Genomic evaluation of male reproductive adaptations and responses to dehydration in Peromyscus eremicus (Cactus mouse)

Lauren Kordonowy

University of New Hampshire, Durham

Follow this and additional works at: https://scholars.unh.edu/dissertation

Recommended Citation
Kordonowy, Lauren, "Genomic evaluation of male reproductive adaptations and responses to dehydration in Peromyscus eremicus (Cactus mouse)" (2017). Doctoral Dissertations. 152.
https://scholars.unh.edu/dissertation/152
Abstract
Research elucidating the genetic architecture of physiological mechanisms enabling survival and reproduction in extreme environments is becoming prominent in evolutionary biology. The desert, in particular, poses numerous challenges for its endemic species, and mammals (and often, rodents) have been the focus for survival adaptations pertaining to water-limitation. However, desert rodent adaptation research has focused predominantly on survival, while potential physiological reproductive adaptations to dehydration have received less attention, aside from research evaluating water as reproductive cue. The fact that we do not know the physiological mechanisms enabling reproduction during dehydration is surprising, as desert rodents must possess adaptations to successfully reproduce in their water-limited habitats. The cactus mouse (Peromyscus eremicus), a desert-specialist in the Southwest United States, is the focus of my genetic exploration of reproductive adaptations to dehydration. My dissertation describes three research studies that 1) characterize male cactus mouse reproductive tissue transcriptomes and find signatures of positive selection in these tissues relative to other rodent species, 2) describe differential expression of genes responding to water-limitation within testes, providing candidate genes for future studies exploring the impacts of acute and chronic drought on P. eremicus reproduction, and 3) generate a seminal vesicle proteome representative of proteins present in hydrated and dehydrated conditions experienced by the cactus mouse. These three studies contribute comprehensive genetic data critical to future research exploring the effects of water-limitation on reproduction as well as the genetic mechanisms for potential male reproductive adaptations in this desert-adapted rodent. Research on desert adaptations is particularly timely, as climate change will result in more frequent stochastic drought events and elevated temperatures. These increasing abiotic shifts will exacerbate clinical challenges for global health; thus, an enhanced understanding of mammalian desert-specialist adaptations may improve our ability to address these physiological demands in humans.

Keywords
Bioinformatics, Cactus Mouse, Evolution, Peromyscus eremicus, Reproduction, Transcriptome, Genetics, Bioinformatics, Biology
Genomic evaluation of male reproductive adaptations and responses to dehydration in *Peromyscus eremicus* (Cactus mouse)

BY

Lauren Kordonowy
B.A. in Biology (with Honors), Kenyon College, 2006
M.Sc. in Biological Sciences, Simon Fraser University, 2009

DISSERTATION

Submitted to the University of New Hampshire in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy
in
Genetics

May, 2017
Genomic evaluation of male reproductive adaptations and responses to dehydration in *Peromyscus eremicus* (Cactus mouse)

BY

Lauren Kordonowy

This dissertation has been examined and approved in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Genetics by:

**Dissertation Advisor**: Matthew MacManes, Ph.D., Assistant Professor of Molecular, Cellular, and Biomedical Sciences (UNH)

David Plachetzki, Ph.D., Assistant Professor of Molecular, Cellular, and Biomedical Sciences (UNH)

Philip Hatcher, Ph.D., Professor of Computer Science (UNH)

W. Kelley Thomas, Ph.D. Hubbard Professor of Genomics and Director for Department of Molecular, Cellular, and Biomedical Sciences (UNH)

David Townson, Ph.D., Department Chair and Professor of Department of Animal and Veterinary Sciences (University of Vermont)

4-28-2017
Date of Defense

Original approval signatures are on file with the University of New Hampshire Graduate School.
Citation and Copyright for Chapter 1:


©Copyright 2016 Kordonowy and MacManes

Distributed under Creative Commons CC-BY 4.0

OPEN ACCESS

Citation for Chapter 2:


bioRxiv DOI: https://doi.org/10.1101/096057; BMC Genomics: in review
ACKNOWLEDGEMENTS

I deeply acknowledge Dr. Matt MacManes for serving as my Ph.D. advisor. He granted me the incredible opportunity to change course in my Ph.D. research at a critical juncture during my degree. Moreover, Matt allowed me the freedom to explore research questions that interested me, and he provided patient and continuous expertise in a novel research area in my education: bioinformatics and genomics. Furthermore, he has fully supported my mentorship of several undergraduates in his lab, which has provided me with invaluable experience. Finally, Matt has embraced my passion for outreach in the sciences, and he has been a valuable partner in our UNH Science Sleuths education program for local pre-school aged children.

I also want to acknowledge Philip Hatcher, Kelley Thomas, David Townson, and David Plachetzki for serving on my dissertation committee. Thank you for meeting with me frequently to aid my research progress. I have always felt like your doors have been open to me, and you have been instrumental in my educational success at UNH.

I further acknowledge my intrepid fellow MacManes lab graduate student, Andrew Lang. Andrew has been a skilled and supportive partner at the lab bench, in coursework, in writing, in teaching, and in education outreach. Thank you for being my best ever lab partner!

I finally acknowledge my husband, Jonathan Cummings, for moving to New Hampshire with me and for his tireless support during my entire Ph.D. Thank you for countless hours of compassion and intellectual questions and for your expertise in R!

I also acknowledge my entire family and all of my friends who have supported me throughout my graduate education and who have made this time enjoyable.

Last, but not least, thank you, Bailey!
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMMITTEE PAGE</td>
<td>ii</td>
</tr>
<tr>
<td>COPYRIGHT</td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>iv</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>viii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>ix</td>
</tr>
<tr>
<td><strong>CHAPTER</strong></td>
<td></td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>I. Characterization of a male reproductive transcriptome for <em>Peromyscus eremicus</em> (Cactus mouse)</td>
<td>15</td>
</tr>
<tr>
<td>Abstract</td>
<td>16</td>
</tr>
<tr>
<td>Introduction</td>
<td>17</td>
</tr>
<tr>
<td>Methods</td>
<td>21</td>
</tr>
<tr>
<td>Results and Discussion</td>
<td>27</td>
</tr>
<tr>
<td>Conclusions</td>
<td>35</td>
</tr>
<tr>
<td>Dryad Data File List</td>
<td>36</td>
</tr>
<tr>
<td>Additional Information and Declarations</td>
<td>37</td>
</tr>
<tr>
<td>Tables and Figures</td>
<td>39</td>
</tr>
<tr>
<td>References</td>
<td>47</td>
</tr>
</tbody>
</table>
II. Characterizing the reproductive transcriptomic correlates of acute dehydration in males in the desert-adapted rodent, *Peromyscus eremicus* .......................................................... 57
  Abstract ........................................................................................................ 58
  Background ................................................................................................. 60
  Methods ....................................................................................................... 64
  Results .......................................................................................................... 71
  Discussion ................................................................................................. 77
  Conclusion ................................................................................................. 83
  Dropbox Data File List ................................................................................ 84
  Declarations .............................................................................................. 85
  Tables and Figures ..................................................................................... 87
  References ................................................................................................. 109
  
III. Characterization of the Seminal Vesicle Proteome of *Peromyscus eremicus* (Cactus Mouse) in hydrated and dehydrated conditions ................................................ 125
  Abstract ...................................................................................................... 126
  Dataset Brief ............................................................................................. 127
  Supplemental Materials ............................................................................. 131
  Figure ........................................................................................................ 132
  References ................................................................................................. 133

APPENDIX: IACUC Approval Letter .............................................................. 137
# LIST OF TABLES

**CHAPTER 1:** Table 1 .................................................................................................................. 39  
**CHAPTER 1:** Table 2 .................................................................................................................. 40  
**CHAPTER 1:** Table 3 .................................................................................................................. 41  
**CHAPTER 2:** Table 1 .................................................................................................................. 87  
**CHAPTER 2:** Table 2 .................................................................................................................. 88  
**CHAPTER 2:** Table 3 .................................................................................................................. 89  
**CHAPTER 2:** Table 4 .................................................................................................................. 92  
**CHAPTER 2:** Table 5 .................................................................................................................. 93  
**CHAPTER 2:** Supplemental Table 1 ......................................................................................... 103  
**CHAPTER 2:** Supplemental Table 2 ......................................................................................... 104
LIST OF FIGURES

CHAPTER 1: Figure 1 ........................................................................................................... 43
CHAPTER 1: Figure 2.......................................................................................................... 44
CHAPTER 1: Figure 3.......................................................................................................... 45
CHAPTER 1: Figure 4.......................................................................................................... 46
CHAPTER 2: Figure 1.......................................................................................................... 94
CHAPTER 2: Figure 2.......................................................................................................... 95
CHAPTER 2: Figure 3.......................................................................................................... 96
CHAPTER 2: Figure 4......................................................................................................... 100
CHAPTER 2: Figure 5......................................................................................................... 102
CHAPTER 2: Supplemental Figure 1.................................................................................. 106
CHAPTER 2: Supplemental Figure 2.................................................................................. 107
CHAPTER 2: Supplemental Figure 2.................................................................................. 108
CHAPTER 3: Figure 1......................................................................................................... 132
ABSTRACT

Genomic evaluation of male reproductive adaptations and responses to dehydration in *Peromyscus eremicus* (Cactus mouse)

By

Lauren Kordonowy

University of New Hampshire,
May, 2017

Research elucidating the genetic architecture of physiological mechanisms enabling survival and reproduction in extreme environments is becoming prominent in evolutionary biology. The desert, in particular, poses numerous challenges for its endemic species, and mammals (and often, rodents) have been the focus for survival adaptations pertaining to water-limitation. However, desert rodent adaptation research has focused predominantly on survival, while potential physiological reproductive adaptations to dehydration have received less attention, aside from research evaluating water as reproductive cue. The fact that we do not know the physiological mechanisms enabling reproduction during dehydration is surprising, as desert rodents must possess adaptations to successfully reproduce in their water-limited habitats. The cactus mouse (*Peromyscus eremicus*), a desert-specialist in the Southwest United States, is the focus of my genetic exploration of reproductive adaptations to dehydration. My dissertation describes three research studies that 1) characterize male cactus mouse reproductive tissue transcriptomes and find signatures of positive selection in these tissues relative to other rodent species, 2) describe differential expression of genes responding to water-limitation within testes, providing candidate genes for future studies exploring the impacts of acute and chronic drought
on *P. eremicus* reproduction, and 3) generate a seminal vesicle proteome representative of proteins present in hydrated and dehydrated conditions experienced by the cactus mouse. These three studies contribute comprehensive genetic data critical to future research exploring the effects of water-limitation on reproduction as well as the genetic mechanisms for potential male reproductive adaptations in this desert-adapted rodent. Research on desert adaptations is particularly timely, as climate change will result in more frequent stochastic drought events and elevated temperatures. These increasing abiotic shifts will exacerbate clinical challenges for global health; thus, an enhanced understanding of mammalian desert-specialist adaptations may improve our ability to address these physiological demands in humans.
INTRODUCTION

Overview and Rationale

There has been an increase in evolutionary biology research exploring the genetic underpinnings of physiological, behavioral and ecological traits. Furthermore, these studies frequently utilize genomic and transcriptomic methods to explore genes under selection, to evaluate differential gene expression levels, and to describe gene ontologies (e.g. Atallah et al., 2013; Plachetzki et al., 2014; Guillen et al., 2015; Bedford and Hoekstra, 2015; Munshi-South and Richardson, 2016; Behringer et al., 2015; Chen et al, 2016; Suarez-Vega et al., 2016; Macrander et al., 2016; Marra et al., 2014, MacManes, 2017). However, this research is often exploratory, as it frequently generates candidate genes for future studies to evaluate. This is because analyses requisite for studies attributing genes with functionality are best executed on model organisms, which already possess a litany of genomic, transcriptomic, and physiological resources. Non-model organisms with environmental adaptations of interest have often been the focus of behavioral, ecological or physiological research, but until recently, they rarely had complete, high-quality genomes or comprehensive transcriptomes or proteomes (Ellegren, 2014).

Thus, it is incumbent upon evolutionary biologists to collect sufficient genetic data to bridge the informational gap between model and non-model species before they can begin to pursue causative adaptive research on genes of interest with functional studies (Storz and Wheat, 2010). Such is the case with my dissertation study species, the cactus mouse (Peromyscus eremicus).

The cactus mouse has been the focus of intermittent biological research for decades (Dewey et al., 1966; summarized in: King, 1968; Veal and Caire, 1979; MacMillen and Garland,
1989); however, evolutionary biologists have only recently begun collecting genetic data on this species for adaptation research (MacManes and Eisen, 2014; Kordonowy and MacManes, 2016; Kordonowy and MacManes, 2017, Kordonowy et al., 2017a, MacManes, 2017). This desert-adapted rodent is endemic to the Southwestern United States, and this species is remarkable for its ability to live entirely without drinking water (Veal and Caire, 1979; Caire, 1999). Recent research has sought to elucidate cactus mouse renal adaptations for survival using genomic (MacManes and Eisen, 2014, MacManes, 2017) and physiological methods (Kordonowy et al., 2017b). However, previous research focused exclusively on cactus mouse survival adaptations and did not address potential physiological reproductive adaptations to combat water-limitation.

My thesis explores potential male reproductive adaptations in the cactus mouse using transcriptomic methodologies and bioinformatic analyses, as well as proteomics. The aims of my dissertation research are to (1) genetically characterize male cactus mouse reproductive tissues utilizing transcriptomics (testes, epididymis and vas deferens: Kordonowy and MacManes, 2016) and proteomics (seminal vesicles: Kordonowy et al., 2017a), (2) to determine cactus mouse testes responses to dehydration (Kordonowy and MacManes, 2017), and (3) to propose and explore our hypothesis for male cactus mouse reproductive physiological adaptations to water-limitation (Kordonowy and MacManes, 2016). In the last three years, I have generated genetic resources for previously uncharacterized male reproductive tissues in the cactus mouse, and I have also conducted comparative and descriptive analyses to determine male reproductive responses to dehydration and to find candidate genes for potential reproductive adaptations. This will allow future studies to evaluate functional causality between the genes demonstrating transcriptomic responses to dehydration with physiological adaptations for acute
water-limitation in the cactus mouse: an ultimate goal in this evolutionary biology research.

The implications of desert adaptation research extend well beyond the realm of scientific inquiry within the field of experimental biology. Indeed, Johnson and colleagues (2017) asserted that studies on adaptations in desert-specialized mammals will be instrumental in our ability to address a growing health crisis due to the mounting effects of dehydration and heat-stress resulting from climate change.

**Background**

Desert mammals have evolved a variety of physiological and behavioral adaptations to live in hot, arid environments, including metabolic water production (MacMillen & Hinds, 1983; reviewed in Walsberg 2000), renal adaptations (Schmidt-Nielsen et al., 1948; Schmidt-Nielsen & Schmidt-Nielsen, 1952; Vimtup & Schmidt-Nielsen, 1952; Dantzler, 1982; Diaz, Ojeda & Rezenda, 2006; Urity et al., 2012), large ear size (Schmidt-Nieslen, 1964; Hill & Veghte, 1976), burrowing (in Vorhies, 1945; Kelt, 2011), nocturnality (Stephens & Tello, 2009; Fuller et al., 2014), and seasonal torpor (Kalabukhov 1960; Geiser, 2010). *Peromyscus eremicus* possesses several of these adaptations; the cactus mouse is nocturnal and enters torpor during the hottest summer months, it uses burrows, and it produces metabolic water (reviewed in Veal and Caire, 1979; MacMillan and Garland, 1989). However, the cactus mouse does not possess the same kidney architecture as the kangaroo rat; namely, *P. eremicus* do not concentrate urine via elongated loops of Henle (Dewey et al., 1966; MacManes 2016, unpublished data). Thus, recent physiological and genetic studies have begun to pursue the currently unknown mechanisms for renal adaptations in the cactus mouse (MacManes and Eisen, 2014; Kordonowy et al., 2017b; MacManes, 2017). However, my dissertation research is the first to hypothesize and explore potential reproductive adaptations in this species (Kordonowy and MacManes, 2016).
The rationale for evaluating potential reproductive adaptations is that desert-adapted organisms are subject to considerable selective pressures for successful reproduction despite water-limitation. Reproductive proteins have been shown to undergo rapid rates of evolution (reviewed in Swanson and Vacquier, 2002; Ramm et al., 2014). For example, among rodents, this has been demonstrated in *Mus musculus* epididymis-specialized genes (Dean et al., 2008), in sperm proteins involved in motility and in sperm-egg interactions for various mouse strains (Vicens et al., 2014), and in *Peromyscus* testes proteins (Turner et al., 2008). Therefore, I predicted that selective habitat forces in deserts are capable of shaping reproductive proteins in response to dehydration in *P. eremicus* (Kordonowy and MacManes, 2016).

Several studies have explored the role of water as a cue for reproduction in desert rodents (but not cactus mouse), as well as the suppressive reproductive effects of dehydration (reviewed in Schwimmer and Haim, 2009; Bales and Hostetler 2011; tested in: Yahr and Kessler, 1975; Breed, 1975, Christian 1979; Breed and Leigh, 2011; Henry and Dubost, 2012; Sarli et al., 2015; Sarli et al., 2016). The physiological mechanisms responsible for these reproductive effects are currently unknown, though some researchers have proposed that arginine vasopressin, an antidiuretic hormone, is involved (reviewed in Schwimmer and Haim, 2009; tested in: Shanus and Haim, 2004; Wube et al., 2008; Bukovetzky et al., 2012a; 2012).

However, prior to evaluating evidence for reproductive adaptations in the male cactus mouse, we needed to produce a strong foundational genetic base to pursue this hypothesis; specifically, we required comprehensive genetic data for the investigated tissues. I chose to generate genetic data for four male reproductive tissues involved in sperm development and movement: the testes, epididymis, vas deferens, and seminal vesicles. Firstly, I sequenced and assembled a transcriptome for the testes, epididymis and vas deferens. I generated a proteome
for the seminal vesicles, because they are extraordinarily protein rich, making them ideal for building a proteome while unconducive for sequencing a transcriptome. The transcriptomic and proteomic data for these four tissues has allowed us to begin to pursue our research question on potential male cactus mouse reproductive adaptations.

My research utilizes bioinformatics to evaluate genetic data from male cactus mouse reproductive tissues. Although the genetic mechanisms responsible for adaptive morphologies among species residing in extreme environments remain largely unknown, this is a rapidly expanding area of research (e.g. Cheviron and Brumfield, 2011; Bedford and Hoekstra, 2015; Herrera, Watanabe & Shank, 2015). This expansion is largely a result of genomic resources allowing researchers to elucidate the ultimate mechanisms of adaptive evolution with higher sensitivity (e.g. Guillen et al., 2015; MacManes and Eisen, 2014). Recent adaptation research in other Peromyscus species has successfully used genomic and transcriptomic approaches (reviewed in: Bedford and Hoekstra, 2015; Munshi-South and Richardson, 2016). Furthermore, several studies have evaluated differential gene expression and genes under positive selection in desert-rodent kidneys (Marra et al., 2012; Marra et al., 2014; MacManes & Eisen, 2014; MacManes, 2017). Differential expression analyses can identify genes with higher or lower expression in species occupying extreme environments, and these candidate genes serve as the focus for additional analyses to determine whether they play a functional role in adaptation. The manuscript (in review, BMC Genomics) in Chapter 2 uses this approach to address my dissertation research question.

Summary of Dissertation Research and Thesis Chapter Descriptions

Genomic, transcriptomic and proteomic resources utilized within the context of evolutionary biology provide incontrovertibly powerful approaches to explore the genetic
mechanisms for organismal adaptations to extreme environments. My dissertation describes the harnessing of genetic resources to pursue the following evolutionary biology research question:

*What reproductive adaptations do desert-specialized male Peromyscus eremicus (cactus mouse) possess to combat dehydration?* In order to address this line of inquiry, we have generated considerable genomic (MacManes, *unpublished data*), transcriptomic, and proteomic data for the relevant male reproductive tissues in this study species.

Chapter 1 is a published manuscript (Kordonowy and MacManes, 2016) characterizing a comprehensive three-tissue male reproductive transcriptome for *P. eremicus*. This transcriptomic data provides an essential genetic resource for reproductive tissue research in this species. Indeed, we leveraged this transcriptomic resource for our differential expression manuscript (described below). This Chapter 1 manuscript also includes a comparative analysis of reproductive tissues across three rodent species to identify genes with signatures of positive selection in cactus mouse. These candidate genes should be of interest to future studies addressing our research question on male cactus mouse reproductive adaptations.

Chapter 2 is a manuscript currently in review at *BMC Genomics* (Kordonowy and MacManes, 2017), which describes a large scale differential gene expression study, wherein we experimentally exposed 22 male cactus mice to either acute dehydrated or hydrated treatment conditions. We performed a robust three-tiered statistical analysis for evaluating differential expression of genes in testes between treatments, and our results surprisingly alluded to reproductive modulation among testes of acutely dehydrated mice. Our work illustrates the need for more research on reproductive effects of chronic and acute dehydration in this species. The identified differentially expressed genes in this manuscript should be evaluated in future research
to causatively link their altered expression levels with reproductive mitigation; however, our findings are unexpected in a species highly adapted for desert-habitation.

