using the genome of M. musculus (GRCm38/mm10) and the high-quality and well-curated GENCODE VM15 as the reference gene/transcript set as well as extensive transcriptome sequencing datasets for P. polionotus corresponding to five tissues (brain, testis, hypothalamus, main olfactory epithelium, and vomeronasal organ) and skin RNA-sequencing (RNA-Seq) data from the prairie deer mouse, P. maniculatus bairdii.

To obtain quantitative measurements of the completeness of the genome assembly, we used BUSCO v.3.0.2 (73) with BLAST v.2.2.28+; HMMER v.3.1b2, and AUGUSTUS v.3.3.2. We used human as species, which specifies the parameters used by AUGUSTUS, and the mammalia and euarchontoglires gene sets for our analyses.

**Population Sequencing, Variant Calling, and Genotype Likelihoods.** For high-coverage WGS of representative beach (P. p. leucocephalus) and mainland (P. p. subgriseus) populations, we extracted DNA from ∼20 mg of liver tissue and generated sequencing libraries using a PCR-free KAPA HT kit. Following enzymatic fragmentation, we used size selection to enrich for a 450-bp insert size and ligated Illuma adapters. We sequenced the resulting libraries using 150-bp paired-end sequencing on an Illumina NovaSeq S4 flow cell to achieve 15 to 20x coverage.

For additional Gulf, Atlantic, and mainland populations, we used a sequence-capture strategy aimed at sequencing both putatively neutral loci and the pigmentation genes Agouti and Mctr. Specifically, we targeted ∼5,000 1.5-kb noncoding regions randomly distributed across the genome as well as 190- and 150-kb regions flanking the Agouti and Mctr loci, respectively (ref. 29 has capture array design details, and ref. 74 has Agouti and Mctr sequencing details). This strategy was applied to five Gulf beach mouse subspecies (Peromyscus p. ammobates, Peromyscus p. alliophys, Peromyscus p. trisyllepsis, Peromyscus p. peninsularis, P. p. leucocephalus), three mainland subspecies (P. p. polionotus, P. p. albifrons, three populations of P. p. subgriseus), and one Atlantic beach subspecies (P. p. niveiventris) (SI Appendix, Table S1). The availability of both high-quality WGS and sequence-capture data for the P. p. leucocephalus and P. p. subgriseus subspecies allowed us to verify that the sequence-capture loci accurately represented each population’s genetic diversity.

For both WGS and sequence-capture data, we converted raw fastq files to unmapped bam files using FastqToSam (Picard toolkit v.2.18.4 (75)) and then, marked Illuma adapters using MarkIllimunAdapters (Picard). Using SamToFastq (Picard), we created interleaved fastq files and clipped adapter sequences.

We mapped sequencing reads to the P. polonotus subgriseus reference genome (see above) using bwa-mem (76), with -p to indicate interleaved paired-end fastq input and -M to mark short split hits as secondary for compatibility with Picard. We then used MergeBamAlignment (Picard) to merge mapped and unmapped bam files to preserve read group information and sequencing duplicates using MarkDuplicates (Picard), with OPTICAL_DUPLICATE_PIXEL_DISTANCE = 2,500 to account for artifacts generated from the patterned flow cell found in the NovaSeq S4.

We then called variants separately for the WGS and sequence-capture data sets to reduce processing time, as they vary significantly in both coverage and sample number. However, the following variant calling and filtering steps were applied equally to both data types. To begin, we used HaploTypoCaller (GATK v.3.8) on the merged bam files with the default heterozygosity prior (het = 0.005) and -ERC GVCF to produce per-sample genomic Variant Calling Format (gVCF) files. For the X chromosome, we specified a prior input ploidy based on a comparison of coverage with the autosomes using samtools depth (samtools v.1.10 (78)). Next, for the WGS data, we generated variant + invariant cohort-level vcf files for each chromosome using GenotypeGVCFs (GATK) with ‘-max-alternate-alleles 4 -all-sites’ for the sequence-capture data, the ‘all-sites’ parameter was removed, and only variants were reported. These raw cohort-level vcf files were split into INDELs and SNPs with SplitVcfs (Picard) and invariant sites with SelectVarians (GATK). We performed filtering on each set independently, excluding SNPs with Variant Confidence/Quality by Depth (QD) < 2.0, FS > 10.0, Root Mean Square (RMS) Mapping Quality (MQ) < 40.0, MQRankSum < −12.5, ReadPosRankSum < −8.0, or SOR > 3.0 and excluding INDELs with QD < 2.0, Phred-scaled p-value using Fisher’s exact test to detect strand bias (FS) > 200.0, ReadPosRankSum < −20.0, or Symmetric Odds Ratio of 2x2 contingency table to detect strand bias/QUAL – quality score (SOR) > 3.0. We also retained invariant sites with QUAL ≥ 20 using bcftools v.1.11-95 (79). These filtering parameters were based on a combination of GATK recommendations for datasets without truth/training sets and visual inspection of the distributions for each metric. We also set individual genotype calls to missing if the read depth at a given site was less than five. Finally, we combined the sequence-capture dataset with the WGS dataset using vcf-merge (vcftools v.0.1.15 (80)).

**Estimation of Population Structure.** To test for population structure, we ran a gPCA using PCAngsd v.0.973, which is specialized for use with low-coverage, high-throughput sequencing data (81). We used beagle genotype likelihood files for all sequence-capture loci as input and ran the program with default parameters. Using the output covariance matrix, we calculated eigenvalues and eigenvectors with the base R function eigen. We estimated population differentiation (FST) for all pairwise population comparisons using the program ANGSD v.0.999-21-g4c6d001 (82). We first calculated the two-population site frequency spectra (2DSFS) using the Site Allele Frequency (SAF) likelihood files generated by ANGSD, running realSFS with default parameters. We then generated the FST index for each population pair with the realSFS fist index, supplying each population’s SAF index and the 2DSFS with default parameters. The resulting FST index file allowed us to estimate global FST as well as FST in sliding windows using realSFS fist stats and realSFS fist stats2, respectively.

