The maintenance of water balance in animals is one of the most important physiologic processes, and is critical for survival. Indeed, humans 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. In contrast, 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. Despite being a well-known ecological phenomenon with obvious implications for human health, we know very little of the underlying mechanisms that allow for survival. **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 non-survivable conditions.** This proposal represents the foundational steps toward developing Peromyscus eremicus as a model system for the study of physiologic water conservation. Indeed, this model offers the scientific community a unique opportunity to gain a deep understanding into the physiology and genomics of osmoregulation in extreme environments – a critically important insight that is impossible using traditional model system like *Mus*. While not a part of this proposal, this project lays the groundwork for *my long-term research goal* – to identify the causal links between phenotype and genotype, using emerging technologies like the CRISPR-Cas9 system.

SPECIFIC AIM 1: Using captive desert-adapted mice, I will link multiple physiological variables, as well as their genomic underpinnings to differences in temperature, relative humidity, and water availability.

> My working hypothesis is that while mice may demonstrate physiological signs of dehydration (e.g. elevated Na), patterns of renal gene expression, isoform use, and methylation will be distinct from those typical of illness.

SPECIFIC AIM 1: Given the transition from the obligate intake of fluids as infants, to it's complete absence later in life, the ontogeny of physiologic water conservation will be elucidated.

> I hypothesize that patterns of renal gene expression during fetal development through weaning will resemble patterns of gene expression, isoform use, and methylation typical of adult mice when water is freely available.

The proposed project aims to integrate studies of physiology, genomics, and computational biology to gain a deep understanding of a fundamental physiological problem – how to conserve water when intake is limited. *Although dehydration is both common and dangerous, a large swath of the biology underlying its physiological effects is currently invisible to researchers using traditional mammalian models of disease that lack the eco-evolutionary history present in desert-adapted mice*. This project will fill a critically important gap in our understanding, which is in support of the specific research aims of the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK).

i. Significance

Dehydration, whether caused by exposure to extreme environmental conditions, water deprivation, or by infection (e.g. diarrheal illnesses) represents a significant threat to human life. When death is avoided, dehydration may lead to chronic health conditions like renal failure. While the mechanisms underlying physiological compromise are well characterized [ref], some animals possess the ability, much unlike humans, to osmoregulate despite extreme heat and a complete lack of extrinsic water intake [ref]. Specifically, highly adapted desert mice may never drink water [ref], produce an extremely viscous urine (or no urine at all) [ref], and excrete urea in the form of uric acid crystals in the feces [ref]. Focusing on osmoregulation, patterns of renal gene expression have been shown to be highly derived in some desert adapted rodents (e.g. *Dipodomys* [ref]), but not in others like (*Notomys*), and therefore the extent to which differences in gene expression underlies phenotype remains unknown. In addition to its' generality, both studies of non-model organisms focus on a limited number of genes, and therefore may not appropriately assay the complexity of the genetics of extreme renal osmoregulation. In contrast with this work, several studies attempting to understand the effects of dehydration in model organisms (e.g. *Mus* and *Rattus*) exist, though while these works benefit from an extensive genomic and physiologic toolbox, that they lack the appropriate phenotype (extreme osmoregulatory abilities) limits insight. In summary, the study of extreme renal osmoregulation is hindered on one hand by a lack of tools, and on the other by a lack of an appropriate phenotype. This limitation may undercut our ability to efficiently develop novel therapies directed at mitigating the untoward effects of 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 non-survivable conditions. Pursuant to this goal, I have developed significant genomic resources in a novel mammalian model, *Peromyscus eremicus* and will apply these tools the understanding of extreme osmoregulation. The contribution is to effectively leverage the power of a sophisticated genomic and physiologic toolset of a model organism against a uniquely adapted rodent, which will allow for a synthetic understanding of extreme osmoregulation, and ultimately novel insights into the cause of - and cure for - dehydration related mortality and morbidity.

ii. Innovation

In spite of modern medicine, millions of people die from dehydration. Millions more are affected by chronic kidney disease, a condition whose underlying causes are diverse, though many resemble the pathophysiology of dehydration (e.g. low renal perfusion pressure). The proposed work recognizes that successful treatment requires an appropriate model, and while traditional models are wonderful, they lack the biology (extreme osmoregulation) upon which more successful interventions may be modeled. The desert-adapted rodent *P. eremicus* retains many of the beneficial characteristics of model organisms, while enhancing opportunity to assay interesting biological phenomenon. In addition to this fundamental innovation, the project it innovative in a number of other ways.