Chapter 3 is a manuscript that will be submitted imminently to Proteomics as a Dataset Brief (Kordonowy et al., 2017a) describing the seminal vesicle proteome for *P. eremicus*. This proteome was generated using three hydrated and three dehydrated cactus mice, because it is essential that it be representative of both hydration states that this species experiences. We describe the gene ontology for the proteome. Our intent is that future research will use this proteome to evaluate protein composition differences between experimentally manipulated differentially hydrated laboratory mice and wild caught cactus mice to explore reproductive protein modulation responses to drought. Such work would both significantly contribute to our limited understanding of dehydration effects on fertility and reproductive success in cactus mouse as well as aid in evaluating our novel male reproductive adaptation hypothesis.

REFERENCES:


Doi: 10.1530/jrf.0.0450273.


http://doi.org/10.1016/j.cbpa.2012.04.007


DOI :10.3389/fpls.2016.01044.


**Fuller A, Hetem RS, Maloney SK, Mitchell D. 2014.** Adaptation to Heat and Water Shortage in Large, Arid-Zone Mammals. Physiology 29(3): 159-167. DOI: 10.1152/physiol.00049.2013

**Geiser F. in** Aestivation: Molecular and Physiological Aspects (Volume 49 in the series Progress in Molecular and Subcellular Biology, Springer, Berlin, 2010, Navas A, Carvalho C, Eduardo J.) 95-111. DOI:10.1007/978-3-642-02421-4_5


Kordonowy L, MacManes L. 2017. Characterizing the reproductive transcriptomic correlates


Doi: 10.14814/phy2.13218


DOI: 10.1093/gbe/evw155

**Marra NJ, Eo SH, Hale MC, Waser PM, DeWoody JA. 2012.** *A priori* and *a posteriori*
approaches for finding genes of evolutionary interest in non-model species: Osmoregulatory
genes in the kidney transcriptome of the desert rodent Dipodomys spectabilis (banner-tailed

osmoregulation in Heteromyid rodents as revealed by RNA-seq. Molecular Ecology 23(11):
2699-2711. DOI: 10.1111/mec.12764.

Munshi-South J, Richardson JL. 2016. Peromyscus transcriptomics: understanding adaptation
and gene expression plasticity within and between species of deer mice. Seminars in Cell &

Plachetzki DC, Pankey MS, Johnson BR, Ronne EJ, Kopp A, Grosberg RK. 2014. Gene co-
expression modules underlying polymorphic and monomorphic zooids in the colonial
Hydrozoan, Hydractinia symbiolongicarpus. Integrative and Comparative Biology 54(2): 276-
283. DOI: 10.1093/icb/icu080

Ramm SA, Scharer L, Ehmcke J, Wistuba J. 2014. Sperm competition and the evolution of

Sarli J, Lutermann H, Alagali AN, Mohammed OB, Bennett NC. 2015. Reproductive
patterns in the Baluchistan gerbil, Gerbillus nanus (Rodentia: Muridae), from western Saudi
http://dx.doi.org/10.1016/j.jaridenv.2014.09.007

Sarli J, Lutermann H, Alagali AN, Mohammed OB, Bennett NC. 2016. Seasonal
reproduction in the Arabian spiny mouse, Acomys dimidiatus (Rodentia: Muridae) from Saudi


Doi: 10.1016/j.cbpa.2008.06.027

CHAPTER 1

Characterization of a male reproductive transcriptome for *Peromyscus eremicus* (Cactus mouse)

Lauren L. Kordonowy* and Matthew D. MacManes*

Department of Molecular, Cellular, and Biomedical Sciences, University of New Hampshire, Durham, NH, United States

* These authors contributed equally to this work.

Submitted 20 April 2016

Accepted 27 September 2016

Published 27 October 2016

Corresponding author: Lauren L. Kordonowy, lauren.kordonowy@unh.edu

Academic editor: Ana Pavasovic

Additional Information and Declarations can be found on page 17

DOI 10.7717/peerj.2617

©Copyright 2016 Kordonowy and MacManes

Distributed under Creative Commons CC-BY 4.0

OPEN ACCESS

How to cite this article: Kordonowy and MacManes (2016), Characterization of a male reproductive transcriptome for *Peromyscus eremicus* (Cactus mouse). *PeerJ* 4:e2617; DOI 10.7717/peerj.2617
ABSTRACT

Rodents of the genus *Peromyscus* have become increasingly utilized models for investigations into adaptive biology. This genus is particularly powerful for research linking genetics with adaptive physiology or behaviors, and recent research has capitalized on the unique opportunities afforded by the ecological diversity of these rodents. Well characterized genomic and transcriptomic data is intrinsic to explorations of the genetic architecture responsible for ecological adaptations. Therefore, this study characterizes the transcriptome of three male reproductive tissues (testes, epididymis and vas deferens) of *Peromyscus eremicus* (cactus mouse), a desert specialist. The transcriptome assembly process was optimized in order to produce a high quality and substantially complete annotated transcriptome. This composite transcriptome was generated to characterize the expressed transcripts in the male reproductive tract of *P. eremicus*, which will serve as a crucial resource for future research investigating our hypothesis that the male cactus mouse possesses an adaptive reproductive phenotype to mitigate water-loss from ejaculate. This study reports genes under positive selection in the male cactus mouse reproductive transcriptome relative to transcriptomes from *Peromyscus maniculatus* (deer mouse) and *Mus musculus*. Thus, this study expands upon existing genetic research in this species, and we provide a high quality transcriptome to enable further explorations of our proposed hypothesis for male cactus mouse reproductive adaptations to minimize seminal fluid loss.

Subjects: Bioinformatics, Evolutionary Studies, Genetics, Genomics, Zoology

Keywords: *Peromyscus eremicus*, Transcriptome, Genomics, Bioinformatics, Adaptation, Desert physiology, Cactus mouse, Reproduction
INTRODUCTION

The rapid infusion of novel bioinformatics approaches in the fields of genomics and transcriptomics has enabled the coalescence of the fields of genetics, physiology and ecology into innovative studies for adaptation in evolutionary biology. Indeed, studies on the biology of adaptation had previously been dominated by research painstakingly documenting morphological shifts associated with ecological gradients (e.g., in Peromyscus: Carleton, 1989; MacMillen & Garland, 1989). However, the discipline of bioinformatics has breathed new life into the field of adaptation biology. Specifically, while the morphological basis as well as the physiological mechanisms of adaptation have been explored for a variety of species in extreme environments, the genetic underpinnings of these adaptations have only recently become a larger area of research (Hoekstra et al., 2006; Cheviron & Brumfield, 2011; Lorenzo et al., 2014; Guillen et al., 2015). High-throughput sequencing technology of model and non-model organisms (Ellegren, 2014) enables evolutionary biologists to conduct genome and transcriptome wide analyses and link patterns of gene selection with functional adaptations.

Studies on the genetic basis of adaptation have included a wide variety of taxa. For example, butterflies in the Heliconius genus have been a particularly effective study system for determining the genetic basis of pigmentation patterns, and there is evidence of interspecific introgression for genes enabling adaptive mimicry patterns (Hines et al., 2012; Heliconius Genome Consortium, 2012). In addition, a population genomic study in three-spine sticklebacks has elucidated many loci responsible for divergent adaptations from marine to freshwater environments (Jones et al., 2012). Furthermore, researchers have developed a list of candidate genes that have evolved in multiple populations of freshwater adapted three-spine
sticklebacks (Hohenlohe et al., 2010). Another active area of adaptation genetics research focuses on species residing in extreme environments. High altitude adaptations to hemoglobin variants have been identified in multiple organisms, including humans (Lorenzo et al., 2014), several species of Andean ducks (Mccracken et al., 2009a; Mccracken et al., 2009b), and deer mice, *Peromyscus maniculatus* (Storz et al., 2010; Natarajan et al., 2015). The genetic pathways responsible for physiological adaptations to desert habitats remain enigmatic; however, considerable progress has been made developing candidate gene sets for future analyses (e.g., Guillen et al., 2015; MacManes & Eisen, 2014; Marra et al., 2012; Marra, Romero & DeWoody, 2014). Functional studies will stem from this foundational research aimed at identifying the genomic underpinnings of adaptations to extreme environments; yet, it is inherently challenging and critically important to demonstrate that specific loci are functionally responsible for adaptations (Storz & Wheat, 2010).

Rodents of the genus *Peromyscus* have been at the forefront of research elucidating the genetic basis for adaptation (reviewed in Bedford & Hoekstra, 2015). This diverse genus has served as an ideal platform for adaptation research spanning from the genetic basis of behavioral adaptations—such as complex burrowing in *Peromyscus polionotus* (Weber & Hoekstra, 2009; Weber, Peterson & Hoekstra, 2013)—to the loci responsible for adaptive morphology—such as coat coloration in *Peromyscus polionotus leucocephalus* (Hoekstra et al., 2006), and including kidney desert adaptations in *Peromyscus eremicus* (MacManes & Eisen, 2014). We are currently using *Peromyscus eremicus* as a model species for investigating the genetic bases of desert adaptations. This paper describes a crucial component of this research aim.

Initial steps toward understanding the genetics of adaptation must include the genomic and transcriptomic characterization of target study species (MacManes & Eisen,
To this end, we assembled and characterized a composite transcriptome for three male reproductive tissues in the desert specialist, *P. eremicus*. This species is an exceptional example of desert adaptation, as individuals may live entirely without water access (*Veal & Caire, 2001*). *MacManes & Eisen (2014)* assembled transcriptomes from kidney, hypothalamus, liver, and testes of this species, and they identified several candidate genes potentially underlying adaptive renal physiology. However, to our knowledge, potential physiological reproductive adaptations to water limitation have not been studied in this species or in other desert rodents. We hypothesize that male *P. eremicus* possess reproductive adaptations to mitigate water loss.

The adaptive kidney physiology in kangaroo rat species (genus: *Dipodomys*), which produce highly concentrated urine via a disproportionately long loop of Henle (*Schmidt-Nielsen et al., 1948; Schmidt-Nielsen & Schmidt-Nielsen, 1952; Vimtrup & Schmidt-Nielsen, 1952; Urity et al., 2012*), has been well established. We propose that there may also be reproductive adaptations to mitigate water loss in desert rodents. Recent findings pertaining to the genetic signatures of adaptive kidney function in desert rodents (*Marra et al., 2012; Marra, Romero & DeWoody, 2014; MacManes & Eisen, 2014*), suggest that such hypothesized reproductive adaptations may be detectable at the genetic level (should they exist) using similar comparative transcriptomic methods. We present this hypothesis in light of the mounting body of research for high rates of reproductive protein evolution (reviewed in *Swanson & Vacquier, 2002; Ramm et al., 2014*), which we propose indicates that reproductive tissues may possess a significant capacity for evolving in response to strong selective pressures.

*Dewsbury (1982)* made the assertion that producing ejaculates incurs a cost for male mammals, which produce relatively high sperm counts, and his analysis utilizes *P.
**maniculatus** as a model. Ejaculation has also been demonstrated to be costly in *Mus musculus domesticus; Ramm & Stockley (2007)* found that mice are able to manipulate the quantity of sperm released in response to varying competition for females. *Peromyscus* and other rodents exhibit rapid evolution of testis-expressed proteins (*Turner, Chuong & Hoekstra, 2008*). In addition, the epididymal transcriptome of *M. musculus* shows evidence for positive selection among epididymis-specialized genes that are secreted, which the authors attribute to their putative evolutionary importance (*Dean, Good & Nachman, 2008*). Moreover, a recent analysis for sperm genes from multiple mouse strains found that sperm proteins involved in both motility and in sperm-egg interactions show signatures of positive selection, potentially facilitating evolutionary mechanisms for sperm-competition and sexual conflict (*Vicens, Luke & Roldan, 2014*). We therefore propose that both the inherently costly nature of producing ejaculate and the rapid evolution of genes in murine testes, epididymis, and sperm would be ideal for desert rodents to evolve ejaculate adaptations to limited water availability.

In order to develop genomic resources that will allow us to begin to test our hypothesis related to reproductive water conservation, we developed a transcriptome comprising three male specific reproductive tissues in *P. eremicus*. Specifically, we assembled a composite reproductive transcriptome for the epididymis, testes, and vas deferens. These tissues were chosen because they constitute numerous physiological roles in spermatogenesis and in the generation and transportation of seminal fluids. We posit that one or more of these reproductive tissues possess phenotypic characteristics to mitigate water loss from seminal fluids. We also propose that any molecular mechanisms responsible for such an adaption should be elucidated through future differential gene expression and comparative transcriptomic studies. However, the initial assembly and characterization of a
comprehensive male reproductive tissue transcriptome in the current manuscript will be essential for research studies exploring this hypothesized male *P. eremicus* reproductive desert adaptation. For example, this transcriptome will be instrumental in further experimental studies investigating differential gene expression in these tissue types in response to variable hydration levels or in large-scale comparative transcriptomic studies spanning numerous desert and non-desert rodents.

Here, we characterize a male cactus mouse tissue-specific transcriptome by presenting a preliminary exploration of transcript abundance and putative homology. We also perform comparative transcriptomic analyses to identify evidence of positive selection in genes potentially related to the hypothesized reproductive desert adaptations. It is beyond the scope of this manuscript to evaluate the functionality of these candidate genes in the context of desert adaptation, much less male specific reproductive adaptations. However, the elucidation of candidate genes in the context of male reproductive tissues will be instrumental for future studies aimed at determining which genes are functionally responsible for the proposed reproductive adaptations to water limitation in male cactus mouse.

**METHODS**

**Tissue samples, RNA extraction, cDNA library preparation and sequencing**

The *Peromyscus eremicus* male used for this study was captive born and descendant from a population from the *Peromyscus* Genetic Stock Center (Columbia, South Carolina). This individual was housed in a facility at the University of New Hampshire designed to mimic desert conditions in the Southwestern United States. Specifically, the temperature increases gradually during the light hours until it peaks at 90° Fahrenheit in the afternoon,
and the temperature decreases during hours of darkness to 75°. Humidity levels are 10% during the daylight hours and 25% during darkness. The photoperiodic cycle in this desert chamber is for long days of photostimulation, with 16 h of light, and 8 h of darkness. The colony includes males and females which are housed within a single room, providing olfactory cues that stimulate reproductive maturity. The photoperiod and shared housing in this colony result in the attainment of reproductive maturity in both sexes. Males are deemed reproductively mature when they are fully scrotal. The males do not undergo seasonal testicular atrophy, as evidenced by their consistent scrotal condition and their year-round successful reproduction.

A single reproductively mature *P. eremicus* male was sacrificed via isoflurane overdose and decapitation. This was done in accordance with University of New Hampshire Animal Care and Use Committee guidelines (protocol number 130902) and guidelines established by the American Society of Mammalogists (*Sikes et al., 2011*). Testes, epididymis, and vas deferens were immediately harvested (within ten minutes of euthanasia), placed in RNAlater (Ambion Life Technologies) and stored at −80 °C until RNA extraction. We used a standard TRizol, chloroform protocol for total RNA extraction (Ambion Life Technologies). We evaluated the quantity and quality of the RNA product with a Qubit 2.0 Fluorometer (Invitrogen) and a Tapestation 2200 (Agilent Technologies, Palo Alto, USA).

We used a TURBO DNAse kit (Ambion) to eliminate DNA from the samples prior to the library preparation. Libraries were made with a TruSeq Stranded mRNA Sample Prep LS Kit (Illumina). Each of the three samples was labeled with a unique hexamer adapter for identification after multiplex single lane sequencing. Following library completion, we
confirmed the quality and quantity of the DNA product with the Qubit and Tapestation. We submitted the multiplexed sample of the libraries for running on a single lane at the New York Genome Center Sequencing Facility (NY, New York). Paired end sequencing reads of length 125 bp were generated on an Illumina 2500 platform. Reads were parsed by tissue type according to their unique hexamer IDs in preparation for transcriptome assembly.

**Reproductive transcriptome assembly**

The composite reproductive transcriptome was assembled with reads from the testes, epididymis and vas deferens using the previously developed Oyster River Protocol for *de novo* transcriptome assembly pipeline (MacManes, 2016). Briefly, the reads were error corrected with Rcorrector v1.0.1 (Song & Florea, 2015). We used the *de novo* transcriptome assembler Trinity v2.1.1 (Haas et al., 2013; Grabherr et al., 2011). Within the Trinity platform, we ran Trimmomatic (Bolger, Lohse & Usadel, 2014) to remove the adapters, and we also trimmed at PHRED < 2, as recommended by MacManes (2014).

Next we evaluated transcriptome assembly quality and completeness using BUSCO v1.1b1 and Transrate v1.0.1. BUSCO (Simão et al., 2015) reports the number of complete, fragmented, and missing orthologs in assembled genomes, transcriptomes, or gene sets relative to compiled ortholog databases. We ran BUSCO on the assembly using the ortholog database for vertebrates, which includes 3,023 genes. The assembly was also analyzed by Transrate using the *Mus musculus* peptide database from Ensembl (downloaded 2/24/16) as a reference. The Transrate score provided a metric of *de novo* transcriptome assembly quality, and the software also generated an improved assembly comprised of highly supported contigs (Smith-Unna et al., 2016). Finally, we re-ran BUSCO on the improved assembly
generated by Transrate to determine if this assembly had similar metric scores for completeness as the original assembly produced by Trinity. As alternatives to the original Trinity assembly and the optimized Transrate assembly, we proceeded with our optimization determinations by filtering out low abundance contigs from the original Trinity assembly. First we calculated the relative abundance of the transcripts with Kallisto v0.42.4 and Salmon v0.5.1. Kallisto utilizes a pseudo-alignment algorithm to map RNA-seq data reads to targets for transcript abundance quantification (Bray et al., 2015). In contrast, Salmon employs a lightweight quasi-alignment method and a high speed streaming algorithm to quantify transcripts (Patro, Duggal & Kingsford, 2015). After determining transcript abundance in both Kallisto and Salmon, we removed contigs with transcripts per million (TPM) estimates of less than 0.5 and of less than 1.0 in two separate optimization trials (as per MacManes, 2016). Finally, we evaluated these two filtered assemblies with Transrate and BUSCO to determine the relative quality and completeness of both assemblies. We chose the optimal assembly version by comparing Transrate and BUSCO metrics and also through careful consideration of total contig numbers across all filtering and optimizing versions. The chosen assembly was the Transrate optimized TPM > 0.5 filtered assembly, and this assembly was used for all subsequent analyses.

**Annotation, transcript abundance, and database searches**

We used dammit v0.2.7.1 (Scott, 2016) to annotate the optimized transcriptome assembly (as per MacManes, 2016). Within the dammit platform, we predicted protein coding regions for each tissue with TransDecoder v2.0.1 (Haas et al., 2013), which was used to find open reading frames (ORFs). Furthermore, dammit utilizes multiple database searches
for annotating transcriptomes. These database searches include searches in Rfam v12.0 to find non-coding RNAs (Nawrocki et al., 2015), searches for protein domains in Pfam-A v29.0 (Sonnhammer, Eddy & Durbin, 1997; Finn et al., 2016), the execution of a LAST search for known proteins in the UniRef90 database (Suzek et al., 2007; Suzek et al., 2015), ortholog matches in the BUSCO database, and orthology searches in OrthoDB (Kriventseva et al., 2015).

We used the assembly annotated by dammit to re-run Kallisto to determine transcript abundance within each of the three tissue types. We used TPM counts of expression for all three tissues to generate counts of transcripts specific to and shared across tissue types. We also downloaded the ncRNA database for Mus musculus from Ensembl (v 2/25/16), and we did a BLASTn (Altschul et al., 1990; Madden, 2002) search for these ncRNAs in our assembly. This database has 16,274 sequences, and we determined the number of transcript ID matches and the number of unique ncRNA sequence matches for our assembly. We also counted how many transcript matches were present in each of the tissues, and we referenced the corresponding Kallisto derived TPM values to determine the number of unique and ubiquitous transcript matches for each tissue.

We searched the annotated assembly for transporter protein matches within the Transporter Classification Database (tcdb.org). This database has 13,846 sequences representing proteins in transmembrane molecular transport systems (Saier et al., 2014). We executed a BLASTx (Altschul et al., 1990; Madden, 2002) search to find the number of transcript ID matches and the number of unique transporter protein matches within the assembly. Next we determined how many transcript ID matches were found in each of the three tissues. As previously described above, we also cross-referenced these matches with
the Kallisto derived TPM values to find the number of ubiquitous and unique transcript matches by tissue type.

