Comparative Biochemistry and Physiology, Part D 7 (2012) 328–339

Contents lists available at SciVerse ScienceDirect

Comparative Biochemistry and Physiology, Part D

journal homepage: www.elsevier.com/locate/cbpd

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 kangaroo rat)

Nicholas J. Marra [a,*], Soo Hyung Eo [a,1], Matthew C. Hale [b], Peter M. Waser [b], J. Andrew DeWoody [a,b]

[a] Department of Forestry & Natural Resources, Purdue University, West Lafayette, IN 47907, USA
[b] Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA

ARTICLE INFO

Article history:
Received 26 March 2012
Received in revised form 30 June 2012
Accepted 2 July 2012
Available online 13 July 2012

Keywords:
Digital gene expression
Molecular evolution
Water balance
454
Next-generation sequencing
RNA-seq

ABSTRACT

One common goal in evolutionary biology is the identification of genes underlying adaptive traits of evolutionary interest. Recently next-generation sequencing techniques have greatly facilitated such evolutionary studies in species otherwise depauperate of genomic resources. Kangaroo rats (*Dipodomys sp.*) serve as exemplars of adaptation in that they inhabit extremely arid environments, yet require no drinking water because of ultra-efficient kidney function and osmoregulation. As a basis for identifying water conservation genes in kangaroo rats, we conducted *a priori* bioinformatics searches in model rodents (*Mus musculus* and *Rattus norvegicus*) to identify candidate genes with known or suspected osmoregulatory function. We then obtained 446,758 reads via 454 pyrosequencing to characterize genes expressed in the kidney of banner-tailed kangaroo rats (*Dipodomys spectabilis*). We also determined candidates *a posteriori* by identifying genes that were overexpressed in the kidney. The kangaroo rat sequences revealed nine different *a priori* candidate genes predicted from our *Mus* and *Rattus* searches, as well as 32 *a posteriori* candidate genes that were overexpressed in kidney. Mutations in two of these genes, *Slc12a1* and *Slc12a3*, cause human renal diseases that result in the inability to concentrate urine. These genes are likely key determinants of physiological water conservation in desert rodents.

1. Introduction

Evolutionary adaptations are myriad, and all have a genetic basis. This genetic basis can be manifested through changes in gene sequence or gene expression that contribute to adaptive phenotypes (Orr and Coyne, 1992; Feder and Mitchell-Olds, 2003). For instance, multiple studies have identified loci associated with the loss of lateral plates and pelvic armor in freshwater threespine stickleback (*Gasterosteus aculeatus*) populations (Cresko et al., 2004; Shapiro et al., 2004). Another example is the role of various melanocortin-1-receptor (*Mc1r*) substitutions in the expression of adaptive color morphs of pocket mice *Chaetodipus intermedius* (Nachman et al., 2003; Nachman, 2005). The search for cognate genes underlying such traits is a major focus of modern evolutionary genetics. In model species such as *Arabidopsis thaliana* and *Drosophila melanogaster*, these searches are usually aided by entire genome sequences and comprehensive genetic maps that aid in population genomic and/or quantitative genetic approaches (Feder and Mitchell-Olds, 2003). In non-model species, the search for such genes proves more difficult (Stinchcombe and Hoekstra, 2007). Fortunately, advances in DNA sequencing provide an avenue by which biologists who study non-model species can begin to identify adaptive genes underlying traits that have been targets of strong natural selection (Hudson, 2008; Wheat, 2010; Ekblom and Galindo, 2011).

One adaptive trait presumably under strong selection is water conservation in desert animals. Water is scarce in deserts and many organisms are forced to satisfy their water requirements by feeding on succulent foods or tolerating drastic dehydration coupled with an increased drinking capacity (Schmidt-Nielsen and Schmidt-Nielsen, 1952; Schmidt-Nielsen, 1964). The maintenance of homeostasis creates numerous avenues for water to be lost, including evaporative water loss during thermal regulation, respiratory water loss, and loss during waste excretion (Schmidt-Nielsen, 1964; Cain III et al., 2006). These sources of water loss impose strong selective pressures on osmoregulation, and both behavioral and physiological adaptations for water conservation have evolved accordingly.

