Review

*Peromyscus* transcriptomics: Understanding adaptation and gene expression plasticity within and between species of deer mice

Jason Munshi-South\(^a\)\(^*,b\), Jonathan L. Richardson\(^b\)

\(^a\) Louis Calder Center—Biological Field Station, Fordham University, 31 Whippoorwill Road, Armonk, NY 10504, USA
\(^b\) Providence College, 1 Cunningham Square, Providence, RI 02918, USA

**ABSTRACT**

Deer mice in the genus *Peromyscus* occupy nearly every terrestrial habitat in North America, and have a long history as subjects of behavioral, ecological, evolutionary, and physiological study. Recent advances in transcriptomics, the study of the complete set of RNA transcripts produced by certain cell types or under certain conditions, have contributed to the development of *Peromyscus* as a model system. We review the recent use of transcriptomics to investigate how natural selection and gene expression plasticity contribute to the existence of deer mice in challenging environments such as highlands, deserts, and cities across North America. Transcriptomics also holds great promise for elucidating the genetic basis of mating systems and other behaviors in *Peromyscus*, but has to date been underutilized for developmental biology and disease studies. Future *Peromyscus* studies should apply robust comparative frameworks to analyze the transcriptomics of multiple populations of the same species across varying environmental conditions, as well as multiple species that vary in traits of interest.

© 2016 Published by Elsevier Ltd.

1. Introduction

The ability to identify the specific genes underlying phenotypic variation has long been the holy grail of studies investigating adaptive evolution [1,2]. While phenotypes are the grist through which evolution occurs via natural selection, variation in nucleotide sequence or expression levels of the relevant genes provides an alternative, and often more sensitive, way of detecting a response to environmental conditions [3]. This is particularly true in non-model species that may not have high-quality reference genomes and may be difficult to raise in laboratory settings to F1 or F2 generations [4,5].

With the recent proliferation of high-throughput sequencing, the ability to sequence transcribed RNA (i.e. transcriptomes) has made it feasible for scientists to measure the relative expression levels of different coding regions of the genome. Such approaches are particularly useful when studying species that do not have sequenced genomes or annotated transcriptomes where gene function is well characterized [6]. RNA sequencing (RNA-Seq) has been used effectively to identify gene expression differences between individuals from different natural environments and individuals exposed to different experimental conditions (Fig. 1). RNA-Seq is now widely recognized as a powerful tool allowing researchers to expand transcriptomics beyond clinical settings and a handful of model species [3].

*Peromyscus* (aka deer mice) are quickly becoming models for studies on evolutionary ecology, physiology and behavior [7]. *Peromyscus* have been the focus of ecological, evolutionary and epidemiological research for more than a century, and there is now abundant information on the natural history of this taxon [8]. In addition, several *Peromyscus* species have continental-wide ranges across North and Central America. These species occur across a wide range of habitat types, and a correspondingly diverse array of natural selection regimes that would promote adaptive divergence between populations experiencing different environmental conditions. In addition to their wild counterparts, deer mice have also been amenable to laboratory rearing and experimentation. As a result, abundant information exists on the basic biology of lab-raised *Peromyscus* [9,10].

Below we detail the important role that RNA sequencing and transcriptomics have played in developing *Peromyscus* as a new model system in biology. We highlight seminal studies on the genes underpinning *Peromyscus* adaptation, including adaptation to high-elevation hypoxia [11], osmotic balance in deserts [12], and tolerance to urbanization [13]. We also detail the design of experimental transcriptomic studies that can be used to evaluate the genes underpinning divergent phenotypes and evolution in *Peromyscus*.