**Comparative analysis for genes under positive selection**

We performed a three-species comparative analysis to identify genes under positive selection in the male reproductive transcriptome for the *P. eremicus* lineage relative to *M. musculus* and *P. maniculatus*. The *M. musculus* nucleotide and protein sequences were downloaded (version GRCm38) from Ensembl (ensembl.org). The *P. maniculatus* nucleotide and protein files were downloaded (version GCF_000500345.1_Pman_1.9) from NCBI (ncbi.nlm.nih.gov). The comparative analysis was conducted as described below. The corresponding nucleotide and protein files for all three species were modified so that the header names for the sequences in each species file pair were concordant. Next, we found orthologous groups of protein sequences using OrthoFinder v0.6.1 (*Emms & Kelly, 2015*). The two OrthoFinder command scripts aligned sequences with MAFFT v7.123b (*Katoh et al., 2002*), built trees with FastTreeMP v1 (*Price, Dehal & Arkin, 2010*), and generated orthogroups based on sequence similarity. Our next script selected the single copy orthologs (SCOs) from among these orthogroups for analysis. Then we selected the cds file transcripts for all three species corresponding to the previously identified single copy orthologs. Finally, we ran a script which aligned the sequences with PRANK v150803 (*Löytynoja, 2014*) and performed the analyses for positive selection for the single copy orthologous sequences with codeml in PAML v4.8 (*Yang, 1997; Yang, 2007*). Specifically, we performed the M2a branch site test (*Zhang, Nielsen & Yang, 2005*) for positive selection after stipulating *P. eremicus* as the foreground lineage. Genes were deemed under positive selection if the omega values
(w) exceeded 1 within the $M2a$ model and if they yielded statistically significant results for the likelihood ratio test (LRT) comparison between the null and alternative models. To perform the LRT, we determined if $2\Delta \ln L$ (two times the difference between the log likelihood values for the alternative—$M2a$—and the null—$M1a$—models) exceeded a chi-square value corresponding to a significant $p$-value ($p < 0.05$) (Yang, 1998) after applying the Benjamini-Hochburg correction for multiple comparisons (Benjamini & Hochberg, 1995). For each of the $P. eremicus$ gene sequences demonstrated to be under positive selection according to the LRT, we selected the $M. musculus$ cds sequence from the SCO group corresponding to the $P. eremicus$ gene sequence. Then we performed a BLASTn search on the $M. musculus$ sequence to find gene matches on NCBI to determine the gene identity for the $P. eremicus$ sequence under positive selection.

Of note, the code for performing all of the above analyses can be found at GitHub (https://github.com/macmanes-lab/peer_reproductive/transcriptome). The data files are available on Dryad (doi: 10.5061/dryad.01c3t).

RESULTS AND DISCUSSION

Reproductive transcriptome assembly

There were 45–94 million paired reads produced for each of the three transcriptome datasets, yielding a total of 415,960,428 reads. The raw reads are available at the European Nucleotide Archive under study accession number PRJEB13364.

We assembled a de novo composite reproductive transcriptome with reads from testes, epididymis and vas deferens. The evaluation of alternative optimized assemblies allowed us to generate a substantially complete transcriptome of high quality. The
alternative assemblies had raw Transrate scores ranging from 0.156–0.194 (Table 1). However, the scores for the improved assemblies generated by Transrate, consisting of only highly supported contigs, ranged between 0.285 and 0.349, which is well above the threshold Transrate score of 0.22 for an acceptable assembly. The BUSCO results indicated that the assemblies were highly complete, with complete matches ranging from 73–90% of vertebrate orthologs (Table 2). These BUSCO benchmark values are consistent with the most complete reported assessments for transcriptomes from other vertebrate taxa (busco.ezlab.org). Furthermore, our BUSCO values exceed that of the only available reported male reproductive tissue (from a coelacanth: Latimeria menadoensis testes), which was 71% complete (Simão et al., 2015). The assembly version with the highest quality in relation to the Transrate metrics was the Transrate optimized Trinity assembly.

Specifically, the optimized Transrate score was 0.3492, and the percent coverage of the reference assembly was also highest, with 45% of the mouse database represented. This assembly was highly competitive for completeness, as indicated by the fact that it contained 85% of vertebrate single copy orthologs. However, this assembly had an exorbitantly high number of contigs (657,952 contigs), which is nearly an order of magnitude more contigs than the next best performing assembly: the Transrate optimized TPM > 0.5 filtered assembly (78,424 contigs). In consideration of the dramatically more realistic contig number for the Transrate optimized TPM > 0.5 filtered assembly, and in light of its second best performance for Transrate score (0.3013), reasonable Transrate mouse reference assembly coverage (37%), and sufficiently high BUSCO completeness (73% orthologs found), we chose this assembly as our optimized transcriptome. Therefore, we proceeded with this optimized assembly version as our finalized transcriptome assembly for our analyses.
Annotation, transcript abundance and database searches

The reproductive transcriptome assembly annotations were produced by dammit, and they are available on Dryad in a gff3 file format. Furthermore, TransDecoder was used to predict coding regions in the assembly. TransDecoder predicted that 49.5% (38,342) of the transcripts (78,424 total) contained ORFs, of which 63.9% (24,808) had complete ORFs containing a start and stop codon. The predicted protein coding regions generated by TransDecoder are reported in five file types, and they are available on Dryad. Furthermore, the Pfam results yielded 30.7% of transcripts (24,107) matching to the protein family database. In contrast, the LAST search found that 75.9% of transcripts (59,503) matched to the UniRef90 database. We have uploaded the homology search results generated by Pfam and UniRef90 matches onto Dryad. In addition, 1.04% (816) of transcripts matched to the Rfam database for ncRNAs, and these results are posted in Dryad. Of note, 80.1% (62,835) of the transcripts were annotated using one or more of the above described methods (the dammit.gff3 file is posted in Dryad), and it is this final annotated assembly that was used for all subsequent analyses (this annotated transcriptome is available on Dryad).

The Kallisto generated TPM counts of expression (available on Dryad) were utilized to determine which transcripts were ubiquitous and specific to the three tissue types, which we have depicted in a Venn diagram format (Fig. 1). The assembly consisted of 78,424 different transcript IDs, of which 64,553 were shared across all three tissues. The number of unique transcripts were as follows: 3,563 in testes, 342 in epididymis, and 502 in vas deferens. The relatively large number of unique transcripts in the testes is consistent with previous findings which describe the testes as the tissue with the highest number of
tissue-enriched genes in the human body (Djureinovic et al., 2014; Uhlén et al., 2015).

However, because this expression data was generated by a single individual, we want to emphasize that these results have no statistical power. Rather, we view these numerical comparisons of unique and ubiquitous transcript counts as an exploratory evaluation of potential relationships between tissue types. Such comparisons across tissue types for a single individual of this species previously indicated that the kidney had relatively higher numbers of unique transcripts than testes (MacManes & Eisen, 2014). Future research with multiple individuals will be necessary to statistically evaluate the relative rates of transcript expression between these reproductive tissues, as well as their relationship with other non-reproductive tissue types.

In addition, we searched for Mus musculus ncRNA sequence matches within our assembly. There were 15,964 transcript matches, which correspond to 2,320 unique ncRNA matches, and they are posted on Dryad. The transcript matches by tissue type were found using the Kallisto TPM determinations, and they were as follows, testes: 15,260, epididymis: 15,552, and vas deferens: 15,558. A Venn Diagram depicts unique and shared transcript matches by tissue type (Fig. 2). The majority of transcript matches were ubiquitous to all three tissues (14,724), and there were far fewer tissue specific matches. The testes had more unique transcript matches (185) than the epididymis (26) or the vas deferens (45). These findings are consistent with our results above regarding the relative numbers of total unique transcripts in the assembly by tissue type. However, these counts for relative transcript matches among tissue types were generated with transcripts from a single individual; therefore, the comparative results across tissue type were not statistically evaluated. It is seemingly probable that the diversity of transcripts for regulation should be
highest in tissues generating relatively diverse proteins, in this case, the testes, which did have the highest number of unique transcript matches. The role of ncRNA in reproductive tissues throughout multiple developmental stages has recently been reviewed in detail (Hale, Yang & Ross, 2014). In addition, ncRNAs have been found to be highly abundant in murine testes (Sun, Lin & Wu, 2013). Furthermore, sperm from humans and mice contain a significant number of ncRNAs (Krawetz et al., 2011; Kowano et al., 2012). However, we are unaware of any research investigating the involvement of ncRNA in desert adaptations; therefore, we cannot speculate on particular ncRNA matches within our dataset that may have potential desert adaptive roles.

Our search for transporter protein matches within the Transporter Classification Database yielded 7,521 different transcript matches, corresponding to 1,373 unique transporter protein matches, and they are posted on Dryad. The number of transcript matches was highly similar between the tissue types (testes 7,025; epididymis 7,115; vas deferens: 7,071). We generated a Venn Diagram to display the numbers of shared and unique transcript matches to the transporter protein sequences (Fig. 3). Most transcript matches were present in all three tissues (6,472), and there were relatively few unique matches in the three tissue types. However, the testes had the highest number of unique transcript matches (215) relative to the epididymis (19) and the vas deferens (37). These comparative results across tissue types represent exploratory findings for a single mouse, and the count data have not been statistically tested, but these preliminary results should be investigated in future studies. Furthermore, our BLASTx search of this transporter protein database yielded transcript matches for multiple solute carrier proteins. We are particularly interested in solute carrier proteins because previous research has found candidate genes in
this protein family for desert adaptations in kidneys of the kangaroo rat (Marra et al., 2012; 
Marra, Romero & DeWoody, 2014) and the cactus mouse (MacManes & Eisen, 2014). In addition, 
we had multiple matches to aquaporins, which are water channels allowing transport across 
cellular membranes. One transcript matched specifically to Aquaporin 3, a sperm water 
channel found in mice and humans, which is essential to maintaining sperm cellular integrity 
in response to the hypotonic environment within the female reproductive tract (Chen et al., 2011).

Comparative analysis for genes under positive selection

To find evidence for genes undergoing positive selection in the male reproductive 
transcriptome of P. eremicus, we compared this species with two other generalist rodents. We 
chose M. musculus as the non-desert adapted outgroup because this species possesses 
transcriptomic resources which are exceptional in their annotation and completeness. The 
widely distributed (Carleton, 1989) habitat generalist deer mouse, P. maniculatus, was chosen 
because it harbors the most complete transcriptomic data available among the Peromyscus 
genus.

There were 3,731 panorthologous groups (single copy orthologs) in our three 
species comparison. The branch test was successfully implemented for SCOs when all 
three sequences aligned adequately with PRANK and when codeml produced both M1a 
and M2a output files for the LRT comparison (n=2,820 in total). The M2a test indicated that 
42 genes were evolving under a model of positive selection in the Cactus mouse (Table 3).

Therefore, we investigated whether previous research on either rodent reproductive 
tissues or on desert specialized rodents documented evidence of positive selection for any of
these genes or gene families. Only one of these 42 genes matched an epididymis-specialized secreted gene undergoing positive selection in another rodent (C57/BL6 mice). Namely, Qsox2 is a match for Qscn6l1, a member of the sulfhydryl oxidase/quiescin-6 family, which is purportedly involved in neuroblastoma apoptosis (Dean, Good & Nachman, 2008). However, we cannot speculate regarding the functionality this gene has in male rodent reproductive tissue, or why it appears to be evolving under a model of positive selection in these two studies. Another of our 42 positively selected genes, Lrrc46, may have some similarity to Lrrc50, a gene under positive selection in testes of Peromyscus (Turner, Chuong & Hoekstra, 2008). Leucine-rich repeat containing (Lrrc) genes have diverse biological roles; therefore, we also will not speculate on any correspondence between these two genes. As expected given our experimental design, there was no concordance between our 42 genes and those found to be under positive selection in a recent study on mouse spermatozoa proteins (Vicens, Luke & Roldan, 2014).

Our search for gene matches from the current study with other desert rodent research revealed notable similarities. Two solute carrier proteins, Slc15a3 and Slc47a1, were found to be under positive selection in our analysis. This finding bears particular significance because another protein in this family, Slc2a9, shows signatures of positive selection in desert rodent kidney transcriptomes in Dipodomys spectabilis and Chaetodipus baileyi (Marra, Romero & DeWoody, 2014) and P. eremicus (MacManes & Eisen, 2014). Solute carriers are a large family of cell membrane proteins that are responsible for transporting solutes (reviewed in Hediger et al., 2004; Hediger et al., 2013; César-Razquin et al., 2015). Furthermore, Marra, Romero & DeWoody (2014), hypothesize that solute carriers are critical for osmoregulation in desert rodents, and they assert that these genes may be under evolutionary pressure in such rodents.
In response to the potential relevance of the two solute carriers under positive selection in our study to desert rodent osmoregulation, we generated STRING (Snel et al., 2000; Szklarczyk et al., 2015) diagrams for their protein-protein interactions (Fig. 4). These diagrams demonstrate multiple connections to other solute carriers for both proteins, thereby suggesting their potential functional roles.

This analysis of genes undergoing positive selection in *P. eremicus* relative to *P. maniculatus* and *M. musculus* provides candidate genes for desert specialization in the cactus mouse which can be the target of future studies focused on ascertaining which genes may be functionally responsible for our hypothesized male reproductive desert adaptation.

However, desert specialization is not the sole difference between *P. eremicus* and *P. maniculatus*, much less between *P. eremicus* and *M. musculus*. *P. maniculatus* is highly promiscuous, while *P. eremicus* is relatively socially monogamous (Wolff, 1989). Indeed, *P. maniculatus* has been the subject of considerable sperm competition research (Dewsbury, 1988; Fisher & Hoekstra, 2010). Differences in reproductive systems, and even potentially in sperm aggregation of *P. maniculatus* (Fisher & Hoekstra, 2010), may manifest themselves as evidence of selection patterns between these two *Peromyscus* species. Therefore, we are not proposing that the 42 genes we found to be under positive selection in the male Cactus mouse are functionally responsible for adaptive desert physiology. Rather, we are proposing that they are interesting candidate genes for future studies investigating the genetic underpinnings of physiological desert adaptations, including our hypothesized male reproductive adaptation, on a functional level. Several of these genes, specifically those in the solute carrier family, seem particularly promising for such work because they are undergoing rapid evolution in multiple desert rodent species.
CONCLUSIONS

Although researchers have determined that renal adaptations are responsible for mitigating water loss in kangaroo rats via the genitourinary tract (Schmidt-Nielsen et al., 1948; Schmidt-Nielsen & Schmidt-Nielsen, 1952; Vimtrup & Schmidt-Nielsen, 1952; Urity et al., 2012), we present the novel hypothesis that there may also be male reproductive adaptations to arid environments that allow desert specialists like the cactus mouse to conserve water during reproduction. Previous efforts to elucidate the genomic basis of desert adaptations have described candidate genes for adaptive renal physiology in some desert specialized rodents (Marra et al., 2012; Marra, Romero & DeWoody, 2014), including the cactus mouse (MacManes & Eisen, 2014). In light of these findings, we propose that if the male cactus mouse possesses an adaptive reproductive phenotype to mitigate water loss via seminal fluids in response to limited-water availability, such an adaptation will be detectable through transcriptomic analyses. The current study generates and characterizes a transcriptome for male reproductive tissues from the cactus mouse as an initial step towards future efforts to explore this hypothesized reproductive adaptation. This study describes a composite transcriptome from three male reproductive tissues in the desert specialist Peromyscus eremicus. Our analyses include quality and completeness assessments of this reproductive assembly, which we generated using reads from testes, epididymis, and vas deferens of a male cactus mouse. We generate annotations and search relevant databases for ncRNAs and transporter protein sequences. We also describe the degree of ubiquity between transcripts among the three tissues as well as identify transcripts unique to those tissues utilizing preliminary (based on a single individual) transcript differences between tissue types. Furthermore, we find genes evolving under a model of positive selection in the P. eremicus
male reproductive transcriptome relative to *P. maniculatus* and *M. musculus* in order to generate a list of candidate genes for future investigations in desert adaption genetics. Our future research will investigate the hypothesized male reproductive physiological adaptation to water limitation in cactus mouse through a differential gene expression study, and the characterization of this reproductive transcriptome will form the foundation of studies along this vein. Moreover, this research contributes transcriptomic materials to a larger body of work in the expanding field of adaptation genetics, which benefits tremendously from enhanced opportunities for comparative analyses.

**DRYAD DATA FILE LIST**

Final Annotated Reproductive Tissue Transcriptome: *reproductive.annotated.fasta* (127 MB)

Transdecoder (Five Files):

*transdecoder.gff3* (49 MB)

*transdecoder.pep* (27 MB)

*transdecoder.cds* (62 MB)

*transdecoder.mRNA* (172 MB)

*transdecoder.bed* (10 MB)

Pfam Annotation: *reproductive.pfam.gff3* (32 MB) Rfam Annotation: *reproductive.rfam.gff3* (191 KB) Dammit Annotation: *reproductive.dammit.gff3* (98 MB) UniRef90 Annotation: *reproductive.uniref.gff3* (9.5 MB)

Kallisto Results for Annotated Transcriptome (Three Files):

*kallisto.testes.tsv* (3.4 MB)

*kallisto.epi.tsv* (3.4 MB)
kallisto.vas.tsv (3.4 MB)

ncRNA Database Matches (three files):
epi.tpm.plus.ncRNA.txt (4.3 MB)
vas.tpm.plus.ncRNA.txt (4.3 MB)
testes.tpm.plus.ncRNA.txt (4.3 MB) tcdb Database Matches (three files):
epi.tpm.plus.tcdb.txt (946 KB)
vas.tpm.plus.tcdb.txt (946 KB)
testes.tpm.plus.tcdb.txt (946 KB)

Comparative analysis for genes under positive selection (three files): PAMLresults.txt (354 KB)

significant_orthogroups.txt (6 KB)
42SeqsM2aPosSel.txt (89 KB).

ADDITIONAL INFORMATION AND DECLARATIONS

Funding
The work was funded by startup funds from the University of New Hampshire. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures
The following grant information was disclosed by the authors: University of New Hampshire.

Competing Interests
The authors declare there are no competing interests.
Author Contributions

Lauren L. Kordonowy conceived and designed the experiments, performed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.

Matthew D. MacManes conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, reviewed drafts of the paper.

Animal Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

University of New Hampshire Animal Care and Use Committee: (protocol number 130902).

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

European Nucleotide Archive (PRJEB13364, ID: 318672).

Data Availability

The following information was supplied regarding data availability: Dryad DOI: 10.5061/dryad.01c3t; Github: https://github.com/macmanes-lab/peer_reproductivetrancriptome.
**Table 1** Transrate results for the reproductive transcriptome assembly produced by different optimization methods.

<table>
<thead>
<tr>
<th>Assembly</th>
<th>Transrate Score</th>
<th>Optimized Score*</th>
<th># Read Pairs (fragments)</th>
<th>Contigs (n_seqs)</th>
<th># Good Contigs</th>
<th>% Good Contigs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trinity Original</td>
<td>0.1944</td>
<td>0.3492</td>
<td>207,980,214</td>
<td>856,711</td>
<td>657,952</td>
<td>0.77</td>
</tr>
<tr>
<td>Filter TPM&lt;0.5</td>
<td>0.1672</td>
<td>0.3013</td>
<td>207,980,214</td>
<td>147,966</td>
<td>78,424</td>
<td>0.53</td>
</tr>
<tr>
<td>Filter TPM&lt;1.0</td>
<td>0.156</td>
<td>0.2854</td>
<td>207,980,214</td>
<td>80,165</td>
<td>54,140</td>
<td>0.68</td>
</tr>
</tbody>
</table>

**NOTES.**

a This is the score of the Transrate optimized assembly in Table 2.
Table 2 BUSCO metrics for the reproductive transcriptome assembly produced by different optimization methods.

<table>
<thead>
<tr>
<th>Assembly</th>
<th>% Complete</th>
<th>% Duplicated</th>
<th>% Fragmented</th>
<th>% Missing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trinity original</td>
<td>90</td>
<td>49</td>
<td>3.4</td>
<td>5.5</td>
</tr>
<tr>
<td>Transrate Optimized</td>
<td>85</td>
<td>44</td>
<td>4.3</td>
<td>9.7</td>
</tr>
<tr>
<td>Filter TPM&lt;0.5</td>
<td>85</td>
<td>38</td>
<td>3.0</td>
<td>11</td>
</tr>
<tr>
<td>Transrate TPM&lt;0.5</td>
<td>73</td>
<td>31</td>
<td>3.9</td>
<td>22</td>
</tr>
<tr>
<td>Filter TPM&lt;1.0</td>
<td>80</td>
<td>28</td>
<td>2.8</td>
<td>16</td>
</tr>
<tr>
<td>Transrate TPM&lt;1.0</td>
<td>74</td>
<td>25</td>
<td>3.4</td>
<td>21</td>
</tr>
</tbody>
</table>
Table 3 The 42 genes reached statistical significance (p < 0.05) after correcting for multiple hypothesis testing for the M2a branch-site test for positive selection in PAML in the male Cactus mouse reproduction transcriptome.