Behavioral adaptations help organisms avoid situations where water loss occurs as a result of thermal regulation or through respiration. Physiological adaptations can further limit water loss. One physiological adaptation for water conservation is to increase water retention during

* Corresponding author. Tel.: +1 201 819 8237.
E-mail addresses: nmarra@purdue.edu (N.J. Marra), eo3@wisc.edu (S.H. Eo), mchale@purdue.edu (M.C. Hale), waserpm@purdue.edu (P.M. Waser), dewoody@purdue.edu (J.A. DeWoody).
[1] Present Address: Department of Zoology, University of Wisconsin, Madison, WI 53706, USA.

1744-117X/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.cbd.2012.07.001

waste production (Schmidt-Nielsen B. et al., 1948; Schmidt-Nielsen and Schmidt-Nielsen, 1952). There is ample evidence to indicate that various desert animals have enhanced kidney efficiency that leads to the production of concentrated urine (Schmidt-Nielsen and Schmidt-Nielsen, 1952; Schmidt-Nielsen, 1964). Enhanced kidney efficiency is realized by the higher relative medullary thickness seen in many mammals from arid and marine environments (Schmidt-Nielsen and O'Dell, 1961; Beuchat, 1996; Al-kahtani et al., 2004), including those in the Heteromyidae. This family of New World rodents largely inhabits arid and semi-arid environments in southwestern North America and includes the kangaroo rats (genus *Dipodomys*) (Eisenberg, 1963; Alexander and Riddle, 2005).

The banner-tailed kangaroo rat (*Dipodomys spectabilis*) has long been noted for its ability to survive in xeric habitats with little to no drinking water (Vorhies and Taylor, 1922; Holdenried, 1957; Schmidt-Nielsen, 1948; Schmidt-Nielsen, 1964). Its diet is dominated by seeds (Vorhies and Taylor, 1922; Holdenried, 1957; Best, 1988) and this species seldom feeds on highly succulent food, so it must efficiently conserve the water it obtains from its diet and the environment. Behaviorally, water conservation is achieved by 1) fossorial denning in large burrows with elevated humidity (Schmidt-Nielsen and Schmidt-Nielsen, 1950, 1952; Holdenried, 1957) and 2) nocturnal activity that limits water loss by reducing exposure to high temperatures that would necessitate excessive thermal regulation (Holdenried, 1957). However, *D. spectabilis* is most famous for its ability to retain water during waste production. In particular, banner-tailed kangaroo rats produce extremely concentrated urine (Schmidt-Nielsen B. et al., 1948; Schmidt-Nielsen et al., 1948a, 1948b) by means of elongated loops of Henle (Howell and Gersh, 1935; Schmidt-Nielsen, 1952; Vimtrup and Schmidt-Nielsen, 1952). Thus far, the genetic basis of this adaptation is unknown.

There are several genetic mechanisms that could explain the urine-concentrating ability of kangaroo rats. For instance, efficient osmoregulation could result from the evolution of one or a few genes by strong positive and/or purifying selection on protein-coding sequence(s) in kangaroo rats. Alternatively, selection might act on regulatory regions to alter the expression of existing genes common to many species. For example, upregulation of a key gene during urine production could produce an abundance of kidney proteins essential for efficient transport of water from the filtrate back into the blood stream. Finally (and perhaps most likely), urine-concentrating ability may have evolved through selection on protein-coding sequences and on regulatory control of gene expression.

We used massively-parallel sequencing and other approaches to identify candidate osmoregulatory genes. First, we identified a list of gene ontology (GO) terms related to water retention from the AmiGO database and downloaded *a priori* candidate *Mus musculus* and *Rattus norvegicus* genes annotated with these terms. Next, we used 454 pyrosequencing (Roche) of cDNA from four individual kangaroo rats to identify the major components of the kidney transcriptome and to identify target genes responsible for efficient osmoregulation. Finally, we identified additional candidate genes *a posteriori* that were significantly overexpressed in kangaroo rat kidney relative to a reference tissue (spleen). Candidate genes were then compared to the remaining genes identified in the *de novo* kidney transcriptome assembly regarding the distribution of GO terms, presence of molecular markers, and molecular measures of natural selection.