**Estimation of Population Relationships.** To estimate the relationships among the sampled subspecies, we constructed a population-level tree using the BEAST2 v.2.6.0 application SNAPP, a multispecies coalescent-based tool that uses biallelic markers as input (83-85). Our input data consisted of genome-wide putatively neutral variants sampled in both the sequence-capture and whole-genome datasets (i.e., excluding the Agouti and Mctr regions). Briefly, we chose the two highest-coverge individuals representing each population, then retained biallelic SNPs with minor allele frequency (MAF) greater than 0.05, excluded variants that violated Hardy-Weinberg equilibrium (P value < 0.001) in four or more populations, and thinned the remaining variants so that none were within 100 bp of each other. The remaining variants were realligned as a phylip file and converted to the xml format required by SNAPP/BEAST2 using the script snapp_prep.rb (https://raw.githubusercontent.com/mmatschiner/snapp_prep/master/snapp_prep.rb). To specify a starting tree constraint (-s), we ran RAxML v.8.2.12 (86) with ascertainment bias correction (-a) and the 2DSFS with default parameters. The resulting RAxML tree file allowed us to estimate global FWT as well as FWT in sliding windows using realSFS fist stats and realSFS fist stats2, respectively.
that the crown divergence of all subgroups, excluding *P. maniculatus nubireae* (outgroup), should approximate a normal distribution with a mean of 8.9 kya and an SD of 1.5 kya. These values were taken from SMC++, estimates of the divergence time between mainland (*P. p. subgriseus*) and beach (*P. p. leucocephalus*) subgroups, assuming a generation time of 4 mo (i.e., three generations per year) (Demographic Inference). Finally, we sampled 1,000 random variants from the remaining dataset to speed up run times and specified 1 million Mar-

**Demographic inference.** The whole-genome, high-density sequencing cover-
age for one mainland (*P. p. subgriseus*) and one beach (*P. p. leucocephalus*) subgroup allowed us to infer demographic histories with high resolution. Spe-
cifically, we used the program SMC++, v1.15.4.dev3+gb53a36d.d20200521 (89) to estimate population divergence times and parameterize population size changes in additional populations. To mask low-quality regions, we followed the SNPable protocol (http://lh3lh3.users.sourceforge.net/snpable.shtml) to identify regions in the assembly with poor mapability using a k-mer size of 150 bp. SNPs that violated Hardy-Weinberg equilibrium (*P < 0.01*) and that had low population coverage (<80% samples genotyped) were also excluded.

We then used the vcft2smc command to create the per-population SMC++
input files, supplying mapability, missingness, and Hardy-Weinberg masks to
exclude low-quality regions in the dataset. The “distinguished individual” (DI),
a key feature of SMC++, was specified as the highest-coverage sample for each population. We generated two-population input files using the same command and input files but with no specified DI (not applicable to multipopulation analy-
sis). For single-population inference, we used cv with the following parameters:
“-folds 4 -timepoints 1e3 5e7 -Nmax 1e8 -Nmax 1e8” and a germline mutation
rate of 5.3e-9 (90). We also ran estimate, an earlier version of smc++ cv, with identical parameters and the UCSC mm39toMm10 chain
coordinates in mm39 to mm10 assembly coordinates using liftOver with default
parameters and the UCSC mm39toMm10 chain file (https://hgdownload.soe.
ucsc.edu/goldenPath/mm39/liftOver/mm39toMm10.over.chain.gz); then, we
converted mm10 coordinates to *P. polonontus* coordinates using the same approach
described above.

**Genome-Wide Association Mapping.** Genotype-phenotype associations were
determined using the mixed-model approach implemented in EMMAX v.beta-
07Mar2010, accounting for population structure/relatedness by incorporating a
Balding-Nichols kinship matrix as a random effect (91). We set the statistical sig-
nificance threshold at *P = 0.05* after correcting (Bonferroni method) for the
number of effective independent tests obtained with Genetic Type I error calcula-
tor v0.2 (92). After excluding samples with more than 50% missing genotypes
from these analyses, we were left with *n = 152* samples. We used both biallelic
SNPs and INDELS for association mapping but excluded markers with >50%
missing data, with an MAF of <0.05, or deviating from Hardy-Weinberg equilib-
rium (*P < 0.001*). We generated Manhattan plots and Quantile-Quantile (QQ)
plots using the qqman v.0.1.4 (93) and snpStats v.1.32.0 (94) R libraries, respec-
tively. Using plink v1.09b6.15, we calculated pairwise linkage disequilibrium
(r2) among SNPs in the focal region (lags: -chr chr4 -from-bp 8920301 -to-bp
9870301 -2 -ld-window 0 -0 -ld-window 1000). Next, we estimated the propor-
tion of variance explained for a given SNP (assuming Hardy-Weinberg equilib-
rium) using genotype-phenotype regressions.

**Sequence Conservation.** To evaluate the nucleotide sequence conservation level of the *Agouti* locus in *P. polonontus*, including the candidate regulatory region, we downloaded all available orthologous rodent *Agouti* sequences from National Center for Biotechnology Information (NCBI; accessed 8 September
2020) using esearch (db gene query “ortholog_gene_4344group” AND
rodent[orgn]”) in combination with esummary and extract from EDBase v1.38.
Next, we manually added 15 kb to each of the start and end coordinates (or the maximum number of base pairs if hitting a scaffold end) using a custom awk
script and retrieved the corresponding nucleotide sequences with efetch. The
sequence of *Nannospalax galili* was removed due to a lack of available flanking
sequence. Finally, we determined sequence conservation between *P. polonontus*
and the remaining 26 rodent species using mVISTA (accessed 8 September
2020 (95)) and phyloP v.1.4.96. For phyloP, we provided a phylogeny of the
27 species based on data from TimeTree (accessed 15 June 2021 (97)).

**Regulatory Database Queries.** To determine if the candidate regulatory region of
*Agouti* contains any known regulatory elements or TF binding sites, we down-
loaded both phastCons60way conserved elements and ORFAnno regulatory
elements from the University of California Santa Cruz (UCSC) genome browser in
*Mus musculus* mm10 coordinates (http://hgdownload.cse.ucsc.edu/;download+ of
mm10/database/). Elements from each database were converted to *P. polonontus*
genomic coordinates using UCSC’s liftOver v.358 (65) and a custom chain file,
with the parameters “multiple -minMatch = 0.70.”

We obtained ENSEMBL regulatory features using the R package biomaRT
v.2.38.0 (98, 99). The mm39 regulatory feature dataset was retrieved with the
function useDataset(), with the parameters “dataset=mmusculus_regulatory_-
feature,” marts=”ENSEMBL_MART_FUNCGEN,” and getBM() used to retrieve
entries from the broader *Agouti* region using the extended *Agouti* coordinates
((2:154785921:155055915) for the mm39 assembly. We directly converted
coordinates in mm39 to mm10 assembly coordinates using liftOver with default
parameters and the UCSC mm39toMm10 chain file (https://hgdownload.soe.
ucsc.edu/goldenPath/mm39/liftOver/mm39toMm10.over.chain.gz); then, we
converted mm10 coordinates to *P. polonontus* coordinates using the same approach
described above.

