

i. Significance

For biologists interested in understanding the relationship between fitness, genotype, and phenotype, modern sequencing technologies provide for an unprecedented opportunity to gain a deep understanding of genome level processes that together, underlie adaptation. This newfound power, combined with a deep understanding of ecology, ontogeny, evolutionary history and fitness benefit represents an extremely rich and novel context within which to better understand classic questions in evolutionary biology - the links between environmental, phenotypic and genetic variation. The proposed research leverages this emerging opportunity to explore adaptation in one of the most harsh environments on Earth - the desert. Here, heat *and* drought provide for powerful selective forces, testing animals' ability to osmoregulate and thus to survive.

The maintenance of water balance in animals is one of the most important physiologic processes, and is critical to desert survival. Indeed, mammals are exquisitely sensitive to changes in osmolality, with slight derangement eliciting physiologic compromise. When the loss of water exceeds dietary intake, dehydration - and in extreme cases, death - can occur. Unlike most mammals, animals living in desert habitats are subjected to long periods of extreme heat and intense drought. As a result, desert animals have evolved mechanisms through which physiologic homeostasis is maintained despite severe and prolonged dehydration. **The proposed research uses a novel approach integrating physiology, evolutionary genomics, and computational biology to better understand how animals survive in what appear to be unsurvivable conditions.** This integrative research program will significantly enhance our understanding of the physiologic processes underlying osmoregulation in extreme environments. Its impact is broad, having implications for studies of human health, conservation, and climate change. This project both leverages and enhances existing ecological, behavioral, and genomic resources available to an active *Peromyscus* research community, ultimately allowing for interesting studies of comparative functional biology not possible in traditional model organisms like *Mus*.

ii. Innovation

I propose to study extreme physiologic water conservation using a captive colony of *Peromyscus eremicus* rodents native to the desert Southwest. These rodent are housed in a specially designed walk-in desert chamber. This chamber will replicate the intense heat and aridity of the natural environment, while preserving the ability to manipulate the relevant variables (temperature, humidity, water availability) to meet experimental needs. For animals exposed to various experimental treatments (Figure 1, n=10 per treatment), I will collect and analyze physiologic data relevant to hydration status (e.g. serum electrolyte levels), and kidney function (e.g. BUN, Creatinine). In addition to this, because water requirements are related to metabolic rates I will collect data related to metabolism using a chamber designed to function in extreme desert conditions. Lastly, though most individuals are functionally anuric, urine will be collected when available and analyzed for specific gravity and osmolality.

In addition to the thorough characterization of physiology related to desert survival, I will characterize the genomic changes that underlie changes in physiology. To do this, I will perform carefully controlled Bisulfite and RNA sequencing experiments using renal medullary tissue to assay the genomic changes accompanying changes in physiology. These data represent a uniquely

rich characterization of both the phenotype (=physiology) and genomics surrounding desert life, and will significantly enhance our understanding of desert survival.

ii. Background

The study of adaptation, or the process through which animals become fitted to their environment has intrigued researchers for decades (Darwin, 1859; Fisher, 1930), though only recently have we had the ability to study the underlying genomic mechanisms. Interestingly, researchers interested in understanding the genetics of adaptation have the ability to ground modern studies of genetics on decades of work aimed at understanding the ecological context within which adaptation occurs. One particularly salient example of the connection between studies of ecology and natural history and modern genomics can be found in the study of physiologic adaptation to desert conditions. Here, remarkable physiologic, morphologic (Dickinson et al., 2007; Huntley et al., 1984; Schmidt-Nielsen and Schmidt-Nielsen, 1950; Schmidt-Nielsen, 1952) and behavioral (Nagy, 1994) adaptation has been studied in the context of desert ecology. These studies provide a rich context for the current work, which aims to define the relationship between physiology and genomics in rodents able to thrive in amongst the most harsh of conditions on Earth.

Though classic research relating morphology (especially renal ultrastructure) and physiology to desert adaptation has been done, we know virtually nothing about how extreme heat and aridity may affect other core physiological and metabolic processes. For instance, the maintenance of normal serum electrolyte concentration is challenging in the context of desert conditions. Indeed, electrolyte derangement is often the ultimate cause dehydration-related death, yet for the vast majority of desert animals, we know nothing of electrolyte balance - not even typical values! Understanding these critical processes is fundamental to our understanding of physiological adaptation, and will be accomplished as part of this project.