2. Adaptive evolution and gene expression in challenging environments

One or more species of deer mice are present in nearly every habitat in North America, and *Peromyscus* is the most speciose genus of North American mammals with 56 currently recognized species [7]. Given their radiation across an impressive range of environments,
Fig. 1. Schematic illustration of the potential for transcriptomic research in *Peromyscus*. Panel A demonstrates how transcriptomics could be used to identify gene expression differences between different species of *Peromyscus* with distinct mating behavior and parental care (e.g. *P. californicus* and *P. maniculatus*). Panel B illustrates the utility of an experimental transcriptomics approach in identifying the genes under selection in divergent habitats. In this example, individuals of the same species are obtained from contrasting habitats (urban vs. rural areas) with divergent natural selection regimes (contaminated vs. uncontaminated soils). Common-garden experiments are performed on individuals acclimated to the common environment, or their F1 offspring. These experiments manipulate environmental variables hypothesized to be agents of natural selection in their natural habitat. Exper-
deer mice have recently become one of the preferred mammalian systems for understanding adaptive responses of wild populations to environmental challenges. Adaptive evolution occurs when advantageous traits increase in population frequency from one generation to the next. The most widespread species, *P. manilatus* and to a lesser extent, *P. leucopus*, occur nearly continent-wide, and thus are particularly useful for intraspecific studies of adaptive evolution and gene expression between populations that occupy highly contrasting environments [13,14]. Habitat specialists, such as *P. eremicus* and other species that are limited to arid regions, are excellent candidates for both intra- and interspecific studies of gene expression differences between habitats [12].

Transcriptomics provides a two-fold value for studies of deer mice in challenging environments because a single, well-designed RNA-Seq experiment can discover 1) both single nucleotide polymorphisms (SNPs) within protein-coding genes that are functionally connected to adaptive phenotypic changes, and 2) changes in gene expression that are associated with physiological or behavioral responses to challenging environments. Variation in gene expression in wild populations will likely be high due to variable environmental conditions and plasticity, and thus large sample sizes will be necessary to categorize the variation of interest. Guidelines for designing robust RNA-Seq experiments in natural habitats are emerging [3]. Once identified in the field, these SNPs or gene expression changes can be functionally validated using captive *Peromyscus* populations that are subjected to a common environment and experimental treatments related to hypotheses about selection pressures in the wild (Fig. 1). Below we review studies on adaptive evolution and gene expression in *Peromyscus* that occupy distinct habitats with strong natural selection differentials: high elevation, light-colored substrates, deserts, and cities.

### 2.1. High altitude: hypoxia and cold

*P. manilatus* is the most widespread deer mouse species, including montane habitats at ≈ 4300 m above sea level in the Rocky Mountains [7]. Mammals at such elevations must deal with substantially less oxygen available for respiration, as well as extreme cold during winter months when non-hibernatory species such as deer mice are still active. Directional selection for adaptive phenotypes or transcriptional plasticity are possible explanations for the enhanced respiration and metabolic heat production that would be necessary for mammalian existence under hypoxia. Several studies have now confirmed that both of these phenomena contribute to the existence of deer mice at such high altitudes.

It has been known for decades that *P. manilatus* exhibit diverse hemoglobin phenotypes due to multiple, polymorphic hemoglobin loci that resulted from a complex history of gene duplications. Furthermore, some of these geneotypes are associated with high-altitude populations and relevant physiological traits [15,16]. In particular, one highland haplotype is associated with higher blood oxygen affinity and oxygen consumption when mice are exposed to cold or aerobic activity [17]. Intense directional selection on the maximal rate of oxygen consumption was documented for wild, highland *P. manilatus*, thus confirming the adaptive value of the highland haplotype [18]. Storz et al. [19] subsequently identified five specific amino acid substitutions in two closely linked, α-globin gene duplicates that underlie adaptive increase in oxygen-binding affinity. Two β-globin genes were also later identified as targets of divergent selection between low- and high-elevation populations [20]. Having identified the relevant sets of paralogs, Storz et al. [21] then characterized the isoform diversity produced by allelic diversity between two altitudinal zones, and recently reported that substantial epistasis between amino acid mutations contributes to increased blood oxygen affinity [22]. These mutations also occur in *P. manilatus* populations from lower elevations and with other closely-related species, suggesting that variation in blood oxygen affinity is influenced by many genes of small effect [23]. Presence of adaptive alleles of small effect in lowland populations also suggests that adaptation to high altitude proceeded from standing variation in *P. manilatus*.