2. Materials and methods

2.1. Sample collection, RNA extraction and cDNA synthesis

Four adult *D. spectabilis* individuals were collected from our long-term field site near Portal, Arizona in December, 2009 (see Busch et al., 2009 for GPS coordinates). These individuals were trapped with Sherman live traps (Jones, 1984), and included two males (♂0828 and ♂0812) and two females (♀0862 and ♀0871). Each individual was euthanized according to IACUC approved protocols and dissected to remove target (kidney) and reference (spleen) tissue. We used a reference tissue to identify distinctive patterns of gene expression in *D. spectabilis* kidney; we chose spleen in an attempt to characterize major histocompatibility complex genes as part of a separate study (Marra et al., in prep). Each tissue was immediately minced and part of the sample was frozen in liquid nitrogen while another section was placed in TRIzol® reagent (Invitrogen) and frozen on dry ice for transport back to Purdue University.

Separate RNA extractions were conducted for kidney and spleen tissue from each individual; thus there were a total of 8 "libraries". We chose kidney to capture candidate genes involved in osmoregulation, and spleen as a control tissue. We used TRIzol® reagent (Invitrogen) for RNA extractions according to the manufacturer's instructions. RNA quality and quantity was assessed via gel electrophoresis and spectrophotometry (Nanodrop 8000; Thermo Scientific). These tests revealed possible genomic DNA contamination in the four spleen RNA extracts, so we treated them with DNaseI (NEB) prior to cDNA synthesis. Subsequently, all extracts were used for separate cDNA synthesis using the ClonTech SMART cDNA synthesis kit (Zhu et al., 2001) with a modified CDS III/3' primer (Hale et al., 2009, 2010; Eo et al., 2012).

Second strand cDNA synthesis was conducted using the ClonTech PCR Advantage II polymerase and a thermal profile that included a 1 min denaturation at 95 °C followed by 25, 28, or 30 cycles of 95 °C for 15 s, 68 °C for 6 min; the number of cycles was optimized for the quality and size distribution of each cDNA library. Each cDNA library was digested with *Sfi*I and subsequently purified with a QIAquick PCR purification kit (Qiagen) to remove oligos. The quality and quantity of each library was checked with spectrophotometry (Nanodrop 8000; Thermo Scientific) and gel electrophoresis, revealing that the main size distribution of these double-stranded cDNA libraries was 500–3000 bp.

2.2. 454 sequencing and transcriptome assembly

Each of the eight cDNA libraries was prepared for 454 pyrosequencing according to standard methods (Margulies et al., 2005). Briefly, the cDNA was sheared by nebulization, ligated to adaptor sequences, captured on individual beads, and then subjected to emulsion PCR and flowed into individual wells of a picotiter plate (PTP device) for sequencing. During this preparation, a library specific Multiplex IDentifier (MID) sequence was ligated to the fragments from each library, which allowed us to sort the reads according to their library of origin. The eight cDNA libraries were sequenced on 1/2 of a PTP device using the Roche GS-FLX (454) platform and Titanium chemistry.

The resulting sequence reads were screened to remove poor quality reads and adaptor sequences, and then trimmed to remove any remnant sequence from the cDNA synthesis primers using custom Perl scripts adapted from Meyer et al. (2009). The program PCAP (Huang et al., 2003) was used to assemble the trimmed reads into unique *de novo* assemblies for each cDNA library as well as tissue specific assemblies. All assemblies were conducted using the default settings for PCAP (minimum identity cutoff of 92%, see Huang et al., 2003). Previous studies utilized this program for successful *de novo* assembly of transcriptome data from other non-model species (Hale et al., 2009, 2010; Eo et al., 2012). For tissue specific assemblies, all reads were assembled together from the four kidney or the four spleen libraries. The contigs from each assembly were given preliminary gene descriptions using a BLASTx search (e-value $\leq 1.00\times10^{-6}$ and bit score ≥ 40) against NCBI's non-redundant (nr) database.