<table>
<thead>
<tr>
<th>Orthogroup ID</th>
<th>p-value</th>
<th>BLASTn Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>OG0010592</td>
<td>1.18E-10</td>
<td>Q6 sulfhydryl oxidase 2 (Qsox2)</td>
</tr>
<tr>
<td>OG0010774</td>
<td>1.04E-03</td>
<td>adenylate kinase 6 (Ak6)</td>
</tr>
<tr>
<td>OG0010833</td>
<td>4.90E-05</td>
<td>plakophilin 2 (Pkp2)</td>
</tr>
<tr>
<td>OG0011177</td>
<td>3.15E-03</td>
<td>1-acylglycerol-3-phosphate O-acyltransferase 2 (Agpa2)</td>
</tr>
<tr>
<td>OG0011272</td>
<td>1.06E-04</td>
<td>cDNA sequence BC089491</td>
</tr>
<tr>
<td>OG0011374</td>
<td>2.83E-03</td>
<td>hepatoma derived growth factor-like 1 (Hdgfl1)</td>
</tr>
<tr>
<td>OG0011384</td>
<td>2.05E-03</td>
<td>chitinase, acidic 1 (Chia1)</td>
</tr>
<tr>
<td>OG0011551</td>
<td>7.05E-04</td>
<td>zinc finger protein 770 (Zfp770)</td>
</tr>
<tr>
<td>OG0011784</td>
<td>1.63E-04</td>
<td>zinc finger, DHHC domain containing 19 (Zdhhc19)</td>
</tr>
<tr>
<td>OG0011914</td>
<td>1.33E-09</td>
<td>fumarylacetoacetate hydrolase domain containing 2A (Fahd2a)</td>
</tr>
<tr>
<td>OG0012115</td>
<td>0.00E+00</td>
<td>NUT midline carcinoma, family member 1 (Nutm1)</td>
</tr>
<tr>
<td>OG0012232</td>
<td>2.06E-03</td>
<td>leucine rich repeat neuronal 2 (Lrrn2)</td>
</tr>
<tr>
<td>OG0012396</td>
<td>1.19E-06</td>
<td>leucine rich repeat containing 46 (Lrrc46)</td>
</tr>
<tr>
<td>OG0012449</td>
<td>2.79E-04</td>
<td>F-box and leucine-rich repeat protein 14 (Fbxl14)</td>
</tr>
<tr>
<td>OG0012511</td>
<td>3.27E-08</td>
<td>SMAD family member 6 (Smad6)</td>
</tr>
<tr>
<td>OG0012690</td>
<td>0.00E+00</td>
<td>solute carrier family 47, member 1 (Slc47a1)</td>
</tr>
<tr>
<td>OG0012869</td>
<td>4.59E-02</td>
<td>bromodomain and WD repeat domain containing 3 (Brwd3)</td>
</tr>
<tr>
<td>OG0013171</td>
<td>5.47E-06</td>
<td>ArfGAP with SH3 domain, ankyrin repeat and PH domain 3 (Asap3)</td>
</tr>
<tr>
<td>OG0013288</td>
<td>6.09E-03</td>
<td>phosphodiesterase 3B, cGMP-inhibited (Pde3b)</td>
</tr>
<tr>
<td>OG0013304</td>
<td>3.42E-02</td>
<td>persephin (Pspn)</td>
</tr>
<tr>
<td>OG0013342</td>
<td>4.11E-02</td>
<td>mitochondrial methionyl-tRNA formyltransferase (Mtfmt)</td>
</tr>
<tr>
<td>OG0013590</td>
<td>1.77E-02</td>
<td>matrix metallopeptidase 15 (Mmp15)</td>
</tr>
<tr>
<td>OG0013771</td>
<td>1.77E-02</td>
<td>neuronal pentraxin 2 (Nptx2)</td>
</tr>
<tr>
<td>OG0013841</td>
<td>4.34E-03</td>
<td>transformation related protein 63 regulated like (Tprgl)</td>
</tr>
<tr>
<td>OG0014000</td>
<td>1.25E-02</td>
<td>alkaline phosphatase, liver/bone/kidney (Alpl)</td>
</tr>
<tr>
<td>OG0014048</td>
<td>0.00E+00</td>
<td>BPI fold containing family A, member 5 (Bpifa5)</td>
</tr>
<tr>
<td>OG0014062</td>
<td>1.42E-02</td>
<td>SPARC related modular calcium binding 2 (Smoc2)</td>
</tr>
<tr>
<td>OG0014193</td>
<td>1.05E-06</td>
<td>carbonic anhydrase 11 (Car11)</td>
</tr>
<tr>
<td>OG0014286</td>
<td>9.65E-03</td>
<td>cDNA 4930550C14 gene</td>
</tr>
<tr>
<td>accession</td>
<td>score</td>
<td>description</td>
</tr>
<tr>
<td>-----------</td>
<td>---------</td>
<td>-----------------------------------------------------------------</td>
</tr>
<tr>
<td>OG0014309</td>
<td>7.40E-03</td>
<td>chymotrypsin-like elastase family, member 1 (Cela1)</td>
</tr>
<tr>
<td>OG0014333</td>
<td>2.11E-02</td>
<td>minichromosome maintenance 9 homologous recombination repair factor (Mcm9)</td>
</tr>
<tr>
<td>OG0014414</td>
<td>1.60E-04</td>
<td>growth arrest specific 6 (Gas6)</td>
</tr>
<tr>
<td>OG0014634</td>
<td>2.62E-03</td>
<td>glutathione peroxidase 2 (Gpx2)</td>
</tr>
<tr>
<td>OG0014827</td>
<td>0.00E+00</td>
<td>secretory carrier membrane protein 5 (Scamp5)</td>
</tr>
<tr>
<td>OG0014913</td>
<td>4.34E-04</td>
<td>STIP1 homology and U-Box containing protein 1 (Stub1)</td>
</tr>
<tr>
<td>OG0015025</td>
<td>3.23E-08</td>
<td>pancreatic lipase-related protein 2 (Pnliprp2)</td>
</tr>
<tr>
<td>OG0015310</td>
<td>1.06E-03</td>
<td>claspin (Clspn)</td>
</tr>
<tr>
<td>OG0015466</td>
<td>7.62E-05</td>
<td>excision repair cross-complementing rodent repair deficiency complementation group 6 like (Ercc6l)</td>
</tr>
<tr>
<td>OG0015627</td>
<td>4.18E-03</td>
<td>solute carrier family 15, member 3 (Slc15a3)</td>
</tr>
<tr>
<td>OG0015662</td>
<td>1.52E-03</td>
<td>epithelial membrane protein 2 (Emp2)</td>
</tr>
<tr>
<td>OG0015704</td>
<td>3.96E-03</td>
<td>cubilin (intrinsic factor-cobalamin receptor) (Cubn)</td>
</tr>
<tr>
<td>OG0015713</td>
<td>4.97E-07</td>
<td>suppressor APC domain containing 1 (Sapcd1)</td>
</tr>
</tbody>
</table>
Figure 1 Venn Diagram of transcript expression differences and similarities between the three reproductive tissues for a single male mouse. The total number of transcripts is 78,424.
Figure 2 Venn Diagram of transcript matches between the three reproductive tissues to ncRNA sequences in *Mus musculus*. The total number of transcript matches across the tissue types is 15,964.
**Figure 3** Venn Diagram of transcript matches between the three reproductive tissues to protein sequences in the Transporter Classification Database. The total number of transcript matches across the tissue types is 7,521.
Figure 4 STRING diagrams of protein interactions for two proteins evolving under a mode of positive selectin in *P. eremicus*: Slc5a3 and Slc47a1.
REFERENCES


Dean MD, Good JM, Nachman MW. 2008. Adaptive evolution of proteins secreted during sperm


PP, Jones TA, Tate J, Finn RD. 2015. Rfam 12.0: updates to the RNA families database. 


DOI 10.1371/journal.pone.0091302.


CHAPTER 2

Characterizing the reproductive transcriptomic correlates of acute dehydration in males in the desert-adapted rodent, *Peromyscus eremicus*

Lauren Kordonowy ¹* and Matthew MacManes ¹†

1. University of New Hampshire Department of Molecular, Cellular and Biomedical Sciences Durham, NH, USA

* lauren.kordonowy@unh.edu
† matthew.macmanes@unh.edu

**Corresponding Author:**

Lauren Kordonowy, lauren.kordonowy@unh.edu

**Citation:**

Kordonoway L, MacManes L. 2017. Characterizing the reproductive transcriptomic correlates of acute dehydration in males in the desert-adapted rodent, *Peromyscus eremicus*

bioRxiv DOI: [https://doi.org/10.1101/096057](https://doi.org/10.1101/096057); *BMC Genomics: in review*
ABSTRACT

The understanding of genomic and physiological mechanisms related to how organisms living in extreme environments survive and reproduce is an outstanding question facing evolutionary and organismal biologists. One interesting example of adaptation is related to the survival of mammals in deserts, where extreme water limitation is common. Research on desert rodent adaptations has focused predominantly on adaptations related to surviving dehydration, while potential reproductive physiology adaptations for acute and chronic dehydration have been relatively neglected. This study aims to explore the reproductive consequences of acute dehydration by utilizing RNAseq data in the desert-specialized cactus mouse (*Peromyscus eremicus*). Specifically, we exposed 22 male cactus mice to either acute dehydration or control (fully hydrated) treatment conditions, quasimapped testes-derived reads to a cactus mouse testes transcriptome, and then evaluated patterns of differential transcript and gene expression. Following statistical evaluation with multiple analytical pipelines, nine genes were consistently differentially expressed between the hydrated and dehydrated mice. We hypothesized that male cactus mice would exhibit minimal reproductive responses to dehydration; therefore, this low number of differentially expressed genes between treatments aligns with current perceptions of this species’ extreme desert specialization. However, these differentially expressed genes include Insulin-like 3 (Insl3), a regulator of male fertility and testes descent, as well as the solute carriers Slc45a3 and Slc38a5, which are membrane transport proteins that may facilitate osmoregulation. Together, these results suggest that in male cactus mice, acute dehydration may be linked to reproductive modulation via Insl3, but not through gene expression differences in the subset of other *a priori* tested reproductive hormones. Although water availability is a reproductive cue in desert-rodents exposed to chronic drought, potential reproductive modification via Insl3 in
response to acute water-limitation is a result which is unexpected in an animal capable of surviving and successfully reproducing year-round without available external water sources. Indeed, this work highlights the critical need for integrative research that examines every facet of organismal adaptation, particularly in light of global climate change, which is predicted, amongst other things, to increase climate variability, thereby exposing desert animals more frequently to the acute drought conditions explored here.

KEYWORDS

adaptation, testes, genetics, transcriptomics, differential expression, reproduction, physiology, dehydration, cactus mouse, *Peromyscus eremicus*
BACKGROUND

For decades, evolutionary biologists have successfully described examples where natural selection has resulted in the exquisite match between organism and environment (e.g. Salinity adaptations in three-spine sticklebacks: Hohenlohe et al., 2010; Jones et al. 2012; high-altitude adaptations for hemoglobin in deer mice and humans: Storz et al., 2010, Lorenzo et al., 2015; and Peromyscus adaptations for multiple environments: Hoekstra et al., 2006; Bedford & Hoekstra, 2015; Munshi-South & Richardson, 2016). The match between organism and environment must be studied in the context of both components of fitness: survival and reproductive success, because both aspects of selection are critical to long term persistence in a given environment. Habitat specialists must possess phenotypes enabling survival and successful reproduction; therefore, cases where environmental selective pressures result in reduced reproductive success (e.g. Martin & Wiebe, 2004; Bolger, Patten & Bostock, 2005; Evans et al., 2010; Wingfield, Kelley & Angelier, 2011), but not survival, demand attention. Species occupying extreme environments are likely more vulnerable to the bifurcation of these two components of fitness. Moreover, long-term events like global climate change are predicted to increase climate variability and may enhance the challenges faced by species living on the fringes of habitable environments (Martin & Wiebe, 2004; Somero, 2010; Wingfield, Kelley & Angelier, 2011; Wingfield, 2013; Asres & Amha, 2014).

Deserts present extraordinary environmental impediments for habitation, including extreme heat, aridity, and solar radiation. Examples of well-described desert mammal behavioral adaptations are seasonal torpor (reviewed in Kalabukhov 1960; Geiser, 2010), nocturnality (e.g. Stephens & Tello, 2009; Fuller et al., 2014) and burrowing (reviewed in Vorhies, 1945; Kelt, 2011) to avoid high temperatures and sun exposure. Desert mammals also exhibit a wide range
of morphological adaptations, including large ears for effective heat dissipation (e.g. Schmidt-Nieslen, 1964; Hill & Veghte, 1976), metabolic water production (e.g. MacMillen & Hinds, 1983; reviewed in Walsberg 2000), and renal adaptations to minimize water-loss (e.g. Schmidt-Nielsen et al., 1948; Dantzler, 1982; Diaz, Ojeda & Rezenda, 2006). Although desert rodents must possess adaptations conferring survival and reproductive benefits, researchers have focused on their physiological adaptations for survival. For example, renal adaptations in species of Kangaroo rats (*Dipodomys* species) have been described and explored for over 60 years (Schmidt-Nielsen et al., 1948; Schmidt-Nielsen and Schmidt-Nielsen, 1952; Marra et al., 2012; Uirty et al., 2012). While early research determined the renal physiology for Kangaroo rats (Schmidt-Nielsen et al., 1948; Schmidt-Nielsen and Schmidt-Nielsen, 1952; Vimtrup and Schmidt-Nielsen), recent research has focused on the genetic underpinnings of this phenotype (Marra et al., 2012; Uirty et al., 2012; Marra, Romero & DeWoody, 2014; Marra et al., 2014), which is indicative of a larger methodological shift in the approach for examining adaptation.

Research in another desert-adapted rodent, *Peromyscus eremicus* (cactus mouse), has followed a somewhat different trajectory; however, it too has only pursued survival oriented physiological mechanisms (but see Kordonowy and MacManes, 2016; Kordonowy et al., 2017; MacManes, 2017). The ecology, physiology and behaviors of the cactus mouse in comparison with other *Peromyscus* species were summarized in 1968 (King, ed.), and the relationships between basal metabolic rate, body mass, and evaporative water loss were reviewed several decades later (MacMillen and Garland, 1989). Known desert adaptations for cactus mouse include nocturnality and torpor (reviewed in Veal and Caire, 1979; Caire, 1999); however, the cactus mouse does not possess the same elaborate kidney structures responsible for renal adaptations in kangaroo rats (Dewey et al., 1966; MacManes 2016, unpublished data).
physiological renal adaptations in *P. eremicus* have not been described in detail, despite considerable explorations of other aspects of this species’ biology (reviewed in Veal and Caire, 1979; Caire, 1999). In order to initially characterize renal function of the cactus mouse, water consumption measurements and electrophysical dehydration effects for this species have also recently been documented (Kordonowy et al., 2017). Because the renal mechanisms for mitigating renal water-loss in *P. eremicus* have not been determined, a comparative genetic approach may be instrumental for characterizing this species’ adaptive kidney phenotype. To this end, MacManes and Eisen (2014) conducted a comparative analysis to find genes expressed in the kidney tissue of cactus mouse that were under positive selection relative to other mammals. MacManes (2017) also recently conducted differential gene expression analyses on cactus mouse kidneys subjected to acute dehydration to explore transcriptomic renal responses. However, the transcriptomic resources available for this species extend considerably beyond renal tissue; transcripts from cactus mouse (as well as numerous other *Peromyscus* species) have been heavily utilized to pursue questions related to multiple aspects of evolutionary biology (reviewed in Bedford and Hoekstra, 2015; Munshi-South and Richardson, 2016). Current investigations into cactus mouse desert-adaptive renal physiology include transcriptomic analyses (MacManes 2017); however, we extended this genetic approach by shifting the focus from adaptions for survival to include physiological adaptations for reproductive success (Kordonowy and MacManes, 2016). The cactus mouse is an ideal system for investigating dehydration effects on reproduction, as well as potential reproductive adaptations for drought, given decades of study of reproductive biology, as well as more recent development of transcriptomic resources that include male reproductive tissues.

Substantial research has been done on the effects of various types of stress on
reproduction (e.g. Wingfield & Sapolsky, 2003; Ahmed et al., 2015; Nargund, 2015; Wingfield, 2013); furthermore, the impacts of dehydration stress on reproduction in desert specialized rodents have been historically explored by studies documenting the impacts of water availability as a reproductive cue (reviewed in Schwimmer and Heim 2009; Bales and Hostetler, 2011). Specifically, some female desert rodents have shown evidence of reproductive attenuation due to water-limitation (Mongolian gerbil: Yahr and Kessler, 1975; hopping mouse: Breed, 1975), and male Mongolian gerbils subject to dehydration had decreased reproductive tissue mass (Yahr and Kessler, 1975). In contrast, Shaw’s jird, an Egyptian desert rodent, did not elicit perceivable reproductive response to water deprivation in either males or females (El, Bakry et al., 1999). Furthermore, water-supplementation studies among wild desert rodents resulted in prolonged breeding seasons in the hairy-footed gerbil and the four-striped grass mouse, but not in the Cape short-eared gerbil (Christian, 1979). Recent research has confirmed the importance of rainfall as a reproductive cue in the Arabian spiny mouse (Sarli et al., 2016), the Baluchistan gerbil (Sarli et al., 2015), Chessman’s gerbil (Henry and Dubost, 2012) and the Spinifex hopping mouse (Breed and Leigh, 2011). The focus of this previous research was to investigate reproductive cues and consequences of water-limitation in desert rodents, namely how species have adapted breeding onset and cessation patterns to respond to water availability. Our current study experimentally tests reproductive responses to acute dehydration using a differential gene expression approach in the cactus mouse, which has not been previously evaluated for reproductive impacts of dehydration.

In nature, wild cactus mice are subjected to both acute and chronic dehydration, and understanding the reproductive effects of dehydration stress is an initial step for fully characterizing the suite of phenotypes enabling their successful reproduction. Given that this
species has evolved in southwestern United States deserts and it breeds continuously throughout the year (Veal and Caire, 1979; Caire, 1999), we predict that neither acute nor chronic water stress, while physiologically demanding, would be associated with reproductive suppression. To evaluate acute water stress reproductive tissue gene expression responses in the current study, we leveraged previous research that characterized the transcriptome of male *P. eremicus* reproductive tissues from functional and comparative perspectives (Kordonowy and MacManes, 2016). We extend upon this work by performing an RNAseq experiment to identify differentially expressed genes in testes between male *P. eremicus* subjected to acute dehydration versus control (fully hydrated) animals in order to determine the impacts, if any, on male reproduction. We hypothesized that male cactus mice would exhibit minimal gene expression level reproductive responses to acute dehydration because they are highly desert-adapted and they breed year-round, including in times of chronic drought. Specifically, we predicted that genes linked to reproductive function would not be differentially expressed in the testes in response to acute dehydration. We pursued this line of research on the effects of dehydration on reproduction in cactus mouse in order to begin to address the need for additional studies focusing on physiological adaptations related to reproductive success in animals living in extreme, and changing, environments.

**METHODS**

**Treatment Groups, Sample Preparation and mRNA Sequencing**

The cactus mice used for this study include only captive born individuals purchased from the *Peromyscus* Genetic Stock Center (Columbia, South Carolina). The animals at the stock center are descendant from individuals originally collected from a hot-desert location in Arizona.
more than 30 years ago. The colony used in this study has been housed since 2013 at the University of New Hampshire in conditions that mimic temperature and humidity levels in southwestern US deserts, as described previously (Kordonowy & MacManes, 2016). Males and females are housed together, which provides olfactory cues to support reproductive maturation. Males do not undergo seasonal testicular atrophy, as indicated by successful reproduction throughout the year. The individuals used in this study were all of the same developmental stage – reproductively mature – which was assessed by observing that the testes had descended into the scrotum from the abdomen, making them visible.

Males that had free access to water prior to euthanasia are labeled as WET mice in our analyses. Mice that were water deprived, which we refer to as DRY mice, were weighed and then water deprived for ~72 hours directly prior to euthanasia. All mice were weighed prior to sacrifice, and DRY mice were evaluated for weight loss during dehydration. Individuals in the study were collected between September 2014 – April 2016.

Cactus mice were sacrificed via isoflurane overdose and decapitation in accordance with University of New Hampshire Animal Care and Use Committee guidelines (protocol number 130902) and guidelines established by the American Society of Mammalogists (Sikes et al., 2016). Trunk blood samples were collected following decapitation for serum electrolyte analyses with an Abaxis Vetscan VS2 using critical care cartridges (Abaxis). The complete methodology and results of the electrolyte study, as well as the reported measures of water consumption and weight loss due to dehydration are described fully elsewhere (Kordonowy et al., 2017). Rather, this study focused on differential gene expression between the testes of 11 WET and 11 DRY mice. Testes were harvested within ten minutes of euthanasia, placed in RINAlater (Ambion Life Technologies), flash-frozen in liquid nitrogen, and stored at -80° degree Celsius. A TRIzol,
chloroform protocol was implemented for RNA extraction (Ambion Life Technologies). Finally, the quantity and quality of the RNA product was evaluated with both a Qubit 2.0 Fluorometer (Invitrogen) and a Tapestation 2200 (Agilent Technologies, Palo Alto, USA).

Libraries were made with a TruSeq Stranded mRNA Sample Prep LT Kit (Illumina), and the quality and quantity of the resultant sequencing libraries were confirmed with the Qubit and Tapestation. Each sample was ligated with a unique adapter for identification in multiplex single lane sequencing. We submitted the multiplexed samples of the libraries for processing on lanes at the New York Genome Center Sequencing Facility (NY, New York). Paired end sequencing reads of length 125bp were generated on an Illumina 2500 platform. Reads were parsed by individual samples according to their unique hexamer IDs in preparation for analysis.