The functional relationship between hemoglobin mutations and oxygen-binding affinity in hypoxic environments is one of the best examples of an adaptive phenotype linked to genotype in *Peromyscus*, or perhaps any free-living mammal. However, other candidate loci and plasticity in gene expression are likely to play roles in adaptive physiological responses to high-altitude environments. Several transcriptomic studies have now employed experimental approaches to examine regulatory plasticity in *P. manilatus*. Cheviron et al. [14] used RNA-Seq to identify gene expression differences between highland and lowland *P. manilatus*, as well as interspecific differences between *P. manilatus* and lowland *P. leucopus*. Genes involved in metabolic pathways were significantly overrepresented compared to the total set of genes that were differentially expressed between *P. manilatus* populations, as well as between *P. manilatus* and *P. leucopus*. In particular, dozens of genes in the fatty acid β-oxidation and oxidative phosphorylation (OXPHOS) pathways were differentially expressed between highland *P. manilatus* and lowland *P. manilatus*/leucopus. More than 75% of these genes were up-regulated in highland *P. manilatus* compared to the other two groups, indicating that enhanced expression in these pathways is associated with adaptive high-altitude phenotypes in deer mice. Cheviron et al. [14] further found that genes involved in glycolysis and carbohydrate metabolism were not overrepresented among genes exhibiting differential expression. They interpreted these patterns as evidence that highland deer mice rely on lipid metabolism to generate heat through shivering. These findings are in contrast to hypotheses for humans that stress the important role of carbohydrate metabolism in acclimation to high altitude [24], indicating that highland deer mice have evolved at least partially along a different physiological trajectory.

To examine the role of transcriptomic plasticity in thermogenesis at high altitude, Cheviron et al. [11] followed up on the work above with a large-scale RNA-Seq experiment examining expression differences between low- (430 m) and high-altitude (4350 m) deer mice within 1–2 days of capture in their native habitat, after 6 weeks in common-garden conditions (360 m), and on F1 progeny reared in common-garden conditions. In common-garden experiments, organisms are transplanted from their native environment to an experimental environment with the same conditions for all groups. These experiments allow researchers to identify differences between groups that are due to heritable rather than environmental factors. Cheviron et al. identified 1353 genes that exhibited expression differences due to experimental treatment, and found that these genes existed in five transcriptional modules (i.e., transcripts co-expressed together with related functions) related to protein metabolism, regulation of skeletal muscle cell differentiation, regulation of erythrocyte differentiation, water transport, carbohydrate and lipid metabolism, muscle and blood vessel development, brown fat cell differentiation, and other more
general regulatory categories. Comparing highland and lowland mice revealed several transcriptional modules associated with thermogenic capacity and endurance in the experimental common garden and F1 mice. As reported previously [14], genes in lipid oxidation and OXPHOS pathways were upregulated in highland mice, but glycolysis and carbohydrate metabolism were not [11]. These results indicate that plasticity in gene expression plays an important role in adaptive phenotypic differences between highland and lowland deer mice, and that flexible regulation of lipid metabolism is a significant functional component of the shivering thermogenesis phenotype in high altitude populations.

Along with thermogenic capacity and oxidation pathways, Scott et al. [25] hypothesized that highland mice exhibit an adaptive skeletal muscle phenotype characterized by greater capillary density and oxidative capacity. To test this prediction, they measured muscle fiber type and surface capillary density of muscle from highland and lowland deer mice, as well as from laboratory F1 progeny as above. They then used RNA-Seq to identify gene expression associated with muscle phenotypes, and add to the broader sets of genes functionally related to adaptation to hypoxia. As predicted, skeletal muscle in highland deer mice had greater capillary density and a more oxidative phenotype. The RNA-Seq analysis identified two transcriptional modules of 68 genes that differed in expression levels between highland and lowland populations, and five modules with 589 genes that differed between wild and lab-raised F1 mice. Population differences in expression were associated with muscle phenotypes and pathways that influence metabolism, muscle fiber composition, and vascularization [25]. Increase in capillary density and oxidative capacity of muscle may enhance oxygen retention and thermogenesis through shivering as well as mitochondrial respiration of muscle. These differences between groups likely resulted from both divergent natural selection and plasticity as previously reported for highland deer mice [11,14].

The same experimental design used to examine muscle phenotypes was employed by Velotta et al. [26] to examine the role of brown adipose tissue in non-shivering thermogenesis in deer mice. Heat is generated by brown adipose tissue through well-characterized pathways involving the uncoupling of oxidative phosphorylation from ATP synthesis. Nonshivering thermogenesis was estimated in deer mice by measuring oxygen consumption after stimulation of brown adipose tissue by injection of an agonist of \( \beta_3 \)-adrenergic receptors. As with the skeletal muscle/shivering phenotypes, highland deer mice exhibited increased nonshivering thermogenesis. RNA-Seq identified transcriptional differences in several co-regulated genes influencing traits such as brown adipose proliferation, oxygen supply to brown adipose tissue, and the heat generation pathway in brown adipose tissue [26]. Adaptive divergence in hemoglobin genes and oxygen binding affinity, and both adaptive divergence and gene expression plasticity in shivering and non-shivering thermogenesis, all explain how deer mouse populations exist in hypoxic conditions over 4000 m above sea level.