**LasZ Reporter Assay.** To determine if the candidate region was capable of reg-
ulatory activity, we assessed whether it could drive expression of the lacZ reporter
gene in the skin of developing mouse embryos (strain FVB/NJ). To identify the
most appropriate sequence length for this experiment, we specified boundaries
that encompassed the three pPC1-significant SNPs, the unique local topology
regions identified by Saguaro (*Local Tree Inference with Saguaro*), and the tract
of relatively high sequence conservation at the 3’ end of the association and
Saguaro regions, resulting in a total sequence length of 2.6 kb (Fig. 48). While
lacZ experiments are particularly useful for verifying that a regulatory locus is
active, comparisons between alleles of the same locus (e.g., light and dark
alleles) can be challenging due to the noise associated with random genomic
integration of the construct. Therefore, presented with two alternative haplotypes
in this region—”light” and “dark”—we decided to use the light haplotype for these
experiments under the assumption that the light allele was less likely to contain
mutations reducing element activity (i.e., *high Agouti* expression is generally
associated with light pigmentation).

We used the lacZ expression vector hsp68lacZ (a gift from T. Capellini, Har-
vard University, Cambridge, MA; Addgene no. 37843). The light haplotype
sequence file and hsp68lacZ vector were provided to Taconic Biosciences, which
synthesized and cloned the sequence upstream of the hsp68 minimal promoter
followed by pronuclear microinjection, collection of E14.5 embryos, genotyping,
and lacZ staining. Stained embryos were photographed, embedded in Optimal
Cutting Temperature compound (OCT), cryosectioned, and imaged in house.

**TF Binding Site Prediction.** To determine if variation in the regulatory element
could be modifying relevant TF binding sites, we examined motif differences
at variant positions across the region. Specifically, we obtained all polymorphic sites
in the regulatory element (chr4: 9,844,852 to 9,847,500 bp) with MAF > 0.05 in
the *P. p. albinus* population. We extracted the region 15 bp upstream and downstream
of each variant (~30-bp sequence) and used vcf-consensus (vcftools)
 ToString database of 251 *Mus musculus* CORE TF position weight matrices
available on JASPAR (downloaded 20 October 2021 (101)).

**Haplotype Homozygosity Tests.** To test for evidence of nonneutral evolution
in patterns of nucleotide variation, we calculated haplotype statistics. We first ran
fastPHASE v1.4.8 to create phased variant calls (102). We converted the input vcf
for all individual genotypes at *Agouti*, McT1, and the sequence-capture loci to
the fastPHASE format with vcf2fastPHASE.pl (https://github.com/lstevison/vcf-
conversion-tools); then, we ran fastPHASE with the following parameters: -T20
-H50 F. We next conversion the phased output back to the vcf format with fast-
PHASE2VCF.pl.
Using the R package reh4 v.3.1.0, we ran a series of haplotype-based tests to
scan for signatures of positive selection on the light and dark haplotypes (103).
For each population, we converted vcf’s to an reh-compatible file with a custom
script (hap2reh.py) using the P. polonius subgriseus reference genome to
polarize alleles. We then converted these files to haplobjects with data2hphlo-
pholm and computed extended haplotype homozygosity (EHH) statistics with
scan_hh, with the parameters “discordiation integration at border = FALSE, maxgap = 2000’
to accommodate the sequence-capture dataset. We then ran ihh2ihs to
calculate HIs using an MAF filter of 0.05 and default allele frequency bin sizes of
0.025.

Local Tree Inference with Saguarow. As a complementary approach to test for
evidence of selection within Agouti, we used the hidden Markov model
(HMM)-based software Saguarow v0.1 to build local phylogenies from sequence
data (33). As input, we used variant calls from the sequence-capture dataset and
the Agouti and Mct extended loci and filtered out variants with MAF < 0.025.
To reduce computational complexity and help with downstream interpretation,
we reduced the sample size to include only the two highest-coverage representa-
tives of each population. In the case of albinos, we included two individuals
homozygous for the dark allele and two for the light allele (as determined by
their genotype at SNP ch4: 9,845,301 bp). We then used VCF2HMMFeature to
transform the variant calls to a Saguarow-compatible input format. We ran
Saguarow for 15 iterations with default parameters. We transformed the resulting
topologies to phylip files with Saguarow2Phyip and obtained HMM transitions from the
LocalTrees.out file.

Sequence Diversity of the Agouti Regulatory Allele. To characterize
sequence diversity in the Agouti regulatory region, we calculated Fst_DN, DND, and
θH using scikit-allel v.1.3.2 (104). Specifically, we classified albinos individuals
by genotype at the top associated Agouti SNP (ch4: 9,845,301): light/light (n =
107), light/dark (n = 37), and dark/dark (n = 2). In 100-bp nonoverlapping
windows, we calculated Hudson’s Fst_DN and θH using the functions all.indowed_hudson_fst(),
all.indowed_divergence(), and all.indowed_divers-
ity(), respectively. Given the sparse nature of the sequence-capture dataset, we
restricted these calculations to 100-bp windows where at least 70% of albinos
individuals had five or more mapped reads. We then used these data to deter-
mine 95, 99, and 99.5% genomic outliers for DND and Fst_DN.

Data Availability. The P. polonius subgriseus reference genome is
archived in the Whole Genome Sequences BioProject Database (accession
number PRJNA449229) (105). Whole-genome and target-capture sequence
data are archived in the NCBI Sequence Read Archive BioProject Database (accession
number PRJNA838959) (106). Scripts and phenotypic data are available in GitHub
(https://github.com/twooldridge/Agouti_enhancer_paper) (107). Genome
assembly data have been deposited in NCBI (accession no. GCA_003704135.2
(108). All other data are included in the article and/or SI Appendix.

Acknowledgments. We thank M. Omura of the Hanard MCZ for assistance in
preparing and accessioning voucher specimens, N. Bedford for collecting
P. p. leucopus specimens for WGS, E. Kingsley for providing P. m. nubiterae
samples, and J. Weber and E. Delaney for assistance in the field. C. Hu designed
the illustration in Fig. 2A, and T. Capellini provided advice on the reporter assays
and feedback on a draft of this manuscript together with S. He, C. Kratochwil,
and M. Manceau. A.F.K. was supported by European Molecular Biology Organiza-
tion (EMBO) Postdoctoral Fellowship ALTF 47-2018 and German Science Founda-
tion Postdoctoral Fellowship KS 3081/1-1. J.-M.L. was supported by EMBO Grant
ALTF 379-2011, Human Frontiers Science Program Grant L0001086/2012,
and the Belgian American Educational Foundation. V.S.D. was supported by the Por-
tuguese Foundation for Science and Technology. Fieldwork was supported by an
MCZ Putnam Grant (to V.S.D. and H.E.H.), and laboratory work was supported by
NIH Grant R35GM133758 (to R.M.) and NSF Grant DEB-0919190 (to H.E.H.).
H.E.H. is an investigator of the Howard Hughes Medical Institute.
Supplementary Information for

An enhancer of Agouti contributes to parallel evolution of cryptically colored beach mice

T. Brock Wooldridge, Andreas F. Kautt, Jean-Marc Lassance, Sade McFadden, Vera S. Domingues, Ricardo Mallarino, Hopi E. Hoekstra