Lastly, the genomic processes related to desert survival have yet to be characterized. The few studies of genetics that have been done have focused on the role of single members of the Aquaporin gene family (but see Bartolo and Donald (2007)), which are large membrane-bound proteins that are critically involved in renal water transport (Kwon et al., 2009; Verkman, 2002; Brown et al., 1995; Nielsen and Agre, 1995). These studies have shown that changes in Aquaporin (AQP) protein abundance and expression may be related to water availability (Böselt et al., 2009; Gallardo et al., 2005; Bozinovic et al., 2003). In addition to changes in expression, another study showed that the AQP4 pathway was completely lost in the desert rodent *Dipodomys merriami merriami* (Huang et al., 2001). Despite these studies, we have a limited understanding of the genomics of renal water and solute regulation in desert animals. While AQPs are functionally important, water and solute balance is extraordinarily complex, and therefore single-gene studies are necessarily limited in their purview. A more complete understanding of this phenotype and its mechanistic underpinnings will require a sophisticated genome-level approach, which will be the outcome of the proposed research.

iii. Experimental Plan and Preliminary Results

Specific Aim 1: To better understand the physiological effects of desert conditions on rodents, I will relate multiple physiological variables to differences in temperature, relative humidity, and water availability. These experiments (and those described under Aim 2) are fundamentally linked to a series of environmental manipulations, described in Figure 1. I hypothesize that, as a result of unique mechanisms related to solute and water balance, that average electrolyte concentrations will remain relatively constant throughout various experimental manipulations, but the variance in measured levels between individuals will increase in the most extreme conditions.

From all animals in experimental treatment groups ($n=10 \times 6 \text{ treatments}=60 \text{ animals}$), I will collect a urine sample, and measure specific gravity and urine osmolality using an Atago UG- α urine refractometer. Serum electrolytes including potassium, sodium, BUN, creatinine, calcium and bicarbonate ion concentration will be measured. Metabolic parameters including carbon dioxide production and oxygen consumption will be measured during a twelve hour period at the end of the experimental manipulation, just prior to euthanasia, using a standard metabolic chamber (Sable Inc.) modified for use in the desert chamber.

Preliminary Data: To date, I have generated a physiology dataset that consists of serum electrolytes and urine osmolality from 5 animals housed in the 'cold/simulated rain' treatment group. This small dataset allowed me to refine and perfect the protocols for data collection and analysis, and has already suggested interesting physiology. In the small subset of animals housed in the cool, wet, simulated rain conditions for a period of 3 weeks, all had serum Potassium levels >8.0 , Creatinine <0.3 , and Sodium >151 , which is significantly different than 'normal' levels for other mammals.

Specific Aim 2: I will understand the genetic response to extreme heat and aridity via a series of bisulfite sequencing and RNAseq experiments, and will link these patterns to individual physiologic state as defined in Aim 1. Differences in renal methylation patterns and mRNA gene expression will be tested using the 10 individuals per treatment group ($n=60$ total, specified above). I hypothesize that genes responsible for water and solute transport will be particularly active in the most extreme conditions.

For analysis of the bisulfite sequence data, an accurate genome assembly is required. Using the existing draft genome (sequenced using startup funds) as a starting point, I will complete the genome assembly of the primary study organism, *Peromyscus eremicus*. This process will be significantly aided by leveraging the existing genome sequence data available from several other *Peromyscus* species. Specifically, an additional 60X coverage Illumina dataset using 150bp paired end and mate-pair sequencing will supplement the current 40X dataset. Simultaneous with this, I will sequence up to 10X coverage using long-read PacBio sequencing. A short-read assembly

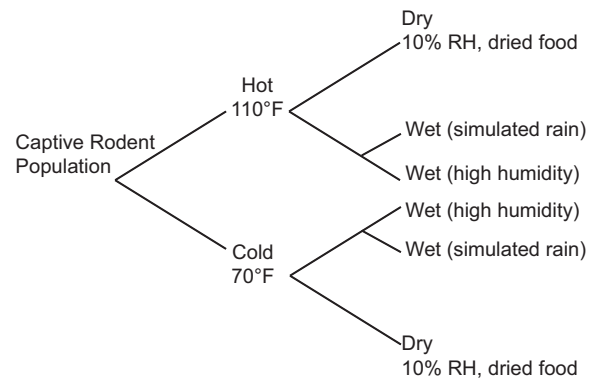


Figure 1: Animals are relegated into either hot or cold treatments. Within treatments ($n=10$ per treatment), animals are exposed to two weeks of varying levels of aridity, from simulated rainfall where water is available *ad libitum*, to dry, where no water is available. RH=relative Humidity