2.2. Deserts, renal physiology, and water stress

A number of Peromyscus species (e.g. P. eremicus, P. crinitus) occupy the deserts of the SW United States and northern Mexico, which are among the harshest environments on Earth. These desert ecosystems are characterized by the highest daytime maximum temperatures in North America, minimal precipitation, and almost no standing water sources, thus posing a severe challenge to maintaining water balance in desert-dwelling organisms. Deserts do not exhibit high mammalian biodiversity, but rodents are relatively common in deserts and thus have likely evolved many adaptations to water stress. Recently, full transcriptomes of the kidney were generated for three non-Peromyscine rodents that inhabit deserts [27,28]. Adaptive divergence and differential expression in genes related to osmoregulation were identified in two of these species [27]. Desert Peromyscus offer several advantages over these other taxa because they are relatively easy to keep in captivity, other closely-related non-desert species are available for interspecific comparisons (e.g. P. californicus), and genomic resources for Peromyscus are under active development to serve the community studying this emerging model group [7].

Recently, a research program to develop P. eremicus as a model system for examining desert adaptations was initiated by characterizing the transcriptome of multiple tissue types (kidney, liver, brain, testes) for one individual using RNA-Seq [12]. A subsequent study added three male reproductive transcriptomes (epididymis, testes, and vas deferens) to transcriptome resources for this species [29]. Nearly four times more unique transcripts were identified in the kidney transcriptome compared to the next highest tissue type (brain), although it is hard to rule out that more rare transcripts were detected by chance in the kidney. Male reproductive transcriptomes were not compared to non-reproductive tissues, but the testes transcriptome exhibited a greater number of unique transcripts than epididymis or vas deferens. Peromyscus genomic resources still lag far behind other model rodents, but these transcriptomes were successfully annotated using Mus musculus resources and the P. maniculatus reference transcriptome. All of the P. eremicus resources generated in these studies are freely available to the Peromyscus research community (Table 1).

To further explore potential desert adaptations, Macmanus & Eisen [12] generated kidney RNA-Seq data for an additional 15 P. eremicus individuals from Palm Desert, California. The top ten values of Tajima’s D calculated for each transcript in this dataset were for genes related to water balance that may be experiencing balancing selection, although the influence of historical demography on these values has not yet been examined. Branch site tests and Tajima’s D for six candidate genes identified in desert heteromyid rodents by Marra et al. [27] revealed one significant transcript in P. eremicus: Slc2a9, a gene involved in solute balance. The other five genes, mostly in the aquaporin gene families, exhibited Tajima’s D values consistent with selection.

As with high-altitude P. maniculatus above, RNA-Seq comparisons of wild-caught P. eremicus from the extreme portions of its range (i.e. most arid and most humid locations) are feasible in this study system. Experimental comparisons of lab-raised P. eremicus and other Peromyscus exposed to simulated desert conditions in the lab will also be productive research areas moving forward. Given the preliminary work at generating high-quality reference transcriptomes, we anticipate significant insight into the molecular mechanisms of adaptation to desert physiology.

Table 1  
Peromyscus transcriptome datasets generated using high-throughput sequencing and currently available on GenBank. na = not applicable.