Correspondence to: hoekstra@oeb.harvard.edu

This file includes:

- Tables S1 to S5
- Figures S1 to S7
- SI References
**Table S1. Sampling information.** List of species, subspecies, and populations included in this study. “Collector/Publications” points to information for the precise sampling method, time, and place. Primary collector initials are as follows: EPK = Evan Kingsley, NLB = Nicole Bedford, VSD = Vera Domingues. “Sequencing strategy” refers to either WGS = whole genome sequencing (highlighted in grey) or Seqcapture = targeted sequence capture array.

| Taxonomy                   | Collectors/Publications | Latitude     | Longitude     | No. samples | Sequencing strategy |
|----------------------------|-------------------------|--------------|---------------|-------------|---------------------|
| *P. polionotus ammobates*  | Mullen et al. 2009 (1)  | 30.229978    | -87.814703    | 15          | Seqcapture           |
| *P. polionotus allophrys*  | Mullen et al. 2009 (1)  | 30.077795    | -85.647814    | 11          | Seqcapture           |
| *P. polionotus leucocephalus* | NLB                  | 30.397536    | -86.729057    | 15          | WGS                 |
| *P. polionotus leucocephalus* | Mullen et al. 2009 (1)  | 30.397536    | -86.729057    | 20          | Seqcapture           |
| *P. maniculatus nubiterrae* | EPK                    | 40.33        | -79.27        | 1           | Seqcapture           |
| *P. polionotus trisyllepsis* | Mullen et al. 2009 (1)  | 30.29371     | -87.463557    | 5           | Seqcapture           |
| *P. polionotus subgriseus* | VSD                    | 29.182833    | -81.795       | 15          | WGS                 |
| *P. polionotus albigrons*  | VSD                    | 30.5411      | -86.075717    | 168         | Seqcapture           |
| *P. polionotus polionotus* | Domingues et al. 2012 (2) | 31.995717    | -85.082967    | 6           | Seqcapture           |
| *P. polionotus subgriseus (A)* | Domingues et al. 2012 (2) | 30.814577    | -84.954529    | 5           | Seqcapture           |
| *P. polionotus subgriseus (T)* | Domingues et al. 2012 (2) | 31.6459      | -84.225       | 5           | Seqcapture           |
| *P. polionotus subgriseus (O)* | Domingues et al. 2012 (2) | 29.207927    | -81.740378    | 5           | Seqcapture           |
| *P. polionotus peninsularis* | Mullen et al. 2009 (1)  | 29.957198    | -85.462412    | 19          | Seqcapture           |
| *P. polionotus niveiventris* | Steiner et al. 2009 (3) | 27.923493    | -80.488186    | 4           | Seqcapture           |
Table S2. Regulatory elements in *Agouti*. *P. polionotus*-based coordinates (HU_Ppol1.3.3) of regulatory features lifted over from the ENSEMBL *Mus musculus* genome (GRCm39). Features found ~10 kb upstream and downstream of the 2-kb candidate region are shown.

| P. polionotus chromosome | P. polionotus start (bp) | P. polionotus end (bp) | M. musculus chromosome | M. musculus start (bp) | M. musculus end (bp) | ENSEMBL ID | ENSEMBL element type |
|-------------------------|--------------------------|------------------------|------------------------|------------------------|------------------------|------------|--------------------|
| chr4                    | 9834661                  | 9834859                | chr2                   | 154872801              | 154873001              | ENSMUSR000000830170 | Predicted enhancer |
| chr4                    | 9842998                  | 9843607                | chr2                   | 154880275              | 154880693              | ENSMUSR000000636464 | Transcription factor binding |
| chr4                    | 9845488                  | 9845887                | chr2                   | 154881887              | 154882302              | ENSMUSR00000830171 | Open chromatin region |
| chr4                    | 9850467                  | 9851000                | chr2                   | 154887001              | 154887600              | ENSMUSR00000830172 | Predicted enhancer |
| chr4                    | 9852071                  | 9852243                | chr2                   | 154888801              | 154889000              | ENSMUSR00000830173 | CTCF Binding Site |
| chr4                    | 9855030                  | 9857063                | chr2                   | 154892000              | 154894001              | ENSMUSR00000636468 | Predicted promoter |
Table S3. Population-level frequency of light allele. Each species is provided with its source location. Frequency of the light allele is calculated as the proportion of haplotypes in a population that have the light-associated allele at chr4:9,845,301 bp.

| Species               | Location           | Frequency of light Agouti allele |
|-----------------------|--------------------|----------------------------------|
| P. p. allophrys       | Gulf Coast Beach   | 1                                |
| P. p. ammobates       | Gulf Coast Beach   | 1                                |
| P. p. leucocephalus   | Gulf Coast Beach   | 1                                |
| P. p. trisylepsis     | Gulf Coast Beach   | 1                                |
| P. p. peninsularis    | Gulf Coast Beach   | 1                                |
| P. p. niveiventris     | Atlantic Coast Beach | 1                          |
| P. m. nubiterra       | Mainland (Outgroup) | 0                              |
| P. p. albifrons       | Mainland           | 0.86                             |
| P. p. polionotus      | Mainland           | 0                                |
| P. p. subgriseus (A)  | Mainland           | 0                                |
| P. p. subgriseus (O)  | Mainland           | 1                                |
| P. p. subgriseus (T)  | Mainland           | 1                                |
Table S4. Transcription factor binding sites overlapping variant positions in the regulatory element of *P. p. albifrons*. “Position” includes only sites on chr4 in the tested regulatory element (9,844,852 bp – 9,847,500 bp) that are variant in *P. p. albifrons* with a minor allele frequency (MAF) > 0.05. Both “Reference allele” and “Alternate allele” are relative to the *P. p. subgriseus* reference genome (no relation to light or dark haplotypes). “Reference-specific TFs” and “Alternate-specific TFs” refer to predicted TF binding sites from JASPAR (4) (see Methods) that are intact in the Reference allele and Alternate allele, respectively. pPC1-associated SNPs are highlighted in gray.