of Illumina data will be done using AllPaths-LG (Maccallum et al., 2009). Illumina contigs will be assembled with the corrected PacBio data using the software package SGA (Simpson and Durbin, 2012) or an overlap-layout-consensus assembler (e.g. Celera, Miller et al. (2008)) to generate a primary *de novo* genome assembly. Assemblies will be validated by use of the CEGMA (Parra et al., 2007) and REAPR (Hunt et al., 2013) algorithms. In general, this approach has been shown to be successful in producing high quality draft assemblies, though multiple assembly methods will be evaluated as per the findings of my previous work (Bradnam et al., 2013).

To annotate the genome, RNA and smRNA sequence data from several tissues, including liver, kidney, and brain will be generated. Sequence data will consist of Illumina strand-specific small-insert 150bp paired end sequencing. Each tissue type will be sequenced to a depth of 100 million reads. Assembly will be accomplished using the Trinity software package (Haas et al., 2013; Grabherr et al., 2011) after error correction (MacManes and Eisen, 2013) and quality trimming (MacManes, 2013). Annotation will be completed using the programs Trinotate (<http://trinotate.sourceforge.net/>) and Maker (Cantarel et al., 2008).

In addition to the assembly and annotation of the *P. eremicus* genome, a secondary result of this work is methods development (**enhancing infrastructure**). To this end, I have already already released a transcriptome assembly pipeline (<http://sourceforge.net/projects/tamrs/>) and automated quality control software (<http://sourceforge.net/projects/qcpro/>). In addition to this, I am an active developer of the transcriptome assembly program Trinity and annotation software Trinotate. Given the popularity of high-throughput sequencing, the demand for these types of tool development programs will likely increase.

Preliminary Data: To date, I have generated a RNAseq dataset that consists of approximately 30M 150nt SE Illumina reads from the same 5 animals housed in the 'cold/simulated rain' treatment group from which I collected physiology data.

Specific Aim 3: Given that desert adapted mice, capable of surviving without water are as neonates dependent on liquid intake, the study of the ontogeny of physiologic water conservation is extremely interesting and relevant to the current work. This phenomenon will be explored using neonate mice in the treatments listed in Aim 1, and methods described in Aim 2. Five neonate mice will be culled per treatment at 3 different timepoints (immediately after birth, mid-lactation, 1 day after weaning). I hypothesize that patterns of gene expression and methylation will resemble those common in conditions where water is available *ab lib*.

iv. Broad Impacts

How desert animals survive without water is an intrinsically interesting example of extreme physiology and adaptation. Indeed, telling people of the remarkable story of a rodent that may never drink water is immediately captivating. Having told hundreds of people, including children, adults, scientifically literate and not, the most common response is something like "Wow! That is amazing. How do they do that?" I intend to **broaden dissemination** by capitalizing on this common response, using traditional and non-traditional methods. While lecture has been the mainstay of scientific communication, its audience tends towards the educated/affluent subset of the population. Given one of my primary goals as a scientists is to broadly share my work, I will use non-traditional modes of communication which will include social media and a blog. All publications are hosted on a public preprint server, and published using open access options. All data and code is freely available using public repositories like Github, Dryad, or Figshare. In addition to this, I will develop an

outreach program that aims to present research findings to school-aged children. These program will be multi-dimensional, with programs geared towards several different age groups.

Broadening participation is an issue about which I care deeply. As a scientist of Native American descent, I strive to be a role model and mentor for underrepresented groups. To this end, I am developing an internship program- in collaboration with the American Indian Higher Education Consortium (<http://aihec.org/>), where qualified Native students may spend a summer working in my lab either in directed or independent research programs. It is my hope that students introduced to my lab via internship may ultimately elect to join the lab. This method of recruitment supplements my general interest in recruiting students from underrepresented groups.