<table>
<thead>
<tr>
<th>Species</th>
<th>GenBank Accession</th>
<th>Cell types</th>
<th>Experimental Treatment</th>
<th>Refs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>P. eremicus</td>
<td>PRJN242486</td>
<td>brain, liver, kidney, testes</td>
<td>na</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td>PRJEB13364</td>
<td>epididymis, testes, vas deferens</td>
<td>na</td>
<td>29</td>
</tr>
<tr>
<td>P. leucopus</td>
<td>PRJNA192723</td>
<td>brain, liver, ovary, testes</td>
<td>urban vs. rural</td>
<td>13, 42, 48</td>
</tr>
<tr>
<td></td>
<td>PRJNA184055</td>
<td>brain, liver</td>
<td>testes, kidney</td>
<td>82</td>
</tr>
<tr>
<td>P. maniculatus</td>
<td>PRJNA240902</td>
<td>liver</td>
<td>na</td>
<td>unpub.</td>
</tr>
<tr>
<td>P. eremicus</td>
<td>PRJNA217920</td>
<td>skeletal muscle</td>
<td>highland vs. lowland, common garden, F1 hybrids</td>
<td>11, 25</td>
</tr>
<tr>
<td></td>
<td>SRP073101</td>
<td>brown adipose tissue</td>
<td>highland vs. lowland, common garden</td>
<td>26</td>
</tr>
</tbody>
</table>
it is likely that such experiments would very quickly improve our knowledge of desert adaptation in *Peromyscus* and develop one or more species as powerful models for functional genomic studies of osmoregulation in mammals. To this end, Kordonowy et al. [30] recently developed an environmental chamber that simulates desert conditions for *P. eremicus*. Preliminary experiments established that *P. eremicus* drink very little water (0.11 mL/day when provided *ad libitum*). Experimental dehydration in the chamber revealed that dehydrated mice lose a substantial percentage of their body mass (nearly 25% in three days), and their serum electrolyte profile changes. Despite these effects, there were no obvious ill effects of water deprivation.

### 2.3. Cities: diet, pollution, immunology

Given the broad geographic range and habitat tolerance of *P. maniculatus* and *P. leucopus*, it seems inevitable in hindsight that these species would become common inhabitants of landscapes constructed by humans. *P. leucopus*, for example, may occur at higher densities near roads than in contiguous forests [31], and small forest fragments may even act as source populations for larger fragments because of ecological release from predators and competitors [32]. Populations fragmented by agriculture maintain high levels of heterozygosity and become only weakly differentiated [33], and white-footed mice have recently expanded their range northward across agricultural areas into Quebec in response to milder and shorter winters [34,35]. *P. leucopus* also occurs in highly-disturbed secondary forests within cities across eastern North America, even in the most densely populated areas of the country such as Central Park in New York City (NYC) [36]. Unlike agricultural landscapes, however, urbanization results in genetic differentiation between populations separated by a few km or less [37]; only narrow, vegetated corridors comprised of rivers, parkways, cemeteries and other “green” infrastructure still maintain any genetic connectivity between NYC populations of white-footed mice [38].

These findings that NYC populations are genetically isolated but still maintain high levels of neutral genetic variation [36–38] indicated that the conditions for local adaptation to urban conditions may be widespread for white-footed mice and other urban *Peromyscus* [39]. Indeed, temporal replacement of mitochondrial haplotypes had previously been reported for Chicago-area *P. leucopus*, suggesting rapid evolution in urban populations [40,41]. However, these earlier studies on urban *P. leucopus* relied on relatively small numbers of microsatellite markers and mitochondrial haplotypes to investigate evolutionary change. In 2013, Harris et al. [13] used RNA-Seq to generate transcriptome data from liver, brain, and gonads from both urban and rural populations around NYC. They identified over 7K gene candidates with functional annotations from *Mus* and *Rattus* reference genomes, as well as hundreds of over-represented gene ontology (GO) terms for individual tissue types. Across the dataset, they also reported over 30K SNPs and 11 candidate genes with elevated ratios of nonsynonymous to synonymous substitutions in urban vs. rural or urban vs. urban comparisons. These candidates had GO terms associated with immune responses, metabolism of xenobiotic compounds, and other functions that could be associated with a priori hypotheses about local adaptation to polluted, densely populated habitat patches in cities.

Harris et al. [42] conducted a much larger RNA-Seq study using liver tissue from urban and rural white-footed mice to build transcriptome resources for this species and identify additional candidate genes under selection in urban environments. This effort produced over 30 Mb of sequence, 40K contigs, 104K SNPs derived from the transcriptome, and 10 genes that were differentially expressed between urban and rural populations. Similar to the candidates identified previously [13], these expression outliers were related to immune responses and transcriptional/translational processes [42]. Scans for selection are often subject to false positive outliers because the influence of historical demography (e.g. population bottlenecks or expansions) may create genomic signatures that mimic the actions of selection [43]. To account for historical demography in this system, Munshi-South et al. [44] generated more than 10K SNPs using double-digest restriction site-associated DNA sequencing (ddRAD-Seq) [45] among 23 white-footed mouse populations spanning an urban-to-rural gradient in the NYC area. They found that levels of genome-wide variation are inversely related to degree of urbanization of populations, and populations within urban areas are more differentiated from each other than populations in rural areas. They concluded that standing genetic variation is lower in urban white-footed mice, but not so low nor genetic drift so strong as to preclude local adaptation [44]. Harris et al. [46] then built a full demographic model for these 23 populations using a recently developed coalescent approach to estimate parameters [47]. This model revealed that a group of mainland populations diverged from Long Island populations after post-glacial colonization events several thousand years ago. However, several contemporary urban populations also diverged from other populations in the last few hundred generations and experienced bottlenecks, indicating that recent urbanization influenced genomic patterns in NYC’s white-footed mice [46].