Assembly of Testes Transcriptome

We assembled a testes transcriptome from a single reproductively mature male using the de novo transcriptome protocol described previously (MacManes, 2016). The testes transcripts were assembled with alternative methodologies utilizing several optimization procedures to produce a high-quality transcriptome; however, the permutations of this assembly process are described extensively elsewhere (MacManes, 2016; Kordonowy and MacManes, 2016). The testes transcriptome we selected was constructed as described below. The raw reads were error corrected using Rcorrector version 1.0.1 (Song and Florea, 2015), then subjected to quality trimming (using a threshold of PHRED <2, as per MacManes, 2014) and adapter removal using Skewer version 0.1.127 (Jiang et al, 2014). These reads were then assembled in the de novo transcriptome assembler BinPacker version 1.0 (Liu et al., 2016). We also reduced sequence redundancy to improve the assembly using the sequence clustering software CD-HIT-EST
version 4.6 (Li & Godzik, 2006; Fu et al., 2012). We further optimized the assembly with Transrate version 1.0.1 (Smith-Unna et al., 2015) by retaining only highly supported contigs (cutoff: 0.02847). We then evaluated the assembly’s structural integrity with Transrate and assessed completeness using the vertebrata database in BUSCO version 1.1b1 (Simão et al., 2015). We quasimapped the raw reads to the assembly with Salmon version 0.7.2 (Patro, et al., 2015) to confirm that mapping rates were high. Finally, the assembly was also annotated in dammit version 0.3.2, which finds open reading frames with TransDecoder and uses five databases (Rfam, Pfam, OrthoDB, BUSCO, and Uniref90) to thoroughly annotate transcripts (https://github.com/camillescott/dammit).

**Differential Gene and Transcript Expression Analyses**

Several recent studies have critically evaluated alternative methodologies for differential transcript and gene expression to determine the relative merits of these approaches (Gierlinski et al., 2015; Schurch et al., 2016; Soneson et al., 2016; Froussios et al., 2016). Soneson and colleagues (2016) demonstrated that differential gene expression (DGE) analyses produce more accurate results than differential transcript expression (DTE) analyses. Furthermore, the differential gene expression approach is more appropriate than differential transcript expression for the scope of our research question, which is true of many evolutionary genomic studies (Soneson et al., 2016). However, because both DTE and DGE approaches are widespread in current literature, we deemed it important to confirm that these methodologies yielded concordant results in the current study.

We utilized edgeR (Robinson, McCarthy & Smith, 2010; McCarthy, Chen & Smith, 2012) as our primary statistical software because Schurch and colleagues (2016) rigorously
tested various packages for analyzing DGE, and edgeR performed optimally within our sample size range. While edgeR is a widely used statistical package for evaluating differential expression, we also confirmed our results with another popular package, DESeq2 (Love et al., 2014), in order to validate our findings.

We performed differential expression analyses with three alternative methodologies. Two analyses were conducted in R version 3.3.1 (R Core Team, 2016) using edgeR version 3.16.1, a Bioconductor package (release 3.4) that evaluates statistical differences in count data between treatment groups (Robinson et al., 2010; McCarthy et al., 2012). Our first method utilized tximport, an R package developed by Soneson and colleagues (2016), which incorporates transcriptome mapping-rate estimates with a gene count matrix to enable downstream DGE analysis. The authors assert that such transcriptome mapping can generate more accurate estimates of DGE than traditional pipelines (Soneson et al., 2016). While our first methodology evaluated differential gene expression, our second analysis used the transcriptome mapped read sets to perform differential transcript expression and identify the corresponding gene matches. The purpose of this second analysis was to evaluate whether the transcript expression results coincided with the gene expression results produced by the same program, edgeR. Finally, our third methodology determined differential gene expression with tximport in conjunction with DESeq2 version 1.14.0 (Love et al., 2014), a Bioconductor package (release 3.4) which also evaluates statistical differences in expression. We performed this alternative DGE analysis with DESeq2 in order to corroborate our DGE results from edgeR. Thus, the results for all three differential expression analyses were evaluated to determine the coincidence among the genes identified as significantly different between the WET and DRY groups. These alternative differential expression methods are described in detail below.
We quasimapped each of the 11 WET and 11 DRY sample read sets to the testes transcriptome with Salmon version 0.7.2 to generate transcript count data. To perform the gene-level analysis in edgeR, we constructed a gene ID to transcript ID mapping file, which was generated by a BLASTn (Altschul et al., 1990; Madden, 2002) search for matches in the Mus musculus transcriptome (Ensembl.org) version 7/11/16 release-85. We then imported the Salmon-generated count data and the gene ID to transcript ID mapping file into R using the tximport package (Soneson et al., 2016) to convert the transcript count data into gene counts. These gene count data were imported into edgeR for differential gene expression analysis (Robinson et al., 2010; McCarthy et al., 2012). We applied TMM normalization to the data, calculated common and tagwise dispersions, and performed exact tests (p < 0.05) adjusting for multiple comparisons with the Benjamini-Hochburg correction (Benjamini and Hochburg, 1995) to find differentially expressed genes, which we identified in Ensembl (ensemble.org).

Next, we performed a transcript-level analysis using edgeR. To accomplish this, the Salmon-generated count data was imported into R and analyzed as was described above for the gene-level analysis in edgeR. After determining which transcript IDs were differentially expressed, we identified the corresponding genes using the gene ID to transcript ID matrix described previously. The significantly expressed transcripts without corresponding gene matches were selected for an additional BLASTn search in the NCBI non-redundant nucleotide database (http://blast.ncbi.nlm.nih.gov/Blast.cgi). However, these results were not subjected to any additional analyses, because these matches were not consistent across all three differential expression analyses. This list of BLASTn search matches is provided in supplementary materials (DTEno-matchBLASTnSequences.md).

The third analysis used DESeq2 to conduct an additional gene-level test, using the same
methods as described for the previous gene-level analysis, with the exception that data were imported into an alternative software package. We determined the significantly differentially expressed genes (p < 0.05) based on normalized counts and using the Benjamini-Hochburg correction (Benjamini and Hochburg, 1995) for multiple comparisons. We only retained genes with a $-1 < \log_2$ fold change $> 1$ in order to filter genes at a conservative threshold for differential expression based on our sample size (Schurch et al., 2016). This filtering was not necessary for either of the edgeR analyses because log$_2$ fold changes exceeded this threshold for the differentially expressed genes and transcripts ($-1.3 < \log_2$ fold change $> 1.4$, in all cases).

We also compared the log$_2$ fold change values (of treatment differences by mapped count) for each gene from the edgeR and DESeq2 gene-level analyses in a linear regression. This statistical test was performed in order to evaluate the degree of concordance between the two DGE analyses. Furthermore, we constructed a list of genes identified as differentially expressed by all three analyses, which were further evaluated for function as well as chromosomal location. These genes were also explored in STRING version 10.0 (string-db.org) to determine their protein-protein interactions (Snel et al., 2000; Szklarczyk et al., 2015).

Lastly, we performed an *a priori* test for DGE in edgeR on a small subset of nine genes encoding hormones and hormone receptors known to be involved in various aspects of reproductive functionality in male rodents. These genes are: steroidogenic acute regulatory protein (StAR), prolactin receptor (Prlr), luteinizing hormone/choriogonadotropin receptor (Lhgc), inhibin (Inha), ghrelin (Ghrl), estrogen related receptor gamma (Essrg), estrogen related receptor alpha (Essra), androgen receptor (Ar), and activin receptor type-2A (Acvr2a). We retrieved the *Mus musculus* genomic sequences for these hormones and receptors from Ensembl (release 88: March 2017) and then executed BLASTn searches for the corresponding
Peromyscus eremicus sequences in the testes transcriptome. The Ensembl gene identifiers (Mus musculus) corresponding to the P. eremicus transcripts were queried from the table of results produced by the edgeR DGE analysis to evaluate treatment differences in expression.

RESULTS

Data and Code Availability

The testes transcriptome was assembled from a 45.8 million paired read data set. Additionally, there were 9-20 million paired reads for each of the 22 testes data sets used for the differential expression analysis (Supplemental Table 1), yielding 304,466,486 reads total for this analysis. The raw reads are available at the European Nucleotide Archive under study accession number PRJEB18655. All data files, including the testes un-annotated transcriptome, the dammit annotated transcriptome, and the data generated by the differential gene expression analysis (described below) are available on DropBox:

https://www.dropbox.com/sh/ffr9xrmjxj9md1m/AACpxjQNn-Jlf25qNdslfRSCa?dl=0.

These files will be posted to Dryad upon manuscript acceptance. All code for these analyses is posted on GitHub (https://github.com/macmanes-lab/testesDGE).

Assembly of Testes Transcriptome

The performance of multiple transcriptome assemblies was evaluated thoroughly, and the selected optimized testes assembly met high quality and completeness standards, and it also contains relatively few contigs and has high read mapping rates (Table 1). Therefore, this transcriptome was used for our differential expression analyses. The transcriptome was also annotated, and the complete statistics for this dammit annotation are provided in Table 1.
Differential Gene and Transcript Expression Analyses

Salmon quasimapping rates of all read datasets to the assembly were sufficiently high (range: 81.46% - 87.02%; mean \(\text{WET} = 84.41\); mean \(\text{DRY} = 83.81\); Supplemental Table 1), indicating the successful generation of transcript count data for our differential expression analyses. The exact test performed for our gene-level analysis in edgeR indicated that fifteen genes reached statistical significance (after adjusting for multiple comparisons) for DGE between the WET and DRY treatment groups (Supplemental Figure 1). Specifically, seven genes were more highly expressed in WET individuals, and eight genes were more highly expressed in DRY individuals (Table 2).

We also performed an alternative transcript-level analysis using the referenced transcriptome mapped reads exclusively with edgeR. The exact test found 66 differentially expressed transcripts (Supplemental Figure 2), 45 of which were more highly expressed in the WET group, and 21 were more highly expressed in the DRY group (Table 3). 10 of these differentially expressed transcripts were consistent with differentially expressed genes from the edgeR DGE analysis. In addition, the significantly expressed transcripts without an Ensembl ID match (nine WET and nine DRY) were retrieved for performing an nt all species BLASTn search (http://blast.ncbi.nlm.nih.gov/Blast.cgi), and these results are in the supplementary materials.

The gene-level analysis conducted in DESeq2 yielded 215 significantly differentially expressed genes (Supplemental Figure 3), 67 of which were more highly expressed in the WET group, while 148 were highly expressed in the DRY group. However, only 20 of these genes remained when we filtered them with a \(-1 < \log_2 \text{fold change} > 1\) to retain genes with a
conservative threshold difference between treatment groups. This list of 20 genes yielded 16 genes more highly expressed in WET mice and four genes highly expressed in DRY mice (Table 4). Nine of these genes overlapped with those found to be significant in the previous two edgeR analyses.

To evaluate the correlation of log₂ fold change results for each gene (Ensembl ID) from the two DGE analyses (edgeR and DESeq2), we performed a regression of these log values, and they were significantly correlated (Figure 1: Adj-R² = 0.6596; F(1,14214) = 2.754x10⁴; p < 2.2x10⁻¹⁶). This further demonstrates the concordance of the DGE analyses in these two software packages.

To evaluate the degree to which the three analyses produced concordant results, we generated a list of genes which were found to be significantly differently expressed by treatment across all three analyses (Supplemental Table 2). There were six genes that were consistently highly-expressed in the WET group and three genes that were highly-expressed in the DRY group. The six highly-expressed WET genes are Insulin-like 3 (Insl3), Free-fatty acid receptor 4 (Ffar4), Solute carrier family 45 member 3 (Slc45a3), Solute carrier family 38 member 5 (Slc38a5), Integrin alpha L (Itgal), and Transferrin (Trf). The three highly-expressed DRY genes are Ras and Rab Interactor 2 (Rin2), Insulin-like growth factory binding protein 3 (Igfbp3), and Connective tissue growth factor (Ctgf). Because the patterns of expression of these nine genes were corroborated by multiple methodologies, we are confident that they are differentially expressed between our treatments. Estimates of expression for these genes generated using the gene-level edgeR analysis are plotted in Figure 2.

The significantly differently expressed genes were evaluated for gene function and chromosomal location (Table 5). These genes occur throughout the genome; namely, they are
located on different chromosomes. The diverse functions of each gene will be described below. In addition, we generated STRING diagrams (string-db.org) to view the protein-protein interactions for each of these nine genes (Snel et al., 2000; Szklarczyk et al., 2015).

Slc38a5 and Slc45a3 are among the highly expressed genes in the WET group (they have lower expression in the DRY group); these two solute carriers are members of a large protein family that is responsible for cross-membrane solute transport (reviewed in Hediger et al., 2004; Hediger et al., 2013; Cesar-Razquin et al., 2015). Slc38a5 is involved sodium-dependent amino-acid transport, while Slc45a3 is purported to transport sugars (Vitavska and Wieczorek, 2013; Schiöth, et al., 2013; http://slc.bioparadigms.org/), thereby playing an important potential role in maintaining water balance via management of oncotic pressures. Slc38a5 (Figure 3a) has interactions with multiple additional solute carriers, including Slc1a5, Slc36a2, Slc36a3, and Slc36a4. Slc38a5 also has an interaction with disintegrin and metalloproteinase domain-containing 7 (Adam7), which is involved in sperm maturation and the acrosome reaction (Oh et al., 2005). In contrast, Slc45a3 (Figure 3b) does not have known protein interactions with other solute carriers; however, this protein does interact with steroidogenic acute regulatory protein (StAR), which is critical in steroidogenesis (Christenson and Strauss III, 2001). Notably, our a priori DGE analysis did not demonstrate treatment differences in expression for StAR.

Insl3 was lower expressed in the DRY group, and this hormone purportedly regulates fertility in male and female mammals by preventing apoptosis of germ cells in reproductive organs of both sexes (Kawamura et al., 2004; Bathgate et al., 2012; Bathgate et al., 2013). In male rodents, Insl3 is critical to development by facilitating testicular descent, and it is also present in testes of adults, where it binds to relaxin family peptide receptor 2 (Rxfp2), also known as Lrg8 (Bathgate et al., 2012; Bathgate et al., 2013). Protein interaction data for Insl3
(Figure 3c) indicate that this hormone interacts with Rxfp2 and Rxfp1, as well as other proteins, including leptin (Lep), a pleiotropic hormone involved in reproduction, immunity, and metabolism (reviewed in Friedman, 2014).

Ffar4 was also down-regulated in the DRY group. Omega-3 fatty acid receptor 1 (O3Far1) is an alias of Ffar4, and it has roles in metabolism and inflammation (Moniri, 2016).

This protein interacts with multiple other free fatty acid receptors and G-protein coupled receptors as well as Stanniocalcin 1 (Stc1) (Figure 3d). Stc1 is involved in phosphate and calcium transportation (Wagner and Dimattia, 2006); however, this protein’s functional role in mice remains enigmatic (Chang et al, 2005).

Another of the lower expressed DRY group genes is Itgal (also known as CDa11a), which has multifaceted roles in lymphocyte-mediated immune responses (Bose et al., 2014). Concordantly, the protein interactions with Itgal (Figure 3e) include numerous proteins integral to immunity, such as Intracellular adhesion molecules (specifically, ICAM1,2,4), which are expressed on the cell surface of immune cells and endothelial cells. Itgal is a receptor for these ICAM glycoproteins, which bind during immune system responses (reviewed in Albelda et al., 1994). However, an additional role of intercellular adhesion molecules has been proposed in spermatogenesis, whereby ICAMs may be integral to transporting non-mobile developing sperm cells through the seminiferous epithelium (Xiao et al., 2013).

The final gene with lower expression levels in the DRY treatment is Trf, which modulates the amount of free-iron in circulation and binds to transferrin receptors on the surface of erythrocyte precursors to deliver iron (reviewed in Gkouvastos et al., 2012). Trf interacts with multiple proteins (Figure 3f) involved in iron transport and uptake, including Steap family member 3 (Steap3), hephaestin (Heph), ceruloplasmin (Cp), Solute carrier protein 40 member 1
(Sle40a1), and several H+ ATPases. Furthermore, Trf is linked to apolipoprotein A-1 (Apoa1), which interacts with immunoglobulin in a complex named sperm activating protein (Spap) to activate the motility of sperm when it inhabits the female genital tract (Akerlof et al., 1991; Leijonhufvud et al., 1997).

One of the highly expressed genes in the DRY group is Rin2, which is involved in endocytosis (reviewed in Doherty and McMahon, 2009) and membrane trafficking through its actions as an effector protein for the GTPases in the Rab family within the Ras superfamily (reviewed in Stenmark and Olkkonen, 2001). Rin2 protein-protein interactions (Figure 4a) include Ras related protein Rab5b and Rab5b, which are involved in vesicle transport as well as vasopressin-regulated water reabsorption. This mechanism for water reabsorption via Aquaporin 2 (Aqp2) in the kidney has been thoroughly reviewed by Boone and Deen (2008) and Kwon and colleagues (2013).

The second gene highly expressed in the DRY group is Igfbp3, which modulates the effects of insulin growth factors. Thus, the protein directly interacts (Figure 4b) with insulin growth factors 1 and 2 (Igf1, Igf2), which are responsible for increasing growth in most tissues (reviewed in le Roth, 1997; Jones and Clemmons, 2008). Ctgf was also highly expressed in the DRY group, and this protein is responsible for increased fibrosis and extracellular matrix formation (Reviewed in Moussad and Brigstock, 2000). The protein interactions for Ctgf (Figure 4c) include many transcription activators in the Hippo signaling pathway, including multiple TEA domain transcription factors (Tead1, 2, 3 and 4), WW domain containing transcription regulator 1 (Wwtr1), as well as Yes-associated protein 1 (Yap1), which is responsible for both increasing apoptosis and preventing cell proliferation to mitigate tumor growth and control organ size (Reviewed in Pan, 2010).
The *a priori* edgeR DGE analysis for the genes encoding nine reproductive hormones and hormone receptors) did not reveal any statistically significant differences between the WET and DRY mice. The log fold change values and corresponding p-values for these genes are in the analysis posted on GitHub. The patterns for these genes by treatment are shown in Figure 5.

**DISCUSSION**

This is the first study to evaluate gene expression levels of a reproductive tissue (testes) in response to acute dehydration in a desert-specialized rodent, *Peromyscus eremicus* (cactus mouse). Our results demonstrate differential expression of Insl3, which is a gene linked to reproduction, but not for a small subset of other reproductive hormone (and hormone receptor) genes. We also found expression differences in two solute carrier proteins, which is consistent with previous findings asserting the importance of this protein family for osmoregulation in desert rodents. Our findings lead us to hypothesize that reproductive function may be modified via Insl3 in acutely dehydrated mice. Any transcriptomic indication of potential reproductive modification in response to acute dehydration is surprising, given that this is not consistent with our understanding of *P. eremicus* as a desert specialist capable of breeding year-round in the wild. However, future studies must determine the physiological effects of decreased Insl3 expression on acutely dehydrated cactus mice. While acute dehydration is less common than chronic dehydration for desert mammals, given their ecology, it is a selective force they must overcome. Indeed, throughout much of the described range of the cactus mouse, rainfall events may occur several times per year. Cactus mice, and many other rodents, are known to rehydrate during these rainfall events (MacManes, *personal observation*). Following rehydration, cactus mice experience acute dehydration, followed by a steady state of chronic dehydration. The
reproductive responses of cactus mice to these acute and chronic dehydration events are unknown; therefore, this study describes the transcriptomic effects of acute dehydration in testes.

InsL3, which is believed to be a hormonal regulator of fertility among mammals of both sexes, inhibits germ line apoptosis in the testes (Kawamura et al., 2004; Bathgate et al., 2012; Bathgate et al., 2013). Within adult rodent testes, luteinizing hormone (LH) stimulates expression of InsL3 in Leydig cells, and InsL3 binds to Lrg8 in seminiferous tubules, which results in inhibited apoptosis of germ-line cells, thus increasing their availability (Kawarmura et al., 2004). In addition, a study using murine Leydig cells demonstrated that InsL3 administration increased testosterone production (Pathirana et al., 2012). The precise mechanistic role of InsL3 in modulating fertility is still being elucidated; however, researchers assert that this hormone is an important regulator of fertility in males and females (reviewed in Bathgate et al., 2012). Indeed, recent research has investigated the utility of InsL3 as an indicator of mammalian fertility (e.g. in humans: Kovac and Lipshultz, 2013; in bulls: Pitia et al., 2016). InsL3 is also critical for the first phase of testicular descent, the transabdominal phase, which occurs during fetal development in rodents; but InsL3 does not appear to be involved in the inguinoscrotal phase which happens in sexually immature or inactive male rodents (reviewed in Hutson et al., 2015).