- Leverage an unprecedented level of control over the experimental environment by use of a desert chamber where natural conditions can be replicated while simultaneously preserving the ability to manipulate water availability.
- Link detailed information on physiology and metabolism to genomic information using a novel analytical tools under active development in the lab.

Aim 1: Using captive desert-adapted mice, I will correlate multiple physiological variables, as well as their genomic underpinnings to differences in temperature, relative humidity, and water availability. I hypothesize that, as a result of unique mechanisms related to solute and water balance, that average serum 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. These differences will be echoed in differences in urine electrolytes and concentration.

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 are fundamentally linked to a series of environmental manipulations, described in Figure 1. The experimental design is fully factorial -- meaning that the focal experimental parameter (e.g. temperature) will be tested in the context of the full range of other conditions (e.g. humidity, water availability). Animal care is standardized between experiments and includes measures to reduce the water content of food and bedding materials. Both of these will be dried in a standard desiccation oven to less than 1% water/volume. Twenty individuals per treatment will be included -- power analyses suggest this sample size will allow me to garner statistical support for patterns with small to medium effect sizes. Together, this design will allow me to tease-apart the physiologic and genomic response to the various conditions.

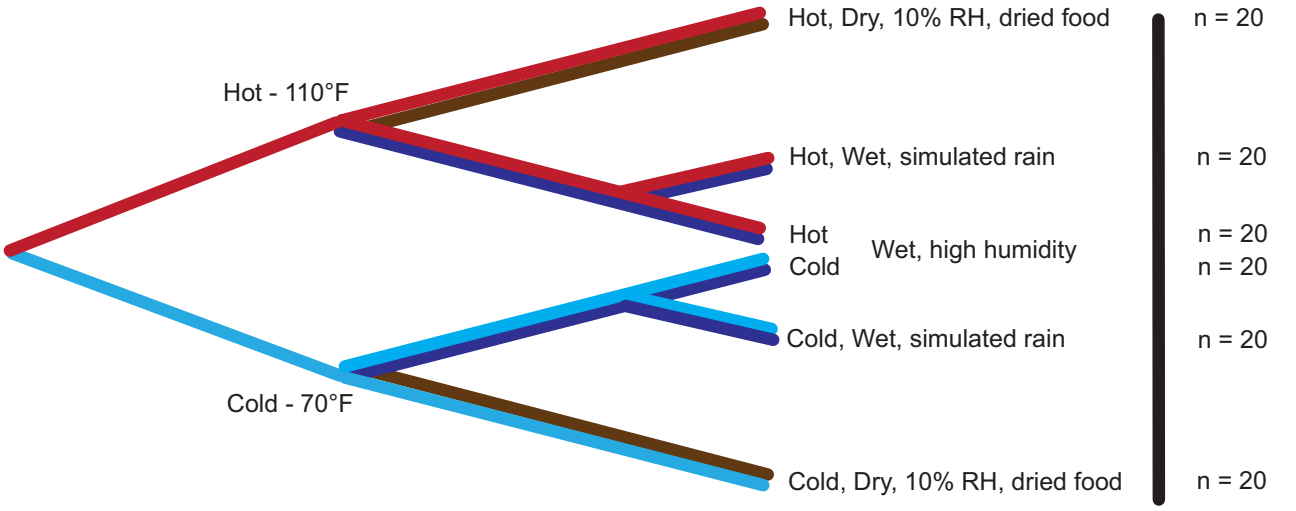

Figure 1: Animals are relegated into either hot or cold treatments. Within treatments (n=20 per treatment), animals are exposed to two weeks of varying levels of aridity, from simulated rainfall where water is available *ad libitum*, to dry, where no water is available. RH=relative Humidity

For each experiment in Aim 1, I will collect information including metabolism, urine concentration and electrolyte content, serum electrolyte content, as well as more basic measures like body temperature and body weight. In the context of limited water intake, how animals achieve electrolyte balance is unknown. Electrolytes are both easy to assay and are critical to physiological well being. Indeed, proper electrolyte balance is fundamental to all other physiological processes like neuronal signal transduction and muscle (including cardiac) contractility. Here, electrolytes will be measured using the VetScan VS2 critical care panel which includes ALT, BUN, Cl, CRE, GLU, K, Na, bicarbonate ion in a 100uL sample volume.