Having inferred the demographic history of urban *P. leucopus* in NYC, Harris and Munshi-South [48] then conducted genome scans using methods that incorporate the demographic history in the null model. From their transcriptome data for 48 individuals from 3 urban and 3 rural populations [42], they identified overlapping subsets of outlier genes using scans based on patterns of differentiation (i.e. *FST*, calibrated using inferred demographic parameters), linkage disequilibrium, and genotype-by-environment associations. The majority of these outlier genes have GO terms associated with metabolic functions, particularly dietary specialization. Many of these genes are in pathways involved in either carbohydrate or lipid metabolism, suggesting that urban white-footed mice in NYC have experienced directional selection for novel diets [48]. These findings will be assessed in the near future by field investigations of diet, a dataset of SNP genotypes from many more than 3 urban and 3 rural populations, and comparisons of SNP allele frequencies between contemporary populations and museum specimens collected over 100 years ago [49].

### 3. Behavior

Closely-related species of *Peromyscus* may exhibit markedly different behaviors that evolved due to natural selection in contrasting social and environmental contexts [7]. If highly heritable, then uncovering the genetic architecture of these behaviors is tractable using laboratory crosses when species are interfertile. Burrow building [50] and mating systems [51,52] are two such complex behaviors that vary between closely-related species and have recently been the subject of transcriptomic studies. *Peromyscus* have also attracted interest as potential alternatives to *Mus* and *Rattus* as laboratory rodent models for human behavior [53]. Behavioral transcriptomics is a young field with high growth potential that is likely to benefit from methodological advances in genomics for years to come [54].

#### 3.1. Burrow construction

One of the best examples of studies linking genotypes to behavioral phenotypes that are relevant to wild organisms is Weber et al.’s [55] analysis of burrow construction in two species of *Peromyscus*. One species, *P. polionotus*, constructs long burrows with both entrance and escape tunnels, whereas its sister species, *P. maniculatus*,
exhibits a more ancestral, relatively short burrow without escape tunnels [56]. Both taxa exhibit their characteristic burrow-building behavior after several generations in captivity, and F1 hybrids exhibit the P. polionotus burrow phenotype because the P. polionotus trait is dominant [50]. Weber et al. [55] extended this work by creating first-generation backcross mice that exhibited a wide range of entrance tunnel lengths and intermediate frequencies of escape tunnels, suggesting that these two traits are influenced by separate genes. A quantitative trait locus (QTL) analysis with hundreds of SNPs genotyped using ddRAD-Seq identified one locus associated with escape tunnels, and three loci associated with entrance tunnel length. They argued that the two burrow traits are influenced by discrete genetic modules, and evolved through the accumulation of alleles that combined in novel ways.

A new Ph.D. dissertation by H. Metz [57] from this same research group further explored the proximate causes of burrow variation in Peromyscus using approaches from developmental biology and transcriptomics. She reports that a locus associated with tunnel length predicts precocious burrowing behavior (typical of P. polionotus) in recombinant F2 backcross hybrids, and that escape tunnels are likely a threshold trait. Using a transcriptomic experiment, it was found that F1 hybrids exhibit brain gene expression that is biased towards P. polionotus alleles. Allele-specific expression also characterized several candidate genes in QTLs that are associated with burrow building behavior [57]. GO terms related to behavior and activity patterns, likely established in functional studies of lab rodents, characterized these candidate genes. Given that several QTLs, candidate genes, and expression differences between burrow phenotypes are now known, the P. polionotus/P. maniculatus species pair would be an ideal system for functional studies using CRISPR-Cas9 [58] or other genomic manipulations. The success of these approaches at elucidating the genomics of a complex behavior in Peromyscus will rightfully inspire behavioral transcriptomic studies in many other wild-living organisms.