Lower InsL3 expression in the testes of acutely dehydrated mice leads us to suggest that fertility may be attenuated due to acute water deprivation. However, future work characterizing the functional consequences of InsL3 down-regulation, including direct measurements of sperm numbers and function, is needed to causatively demonstrate reproductive attenuation. Specifically, does the number or quality of sperm decrease, and does this decrease reduce the probability of successful fertilization? Moreover, what are the temporal dynamics of reproductive suppression? Logically, species with core reproductive functions that are
suppressed by dehydration seem likely to be rapidly outcompeted by others lacking such limitations. Given this assertion, research characterizing the reproductive correlates of chronic dehydration is a logical extension of this work, although doing so is beyond the scope of this study.

Solute carrier proteins, specifically Slc45a3 and Slc38a5, are downregulated in acute dehydration. These genes are part of a large family essential for transferring solutes across membranes (reviewed in Hediger et al., 2004; Hediger et al., 2013; Cesar-Razquin et al., 2015). Another member of this family, Solute carrier family 2 member 9 (Slc2a9), has been found to be undergoing positive selection in studies on kidney transcriptomes of cactus mouse (MacManes & Eisen, 2014) and of other desert rodents (Marra et al., 2014). Our previous work with the male reproductive transcriptome of cactus mouse found evidence for positive selection in two additional solute carrier proteins: Slc15a3 and Slc47a1 (Kordonowy and MacManes, 2016). A recent differential gene expression study in cactus mouse kidneys found that Slc2a1 and Slc8a1 also showed responses to acute dehydration (MacManes, 2017). Therefore, our current findings that two solute carrier proteins are lower expressed in the DRY treatment group is consistent with previous research in the kidney and male reproductive transcriptomes for this species. This leads us to further support the hypothesis originally proposed by Marra and colleagues (2014) that this protein family is intrinsic to osmoregulation in desert rodents. Indeed, the findings of MacManes and Eisen (2014), Kordonowy and MacManes (2016), and MacManes (2017) also lend support to the essential role of solute carrier proteins for maintaining homeostasis in the desert specialized cactus mouse.

In addition to their well characterized role in the maintenance of water and electrolyte balance, the differential expression of solute carrier proteins may have important reproductive
consequences, particularly as they relate to hormone secretion. Indeed, the interaction between Slc38a5 and Adam7 is relevant, because Adam7 is involved in sperm maturation and the acrosome reaction (Oh et al., 2005). Furthermore, the protein-protein interactions between Slc45a3 with StAR and between Insl3 and Lep are of particular interest because both StAR and Lep are integral to reproduction, as well as to homeostasis (reviewed in Christenson and Strauss III, 2001; Anuka et al., 2013; Friedman, 2014; Allison and Myers, 2014). However, our a priori DGE analysis evaluating StAR, and other reproductive hormones, did not show evidence of expression changes. Thus, the protein interactions with reproductive implications are not restricted to solute carrier proteins. The protein relationships between Itgal and intercellular adhesion molecules are also noteworthy with respect to research hypothesizing an integral role for ICAMs in spermatogenesis (Xiao et al., 2013). Furthermore, Trf is linked to Apoa1, which is a critical component of sperm activating protein (Akerlof et al., 1991; Leijonhufvud et al., 1997). While the relationship between these differentially expressed genes and the hormones involved in reproductive function are currently poorly-characterized, our findings that genes integral to sperm development and activation interact with genes differentially expressed in acute dehydration may indicate that, contrary to our expectations, acute dehydration is linked to reproductive modulation in the cactus mouse. However, functional studies will be necessary to elucidate the connection between these genes and physiological responses to dehydration. This is particularly important because many hormones have pleotropic effects, and further mechanisms of action unrelated to reproduction may be elucidated for these proteins in Peromyscus eremicus.

In contrast to genes that are down-regulated in dehydration, the genes that were upregulated in the DRY group are known to be responsible for water homeostasis and cellular
growth. The significance of Rin2 is notable, because this protein is an effector for Rab5, which is a GTPase involved in vasopressin-regulated water reabsorption, a critical homeostatic process mediated through the Aqp2 water channel in kidneys (Boone and Deen, 2008; Kwon et al., 2013). It is not surprising that genes in addition to solute carrier proteins, which are implicated in alternative processes for water homeostasis, are differentially expressed in response to water limitation. The other two genes that are up-regulated in the DRY treatment are indicative of modulated growth due to water limitation. Specifically, Igfb3 interacts directly with insulin growth factors responsible for tissue growth (Le Roth 1997; Jones and Clemmons, 2008), and Ctgf is linked with numerous transcription factors in the Hippo signaling pathway, which modulates apoptosis, proliferation, and organ size control (Pan, 2010).

To complement our male centric research, future studies should evaluate dehydration induced gene expression differences in female reproductive tissues, particularly in the uterus and ovaries during various reproductive stages. Indeed, given that the physiological demands of reproduction are purportedly greater in females, though this is controversial, (Bateman’s Principle: proposed in Bateman, 1948; addressed in Trivers, 1972; reviewed in Knight, 2002; tested in Jones et al., 2002; 2005; Collet et al., 2014), we would expect to see a greater degree of reproductive suppression in females. While such work is beyond the scope of this manuscript, we hope that future research will evaluate female cactus mouse reproductive responses to dehydration.

Our findings are pertinent to physiological research in other desert-rodents showing reproduction suppression in response to water limitation (reviewed in Bales and Hostetler, 2011), specifically, in male and female Mongolian gerbils (Yahr and Kessler, 1975) and female hoping mice (Breed, 1975). The integral role of water as a reproductive cue for desert-rodents has also
been demonstrated in water-supplementation studies (reviewed in Bales and Hostetler, 2011; Christian, 1979) as well as research on the effects of desert rainfall (Breed and Leigh, 2011; Henry and Dubost, 2012; Sarli et al., 2015; Sarli et al., 2016). Thus, Schwimmer and Haim (2009) asserted that reproductive timing is the most evolutionarily important adaptation for desert rodents. Furthermore, desert rodent research supporting a dehydration driven reproductive suppressive pathway mediated by arginine vasopressin (reviewed in Schwimmer and Haim, 1999; tested in Tahri-Joutei and Pointis, 1988a; 1988b; Shanas and Haim, 2004; Wube et al., 2008; Bukovetzky et al., 2012a; Bukovetzky et al., 2012b) is somewhat analogous to our study linking decreased Insl3 expression in testes with dehydration, in that both findings represent non-traditional hormonal modulation of reproduction. We propose that future studies thoroughly explore physiological consequences for non-traditional hormonal pathways in response to dehydration in desert rodents, as well as well-established reproductive modulatory hormones in the hypothalamic-pituitary-gonadal axis.

Emerging from this work is a hypothesis related to the reproductive response to water stress in the cactus mouse, and perhaps other desert rodents. Specifically, we hypothesize that acute dehydration may be related to reproductive mitigation; however, we hypothesize that chronic dehydration is not. Indeed, it is virtually oxymoronic to suggest that chronic dehydration, which is the baseline condition in desert animals, has negative consequences for reproductive success. Indeed, desert rodents dynamically respond to water-availability to initiate and cease reproductive function. Generating an integrative, systems-level understanding of the reproductive responses to both acute and chronic dehydration across desert-adapted rodent is required for testing our hypothesis. While understanding the renal response to dehydration is critical for making predictions about survival, understanding the reproductive correlates is
perhaps even more relevant to evolutionary fitness. This study, to the best of our knowledge, is the first to describe the reproductive correlates of water-limitation in the cactus mouse, and the first to use a differential gene expression approach to evaluate reproductive tissue responses to drought. Furthermore, this study contributes to a research aim to determine whether novel physiological reproductive adaptations are present in male cactus mouse (Kordonowy and MacManes, 2016). Developing a comprehensive understanding of reproductive responses to drought, and also the mechanisms underlying potential physiological adaptations, is necessary if we are to understand how increasing environmental variability due to climate change may modify the distribution of extant organisms.

CONCLUSIONS

The genetic mechanisms responsible for physiological adaptations for survival and reproduction in deserts remain enigmatic. Desert rodent research has focused primarily on physiological adaptations related to survival, specifically on renal adaptations to combat extreme water-limitation. In contrast, while previous studies have investigated reproductive effects of water-limitation in desert rodents, the underlying mechanisms for physiological adaptations for reproduction during acute and chronic dehydration are unknown. Furthermore, ours is the first study to evaluate reproductive transcriptomic responses to water limitation in a desert-rodent, the cactus mouse. To this end, we characterized the reproductive correlates of acute dehydration in this desert-specialized rodent using a highly replicated RNAseq experiment. In contrast to expectations, we describe a potential signal of reproductive modulation in dehydrated male cactus mouse testes. Specifically, dehydrated mice demonstrated significantly lower expression
of Insl3, which is a canonical regulator of fertility (and testes descent). Lower expression was also found in Slc45a3 and Slc38a5, lending further credence to the important role of solute carrier proteins for osmoregulation in the cactus mouse. While the low number of differentially expressed genes between acutely dehydrated and control mice might otherwise have suggested that this species is relatively unaffected by acute water-limitation, the diminished expression of Insl3 in dehydrated mice leads us to propose that acute dehydration may compromise reproductive function via decreased fertility. Indeed, we hypothesize that non-traditional reproductive hormone pathways, such as those involving Insl3 or Avp (which has elicited suppressive reproductive responses in other desert rodent research), warrant further investigation in studies evaluating the reproductive effects of acute and chronic dehydration. Although future research must experimentally evaluate the potential functional relationship between Insl3 expression pattern and reproductive function and fertility, our findings that acute-dehydration alters Insl3 expression may be concerning, particularly with respect to global climate change. Climate change driven increased variabilities in weather patterns may result in a greater frequency of acute water-stress, which could result in reduced reproductive function for the cactus mouse. In addition, because global climate change is predicted to shift habitats toward extremes in temperature, salinity, and aridity, and to alter species ranges, an enhanced understanding of the reproductive consequences of these changes, and of the potential for organisms to rapidly adapt, may enable us to effectively conserve innumerable species facing dramatic habitat changes.

**Supplemental DropBox Files (will be submitted to Dryad upon acceptance):**

Optimized final un-annotated transcriptome (good.BINPACKER.cdhit.fasta)
Annotated transcriptome (good.BINPACKER.cdhit.fasta.dammit.fasta)

Dammit gff3 file of annotation (good.BINPACKER.cdhit.fasta.dammit.gff3)

Salmon folder including salmon quant outputs for 22 individuals (salmon)

Salmon merged quant file (NEWmergedcounts.txt)

Gene ID by Transcript ID matrix (NEWESTfinalMUS.txt)

Transcripts without matches from edgeR DTE analysis (DTEnomatchBLASTnSequences.md)

**List of abbreviations**

Differential Gene Expression (DGE)

Differential Transcript Expression (DTE)

Dehydrated Treatment Group (DRY)

Control Treatment Group (WET)

*All gene identities are defined in the text, but we can include them upon request here as well.*

**Declarations**

_Ethics approval and consent to participate_

All cactus mice were sacrificed via isoflurane overdose and decapitation in accordance with University of New Hampshire Animal Care and Use Committee guidelines (protocol number 130902) and guidelines established by the American Society of Mammalogists (Sikes et al., 2016).

_Consent for publication_

Not applicable.

_Availability of data and materials:_
The raw reads are available at the European Nucleotide Archive under study accession number PRJEB18655. All data files, including the testes un-annotated transcriptome, the dammit annotated transcriptome, and the data generated by the differential gene expression analysis (described below) are available on DropBox (https://www.dropbox.com/sh/ffr9xrmjxj9md1m/AACpxjQNn-Jlf25qNdslfRSCa?dl=0). These files will be posted to Dryad upon manuscript acceptance. All code for these analyses is posted on GitHub (https://github.com/macmanes-lab/testesDGE).

Competing Interests
The authors declare that they have no competing interests

Funding
This work was supported by a National Science Foundation award to Dr. Matthew MacManes (NSF IOS 1455960).

Authors’ Contributions
LK and MM both contributed to the data collection, data generation, bioinformatics, analyses, interpretation, and writing of this manuscript. MM is the Ph.D. supervisor for LK.

Acknowledgments
We would like to acknowledge the MacManes laboratory, including graduate student Andrew Lang, and the undergraduate students in the laboratory. We also acknowledge Dr. Paul Tsang for advice on the reproductive endocrinology described in the discussion.
Table 1: Transcriptome assembly (BinPacker CD-hit-est Transrate Corrected) performance metrics for: contig number, TransRate score (Score), BUSCO indices: % single copy orthologs (% SCO), % duplicated copy orthologs (% DCO), % fragmented (% frag), and % missing (% miss), as well as Salmon mapping rates (% mapping) for the optimized testes assembly. Dammit transcriptome assembly annotation statistics, including searches in the program TransDecoder for open reading frames (ORFs) and searches for homologous sequences in five databases: Rfam, Pfam-A, Uniref90, OrthoDB, and BUSCO. Percentages were calculated from the count number of each parameter divided by the total number of contigs in the transcriptome (155,134). The only exception to this calculation is for complete ORFs, which were calculated as a percentage of the total ORFs (75,482). The BUSCO results for the annotated assembly are not shown here as they are identical to those for the un-annotated assembly.

<table>
<thead>
<tr>
<th>Contig #</th>
<th>Score</th>
<th>% SCO</th>
<th>% DCO</th>
<th>% frag</th>
<th>% miss</th>
<th>% mapping</th>
</tr>
</thead>
<tbody>
<tr>
<td>155,134</td>
<td>0.335</td>
<td>77</td>
<td>27</td>
<td>5.9</td>
<td>16</td>
<td>92.14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Search Type</th>
<th>TransDecoder</th>
<th>Rfam</th>
<th>Pfam-A</th>
<th>Uniref90</th>
<th>OrthoDB</th>
<th>Dammit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter</td>
<td>Total ORFs</td>
<td>Complete ORFs</td>
<td>ncRNAs</td>
<td>Protein Domains</td>
<td>Proteins</td>
<td>Orthologs</td>
</tr>
<tr>
<td>Count</td>
<td>75,482</td>
<td>43,028</td>
<td>937</td>
<td>25,675</td>
<td>62,865</td>
<td>51,806</td>
</tr>
<tr>
<td>Percentage</td>
<td>48.7%</td>
<td>57.0 %</td>
<td>0.6 %</td>
<td>16.6 %</td>
<td>40.5 %</td>
<td>33.4 %</td>
</tr>
</tbody>
</table>
Table 2: EdgeR determined significantly differentially expressed genes by treatment group in *P. eremicus* testes. Of the 15 DGE, seven were significantly more highly expressed in WET mice (High in WET) and eight were more highly expressed in DRY mice (High in DRY).

<table>
<thead>
<tr>
<th>Ensembl ID</th>
<th>log(_2)FC</th>
<th>logCPM</th>
<th>FDR</th>
<th>Gene ID</th>
<th>HIGH</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSMUSG00000079019.2</td>
<td>-4.354</td>
<td>1.650</td>
<td>5.82E-09</td>
<td>Ins13</td>
<td>WET</td>
</tr>
<tr>
<td>ENSMUSG00000054200.6</td>
<td>-3.734</td>
<td>0.619</td>
<td>1.82E-06</td>
<td>Ffar4</td>
<td>WET</td>
</tr>
<tr>
<td>ENSMUSG00000026435.15</td>
<td>-2.448</td>
<td>2.447</td>
<td>1.13E-03</td>
<td>Slc45a3</td>
<td>WET</td>
</tr>
<tr>
<td>ENSMUSG00000025020.11</td>
<td>-2.231</td>
<td>1.770</td>
<td>1.13E-03</td>
<td>Slit1</td>
<td>WET</td>
</tr>
<tr>
<td>ENSMUSG00000031170.14</td>
<td>-2.421</td>
<td>2.578</td>
<td>1.13E-03</td>
<td>Slc38a5</td>
<td>WET</td>
</tr>
<tr>
<td>ENSMUSG00000030830.18</td>
<td>-2.180</td>
<td>1.666</td>
<td>3.37E-02</td>
<td>Itgal</td>
<td>WET</td>
</tr>
<tr>
<td>ENSMUSG00000032554.15</td>
<td>-2.066</td>
<td>3.287</td>
<td>4.85E-02</td>
<td>Trf</td>
<td>WET</td>
</tr>
<tr>
<td>ENSMUSG0000001768.15</td>
<td>3.086</td>
<td>1.006</td>
<td>1.46E-07</td>
<td>Rin2</td>
<td>DRY</td>
</tr>
<tr>
<td>ENSMUSG00000025479.9</td>
<td>2.971</td>
<td>3.001</td>
<td>7.97E-05</td>
<td>Cyp2e1</td>
<td>DRY</td>
</tr>
<tr>
<td>ENSMUSG00000020427.11</td>
<td>2.681</td>
<td>3.887</td>
<td>1.13E-03</td>
<td>Igfbp3</td>
<td>DRY</td>
</tr>
<tr>
<td>ENSMUSG00000019997.11</td>
<td>2.314</td>
<td>3.235</td>
<td>1.13E-03</td>
<td>Ctgf</td>
<td>DRY</td>
</tr>
<tr>
<td>ENSMUSG00000040170.13</td>
<td>1.951</td>
<td>0.753</td>
<td>1.72E-03</td>
<td>Fmo2</td>
<td>DRY</td>
</tr>
<tr>
<td>ENSMUSG00000023915.4</td>
<td>1.534</td>
<td>1.290</td>
<td>2.02E-02</td>
<td>Tnfrsf21</td>
<td>DRY</td>
</tr>
<tr>
<td>ENSMUSG00000052974.8</td>
<td>2.077</td>
<td>0.647</td>
<td>2.26E-02</td>
<td>Cyp2f2</td>
<td>DRY</td>
</tr>
<tr>
<td>ENSMUSG00000027901.12</td>
<td>2.492</td>
<td>-0.620</td>
<td>4.78E-02</td>
<td>Dennd2d</td>
<td>DRY</td>
</tr>
</tbody>
</table>
Table 3: EdgeR determined significantly differentially expressed transcripts by treatment group in *P. eremicus* testes. Of the 66 total DTE, 45 were significantly more highly expressed in WET mice (High in WET) and 21 were more highly expressed in DRY mice (High in DRY). BLASTn matches to Ensembl IDs and corresponding Gene IDs.