In addition to assaying electrolytes themselves, I will conduct measures of urine electrolytes and specific gravity, as the urinary system represents that major pathway through which these chemicals are lost. These parameters will be measured using an Atago UG-α urine refractometer and tests conducted at the IDEXX reference lab. Lastly, I will weigh animals to the nearest 0.1gm every other day, including the day of sacrifice. Body temperature will be assayed with weighing using a digital thermometer and probe designed by World Precision Instruments (Sarasota, FL). In connection with this, feces will be collected and water content will be assayed using standard methods.

I will collect key metabolic parameters such as carbon dioxide production and oxygen consumption that may influence water consumption. These tests will be measured during a twenty four hour period at the end of the experimental manipulation, just prior to euthanasia, using a metabolic chamber (Sable Inc.) modified for use in the desert chamber. Together, these data will represent a uniquely rich characterization of the physiological state of a desert rodent held in captivity but importantly, exposed to conditions typical of the natural environment.

Aim 1a-1: Determine the physiologic response to water deprivation in the desert-adapted rodent *P. eremicus*.

Background: Water availability is thought to be one of the key attributes that determines the hostility of a given environment. For instance, ecosystems like deserts present unique challenges to the animals that inhabit them -- the ways in which animals overcome these challenges may allow researchers unique insights with biomedical importance. Importantly, while we have an understanding of how dehydration looks in the context of illness in non-desert-adapted animals, we have no idea what this looks like in health desert-adapted animals.

Research Plan:To accomplish this aim, I will analyze physiological data from animals help with and without water, factorial with respect to the other conditions. The specific experiments and the types of data collected are described above. Patterns of electrolyte levels and urine concentration that are related to this contrast (wet versus dry) will allow my to understand the how water deprivation may be related to water.

Outcome: These data will be compiled into a large data matrix, with statistical analyses focusing on the identifying patterns related to differences in water availability. In addition, because behavior related to desert survival will be characterized as part of a separate project, changes in electrolytes will be linked to individual level differences in behavior. Ultimately, these physiology data will be combined with the genomic data to understand the relationship between physiology and genomics.

Preliminary data: I have characterized the electrolyte profile of 2 individuals housed at 70F, 50% RH, water *ad lib* and two individuals housed in identical conditions except that drinking water was withheld. Despite being housed in typical laboratory conditions, these animals have remarkably unusual electrolyte panel. For instance, mean serum potassium in an un-hemolyzed sample is unusually high at 8.1mg/dL, while Creatinine is low, with a mean measurement of 0.25mg/dL. mean blood urea nitrogen (BUN) is 47mg/dL. In contrast, animals without *ad lib* water were essentially...

Aim 1a-2: Determine the physiologic response to temperature in the desert-adapted rodent *P. eremicus*.

Background: While water stress is obviously important to the survival of desert rodents - a phenotype which is relevant to human health and wellness, extreme temperatures represent another way in which physiological processes may be challenged. While desert animals may thrive in extreme heat, humans cannot. Understanding the physiological response is characterized in humans, but not in other animals evolved in these conditions. Genes like the heat-shock proteins are protective in humans, but no record of their activity on desert rodents is known.

Research Plan:

Outcome:

Aim 1a-3: Determine the physiologic response to relative humidity in the desert-adapted rodent *P. eremicus*.

- Describe role of serum electrolytes.
- How are these things regulated in the kidney (link to genes)
- typical dehydration panel
- What I'm going to measure
- expected outcome

Aim 1b: Define patterns of gene expression | isoform use | methylation given differences environmental condition. 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. I hypothesize that genes responsible for water and solute transport will be particularly active in the most extreme conditions.

Background: 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.

Research Plan: 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 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).

To identify patterns of differential isoform use, I will generate normalized cDNA libraries using X, and sequence using the single molecule Pacific Biosystems technology. This approach offers significant advantage over short read high-throughput sequencing technologies (e.g. Illumina, Ion Torrent) where accurate assembly of isoforms is extremely problematic (**?**).

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.

Outcome: This work will allow me to look deeply into the genomic processes that underlie adaptation.

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

- Describe typical renal gene expression profile
- link to electrolytes
- what I'm going to measure and how

- BiSeq, RNAseq, PacBio transcriptome to get at isoforms
- expected outcome

Aim 1c: Correlate physiology (=phenotype) with genomic patterns.

- Diff gene expression
- coexpression networks
- Patterns of isoform use
- diff methylation

Aim 2: Given the transition from the obligate intake of fluids as infants, to it's complete absence later in life, the ontogeny of physiologic water conservation will be elucidated.

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*.

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., Römpler, 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.