2.3. A priori candidate genes identified via model species

We presumed that *D. spectabilis* genes integral to efficient osmoregulation might include sequences previously identified in other

rodent genomes (namely *M. musculus* and *R. norvegicus*), but that such a search might not be straightforward given that these three species diverged from one another over 60 Ma ago (Meredith et al., 2011). Towards this end, we identified 38 candidate genes (Supplementary Table 1; GO terms current as of GO database release on 6/25/2011) by searching the gene ontology database for *M. musculus* and *R. norvegicus* genes annotated with the direct children GO terms of Renal System Process (GO:0003014) using the AmiGO browser (Carbon et al., 2009, version 1.8). See Table 1 for a list of the GO terms used and the number of candidate genes that possessed this annotation.

To determine if the *M. musculus* or *R. norvegicus* candidate genes we identified *a priori* matched any of our expressed sequence tags (ESTs) from kangaroo rats, we downloaded amino acid sequences from the Mouse Genome Informatics (MGI) database for each of the proteins encoded by the genes listed in Supplementary Table 1. We queried these mouse and rat sequences with the contigs from the pooled kangaroo rat kidney assembly using standalone BLAST (Altschul et al., 1990) with the BLASTx program (e-value $\leq 1.00\times10^{-10}$ and bit score ≥ 40), which translates sequences in all six reading frames before comparing to the collection of amino acid sequences. We considered a contig as a match to a candidate gene only when the candidate also matched the contig's best hit from the non-redundant (nr) database or when the standalone BLAST search gave a better match (lower e-value than the match in nr).

2.4. A posteriori candidate genes identified via differential expression

In native libraries (i.e., those that have not been normalized), the number of sequence reads per gene can serve as a proxy for relative levels of gene expression (i.e., digital transcriptomics; Murray et al., 2007; 't Hoen et al., 2008; Hale et al., 2009). We reasoned that genes underlying water conservation might be specific to kidney and not expressed at high levels in other tissues. To identify *a posteriori* candidates we tested for transcripts that were significantly overexpressed in kidney relative to spleen tissue. A second BLASTx search against the Swissprot database was conducted for all contigs and singletons from each of the 8 individual libraries with a more conservative threshold of e-value $\leq 1.0\times10^{-10}$. We assigned corresponding gene names of the Swissprot blast hits to contigs and then summed the read counts that comprised each contig with an identical gene name. Read counts for each gene also included singletons ascribed to the same gene (e.g., those spanning a gene region not covered by a contig).

The use of four individuals for each tissue allowed us to obtain replicate read counts for each gene. Thus there were two treatment types (K for Kidney and S for Spleen) and four libraries in each treatment. The program DESeq (Anders and Huber, 2010) was used to test for differential expression between these two treatments by comparing the mean read counts for each gene. DESeq is an R/Bioconductor package that infers differential expression between groups of biological replicates by modeling count data with a negative binomial distribution, estimating variance and means from the data. Additionally the package controls for variance due to differences in library size through linear scaling (Anders and Huber, 2010). Differentially expressed genes were identified after a Benjamini–Hochberg procedure was utilized in order to account for multiple testing at a 5% false discovery rate.

2.5. GO terms and molecular markers present in candidates

Blast2GO® (Götz et al., 2008) was used to assign GO terms and annotation for sequences with top BLAST hits that had an e-value $\leq 1\times10^{-6}$. We identified GO terms that were significantly overrepresented in annotations of the *a priori* candidates as well as those *a posteriori* DE genes that were overexpressed in kidney. To do so we used the GOSSIP (Blüthgen et al., 2005) package in Blast2GO to run a Fisher's Exact Test. This tested for GO terms that were more frequently present in the annotation of the genes in our subset of candidates relative to annotation of the rest of the genes in the kidney data set. To correct for multiple testing, we allowed the program to employ a false discovery rate (FDR) correction and filtered the results for the most specific terms at a FDR$<$0.05.