| Position | Reference allele | Alternate allele | Alternate allele frequency | Reference-specific TFs | Alternate-specific TFs |
|----------|-----------------|-----------------|---------------------------|------------------------|------------------------|
| 9844852  | G               | A               | 0.0648148                 | Foxo3;Neurog1          | Arid5a                 |
| 9844869  | GC              | G               | 0.168712                  | Zfp335                 | Rhox11                 |
| 9844930  | G               | A               | 0.171975                  | --                     | Rhox11                 |
| 9844954  | A               | G               | 0.159091                  | --                     | Mitf                   |
| 9844955  | G               | GTC             | 0.155844                  | --                     | Mitf                   |
| 9845040  | T               | A               | 0.154762                  | --                     | Nkx3-2                 |
| 9845116  | C               | T               | 0.140741                  | Dlx5                   | Nkx2-5                 |
| 9845136  | G               | GA              | 0.142336                  | FoXn1                  | --                     |
| 9845152  | G               | A               | 0.150735                  | Armt::Hif1a;Hes1       | Pax2                   |
| 9845201  | C               | T               | 0.140411                  | --                     | Nkx3-2                 |
| 9845470  | C               | T               | 0.127049                  | --                     | --                     |
| 9846196  | C               | T               | 0.208333                  | Gfl1b;Klf4             | Barhl1;Bcl11b         |
| 9846254  | G               | A               | 0.0620915                 | --                     | Atoh1;Bhlha15;Msgn1;Myc;Npas2;Twist2 |
| 9846286  | A               | G               | 0.123333                  | Atoh1;Bhlha15;Bhlhe40;Fos::Jun;Foxj2;Foxo1;Foxo3;Mitf;Mycnpas2;Pax2;Twist2 Atoh1;Bhlha15;Bhlhe40;Fos::Jun;Foxj2;Foxo1;Foxo3;MitfMycnpas2;Pax2;Twist2 Atoh1;Bhlha15;Bhlhe40;Fos::Jun;Foxj2;Foxo1;Foxo3;MitfMycnpas2;Pax2;Twist2 -- |
| 9846292  | T               | C               | 0.125                     | Dlx5                   | Nkx2-5                 |
| 9846293  | G               | A               | 0.121622                  | --                     | Nkx2-5                 |
| 9846308  | T               | C               | 0.127586                  | Atoh1                  | Hes2                   |
| 9846815  | G               | A               | 0.143357                  | --                     | Nkx2-5                 |
| 9846908  | A               | G               | 0.171975                  | Cebp;Mafb             | Elf5;Hand1::Tcf3;Hic1  |
| 9846919  | C               | T               | 0.0796178                 | --                     | --                     |
| 9846954  | T               | G               | 0.102564                  | --                     | --                     |
| 9846977  | C               | T               | 0.115385                  | --                     | Barhl1;Neurog1;Nkx2-5 |
| 9847023  | T               | G               | 0.153846                  | Elf5;Gata1            | Rfx6                   |
| 9847042  | T               | G               | 0.730263                  | --                     | Hand1::Tcf3            |
| 9847150  | A               | G               | 0.788591                  | Barhl1;Nkx2-5         | Myb;Stat6;Tcf3         |
Table S5. Museum accession information for all samples included in study. “Sample ID” refers to ID used by authors in this study, while “Collection ID” refers to the ID used by the institution where samples are accessioned. “Collection” refers to either MCZ = Harvard Museum of Comparative Zoology, or Lab = internal Hoekstra Lab specimen collection (samples available upon request).