<table>
<thead>
<tr>
<th>Transcript ID</th>
<th>log$_2$FC</th>
<th>logCPM</th>
<th>FDR</th>
<th>Ensembl ID</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>BINPACKER.15365.1</td>
<td>-3.703</td>
<td>0.047</td>
<td>5.31E-11</td>
<td>ENSMUSG00000054200.6</td>
<td>Ffar4</td>
</tr>
<tr>
<td>BINPACKER.2960.1</td>
<td>-4.268</td>
<td>1.147</td>
<td>2.06E-09</td>
<td>ENSMUSG00000079019.2</td>
<td>Insl3</td>
</tr>
<tr>
<td>BINPACKER.17981.2</td>
<td>-2.975</td>
<td>0.436</td>
<td>6.29E-08</td>
<td>ENSMUSG00000026435.15</td>
<td>Slc45a3</td>
</tr>
<tr>
<td>BINPACKER.9961.2</td>
<td>-2.426</td>
<td>1.998</td>
<td>7.50E-07</td>
<td>ENSMUSG00000031170.14</td>
<td>Slc38a5</td>
</tr>
<tr>
<td>BINPACKER.3452.1</td>
<td>-2.507</td>
<td>-0.140</td>
<td>3.56E-06</td>
<td>no match</td>
<td></td>
</tr>
<tr>
<td>BINPACKER.724.4</td>
<td>-2.162</td>
<td>2.667</td>
<td>8.32E-06</td>
<td>ENSMUSG00000032554.15</td>
<td>Trf</td>
</tr>
<tr>
<td>BINPACKER.9604.1</td>
<td>-2.582</td>
<td>0.547</td>
<td>7.87E-05</td>
<td>no match</td>
<td></td>
</tr>
<tr>
<td>BINPACKER.31087.1</td>
<td>-2.908</td>
<td>-0.858</td>
<td>9.74E-05</td>
<td>no match</td>
<td></td>
</tr>
<tr>
<td>BINPACKER.24398.1</td>
<td>-2.440</td>
<td>-0.689</td>
<td>9.74E-05</td>
<td>ENSMUSG00000036596.6</td>
<td>Cpz</td>
</tr>
<tr>
<td>BINPACKER.9726.1</td>
<td>-3.474</td>
<td>-0.107</td>
<td>2.38E-04</td>
<td>ENSMUSG00000026435.15</td>
<td>Slc45a3</td>
</tr>
<tr>
<td>BINPACKER.9218.3</td>
<td>-1.578</td>
<td>1.525</td>
<td>2.76E-04</td>
<td>ENSMUSG00000021253.6</td>
<td>Tgfb3</td>
</tr>
<tr>
<td>BINPACKER.18534.1</td>
<td>-2.332</td>
<td>1.346</td>
<td>4.85E-04</td>
<td>ENSMUSG00000025020.11</td>
<td>Slit1</td>
</tr>
<tr>
<td>BINPACKER.17022.3</td>
<td>-2.899</td>
<td>-0.561</td>
<td>1.00E-03</td>
<td>no match</td>
<td></td>
</tr>
<tr>
<td>BINPACKER.13806.1</td>
<td>-2.442</td>
<td>-0.381</td>
<td>1.13E-03</td>
<td>ENSMUSG00000025172.2</td>
<td>Ankrd2</td>
</tr>
<tr>
<td>BINPACKER.7740.1</td>
<td>-2.790</td>
<td>1.095</td>
<td>1.13E-03</td>
<td>ENSMUSG00000057074.6</td>
<td>Ces1g</td>
</tr>
<tr>
<td>BINPACKER.10034.2</td>
<td>-4.420</td>
<td>0.387</td>
<td>1.23E-03</td>
<td>ENSMUSG00000026516.8</td>
<td>Nvl</td>
</tr>
<tr>
<td>BINPACKER.11560.2</td>
<td>-1.465</td>
<td>2.050</td>
<td>1.66E-03</td>
<td>ENSMUSG00000021913.7</td>
<td>Ogdhl</td>
</tr>
<tr>
<td>BINPACKER.13701.1</td>
<td>-1.312</td>
<td>1.804</td>
<td>2.28E-03</td>
<td>ENSMUSG00000025648.17</td>
<td>Pfkfb4</td>
</tr>
<tr>
<td>BINPACKER.3510.3</td>
<td>-2.163</td>
<td>0.906</td>
<td>2.95E-03</td>
<td>ENSMUSG00000027822.16</td>
<td>Slc33a1</td>
</tr>
<tr>
<td>BINPACKER.15806.1</td>
<td>-1.700</td>
<td>1.062</td>
<td>3.39E-03</td>
<td>ENSMUSG00000015702.13</td>
<td>Anxa9</td>
</tr>
<tr>
<td>BINPACKER.17992.1</td>
<td>-2.542</td>
<td>0.653</td>
<td>3.39E-03</td>
<td>ENSMUSG00000030830.18</td>
<td>Itgal</td>
</tr>
<tr>
<td>BINPACKER.9726.2</td>
<td>-2.119</td>
<td>0.560</td>
<td>3.48E-03</td>
<td>ENSMUSG00000026435.15</td>
<td>Slc45a3</td>
</tr>
<tr>
<td>BINPACKER.6383.3</td>
<td>-2.093</td>
<td>1.270</td>
<td>4.16E-03</td>
<td>ENSMUSG0000002109.14</td>
<td>Ddb2</td>
</tr>
<tr>
<td>Transcript ID</td>
<td>log2FC</td>
<td>logCPM</td>
<td>FDR</td>
<td>Ensembl ID</td>
<td>Gene</td>
</tr>
<tr>
<td>---------------</td>
<td>--------</td>
<td>--------</td>
<td>-------</td>
<td>------------</td>
<td>------</td>
</tr>
<tr>
<td>BINPACKER.20716.2</td>
<td>-4.204</td>
<td>-0.566</td>
<td>5.75E-03</td>
<td>ENSMUSG00000013846.9</td>
<td>St3gal1</td>
</tr>
<tr>
<td>BINPACKER.20114.1</td>
<td>-1.661</td>
<td>0.501</td>
<td>5.97E-03</td>
<td>ENSMUSG00000030972.6</td>
<td>Acsm5</td>
</tr>
<tr>
<td>BINPACKER.18622.1</td>
<td>-1.645</td>
<td>1.704</td>
<td>6.36E-03</td>
<td>no match</td>
<td>-</td>
</tr>
<tr>
<td>BINPACKER.24914.1</td>
<td>-2.211</td>
<td>-0.159</td>
<td>9.83E-03</td>
<td>ENSMUSG00000003555.7</td>
<td>Cyp17a1</td>
</tr>
<tr>
<td>BINPACKER.31815.1</td>
<td>-1.905</td>
<td>-0.770</td>
<td>9.83E-03</td>
<td>no match</td>
<td>-</td>
</tr>
<tr>
<td>BINPACKER.6740.3</td>
<td>-3.090</td>
<td>-0.434</td>
<td>1.04E-02</td>
<td>no match</td>
<td>-</td>
</tr>
<tr>
<td>BINPACKER.20530.1</td>
<td>-1.626</td>
<td>0.545</td>
<td>1.12E-02</td>
<td>ENSMUSG00000038463.8</td>
<td>Olfml2b</td>
</tr>
<tr>
<td>BINPACKER.20656.1</td>
<td>-1.910</td>
<td>-0.531</td>
<td>1.22E-02</td>
<td>ENSMUSG00000029373.7</td>
<td>Pf4</td>
</tr>
<tr>
<td>BINPACKER.4855.1</td>
<td>-1.340</td>
<td>4.025</td>
<td>1.23E-02</td>
<td>ENSMUSG00000059991.7</td>
<td>Nptx2</td>
</tr>
<tr>
<td>BINPACKER.1846.1</td>
<td>-3.280</td>
<td>-0.792</td>
<td>1.23E-02</td>
<td>no match</td>
<td>-</td>
</tr>
<tr>
<td>BINPACKER.6494.2</td>
<td>-3.363</td>
<td>0.029</td>
<td>1.26E-02</td>
<td>ENSMUSG00000052861.13</td>
<td>Dnah6</td>
</tr>
<tr>
<td>BINPACKER.1818.1</td>
<td>-1.713</td>
<td>3.289</td>
<td>2.03E-02</td>
<td>ENSMUSG00000024125.1</td>
<td>Sbpl</td>
</tr>
<tr>
<td>BINPACKER.10743.2</td>
<td>-1.915</td>
<td>-0.525</td>
<td>2.06E-02</td>
<td>ENSMUSG00000041607.16</td>
<td>Mbp</td>
</tr>
<tr>
<td>BINPACKER.13054.2</td>
<td>-1.147</td>
<td>2.697</td>
<td>2.06E-02</td>
<td>ENSMUSG00000022994.8</td>
<td>Adcy6</td>
</tr>
<tr>
<td>BINPACKER.6807.1</td>
<td>-1.330</td>
<td>2.106</td>
<td>2.13E-02</td>
<td>ENSMUSG00000046687.5</td>
<td>Gm5424</td>
</tr>
<tr>
<td>BINPACKER.14160.1</td>
<td>-2.051</td>
<td>0.603</td>
<td>2.86E-02</td>
<td>ENSMUSG00000041556.8</td>
<td>Fbxo2</td>
</tr>
<tr>
<td>BINPACKER.16191.1</td>
<td>-1.431</td>
<td>0.926</td>
<td>3.42E-02</td>
<td>ENSMUSG00000028654.13</td>
<td>Mycl</td>
</tr>
<tr>
<td>BINPACKER.10141.3</td>
<td>-3.283</td>
<td>-1.191</td>
<td>3.68E-02</td>
<td>ENSMUSG00000024132.5</td>
<td>Eci1</td>
</tr>
<tr>
<td>BINPACKER.23790.1</td>
<td>-1.756</td>
<td>-0.275</td>
<td>4.51E-02</td>
<td>ENSMUSG00000001119.7</td>
<td>Col6a1</td>
</tr>
<tr>
<td>BINPACKER.22521.1</td>
<td>-1.841</td>
<td>-0.056</td>
<td>4.52E-02</td>
<td>ENSMUSG00000054083.8</td>
<td>Capn12</td>
</tr>
<tr>
<td>BINPACKER.1061.6</td>
<td>-1.807</td>
<td>1.943</td>
<td>4.93E-02</td>
<td>no match</td>
<td>-</td>
</tr>
<tr>
<td>BINPACKER.17734.1</td>
<td>-1.660</td>
<td>2.109</td>
<td>4.94E-02</td>
<td>ENSMUSG00000049608.8</td>
<td>Gpr55</td>
</tr>
</tbody>
</table>

**HIGH: DRY**

<table>
<thead>
<tr>
<th>Transcript ID</th>
<th>log2FC</th>
<th>logCPM</th>
<th>FDR</th>
<th>Ensembl ID</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>BINPACKER.21794.1</td>
<td>2.434</td>
<td>3.117</td>
<td>4.41E-08</td>
<td>ENSMUSG00000020427.11</td>
<td>Igfbp3</td>
</tr>
<tr>
<td>BINPACKER.28731.1</td>
<td>2.484</td>
<td>1.634</td>
<td>4.41E-08</td>
<td>no match</td>
<td>-</td>
</tr>
<tr>
<td>BINPACKER.5662.4</td>
<td>2.061</td>
<td>2.419</td>
<td>1.32E-07</td>
<td>ENSMUSG00000019997.11</td>
<td>Ctgf</td>
</tr>
<tr>
<td>BINPACKER.87639.1</td>
<td>2.682</td>
<td>0.345</td>
<td>1.96E-07</td>
<td>ENSMUSG0000001768.15</td>
<td>Rin2</td>
</tr>
<tr>
<td>BINPACKER.35470.1</td>
<td>2.367</td>
<td>1.786</td>
<td>1.89E-04</td>
<td>no match</td>
<td>-</td>
</tr>
<tr>
<td>BINPACKER.52106.1</td>
<td>2.096</td>
<td>-0.542</td>
<td>6.83E-04</td>
<td>no match</td>
<td>-</td>
</tr>
<tr>
<td>BINPACKER.3957.3</td>
<td>6.309</td>
<td>1.579</td>
<td>1.02E-03</td>
<td>ENSMUSG00000019988.6</td>
<td>Nedd1</td>
</tr>
<tr>
<td>------------------</td>
<td>-------</td>
<td>-------</td>
<td>-------------</td>
<td>----------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>BINPACKER.116235.1</td>
<td>2.212</td>
<td>0.301</td>
<td>3.94E-03</td>
<td>no match</td>
<td>-</td>
</tr>
<tr>
<td>BINPACKER.4449.4</td>
<td>3.428</td>
<td>-0.538</td>
<td>6.74E-03</td>
<td>ENSMUSG00000005150.16</td>
<td>Wdr83</td>
</tr>
<tr>
<td>BINPACKER.28.2</td>
<td>4.183</td>
<td>2.295</td>
<td>1.05E-02</td>
<td>ENSMUSG00000075706.10</td>
<td>Gpx4</td>
</tr>
<tr>
<td>BINPACKER.56553.1</td>
<td>1.472</td>
<td>0.172</td>
<td>1.46E-02</td>
<td>no match</td>
<td>-</td>
</tr>
<tr>
<td>BINPACKER.93518.1</td>
<td>1.711</td>
<td>-0.793</td>
<td>1.57E-02</td>
<td>no match</td>
<td>-</td>
</tr>
<tr>
<td>BINPACKER.11512.1</td>
<td>1.187</td>
<td>3.654</td>
<td>1.70E-02</td>
<td>ENSMUSG00000031591.14</td>
<td>Asah1</td>
</tr>
<tr>
<td>BINPACKER.66588.1</td>
<td>1.851</td>
<td>-0.347</td>
<td>1.71E-02</td>
<td>no match</td>
<td>-</td>
</tr>
<tr>
<td>BINPACKER.42718.1</td>
<td>1.542</td>
<td>0.507</td>
<td>2.06E-02</td>
<td>ENSMUSG00000030790.15</td>
<td>Adm</td>
</tr>
<tr>
<td>BINPACKER.49203.1</td>
<td>1.639</td>
<td>-0.035</td>
<td>2.44E-02</td>
<td>no match</td>
<td>-</td>
</tr>
<tr>
<td>BINPACKER.147548.1</td>
<td>1.744</td>
<td>-0.007</td>
<td>2.99E-02</td>
<td>ENSMUSG00000042757.15</td>
<td>Tmem108</td>
</tr>
<tr>
<td>BINPACKER.23756.2</td>
<td>1.265</td>
<td>3.468</td>
<td>3.01E-02</td>
<td>ENSMUSG00000022061.8</td>
<td>Nkx3-1</td>
</tr>
<tr>
<td>BINPACKER.12709.1</td>
<td>3.906</td>
<td>2.611</td>
<td>3.01E-02</td>
<td>ENSMUSG00000028639.14</td>
<td>Ybx1</td>
</tr>
<tr>
<td>BINPACKER.5280.2</td>
<td>3.874</td>
<td>0.257</td>
<td>3.76E-02</td>
<td>ENSMUSG00000074582.10</td>
<td>Arfgef2</td>
</tr>
<tr>
<td>BINPACKER.58702.1</td>
<td>1.780</td>
<td>-0.500</td>
<td>4.93E-02</td>
<td>no match</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 4: DESeq2 determined significantly differentially expressed genes by treatment group in *P. eremicus* testes. Of the 20 DGE with a $-1 < \log_2$ fold change > 1, 16 were significantly more highly expressed in WET mice (High in WET) and four were more highly expressed in DRY mice (High in DRY).

<table>
<thead>
<tr>
<th>Ensembl ID</th>
<th>baseMean</th>
<th>log2FC</th>
<th>p-adjusted</th>
<th>Gene ID</th>
<th>HIGH</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENSMUSG00000054200.6</td>
<td>8.77721485</td>
<td>-2.2659204</td>
<td>1.24E-27</td>
<td>Ffar4</td>
<td>WET</td>
</tr>
<tr>
<td>ENSMUSG00000026435.15</td>
<td>38.7630267</td>
<td>-2.2184407</td>
<td>1.16E-42</td>
<td>Slc45a3</td>
<td>WET</td>
</tr>
<tr>
<td>ENSMUSG00000079019.2</td>
<td>24.7158409</td>
<td>-1.6454793</td>
<td>4.55E-13</td>
<td>Insl3</td>
<td>WET</td>
</tr>
<tr>
<td>ENSMUSG00000031170.14</td>
<td>42.2322119</td>
<td>-1.6434261</td>
<td>6.64E-15</td>
<td>Slc38a5</td>
<td>WET</td>
</tr>
<tr>
<td>ENSMUSG00000038463.8</td>
<td>16.2605998</td>
<td>-1.4619721</td>
<td>3.55E-12</td>
<td>Olfml2b</td>
<td>WET</td>
</tr>
<tr>
<td>ENSMUSG00000030830.18</td>
<td>22.0478661</td>
<td>-1.4358002</td>
<td>3.41E-10</td>
<td>Itgal</td>
<td>WET</td>
</tr>
<tr>
<td>ENSMUSG00000032554.15</td>
<td>67.5197473</td>
<td>-1.3762549</td>
<td>7.26E-10</td>
<td>Trf</td>
<td>WET</td>
</tr>
<tr>
<td>ENSMUSG00000021253.6</td>
<td>31.2493344</td>
<td>-1.3551661</td>
<td>7.02E-14</td>
<td>Tgfβ3</td>
<td>WET</td>
</tr>
<tr>
<td>ENSMUSG00000030972.6</td>
<td>13.8934534</td>
<td>-1.1709964</td>
<td>2.37E-07</td>
<td>Acsm5</td>
<td>WET</td>
</tr>
<tr>
<td>ENSMUSG00000059991.7</td>
<td>173.025492</td>
<td>-1.1528314</td>
<td>5.12E-11</td>
<td>Nptx2</td>
<td>WET</td>
</tr>
<tr>
<td>ENSMUSG00000046687.5</td>
<td>44.9527785</td>
<td>-1.0989949</td>
<td>8.31E-09</td>
<td>Gm5424</td>
<td>WET</td>
</tr>
<tr>
<td>ENSMUSG00000024125.1</td>
<td>101.5876</td>
<td>-1.0962074</td>
<td>9.77E-06</td>
<td>Sbpl</td>
<td>WET</td>
</tr>
<tr>
<td>ENSMUSG00000021913.7</td>
<td>46.5401886</td>
<td>-1.0876018</td>
<td>8.70E-07</td>
<td>Ogdh1</td>
<td>WET</td>
</tr>
<tr>
<td>ENSMUSG00000015702.13</td>
<td>27.7002506</td>
<td>-1.0603879</td>
<td>1.95E-05</td>
<td>Anxa9</td>
<td>WET</td>
</tr>
<tr>
<td>ENSMUSG00000036596.6</td>
<td>6.6698922</td>
<td>-1.0243046</td>
<td>9.04E-05</td>
<td>Cpz</td>
<td>WET</td>
</tr>
<tr>
<td>ENSMUSG00000025172.2</td>
<td>13.2622565</td>
<td>-1.0138171</td>
<td>0.00013318</td>
<td>Ankrd2</td>
<td>WET</td>
</tr>
<tr>
<td>ENSMUSG00000042757.15</td>
<td>14.5676529</td>
<td>1.00643936</td>
<td>0.00019556</td>
<td>Tmem108</td>
<td>DRY</td>
</tr>
<tr>
<td>ENSMUSG00000019997.11</td>
<td>64.49614</td>
<td>1.03331405</td>
<td>7.67E-05</td>
<td>Ctgf</td>
<td>DRY</td>
</tr>
<tr>
<td>ENSMUSG00000020427.11</td>
<td>92.3763518</td>
<td>1.56656207</td>
<td>4.55E-13</td>
<td>Igfbp3</td>
<td>DRY</td>
</tr>
<tr>
<td>ENSMUSG0000001768.15</td>
<td>12.3794312</td>
<td>1.72433255</td>
<td>8.16E-16</td>
<td>Rin2</td>
<td>DRY</td>
</tr>
</tbody>
</table>
Table 5: Functional information and chromosome (CHR) locations (*Mus musculus*) for the nine genes differentially expressed across all three analyses in *P. eremicus* testes by treatment group.

<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Gene ID</th>
<th>Gene Function</th>
<th>CHR</th>
<th>HIGH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insulin-like 3</td>
<td>Insl3</td>
<td>testicular function and testicular development</td>
<td>8</td>
<td>WET</td>
</tr>
<tr>
<td>Free-fatty acid receptor 4</td>
<td>Ffar4</td>
<td>metabolism and inflammation</td>
<td>19</td>
<td>WET</td>
</tr>
<tr>
<td>Solute carrier family 45 member 3</td>
<td>Slc45a3</td>
<td>sugar transport</td>
<td>1</td>
<td>WET</td>
</tr>
<tr>
<td>Solute carrier family 38 member 5</td>
<td>Slc38a5</td>
<td>sodium-dependent amino acid transport</td>
<td>X</td>
<td>WET</td>
</tr>
<tr>
<td>Integrin alpha L</td>
<td>Itgal</td>
<td>lymphocyte-mediated immune responses</td>
<td>7</td>
<td>WET</td>
</tr>
<tr>
<td>Transferrin</td>
<td>Trf</td>
<td>iron transport and delivery to erythrocytes</td>
<td>9</td>
<td>WET</td>
</tr>
<tr>
<td>Ras and Rab Interactor 2</td>
<td>Rin2</td>
<td>endocytosis and membrane trafficking</td>
<td>2</td>
<td>DRY</td>
</tr>
<tr>
<td>Insulin-like growth factor binding protein 3</td>
<td>Igfbp3</td>
<td>modulates effects of insulin growth factors</td>
<td>11</td>
<td>DRY</td>
</tr>
<tr>
<td>Connective tissue growth factor</td>
<td>Ctgf</td>
<td>fibrosis and extracellular matrix formation</td>
<td>10</td>
<td>DRY</td>
</tr>
</tbody>
</table>
Figure 1: Correlation of log$_2$ fold change results for all Ensembl ID gene matches from DESeq2 and edgeR DGE analyses (Adj-R$_2$ = 0.6596; F(1,14214) = 2.754x10$^4$; p < 2.2x10$^{-16}$).
Figure 2: Box plots of edgeR analyzed differences in gene expression by treatment for the nine genes significantly differentially expressed in all three analyses. Counts per million (cpms) for both treatments (WET and DRY) are indicated.
Figure 3: STRING diagrams of protein-protein interactions for genes significantly differentially expressed (highly expressed) in the WET treatment group. These six genes are (a) Slc38a5, (b) Slc45a3, (c) Insl3, (d) Ffar4 (also known as O3far1), (e) Itgal, and (f) Trf. Different colored circles stipulate different proteins interacting with the target proteins, small circles are proteins with unknown 3D structure, while larger circles are proteins with some degree of known or predicted 3D structure. Different colors of connecting lines represent different types of interactions between proteins. For fully interactive diagrams of the genes, view the provided links to string-db in the GitHub repository (StringDBlinks.md)

(a)
Figure 4: STRING diagrams of protein-protein interactions for genes significantly differentially expressed (highly expressed) in the DRY treatment group. These three genes are (a) Rin2, (b) Igfbp3, and (c) Ctgf. Different colored circles stipulate different proteins interacting with the target proteins, small circles are proteins with unknown 3D structure, while larger circles are proteins with some degree of known or predicted 3D structure. Different colors of connecting lines represent different types of interactions between proteins. For fully interactive diagrams of the genes, view the provided links to string-db in the GitHub repository (StringDBlinks.md).

(a)
Figure 5: Box plots of edgeR analyzed differences in gene expression by treatment for the nine \emph{a priori} tested reproductive hormone and hormone receptor genes. Counts per million (cpms) for both treatments (WET and DRY) are indicated.
Supplemental Table 1: Testes read data statistics, including sample identification (Mouse ID), number of reads (# Reads), percent reads mapped to transcriptome (% Mapping), and treatment group (TRT). Mouse ID 335T* is the dataset which was used to assemble the testes transcriptome; therefore, these reads were not used for the differential expression analysis.