Molecular markers such as single nucleotide polymorphisms and microsatellites within these genes could prove informative for researchers conducting population genetic studies within this or related species. Thus we scanned our data for single nucleotide polymorphisms (SNPs) and SNPs were detected in both pooled assemblies. SNPs called by PCAP were discarded if they were encompassed by fewer than 4 reads, if the minor allele was present in only a single read, or if the minor allele frequency was $<$0.05. In addition to SNPs, we identified microsatellites using Msatcommander (Faircloth, 2008) with a minimum repeat length of six for di-nucleotide repeats and four for tri-, tetra-, penta-, and hexa-nucleotide repeats.

2.6. Testing for positive selection

To test for positive or purifying selection on *D. spectabilis* genes, we conducted comparisons between the *D. spectabilis* sequences and homologous sequences from *M. musculus* and *R. norvegicus*. We evaluated any *a priori* defined candidate genes found in the kidney assembly, and the *a posteriori* candidates that were differentially expressed (DE). The tested set of DE genes included only those that were overexpressed in the kidney relative to spleen and had orthologous sequences in both *M. musculus* and *R. norvegicus*. The coding sequence (CDS) of each *M. musculus* gene was obtained from Ensembl (Release 64). We further limited our comparisons to cases where Ensembl indicated a one-to-one orthologous relationship between gene scaffolds of another kangaroo rat, *Dipodomys ordii*, *M. musculus*, and *R. norvegicus*. This was done in order to avoid cases where recent gene duplications (after splitting of these taxa) would confound the underlying phylogenetic relationships. Below we describe the methods for aligning *D. spectabilis*, *M. musculus*, and *R. norvegicus* sequences followed by a three taxon test

Table 1
List of GO terms used to identify the *a priori* candidate genes from *Mus musculus* and *Rattus norvegicus*.

Gene ontology term	Accession number	Candidate genes annotated with term
Glomerular filtration	GO:0003094	*Aqp 1*, *Adora 1*, *Ednra*, *Uts2r*
Micturition	GO:00060073	*Cacna1c*, *Chrna3*, *Chrnb2*, *Chrnb4*, *Tacr1*, *Kcnma1*, *Trpv1*
Regulation of urine volume	GO:0035809	*Adrb1*, *Adrb2*, *Oxt*, *Avpr2*
Renal absorption	GO:0070293	*Aqp 4*, *Aqp 1*, *Aqp 7*, *Aqp 3*, *Slc9a3r1*, *Hnf1a*, *Slc12a1*, *Slc12a3*
Renal sodium excretion	GO:0035812	*Tacr1*, *Adora1*, *Uts2r*, *Agt*, *Agtr2*, *Oxt*, *Avpr2*, *Avp*, *Anpep*, *Avpr1a*
Renal sodium ion transport	GO:0003096	*Slc9a3r1*
Renal system process involved in regulation of systemic arterial blood pressure	GO:0003071	*Adora1*, *Uts2r*, *Cyba*, *Agtr1a*, *Agtr1b*, *Cyp11b2*, *G6pd*, *Hsd11b2*, *Ren*, *Pcsk5*, *Cyp11b3*
Renal water homeostasis	GO:0003091	*Aqp 4*, *Aqp 1*, *Aqp 2*, *Aqp 7*, *Aqp 3*, *Wfs1*
Renal water transport	GO:0003097	*Aqp 4*, *Aqp1*, *Aqp 2*, *Aqp 7*, *Aqp 3*

for identifying selection in the heteromyid lineage relative to muroid rodents following a methodology similar to Clark et al. (2003).