| Sample ID | Species       | Collection | Collection ID |
|-----------|---------------|------------|--------------|
| 01_NB_F_EPK04 | P. m. nubiterrae | not accessioned | NA           |
| VSD142    | P. p. albifrons | MCZ        | 68140        |
| VSD143    | P. p. albifrons | MCZ        | 68141        |
| VSD195    | P. p. albifrons | MCZ        | 68192        |
| VSD196    | P. p. albifrons | MCZ        | 68193        |
| VSD197    | P. p. albifrons | MCZ        | 68194        |
| VSD198    | P. p. albifrons | MCZ        | 68195        |
| VSD199    | P. p. albifrons | MCZ        | 68196        |
| VSD200    | P. p. albifrons | MCZ        | 68197        |
| VSD201    | P. p. albifrons | MCZ        | 68198        |
| VSD202    | P. p. albifrons | MCZ        | 68199        |
| VSD203    | P. p. albifrons | MCZ        | 68200        |
| VSD204    | P. p. albifrons | MCZ        | 68201        |
| VSD205    | P. p. albifrons | MCZ        | 68202        |
| VSD207    | P. p. albifrons | MCZ        | 68204        |
| VSD208    | P. p. albifrons | MCZ        | 68205        |
| VSD209    | P. p. albifrons | MCZ        | 68206        |
| VSD210    | P. p. albifrons | MCZ        | 68207        |
| VSD211    | P. p. albifrons | MCZ        | 68208        |
| VSD212    | P. p. albifrons | MCZ        | 68209        |
| VSD214    | P. p. albifrons | MCZ        | 68211        |
| VSD215    | P. p. albifrons | MCZ        | 68212        |
| VSD216    | P. p. albifrons | MCZ        | 68213        |
| VSD217    | P. p. albifrons | MCZ        | 68214        |
| VSD218    | P. p. albifrons | MCZ        | 68215        |
| VSD219    | P. p. albifrons | MCZ        | 68216        |
| VSD220    | P. p. albifrons | MCZ        | 68217        |
| VSD221    | P. p. albifrons | MCZ        | 68218        |
| VSD222    | P. p. albifrons | MCZ        | 68219        |
| VSD223    | P. p. albifrons | MCZ        | 68220        |
| VSD224    | P. p. albifrons | MCZ        | 68221        |
| VSD225    | P. p. albifrons | MCZ        | 68222        |
| VSD226    | P. p. albifrons | MCZ        | 68223        |
| VSD227    | P. p. albifrons | MCZ        | 68224        |
| VSD228    | P. p. albifrons | MCZ        | 68225        |
| VSD229    | P. p. albifrons | MCZ        | 68226        |
| VSD230    | P. p. albifrons | MCZ        | 68227        |
| VSD231    | P. p. albifrons | MCZ        | 68228        |
| VSD233    | P. p. albifrons | MCZ        | 68230        |
| VSD234    | P. p. albifrons | MCZ        | 68231        |
| VSD235    | P. p. albifrons | MCZ        | 68232        |
| VSD236    | P. p. albifrons | MCZ        | 68233        |
| VSD237    | P. p. albifrons | MCZ        | 68234        |
| VSD238    | P. p. albifrons | MCZ        | 68235        |
| VSD239    | P. p. albifrons | MCZ        | 68236        |
| VSD240    | P. p. albifrons | MCZ        | 68237        |
| VSD241    | P. p. albifrons | MCZ        | 68238        |
| VSD242    | P. p. albifrons | MCZ        | 68239        |
| VSD243    | P. p. albifrons | MCZ        | 68240        |
| VSD244    | P. p. albifrons | MCZ        | 68241        |
| VSD245    | P. p. albifrons | MCZ        | 68242        |
| VSD246    | P. p. albifrons | MCZ        | 68243        |
| VSD247    | P. p. albifrons | MCZ        | 68244        |
| VSD248    | P. p. albifrons | MCZ        | 68245        |
| VSD249    | P. p. albifrons | MCZ        | 68246        |
| VSD250    | P. p. albifrons | MCZ        | 68247        |
| VSD251    | P. p. albifrons | MCZ        | 68248        |
| VSD252    | P. p. albifrons | MCZ        | 68249        |
| VSD253    | P. p. albifrons | MCZ        | 68250        |
| VSD254    | P. p. albifrons | MCZ        | 68251        |
| VSD256    | P. p. albifrons | MCZ        | 68253        |
| VSD257 | P. p. albifrons | MCZ | 68254 |
| VSD259 | P. p. albifrons | MCZ | 68256 |
| VSD260 | P. p. albifrons | MCZ | 68257 |
| VSD261 | P. p. albifrons | MCZ | 68258 |
| VSD262 | P. p. albifrons | MCZ | 68259 |
| VSD263 | P. p. albifrons | MCZ | 68260 |
| VSD264 | P. p. albifrons | MCZ | 68261 |
| VSD265 | P. p. albifrons | MCZ | 68262 |
| VSD266 | P. p. albifrons | MCZ | 68263 |
| VSD267 | P. p. albifrons | MCZ | 68264 |
| VSD268 | P. p. albifrons | MCZ | 68265 |
| VSD269 | P. p. albifrons | MCZ | 68266 |
| VSD270 | P. p. albifrons | MCZ | 68267 |
| VSD271 | P. p. albifrons | MCZ | 68268 |
| VSD272 | P. p. albifrons | MCZ | 68269 |
| VSD273 | P. p. albifrons | MCZ | 68270 |
| VSD274 | P. p. albifrons | MCZ | 68271 |
| VSD275 | P. p. albifrons | MCZ | 68272 |
| VSD276 | P. p. albifrons | MCZ | 68273 |
| VSD278 | P. p. albifrons | MCZ | 68275 |
| VSD279 | P. p. albifrons | MCZ | 68276 |
| VSD280 | P. p. albifrons | MCZ | 68277 |
| VSD281 | P. p. albifrons | MCZ | 68278 |
| VSD282 | P. p. albifrons | MCZ | 68279 |
| VSD283 | P. p. albifrons | MCZ | 68280 |
| VSD284 | P. p. albifrons | MCZ | 68281 |
| VSD285 | P. p. albifrons | MCZ | 68282 |
| VSD286 | P. p. albifrons | MCZ | 68283 |
| VSD287 | P. p. albifrons | MCZ | 68284 |
| VSD288 | P. p. albifrons | MCZ | 68285 |
| VSD289 | P. p. albifrons | MCZ | 68286 |
| VSD292 | P. p. albifrons | MCZ | 68289 |
| VSD293 | P. p. albifrons | MCZ | 68290 |
| VSD294 | P. p. albifrons | MCZ | 68291 |
| VSD295 | P. p. albifrons | MCZ | 68292 |
| VSD296 | P. p. albifrons | MCZ | 68293 |
| VSD298 | P. p. albifrons | MCZ | 68295 |
| VSD299 | P. p. albifrons | MCZ | 68296 |
| VSD307 | P. p. albifrons | MCZ | 68304 |
| VSD308 | P. p. albifrons | MCZ | 68305 |
| VSD309 | P. p. albifrons | MCZ | 68306 |
| VSD310 | P. p. albifrons | MCZ | 68307 |
| VSD311 | P. p. albifrons | MCZ | 68308 |
| VSD312 | P. p. albifrons | MCZ | 68309 |
| VSD313 | P. p. albifrons | MCZ | 68310 |
| VSD314 | P. p. albifrons | MCZ | 68311 |
| VSD315 | P. p. albifrons | MCZ | 68312 |
| VSD316 | P. p. albifrons | MCZ | 68313 |
| VSD317 | P. p. albifrons | MCZ | 68314 |
| VSD318 | P. p. albifrons | MCZ | 68315 |
| VSD319 | P. p. albifrons | MCZ | 68316 |
| VSD320 | P. p. albifrons | MCZ | 68317 |
| VSD321 | P. p. albifrons | MCZ | 68318 |
| VSD322 | P. p. albifrons | MCZ | 68319 |
| VSD323 | P. p. albifrons | MCZ | 68320 |
| VSD324 | P. p. albifrons | MCZ | 68321 |
| VSD331 | P. p. albifrons | MCZ | 68328 |
| VSD332 | P. p. albifrons | MCZ | 68329 |
| VSD333 | P. p. albifrons | MCZ | 68330 |
| VSD334 | P. p. albifrons | MCZ | 68331 |
| VSD335 | P. p. albifrons | MCZ | 68332 |
| VSD336 | P. p. albifrons | MCZ | 68333 |
| VSD337 | P. p. albifrons | MCZ | 68334 |
| VSD338 | P. p. albifrons | MCZ | 68335 |
| VSD339 | P. p. albifrons | MCZ | 68336 |
| VSD340 | P. p. albifrons | MCZ | 68337 |
| VSD341 | P. p. albifrons | MCZ | 68338 |
| VSD342 | P. p. albifrons | MCZ | 68339 |
| VSD343 | P. p. albifrons | MCZ | 68340 |
| VSD344 | P. p. albifrons | MCZ | 68341 |
| Code   | Species          | Museum | Catalogue |
|--------|------------------|--------|-----------|
| VSD345 | P. p. albifrons  | MCZ    | 68342     |
| VSD346 | P. p. albifrons  | MCZ    | 68343     |
| VSD347 | P. p. albifrons  | MCZ    | 68344     |
| VSD348 | P. p. albifrons  | MCZ    | 68345     |
| VSD349 | P. p. albifrons  | MCZ    | 68346     |
| VSD350 | P. p. albifrons  | MCZ    | 68347     |
| VSD351 | P. p. albifrons  | MCZ    | 68348     |
| VSD352 | P. p. albifrons  | MCZ    | 68349     |
| VSD353 | P. p. albifrons  | MCZ    | 68350     |
| VSD354 | P. p. albifrons  | MCZ    | 68351     |
| VSD355 | P. p. albifrons  | MCZ    | 68352     |
| VSD356 | P. p. albifrons  | MCZ    | 68353     |
| VSD357 | P. p. albifrons  | MCZ    | 68354     |
| VSD358 | P. p. albifrons  | MCZ    | 68355     |
| VSD63  | P. p. albifrons  | MCZ    | 68061     |
| VSD65  | P. p. albifrons  | MCZ    | 68063     |
| VSD66  | P. p. albifrons  | MCZ    | 68064     |
| VSD67  | P. p. albifrons  | MCZ    | 68065     |
| VSD69  | P. p. albifrons  | MCZ    | 68067     |
| VSD70  | P. p. albifrons  | MCZ    | 68068     |
| VSD71  | P. p. albifrons  | MCZ    | 68069     |
| VSD72  | P. p. albifrons  | MCZ    | 68070     |
| VSD73  | P. p. albifrons  | MCZ    | 68071     |
| VSD74  | P. p. albifrons  | MCZ    | 68072     |
| VSD75  | P. p. albifrons  | MCZ    | 68073     |
| VSD76  | P. p. albifrons  | MCZ    | 68074     |
| VSD77  | P. p. albifrons  | MCZ    | 68075     |
| VSD78  | P. p. albifrons  | MCZ    | 68076     |
| VSD79  | P. p. albifrons  | MCZ    | 68077     |
| VSD80  | P. p. albifrons  | MCZ    | 68078     |
| VSD81  | P. p. albifrons  | MCZ    | 68079     |
| VSD82  | P. p. albifrons  | MCZ    | 68080     |
| VSD83  | P. p. albifrons  | MCZ    | 68081     |
| VSD84  | P. p. albifrons  | MCZ    | 68082     |
| VSD85  | P. p. albifrons  | MCZ    | 68083     |
| VSD86  | P. p. albifrons  | MCZ    | 68084     |
| VSD87  | P. p. albifrons  | MCZ    | 68085     |
| VSD88  | P. p. albifrons  | MCZ    | 68086     |
| CBM2000| P. p. allophrys  | Lab     |           |
| CBM245 | P. p. allophrys  | Lab     |           |
| CBM259 | P. p. allophrys  | Lab     |           |
| CBM264 | P. p. allophrys  | Lab     |           |
| CBM310 | P. p. allophrys  | Lab     |           |