3.2. Social behavior and mating systems

A substantial majority of mammals, including most Peromyscus, exhibit polygynous or promiscuous mating systems [59–61]. However, at least two deer mice, P. californicus [51] and P. polionotus [52], have evolved genetic monogamy and biparental care [62] (although to a more modest degree in the latter) [63]. This variation in mating behavior allows for robust genomic comparisons between closely-related taxa that exhibit interesting behavioral differences. Comparison of vaginal bacterial communities from promiscuous P. maniculatus and monogamous P. californicus captured from the same site revealed that the promiscuous species has greater microbially diverse [64]. A prediction that could then be tested with transcriptomic comparisons is that promiscuous species may exhibit unregulated gene expression, or signatures of selection, in genes associated with the immune system. Indeed, MacManus and Lacey [65] reported higher levels of diversity and enhanced selection in the major histocompatibility complex of P. maniculatus vs. P. californicus, but many other genes are likely to exhibit differences related to mating system variation. Gene expression in the brain is also likely to vary in relation to social behavior and mating system [66]. For example, estrogen receptor alpha (ERα) is expressed at higher levels in male brains of monogamous P. californicus and P. polionotus compared to polygynous P. maniculatus and P. leucopus [67]. ERα was already known to be associated with social monogamy in some voles (Microtus spp.) and dwarf hamsters (Phodopus spp.). There are many other candidate genes that have been identified after decades of research in monogamous Microtus [68,69] that could be examined in Peromyscus [70]. This candidate gene strategy would not identify genes that are important in Peromyscus but not Microtus, and thus transcriptomic studies could develop deer mice as an alternative model system for the genetic basis of mating system variation. For example, ERα expression is not sexually dimorphic in all studied regions of the Peromyscus brain whereas it is in murid rodents, voles, and dwarf hamsters [67].

Morphological, physiological, and life history traits associated with mating behavior are also likely to exhibit genetic differences between monogamous and promiscuous species. For example, sperm competition has driven the evolution of cooperative sperm aggregation in P. maniculatus, but not in the monogamous P. polionotus [71]. This trait would be beneficial for post-copulatory sexual selection when a female has serially mated with multiple males in a short period of time. Faster reproductive development is also likely to be favored in promiscuous vs. monogamous species [72], especially in short-lived taxa where the temporal window of reproductive opportunities may be small. Jacobs-Palmer [73] found that the onset of spermatogenesis occurs earlier in male P. maniculatus than in male P. polionotus, in line with predictions based on their respective mating systems. To examine the influence of gene expression on sexual development, she then created F1 hybrids of these two species and conducted RNA-Seq on testis tissues collected at 10 postnatal periods from 16 to 64 days. These sequence data were then filtered for genes that fit experimental prerequisites (e.g. fixed SNPs between the two species) and had adequate read counts for examining allele-specific expression. She reported that 702/4337 (16.2%) loci exhibited two-fold or greater differences in expression between P. maniculatus and P. polionotus alleles for at least one time point. These differentially expressed genes are related to spermatogenesis, testis development, and sperm phenotype, suggesting that accelerated onset of spermatogenesis in P. maniculatus is influenced by cis-regulatory changes [73].

4. Development

Transcriptomics has been particularly transformational for developmental biology, given the crucial role of gene expression regulation in tissue differentiation. Laboratory Mus have been the workhorse of mammalian developmental biology; the mouse genome [74], encyclopedias of mouse DNA elements [75,76], and other resources [77] for studying gene function and regulation are unparalleled among non-human mammals. However, gene expression patterns are not very similar between humans and Mus [78], prompting calls for efforts to develop nonmodel organisms for developmental study [79]. Such efforts could examine developmental phenomena in broader ecological and evolutionary contexts, and these goals are achievable using RNA-Seq for even highly complex developmental phenotypes.