We collected all *D. spectabilis* contigs which had significant BLAST hits to the *a priori* candidate or DE gene and created a consensus *D. spectabilis* sequence. Sequencher version 5.0 (Genecodes) was used to align each *D. spectabilis* contig in the proper orientation to the presumptive *M. musculus* orthologue by eye and to join contigs that aligned to overlapping or adjacent portions of the gene. This was accomplished by comparing BLASTx searches of each contig against the *M. musculus* transcript. The consensus *D. spectabilis* sequence was then aligned to *M. musculus* and *R. norvegicus* transcripts using DIALIGN-TX (Subramanian et al., 2008). The reference frame for the translation of the *D. spectabilis* sequence was inferred from the BLASTx search.

These alignments started with the region of similarity from the BLASTx hit and continued until the end of the *M. musculus* transcript or until a stop codon was encountered in the *D. spectabilis* sequence. In the latter scenario we encountered cases where the BLASTx search indicated additional regions of similarity between *D. spectabilis* and *M. musculus* separated by gaps in the *D. spectabilis* consensus sequence or in different reading frames. Shifts in reading frame were often preceded by a poor quality base or homopolymer run and thus assumed to be due to sequencing errors. Meanwhile, gaps in the *D. spectabilis* sequence were often the result of incomplete coverage of long transcripts (e.g., one contig covering the 5′ end of the transcript and a second contig at the 3′ end of the transcript but lacking overlapping reads). See Supplemental Fig. 1 for alignment methods in these scenarios. Alignments were constructed for each separate stretch of continuous overlap between the *D. spectabilis* contigs, *M. musculus* sequence, and *R. norvegicus* sequence (e.g., if there was overlap for the first 300 bp, followed by a gap of 50 bp and another overlap of 300 bp, then two separate 300 bp alignments were made). Thus for each gene, we were left with one to several alignments of *D. spectabilis* sequence to different portions of the corresponding *M. musculus* transcript.

Maximum likelihood trees were constructed for each alignment using PhyML 3.0 (Guindon and Gascuel, 2003) following model selection with MrAIC (Nylander, 2004). Each alignment and tree was subsequently used to test for selection along the kangaroo rat lineage relative to the mouse and rat lineages using the dn/ds ratio (ω) along each branch of the tree. In general, $\omega = 1, < 1$, and > 1 are indicative of neutral evolution, purifying selection, and positive selection, respectively. The codeml program in PAML version 4.4 (Yang, 1997, 2007) was used to calculate ω using a three taxa comparison as in the model 1 test of Clark et al. (2003); we used the M0 and M1 models to fix ω or to allow ω to vary along each branch of the tree. A likelihood ratio test was used to evaluate whether the model used to calculate ω was significantly different from the M0 model (where ω is fixed). Positive selection was inferred when $\omega > 1$ and was elevated along the *D. spectabilis* branch relative to the *M. musculus* and *R. norvegicus* branches. When multiple alignments corresponded to a single gene, the analysis was first conducted for each separate alignment and then for an overall gene estimate. For the overall gene estimates, separate alignments were concatenated prior to analysis.

3. Results

3.1. Sequence, assembly, and BLAST annotation

We obtained 446,758 reads spanning 143.75 Mb. After quality control to remove adaptors, cDNA synthesis primers, and reads with poor quality scores, we were left with 433,395 trimmed reads spanning 127.33 Mb. Table 2 shows the distribution of these trimmed reads among the eight libraries as well as the PCAP assembly data for each library. The pooled kidney assembly was derived from 230,299 reads that had a mean length of 296 bases (after trimming). About 58% of these reads assembled into 20,484 contigs with a mean length of 464 bp. Of the 203,096 spleen reads (mean trimmed length of 292 bases), 115,653 assembled into 23,376 contigs (mean length of 435 bp). Thus, the two pooled libraries were similar with respect to the number and length of both reads and contigs.

The BLASTx analysis revealed 9129 contigs in the kidney assembly (44.6% of contigs) and 8237 contigs in the pooled spleen assembly (35.2%) that had significant hits to sequences in the non-redundant (nr) database. These hits were to 6314 (kidney) and 5992 (spleen) unique proteins. This indicates that in some cases, multiple contigs are derived from the same gene, either from alternative splice variants or from different regions of the same transcript for which intervening sequence is lacking.