<table>
<thead>
<tr>
<th>Mouse ID</th>
<th># Reads</th>
<th>% Mapping</th>
<th>TRT</th>
</tr>
</thead>
<tbody>
<tr>
<td>335T*</td>
<td>45759114</td>
<td>85.46</td>
<td>wet</td>
</tr>
<tr>
<td>3333T</td>
<td>15135923</td>
<td>82.56</td>
<td>Wet</td>
</tr>
<tr>
<td>2322T</td>
<td>12584407</td>
<td>82.37</td>
<td>Dry</td>
</tr>
<tr>
<td>382T</td>
<td>14305186</td>
<td>83.87</td>
<td>Dry</td>
</tr>
<tr>
<td>381T</td>
<td>14178847</td>
<td>83.23</td>
<td>Wet</td>
</tr>
<tr>
<td>376T</td>
<td>14588175</td>
<td>82.56</td>
<td>Dry</td>
</tr>
<tr>
<td>366T</td>
<td>13641731</td>
<td>82.95</td>
<td>Wet</td>
</tr>
<tr>
<td>349T</td>
<td>17289781</td>
<td>85.93</td>
<td>Wet</td>
</tr>
<tr>
<td>209T</td>
<td>11724617</td>
<td>84.02</td>
<td>Dry</td>
</tr>
<tr>
<td>265T</td>
<td>11536510</td>
<td>84.17</td>
<td>Dry</td>
</tr>
<tr>
<td>383T</td>
<td>13250034</td>
<td>81.46</td>
<td>Dry</td>
</tr>
<tr>
<td>384T</td>
<td>12152820</td>
<td>82.75</td>
<td>Dry</td>
</tr>
<tr>
<td>102T</td>
<td>11131941</td>
<td>84.84</td>
<td>Wet</td>
</tr>
<tr>
<td>400T</td>
<td>13259393</td>
<td>83.98</td>
<td>Wet</td>
</tr>
<tr>
<td>1357T</td>
<td>20603232</td>
<td>82.32</td>
<td>Wet</td>
</tr>
<tr>
<td>1358T</td>
<td>12240814</td>
<td>86.58</td>
<td>Wet</td>
</tr>
<tr>
<td>1359T</td>
<td>11144962</td>
<td>85.54</td>
<td>Wet</td>
</tr>
<tr>
<td>13T</td>
<td>11075885</td>
<td>83.55</td>
<td>Dry</td>
</tr>
<tr>
<td>343T</td>
<td>9423867</td>
<td>83.58</td>
<td>Dry</td>
</tr>
<tr>
<td>344T</td>
<td>17146134</td>
<td>85.36</td>
<td>Wet</td>
</tr>
<tr>
<td>355T</td>
<td>13948415</td>
<td>85.21</td>
<td>Wet</td>
</tr>
<tr>
<td>888T</td>
<td>18890387</td>
<td>86.52</td>
<td>Dry</td>
</tr>
<tr>
<td>999T</td>
<td>15213425</td>
<td>87.02</td>
<td>Dry</td>
</tr>
</tbody>
</table>
Supplemental Table 2: Significantly differentially expressed genes identified in the three analyses (DGE in edgeR, DTE in edgeR, and DGE in DESeq2) by treatment group in *P. eremeicus* testes. Of the 34 different genes which were more highly expressed in WET mice, six were significant across all three analyses (Gene IDs are italicized). Of the 17 genes which were more highly expressed in DRY mice, three were significant across all three analyses (Gene IDs are italicized).

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>DGE edgeR</th>
<th>DTE edgeR</th>
<th>DGE DESeq2</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Insl3</em></td>
<td>x</td>
<td>X</td>
<td>x</td>
</tr>
<tr>
<td><em>Ffar4</em></td>
<td>x</td>
<td>X</td>
<td>x</td>
</tr>
<tr>
<td><em>Slc45a3</em></td>
<td>x</td>
<td>X</td>
<td>x</td>
</tr>
<tr>
<td><em>Slc38a5</em></td>
<td>x</td>
<td>X</td>
<td>x</td>
</tr>
<tr>
<td><em>Itgal</em></td>
<td>x</td>
<td>X</td>
<td>x</td>
</tr>
<tr>
<td><em>Trf</em></td>
<td>x</td>
<td>X</td>
<td>x</td>
</tr>
<tr>
<td>Slit1</td>
<td>x</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Cpz</td>
<td></td>
<td>X</td>
<td>x</td>
</tr>
<tr>
<td>Tgfb3</td>
<td></td>
<td>X</td>
<td>x</td>
</tr>
<tr>
<td>Ces1g</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Ankrd2</td>
<td></td>
<td>X</td>
<td>x</td>
</tr>
<tr>
<td>Nvl</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Ogdhl</td>
<td></td>
<td>X</td>
<td>x</td>
</tr>
<tr>
<td>Pfkfb4</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Slc33a1</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Anxa9</td>
<td></td>
<td>X</td>
<td>x</td>
</tr>
<tr>
<td>Ddb2</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>St3gal1</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Acsm5</td>
<td></td>
<td>X</td>
<td>x</td>
</tr>
<tr>
<td>Cyp17a1</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Olfm12b</td>
<td></td>
<td>X</td>
<td>x</td>
</tr>
<tr>
<td>Pf4</td>
<td></td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Nptx2</td>
<td></td>
<td>X</td>
<td>x</td>
</tr>
<tr>
<td>Gene ID</td>
<td>DGE edgeR</td>
<td>DTE edgeR</td>
<td>DGE DESeq2</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
<td>-----------</td>
<td>------------</td>
</tr>
<tr>
<td><strong>105</strong></td>
<td>X</td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>DGE edgeR</th>
<th>DTE edgeR</th>
<th>DGE DESeq2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dnah6</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sbpl</td>
<td>X</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Adcy6</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gm5424</td>
<td>X</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Mbp</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fbxo2</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mycl</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eci1</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capn12</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col6a1</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gpr55</td>
<td>x</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**HIGH: DRY**

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>DGE edgeR</th>
<th>DTE edgeR</th>
<th>DGE DESeq2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rin2</td>
<td>x</td>
<td>X</td>
<td>x</td>
</tr>
<tr>
<td>Igfbp3</td>
<td>x</td>
<td>X</td>
<td>x</td>
</tr>
<tr>
<td>Ctgf</td>
<td>x</td>
<td>X</td>
<td>x</td>
</tr>
<tr>
<td>Cyp2e1</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fmo2</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tnfrsf21</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyp2f2</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dennd2d</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nedd1</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Wdr83</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Gpx4</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Asah1</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Adm</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Tmem108</td>
<td></td>
<td>X</td>
<td>x</td>
</tr>
<tr>
<td>Nkx3-1</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Ybx1</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Arfggef2</td>
<td></td>
<td>x</td>
<td></td>
</tr>
</tbody>
</table>
Supplemental Figure 1: Plot of edgeR determined differentially expressed genes. The 15 significant genes are in red, with positive values indicating increased expression in the DRY group, and negative values depicting increased expression in the WET group.
Supplemental Figure 2: Plot of edgeR determined differentially expressed transcripts. The 66 significant transcripts are in red, with positive values indicating increased expression in the DRY group, and negative values depicting increased expression in the WET group.
Supplemental Figure 3: Plot of DESeq2 determined differentially expressed transcripts. The 215 significant transcripts are in red, with positive values indicating increased expression in the DRY group, and negative values depicting increased expression in the WET group.
REFERENCES:


**Boone M, Deen PMT. 2008.** Physiology and pathophysiology of the vasopressin-regulated renal water reabsorption. Pflugers Archiv 456(6): 1005-1024. DOI: 10.1007/s00424-008-0498-1


**Breed WG, Leigh CM. 2011.** Reproductive biology of an old endemic murid rodent of Australia, the Spinifex hopping mouse, *Notomys alexis*: adaptations for life in the arid zone.


MacManes MD. 2016. Establishing evidenced-based best practice for the de novo assembly and
evaluation of transcriptomes from non-model organisms. bioRxiv. DOI: 10.1101/035642

MacManes MD, Eisen MB. 2014. Characterization of the transcriptome, nucleotide sequence
polymorphism, and natural selection in the desert adapted mouse *Peromyscus eremicus*. PeerJ 2:
e642. doi.org/10.7717/peerj.642.

MacManes MD. 2017. Severe acute dehydration in a desert rodent elicits a transcriptional
response that effectively prevents kidney injury. American Journal of Physiology-Renal

MacMillen RE, Garland T Jr. in *Advances in the Study of Peromyscus (Rodentia)* (Texas Tech


Madden T. 2002 Oct 9 [Updated 2003 Aug 13]. The BLAST Sequence Analysis Tool. In:
McEntyre J, Ostell J, editors. The NCBI Handbook [Internet]. Bethesda (MD): National Center
for Biotechnology Information (US); 2002-. Chapter 16. Available from:

Marra NJ, Eo SH, Hale MC, Waser PM, DeWoody JA. 2012. *A priori* and *a posteriori*
approaches for finding genes of evolutionary interest in non-model species: Osmoregulatory
genes in the kidney transcriptome of the desert rodent *Dipodomys spectabilis* (banner-tailed

osmoregulation in heteromyid rodents as revealed by RNA-seq. Molecular Ecology 23(11):
2699-2711. DOI: 10.1111/mec.12764.

**Martin K, Wiebe KL. 2004.** Coping mechanisms of alpine and arctic breeding birds: extreme weather and limitations to reproductive resilience. Integrative and Comparative Biology 44(2):177-185. DOI: 10.1093/icb/44.2.177.


http://doi.org.libproxy.unh.edu/10.1016/j.regpep.2012.07.003


Doi: 10.1016/0024-3205(88)90295-0


https://doi.org/10.1095/biolreprod12.2.249
CHAPTER 3

Characterization of the Seminal Vesicle Proteome of *Peromyscus eremicus* (Cactus mouse) in hydrated and dehydrated conditions

Lauren Kordonowy 1* Donna Hogan 1 Feixia Chu 1 Matthew MacManes 1+

1. University of New Hampshire Department of Molecular, Cellular and Biomedical Sciences
Durham, NH, USA

* [lauren.kordonowy@unh.edu](mailto:lauren.kordonowy@unh.edu)
+ [matthew.macmanes@unh.edu](mailto:matthew.macmanes@unh.edu)

**Corresponding Author:**

Lauren Kordonowy,

[lauren.kordonowy@unh.edu](mailto:lauren.kordonowy@unh.edu)

*Submission Pending* for *Proteomics* as a *Dataset Brief*
ABSTRACT

The cactus mouse (*Peromyscus eremicus*) is a desert-specialized rodent that experiences both chronic and acute dehydration in the Southwestern United States. Our previous research has generated substantial transcriptomic data on *P. eremicus* kidneys, testes, epididymis, and vas deferens in individuals exposed to hydrated and dehydrated conditions; however, the study described here is the first to describe a seminal vesicle proteome for this species. We have produced a seminal vesicle proteome from *P. eremicus* with free access to water and mice that were acutely dehydrated to generate a dataset that is comprehensive for both alternative water-availability states experienced by this species. We have also provided gene ontologies for this proteome using PANTHER. This proteome will provide a crucial resource for future studies characterizing the genetic and proteomic responses of reproductive tissues to drought in this desert-specialized rodent. Furthermore, an enhanced understanding of survival and reproductive responses (and adaptations) to dehydration is particularly relevant to clinical work aimed at minimizing adverse human impacts, as climate change continues to increase the incidence of drought.
DATASET BRIEF

The cactus mouse (*Peromyscus eremicus*), is a desert-specialized rodent endemic to the Southwestern United States that experiences both acute and chronic dehydration and can live entirely without water (Veal and Caire, 1979; Caire, 1999). Previous research has characterized the physiological (kidney - Kordonowy et al., 2017) and transcriptomic (kidney - MacManes and Eisen, 2014; MacManes, 2017, testes - Kordonowy and MacManes, 2017) effects of water limitation, while other studies have characterized cactus mouse reproductive tissue transcriptomes (testes, vas deferens, and epididymis - Kordonowy and MacManes, 2016). The study described here contributes to this desert adaptation research focus by describing a seminal vesicle proteome from cactus mice with access to water and mice that were acutely dehydrated.

Reproductively mature male *Peromyscus eremicus* (described in Kordonowy and MacManes, 2016) were maintained in a laboratory colony at University of New Hampshire in a desert chamber mimicking temperature and precipitation conditions in southwestern US deserts. Three mice were deprived of water for ~72 hours (DRY group) immediately prior to sacrifice; whereas another three mice were provided water *ad libidum* (WET group). Seminal vesicles were harvested from mice within ten minutes of euthanasia (via isoflurane overdose and subsequent decapitation). Proteins from the homogenized seminal vesicles were extracted with a Qproteome Mammalian Prep kit (Qiagen) and frozen in liquid nitrogen within one hour of sacrifice. All research was done in compliance with University of New Hampshire Animal Care and Use Committee guidelines (protocol number 130902) and in accordance with guidelines by the American Society of Mammalogists (Sikes, 2016).

Protein content of the samples was analyzed with a Qubit 2.0 Fluorometer (Invitrogen), and similar protein quantities from each sample were run on different lanes on a 4-15% gradient.
SDS page gel (BioRad) and electrophoresed constantly at 0.02 amps to separate proteins by molecular weight. Each sample lane was discretely sub-divided, and these gel fragments were digested with trypsin, in keeping with current procedures (reviewed in Deutsch et al., 2008; Cappadona et al., 2012).

The current methodology for mass spectrometry data generation for proteomic analysis has been thoroughly described (e.g. Deutsch et al., 2008; Cappadona et al., 2012; Schmidt et al., 2014). This study utilizes shotgun proteomics, wherein MS data is leveraged to identify proteins from their component peptide compositions (reviewed in Duncan et al., 2010; Cappadona et al., 2012). All samples were analyzed by LC-MS/MS as described by Zeng-Elmore and colleagues (2014). For the HPLC, 1 uL aliquots of each digested protein sample were continuously injected by the UltiMate 3000 Autosampler (Dionex Corp) at a flow rate of 450 nL/min into the 150 um x 10 cm reverse-phase capillary column containing C18 resin. The resulting peptides were introduced to the nanoelectrospray ionization source of the LTQ Orbitrap XL mass spectrometer (Thermo Scientific). All of the LC-MS/MS data was collected using an information-dependent mode setting, and MS scans in the Orbitrap (with an m/z range between 315-1800) were alternated with low-energy CID analysis for further characterization.

PAVA (Guan et al., 2011) converted the raw data into peak lists, which were uploaded into Protein Prospector version 5.18.22 (Chalkley et al., 2008). We searched for protein matches with Protein Prospector for each biological sample using a database specific to Peromyscus eremicus, which contains 24,425 Maker-derived protein sequences (MAKER version 2.31.9: Holt and Yandell, 2011). Within Batch-Tag Web, the custom database was searched using the following parameters: the taxonomy was Mus musculus, 1 was the Max. Missed Cleavage, Carbamidomethyl (C) was the constant modification, 0.7 Da was the parent tolerance, 20 ppm
was the fragment tolerance, 2-5 was the precursor charge range, and the remainder of the parameters were the default settings. Specifically, this included a Max E Value for Proteins of 0.01 and for Peptides of 0.05, and a Min Score for Proteins of 22.0 and for Peptides of 15.0. The .txt results files from these Protein Prospector searches were used to compile the 1,142 contig ID matches for building the proteome (described below).

R. Chalkley also generated a random concatenated database (RCD) in Protein Prospector with the *P. eremicus* database in order to determine the false discovery rates (FDRs) of each sample. The proteins within the RCD database are the same size as those in the original *P. eremicus* database, but the protein sequences are randomized (as per Knudsen & Chalkley, 2011). To execute the RCD search, the results generated by each of the previous searches with the *P. eremicus* database were re-submitted for a new search with the RCD *P. eremicus* database to Batch-Tag Web after removing the accession number matches from the Pre-Search Parameters (using the same parameters previously described). The number and percent of matches within the RCD relative to the complete *P. eremicus* database were determined for each sample to calculate FDRs. The FDRs were all <5.25% for each sample of our protein level analysis.

The proteome was generated by including *P. eremicus* protein sequence matches found in Protein Prospector (identified 1,142 contig IDs within the *P. eremicus* database) which were supported by one or more unique peptides. All protein matches (matched contig IDs) were retrieved from the *P. eremicus* database file and compiled. This set of *P. eremicus* protein sequences comprises our un-annotated seminal vesicle proteome. This group of *P. eremicus* sequences was also used to execute a BLASTp (Altschul et al. 1990; Madden, 2002) search to find corresponding proteins in a downloaded (Ensembl.org) *Mus musculus* peptide database (version updated 11/24/16 release-87). This search yielded 1,084 different *Mus* protein matches,
which constitutes our *Mus musculus* annotated *P. eremicus* contig IDs.

All *Mus* protein accession number search matches were extracted from the search results and uploaded into Panther (pantherdb.org), a gene ontology search platform (Mi et al. 2013; 2017), where we conducted our GO search within *Mus musculus*. The Panther GO analysis resulted in 777 total gene matches, which were categorized into five viewable formats: *Biological Process, Molecular Function, Cellular Component, Protein Class,* and *Pathway*. In our supplementary materials we have included the GO results for this list of genes as well as the accession number file we uploaded to generate reproducible results for each of these five formats. However, we present the *Biological Process* results here (Figure 1) to display the relative number of GO results in each category within this format. For example, within *Biological Process*, five genes were grouped in the *Reproduction* GO. Furthermore, among these five genes: three were placed within the *gamete generation* GO, one within the *fertilization* GO, and one was un-classified.

In summary, we generated a seminal vesicle proteome from dehydrated and hydrated cactus mice. Moreover, the PANTHER gene ontologies provide a thorough characterization of the proteins found in this tissue. This study describes a new reproductive tissue proteome in a desert-specialized rodent, which adds to a growing body of research focused on drought tolerance in a wide range of taxa (*e.g. plants*: Behringer et al., 2015; Chen et al., 2016; *insects*: Guillen et al., 2014; *mammals*: Diaz et al, 2006; Fuller et al., 2014). Indeed, understanding organismal responses to water imitation is particularly relevant as current projections predict dramatic climate change, and we may garner clinically relevant information from desert species adapted to combat dehydration (Johnson et al., 2016).

All three of our data-types (raw LC-MS, Protein Prospector protein identification results,
and the proteome sequences) are publically available in the appropriate repositories as per recommendations by Perez-Riverol et al. (2015) and are provided in Supplemental Materials.

**Supplemental Materials are available on GitHub, unless otherwise stipulated:**

SVproteomeMethods markdown file (containing all code and analyses)

*Peromyscus eremicus* seminal vesicle proteome: in UniProt (The UniProt Consortium, 2017)

*Peromyscus eremicus* amino acid sequence database file for Protein Prospector search

*Peromyscus eremicus* FDR database file for Protein Prospector search

*Peromyscus eremicus* proteome BLASTp matches to *Mus musculus* database (with scores)

Panther Gene Ontology Analysis: accession number file for upload and GO results file

Raw LC-MS data: available in PRIDE (Vizcaino et al., 2016)

Protein Prospector data: available in PRIDE

All MS/MS spectrum annotated with masses observed as well as fragment assignments (SKYLINE)
Figure 1 Biological Process GO analysis produced by PANTHER (re-graphed in R)
REFERENCES:


DOI: 10.1152/physiolgenomics.00298.2007

Duncan MW, Aebersold R, Caprioli RM. 2010. The pros and cons of peptide-centric
proteomics. Nature Biotechnology 28(7): 659-664. DOI: 10.1038/nbt0710-659


Knudsen GM, Chalkley RJ. 2011 The Effect of Using an Inappropriate Protein Database for Proteomic Data Analysis. PLoS ONE6(6): e20873. DOI:10.1371/journal.pone.0020873


IACUC #: 160904  
Project: Using Genomics to Understand Adaptation to Desert Life in Rodents  
Approval Date: 15-Sep-2016

The Institutional Animal Care and Use Committee (IACUC) reviewed and approved the protocol submitted for this study under Category C on Page 5 of the Application for Review of Vertebrate Animal Use in Research or Instruction - the research potentially involves minor short-term pain, discomfort or distress which will be treated with appropriate anesthetics/analgesics or other assessments.

Approval is granted for a period of three years from the approval date above. Continued approval throughout the three year period is contingent upon completion of annual reports on the use of animals. At the end of the three year approval period you may submit a new application and request for extension to continue this project. Requests for extension must be filed prior to the expiration of the original approval.

Please Note:

1. All cage, pen, or other animal identification records must include your ACUC # listed above.
2. Use of animals in research and instruction is approved contingent upon participation in the UNH Occupational Health Program for persons handling animals. Participation is mandatory for all principal investigators and their affiliated personnel, employees of the University and students alike. Information about the program, including forms, is available at http://unh.edu/research/occupational-health-program-animal-handlers.

If you have any questions, please contact either Dean Elder at 862-4629 or Julie Simpson at 862-2003.

For the IACUC,  
Jill A. McGaughy, Ph.D.  
Chair

cc: File