Vrana et al. [80] recently published a thorough review of the use of Peromyscus as developmental models. In brief, they outlined the availability of wild-derived populations available through the Peromyscus Genetic Stock Center (http://stkctr.biol.sc.edu/), which includes five species and F1 hybrids of P. maniculatus x P. polionotus at the time of writing this review. Transcriptomic studies of contrasting populations of many of these species and stocks are detailed above; P. leucopus have also been proposed as a useful model for understanding the evolution and development of long lifespan relative to other small mammals [81,82]. Several mutant strains of P. maniculatus are also available for examining the development of pelage color, hairlessness, syndactyly, tremors, epilepsy, and juvenile ataxia. P. maniculatus x P. polionotus hybrids also exhibit parent-of-origin effects on growth and development, with the male P. maniculatus x female P. polionotus cross producing typically invivable offspring. Thus, these two species have been extremely useful for studying hybrid dysgenesis [83,84]. Transcriptomic experiments to examine in-
tra- and interspecific differences in development have not yet been conducted using these Peromyscus resources. The availability of wild-derived and mutant strains, as well as emerging resources such as descriptions of embryonic stages in Peromyscus [85], have set the stage for a productive research program of great value to evolutionary developmental biology.

5. Pathogens and disease

Transcriptomics also holds great promise for understanding the role of Peromyscus in the etiology of several human diseases. P. maniculatus is the primary reservoir host of Sin Nombre hantavirus [86], and P. leucopus is an important reservoir of several tick-borne pathogens including Borellia [87,88], Anaplasma [89,90], and Babesia [91,92]. Lyme disease from Borellia infection is the most common tick-borne disease in humans. Recent findings indicate that both tick vectors and P. leucopus are often co-infected with two or more of these pathogens [93,94], enhancing the risk that humans will be infected by more than one pathogen at a time by a single tick bite. However, the host factors that promote co-infection and disease severity in humans are poorly understood. P. leucopus exhibit a broad range of responses to experimental infections, and transcriptomic studies could shed light on the genetic architecture and gene expression factors related to their immune traits [95]. Candidate gene approaches have been used to examine immunity in wild populations [96], with a particular focus on Toll-like receptors in rodents [97–99]. Wild bank voles (Myodes glareolus) in Europe with certain Toll-like receptor 2 haplotypes are much less likely to be infected with Borellia [100], but it is likely that some pathways of the innate and adaptive immune responses are involved in variability in host infection with tick-borne pathogens. A recent transcriptomic analysis in humans infected with Borellia found many differentially expressed genes that are unique to this infection [101]; similar studies in Peromyscus are needed to fully understand the ecology of Borellia and its primary host.

6. Conclusions & future directions

Transcriptomic and RNA-Seq approaches have demonstrated their value in revealing how natural selection and gene expression plasticity contribute to the existence of deer mice in challenging environments across North America. Transcriptomics holds great promise for elucidating the genetic basis of mating systems and other behaviors in Peromyscus, but has been underutilized for developmental biology and disease studies.

In 2001 Dewey and Dawson [102] argued that Peromyscus could be the “Drosophila of North American Mammalogy”, and began their review with reference to the success of the human and laboratory mouse genome projects. The development of much greater genomic resources, particularly deeply sequenced, assembled, and annotated reference genomes and transcriptomes for multiple species, will be necessary to move Peromyscus transcriptomics forward. Currently only a single Peromyscus genome assembly is available on GenBank (P. maniculatus, GCA_000500345.1 Pman_1.0), along with genome sequencing reads for P. leucopus, P. polionotus, and P. californicus (GenBank BioProject Accession PRJNA170428). Transcriptomes are currently available from only three species (Table 1). Bedford and Hoekstra [7] also identified the need for improved genomic resources, as well as phylogenetic revision, natural history of poorly known Peromyscus species, and use of deer mice in biomedical studies as priorities for research. RNA-Seq applied to many more Peromyscus taxa could be very useful for identifying genes for use in phylogenetic analyses of the group. Range-wide surveys of genetic, phenotypic, and ecological parameters, especially for the widespread species, would likely uncover as yet unknown diversity. The power of the Peromyscus system will not be fully realized until more studies apply robust comparative frameworks to analyze populations in multiple habitats/study sites, multiple populations of the same species, and/or multiple species.

Acknowledgements

Work on this publication was supported by the National Science Foundation under award number DEB 1457523 and the National Institute of General Medical Sciences of the National Institutes of Health under award number R15GM099055 to Jason Munshi-South. The content is solely the responsibility of the authors and does not represent the official views of the National Science Foundation or National Institutes of Health. The authors also thank the Editor of this issue, Hippokratis Kiaris, and one anonymous reviewer for their helpful comments on this manuscript.

References