| CBM415 | P. p. allophrys  | Lab     |           |
| CBM6000| P. p. allophrys  | Lab     |           |
| CBM973 | P. p. allophrys  | Lab     |           |
| MCZ65946| P. p. allophrys| MCZ    | 65946     |
| MCZ65947| P. p. allophrys| MCZ    | 65947     |
| allophrys_MCZ65945| P. p. allophrys| MCZ    | 65945     |
| ABM101 | P. p. ammobates  | Lab     |           |
| ABM105 | P. p. ammobates  | Lab     |           |
| ABM2298| P. p. ammobates  | Lab     |           |
| ABM2618| P. p. ammobates  | Lab     |           |
| ABM2639| P. p. ammobates  | Lab     |           |
| ABM2766| P. p. ammobates  | Lab     |           |
| ABM3   | P. p. ammobates  | Lab     |           |
| ABM4240| P. p. ammobates  | Lab     |           |
| ABM4637| P. p. ammobates  | Lab     |           |
| ABM473 | P. p. ammobates  | Lab     |           |
| ABM5   | P. p. ammobates  | Lab     |           |
| ABM501 | P. p. ammobates  | Lab     |           |
| ABM6   | P. p. ammobates  | Lab     |           |
| MCZ65932| P. p. ammobates| MCZ    | 65932     |
| ammobates_MCZ65930| P. p. ammobates| MCZ    | 65930     |
| 4434   | P. p. leucocephalus| MCZ        | 4434    |
| 69677  | P. p. leucocephalus| MCZ        | 69677   |
| 69679  | P. p. leucocephalus| MCZ        | 69679   |
| 69686  | P. p. leucocephalus| MCZ        | 69686   |
| 7817   | P. p. leucocephalus| MCZ        | 7817    |
| 7819   | P. p. leucocephalus| MCZ        | 7819    |
|    | Species          | Location | Number |
|----|------------------|----------|--------|
| 7820 | P. p. leucocephalus | MCZ      | 7820   |
| 7824 | P. p. leucocephalus | MCZ      | 7824   |
| 7827 | P. p. leucocephalus | MCZ      | 7827   |
| 7828 | P. p. leucocephalus | MCZ      | 7828   |
| 7830 | P. p. leucocephalus | MCZ      | 7830   |
| 7831 | P. p. leucocephalus | MCZ      | 7831   |
| 7832 | P. p. leucocephalus | MCZ      | 7832   |
| 7835 | P. p. leucocephalus | MCZ      | 7835   |
| 7838 | P. p. leucocephalus | MCZ      | 7838   |
| SRIBM110 | P. p. leucocephalus | Lab      |        |
| SRIBM1400 | P. p. leucocephalus | Lab      |        |
| SRIBM160 | P. p. leucocephalus | Lab      |        |
| SRIBM222 | P. p. leucocephalus | Lab      |        |
| SRIBM333 | P. p. leucocephalus | Lab      |        |
| SRIBM334 | P. p. leucocephalus | Lab      |        |
| SRIBM335 | P. p. leucocephalus | Lab      |        |
| SRIBM336 | P. p. leucocephalus | Lab      |        |
| SRIBM337 | P. p. leucocephalus | Lab      |        |
| SRIBM338 | P. p. leucocephalus | Lab      |        |
| SRIBM339 | P. p. leucocephalus | Lab      |        |
| SRIBM430 | P. p. leucocephalus | Lab      |        |
| SRIBM440 | P. p. leucocephalus | Lab      |        |
| SRIBM520 | P. p. leucocephalus | Lab      |        |
| SRIBM531 | P. p. leucocephalus | Lab      |        |
| SRIBM600 | P. p. leucocephalus | Lab      |        |
| SRIBM688 | P. p. leucocephalus | Lab      |        |
| SRIBM734 | P. p. leucocephalus | Lab      |        |
| SRIBM822 | P. p. leucocephalus | Lab      |        |
| SRIBM920 | P. p. leucocephalus | Lab      |        |
| SRIBM66104 | P. p. niveiventris | MCZ 66104 |        |
| SRIBM66105 | P. p. niveiventris | MCZ 66105 |        |
| SRIBM66107 | P. p. niveiventris | MCZ 66107 |        |
| MCZ66106 | P. p. niveiventris | MCZ 66106 |        |
| SABM11 | P. p. peninsularis | Lab      |        |
| SABM12 | P. p. peninsularis | Lab      |        |
| SABM137 | P. p. peninsularis | Lab      |        |
| SABM142 | P. p. peninsularis | Lab      |        |
| SABM145 | P. p. peninsularis | Lab      |        |
| SABM146 | P. p. peninsularis | Lab      |        |
| SABM159 | P. p. peninsularis | Lab      |        |
| SABM175 | P. p. peninsularis | Lab      |        |
| SABM20 | P. p. peninsularis | Lab      |        |
| SABM21 | P. p. peninsularis | Lab      |        |
| SABM22 | P. p. peninsularis | Lab      |        |
| SABM24 | P. p. peninsularis | Lab      |        |
| SABM248 | P. p. peninsularis | Lab      |        |
| SABM5  | P. p. peninsularis | Lab      |        |
| SABM174 | P. p. peninsularis | Lab      |        |
| SABM_ECI154 | P. p. peninsularis | Lab 65948 |        |
| sabm247 | P. p. peninsularis | Lab      |        |
| sabm3  | P. p. peninsularis | Lab      |        |
| JNW49 | P. p. polionotus | MCZ 64653 |        |
| JNW50 | P. p. polionotus | MCZ 64654 |        |
| VSD176 | P. p. polionotus | MCZ 68174 |        |
| VSD182 | P. p. polionotus | MCZ 68180 |        |
| VSD186 | P. p. polionotus | MCZ 68184 |        |
| VSD188 | P. p. polionotus | MCZ 68186 |        |
| VSD2  | P. p. subgriseus (A) | MCZ 68002 |        |
| VSD4  | P. p. subgriseus (A) | MCZ 68004 |        |
| VSD5  | P. p. subgriseus (A) | MCZ 68005 |        |
| VSD7  | P. p. subgriseus (A) | MCZ 68007 |        |
| VSD9  | P. p. subgriseus (A) | MCZ 68009 |        |
| 68144 | P. p. subgriseus (O) | MCZ 68144 |        |
| 68146 | P. p. subgriseus (O) | MCZ 68146 |        |
| 68147 | P. p. subgriseus (O) | MCZ 68147 |        |
| 68148 | P. p. subgriseus (O) | MCZ 68148 |        |
| 68149 | P. p. subgriseus (O) | MCZ 68149 |        |
| 68150 | P. p. subgriseus (O) | MCZ 68150 |        |
| 68151 | P. p. subgriseus (O) | MCZ 68151 |        |
|   | P. p. subgriseus (O) |   |   |
|---|---------------------|---|---|
| 68152 |   | MCZ | 68152 |
| 68153 |   | MCZ | 68153 |
| 68154 |   | MCZ | 68154 |
| 68159 |   | MCZ | 68159 |
| 68160 |   | MCZ | 68160 |
| 68161 |   | MCZ | 68161 |
| 68162 |   | MCZ | 68162 |
| 68163 |   | MCZ | 68163 |
| VSD148 |   | MCZ | 68146 |
| VSD150 |   | MCZ | 68148 |
| VSD154 |   | MCZ | 68152 |
| VSD156 |   | MCZ | 68154 |
| VSD163 |   | MCZ | 68161 |
| VSD118 |   | MCZ | 68116 |
| VSD120 |   | MCZ | 68118 |
| VSD122 |   | MCZ | 68120 |
| VSD123 |   | MCZ | 68121 |
| VSD127 |   | MCZ | 68125 |
| PKBM104 | P. p. trisyllepsis | Lab |
| PKBM1042 | P. p. trisyllepsis | Lab |
| PKBM1077 | P. p. trisyllepsis | Lab |
| PKBM1112 | P. p. trisyllepsis | Lab |
| PKBM145 | P. p. trisyllepsis | Lab |
Figure S1. Pigment trait loadings on phenotypic Principal Component Analysis (pPCA). A. pPCA biplot shows contributions of pigment traits to the first two phenotypic PCs. Percentage values in parentheses correspond to the percent variance explained by each PC. B. Quantile-quantile plot of empirical vs. expected GWAS p-value distributions, indicating no signs of overdispersion or abnormal behavior.
Figure S2. Genetic principal components analysis of the *P. p. albifrons* population. Each dot represents an individual (N=168). Color approximates phenotypic PC1 value. Percentage value on each axis corresponds to the percent variance explained by each genetic PC.
Figure S3. Manhattan plots showing association between phenotype and variation across the *Agouti* locus. Data for phenotypic PCs (pPCs) 1 to 5 are shown. Dashed red lines indicate genome-wide significant threshold, corrected for number of independent tests (see Methods). Gray bar denotes boundaries of peak association found for pPC1. No other pPCs show a significant association with variants in *Agouti*. 
Figure S4. Location of the identified regulatory region in relation to previously implicated regions. The top-associated SNP (chr4: 9,845,301) is shown in red. Vertical bars indicate regions of Agouti that are significantly associated with pigmentation traits in P. maniculatus, from (5).
Figure S5. **Sequence conservation in Agouti among 27 rodent species.** 10-kb region encompassing the SNP with the strongest association to pPC1 (chr4:9,845,301) denoted in gold. Conserved regions are shown in pink. ‘Focal region cacti’ (light blue) indicate the regions identified by Saguaro (6) (see Methods; Fig. 5) with a unique topology relative to the rest of the Agouti locus. The 2.6-kb region used in the lacZ reporter assay (grey) includes the cacti region (light blue), the top associated SNP (gold) and a conserved region (pink). One non-coding exon, 1E, is shown as a landmark (dark blue).
Figure S6. Transgenic embryos with visible skin \textit{lacZ} expression produced by pronuclear injection of the candidate region-\textit{lacZ} vector. All embryos are at stage E14.5, PCR-positive for the \textit{lacZ} vector, and each represents an independent genomic integration event. Blue staining shows \textit{lacZ} expression and regulatory element activity.
Figure S7. Sequence divergence among *Agouti* regulatory allele genotypes in the *P. p. albinorons* population. Left panels show comparisons between *albifrons* individuals homozygous for the light *Agouti* allele (light/light; n = 107) and heterozygous light/dark (n = 37) at the top associated SNP (chr4:9,845,301). Right panels compare light/light (n = 107) and dark/dark (n = 2) genotypes. Despite unbalanced sample sizes, signals of elevated $F_{ST}$ and $D_{XY}$ at the focal regulatory element (denoted by vertical red dotted lines corresponding to the 3 SNPs associated with phenotypic PC1) are seen in both comparisons. All points correspond to 100-bp non-overlapping windows. A. Hudson’s $F_{ST}$. B. $D_{XY}$. C. $\theta_1$, split by genotype group.
References