Lastly, my efforts aimed at broadening dissemination and participation are all within the context of a research program with substantial **benefits to society**. First, Earth is becoming both warmer and drier, and accurate prediction of animals' response to a changing climate requires knowledge of how animals currently living in dry/hot climates survive - this project will deepen the requisite knowledge. Next, kidney disease effects millions of Americans (http://www.cdc.gov/diabetes/projects/pdfs/ckd_factsheet.pdf). Though the causes are diverse, the pathophysiology of many resemble dehydration (e.g. low renal perfusion pressure). Thus, having a better understanding of how desert rodents endure water stress seemingly without complication may enhance our ability to explain why humans and most other mammals cannot.

v. Summary & Future Directions

The maintenance of water balance in animals is one of the most important physiologic processes, and is critical to survival. Osmoregulation in animals living in desert environments is particularly challenging, as extreme heat and aridity is common. Gaining a deeper understanding of the physiological adaptations that allow for desert survival is important and has obvious implications for climate change science, conservation, and human health and medicine. The proposed research aims to generate a uniquely rich dataset, leveraging cutting edge genomic techniques against careful characterization of physiology, all within an ecological context of desert life. A carefully constructed plan for broadening dissemination and participation ensures that the scientific process as well as its results will be available to all interested parties.

This proposal represents the foundational steps toward developing *P. eremicus* as a model system for the study of desert adaptation. Indeed, this model offers the scientific community a unique opportunity to gain a deep understanding into the physiology and genomics of desert adaptation- an insight which is impossible to achieve using traditional model system. While not a part of this proposal, this work lays the groundwork for future studies where causal links are made between phenotype and genotype, using technologies like CRISPR-Cas9 and RNAi.

References

- Bartolo, R.C., Donald, J.A., 2007. The distribution of renal hyaluronan and the expression of hyaluronan synthases during water deprivation in the Spinifex hopping mouse, *Notomys alexis*. Comparative biochemistry and physiology Part A, Molecular & integrative physiology 148, 853-860.
- Böselt, I., Römler, H., Hermsdorf, T., Thor, D., Busch, W., Schulz, A., Schöneberg, T., 2009. Involvement of the V2 vasopressin receptor in adaptation to limited water supply. PLOS ONE 4, e5573.
- Bozinovic, F., Gallardo, P.A., Visser, G.H., Cortés, A., 2003. Seasonal acclimatization in water flux rate, urine osmolality and kidney water channels in free-living degus: molecular mechanisms, physiological processes and ecological implications. Journal of Experimental Biology 206, 2959-2966.
- Bradnam, K.R., Fass, J.N., Alexandrov, A., Baranay, P., Bechner, M., Birol, I., Boisvert, S., Chapman, J.A., Chapuis, G., Chikhi, R., Chitsaz, H., Chou, W.C., Corbeil, J., Del Fabbro, C., Docking, T.R., Durbin, R., Earl, D., Emrich, S., Fedotov, P., Fonseca, N.A., Ganapathy, G., Gibbs, R.A., Gnerre, S., Godzaridis, E., Goldstein, S., Haimel, M., Hall, G., Haussler, D., Hiatt, J.B., Ho, I.Y., Howard, J., Hunt, M., Jackman, S.D., Jaffe, D.B., Jarvis, E., Jiang, H., Kazakov, S., Kersey, P.J., Kitzman, J.O., Knight, J.R., Koren, S., Lam, T.W., Lavenier, D., Laviolette, F., Li, Y., Li, Z., Liu, B., Liu, Y., Luo, R., Maccallum, I., **MacManes**, M.D., Maillet, N., Melnikov, S., Naquin, D., Ning, Z., Otto, T.D., Paten, B., Paulo, O.S., Phillippy, A.M., Pina-Martins, F., Place, M., Przybylski, D., Qin, X., Qu, C., Ribeiro, F.J., Richards, S., Rokhsar, D.S., Ruby, J.G., Scalabrin, S., Schatz, M.C., Schwartz, D.C., Sergushichev, A., Sharpe, T., Shaw, T.I., Shendure, J., Shi, Y., Simpson, J.T., Song, H., Tsarev, F., Vezzi, F., Vicedomini, R., Vieira, B.M., Wang, J., Worley, K.C., Yin, S., Yiu, S.M., Yuan, J., Zhang, G., Zhang, H., Zhou, S., Korf, I.F., 2013. Assemblathon 2: evaluating *de novo* methods of genome assembly in three vertebrate species. GigaScience 2, 10.
- Brown, D., Katsura, T., Kawashima, M., Verkman, A.S., Sabolic, I., 1995. Cellular distribution of the aquaporins: A family of water channel proteins. Histochemistry and Cell Biology 104, 1--9.
- Cantarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B., Holt, C., Sánchez Alvarado, A., Yandell, M., 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Research 18, 188--196.
- Darwin, C., 1859. On the Origin of Species by Means of Natural Selection or the Preservation of Favored Races in the Struggle for Life. 1st ed., J. Murray, London.
- Dickinson, H., Moritz, K., Wintour, E.M., Walker, D.W., Kett, M.M., 2007. A comparative study of renal function in the desert-adapted spiny mouse and the laboratory-adapted C57BL/6 mouse: response to dietary salt load. American Journal of Physiology-Renal Physiology 293, F1093--8.
- Fisher, R., 1930. The Genetical Theory of Natural Selection. Oxford University, London.
- Gallardo, P.A., Cortés, A., Bozinovic, F., 2005. Phenotypic flexibility at the molecular and organismal level allows desert-dwelling rodents to cope with seasonal water availability. Physiological and Biochemical Zoology 78, 145--152.

- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, a., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29, 644--652.
- Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., **MacManes**, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc, R.D., Friedman, N., Regev, A., 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols* 8, 1494--1512.
- Huang, Y., Tracy, R., Walsberg, G.E., Makkinje, A., Fang, P., Brown, D., Van Hoek, A.N., 2001. Absence of aquaporin-4 water channels from kidneys of the desert rodent *Dipodomys merriami merriami*. *American Journal of Physiology-Renal Physiology* 280, F794--F802.
- Hunt, M., Kikuchi, T., Sanders, M., Newbold, C., Berriman, M., Otto, T.D., 2013. REAPR: a universal tool for genome assembly evaluation. *Genome Biology* 14, R47.
- Huntley, A.C., Costa, D.P., Rubin, R.D., 1984. The contribution of nasal countercurrent heat exchange to water balance in the northern elephant seal, *Mirounga angustirostris*. *The Journal of experimental biology* 113, 447--454.
- Kwon, T.H., Nielsen, J., Nielsen, J., Møller, H.B., Møller, H.B., Fenton, R.A., Fenton, R.A., Nielsen, S., Frøkiaer, J., 2009. Aquaporins in the Kidney, in: *Aquaporins*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 95--132.
- Maccallum, I., Przybylski, D., Gnerre, S., Burton, J., Shlyakhter, I., Gnirke, a., Malek, J., McKernan, K., Ranade, S., Shea, T.P., Williams, L., Young, S., Nusbaum, C., Jaffe, D.B., 2009. ALLPATHS 2: small genomes assembled accurately and with high continuity from short paired reads. *Genome Biology* 10, R103.
- MacManes, M.D., 2013. On the optimal trimming of high-throughput mRNA sequence data. Technical Report.
- MacManes, M.D., Eisen, M.B., 2013. Improving transcriptome assembly through error correction of high-throughput sequence reads. *PeerJ* 1, e113.
- Miller, J.R., Delcher, A.L., Koren, S., Venter, E., Walenz, B.P., Brownley, A., Johnson, J., Li, K., Mobarry, C., Sutton, G., 2008. Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics (Oxford, England)* 24, 2818--2824.
- Nagy, K.A., 1994. Seasonal Water, Energy and Food Use by Free-Living, Arid-Habitat Mammals. *Australian Journal of Zoology* 42, 55--63.
- Nielsen, S., Agre, P., 1995. The aquaporin family of water channels in kidney. *Kidney international* .
- Parra, G., Bradnam, K., Korf, I., 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics (Oxford, England)* 23, 1061--1067.
- Schmidt-Nielsen, B., 1952. Renal Tubular Excretion of Urea in Kangaroo Rats. *American Journal of Physiology* 170, 45--56.

- Schmidt-Nielsen, B., Schmidt-Nielsen, K., 1950. Pulmonary Water Loss in Desert Rodents. American Journal of Physiology 162, 31--36.
- Simpson, J.T., Durbin, R., 2012. Efficient *de novo* assembly of large genomes using compressed data structures. Genome Research 22, 549--556.
- Verkman, A.S., 2002. Physiological importance of aquaporin water channels. Annals of medicine 34, 192--200.