1. Mullen LM, Vignieri SN, Gore JA, Hoekstra HE. 2009. Adaptive basis of geographic variation: genetic, phenotypic and environmental differences among beach mouse populations. Proc Royal Soc B 276:3809–3818.
2. Domingues VS, Poh Y, Peterson BK, Pennings PS, Jensen JD, Hoekstra HE. 2012. Evidence of adaptation from ancestral variation in young populations of beach mice. Evolution 66:3209–3223.
3. Steiner CC, Rompler H, Boettger LM, Schoneberg T, Hoekstra HE. 2009. The genetic basis of phenotypic convergence in beach mice: similar pigment patterns but different genes. Mol Biol Evol 26:35–45.
4. Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, Modi BP, Correard S, Gheorghe M, Baranašić D, Santana-Garcia W, Tan G, Chèneby J, Ballester B, Parcy F, Sandelin A, Lenhard B, Wasserman WW, Mathelier A. 2020. JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res 48:D87–D92.
5. Linnen CR, Poh Y-P, Peterson BK, Barrett RDH, Larson JG, Jensen JD, Hoekstra HE. 2013. Adaptive evolution of multiple traits through multiple mutations at a single gene. Science 339:1312–1316.
6. Zamani N, Russell P, Lantz H, Hoeppner MP, Meadows JR, Vijay N, Mauceli E, di Palma F, Lindblad-Toh K, Jern P, Grabherr MG. 2013. Unsupervised genome-wide recognition of local relationship patterns. BMC Genomics 14:347.