3.2. A priori candidate genes identified via model species

Our BLASTx search found that 31 kangaroo rat contigs from the pooled kidney assembly had a similarity match to one of our 38 candidate genes. For 13 of these 31 kangaroo rat contigs (42%), the candidate murine gene was also the top BLAST hit recovered from the non-redundant (nr) database. These 13 contigs matched 9 of our candidate genes (i.e., two murine genes matched multiple kangaroo rat contigs; Table 3 lists these contigs and their candidate gene match). Thus, the kangaroo rat sequence reads include at least 9 genes known to also be expressed in murine kidneys.

3.3. A posteriori candidate genes identified via differential expression

The DESeq analysis revealed 59 differentially expressed (DE) genes between kidney ($n = 4$) and spleen ($n = 4$) at a 5% false discovery rate (see Fig. 1; points in red represent genes with significant DE). Fig. 2 is an expression heat map of the top 59 DE genes. Of these DE genes, 32 were overexpressed in kidney tissue (Table 4). Interestingly one of these 32 is *Slc12a1*, which is also one of our *a priori* candidate genes. *Slc12a1* (solute carrier family 12 member 1) is a chlorine coupled sodium and potassium transporter that acts in renal salt reabsorption (Herbert, 1998).

Fig. 3 illustrates the variance functions for read counts within each treatment (i.e., kidney or spleen). Each dashed line corresponds to the variance of an individual sample (e.g., male 0828 kidney). The

Table 2
Descriptive statistics of the 10 PCAP assemblies (8 individual libraries and 2 pooled). Trimmed reads refer to the number of reads after quality control and removal of both sequencing and cDNA synthesis primers.

Library	Trimmed reads	Mean read length	Contigs	Mean contig length	Mean contig depth	Assembled reads
Male 0828 kidney	41538	289	5051	415	4.87	24589
Male 0812 kidney	30626	307	3518	439	5.27	18523
Female 0862 kidney	36931	286	4325	412	5.02	21700
Female 0871 kidney	121204	298	12439	457	5.94	73924
Total kidney assembly	230299	296	20484	464	6.51	133283
Male 0828 spleen	19146	277	2353	392	4.72	11102
Male 0812 spleen	17192	280	2071	389	5.03	10425
Female 0862 spleen	19752	264	1854	383	5.99	11109
Female 0871 spleen	147006	298	19566	433	4.46	87322
Total spleen assembly	203096	292	23376	435	4.95	115653

Table 3
A priori murine candidate genes present in the pooled kangaroo rat kidney data set.

Contigs numbers*	Gene symbol	Gene name	Reads in contigs
4857.1	*Agt*	Angiotensinogen (serpin peptidase inhibitor, clade A, member 8)	5
1042.1	*Agtr1a*	Angiotensin II receptor, type 1a	22
1069.1	*Aqp2*	Aquaporin 2	22
453.1	*Cyba*	Cytochrome b-245, alpha polypeptide	44
19505.1	*Hsd11b2*	Hydroxysteroid 11-beta dehydrogenase 2	2
14973.1	*Pcsk5*	Proprotein convertase subtilisin/kexin type 5	2
0.1, 8266.1, 8490.1, 9053.1	*Slc12a1*	Solute carrier 12 (sodium/potassium/chloride transporters), member 1	304, 3, 3, 3, respectively
2746.1	*Slc12a3*	Solute carrier 12 (sodium/potassium/chloride transporters), member 3	9
7591.1, 18299.1	*Slc9a3r1*	Solute carrier family 9 (sodium/hydrogen exchanger), member 3 regulator 1	4, 2, respectively

* PCAP naming convention starts with the first contig as contig 0.1, the second contig is 1.1, and so on. We retained this convention for consistency.

variance is composed of the shot noise, defined as the noise inherent to read counts, and the raw variance, which is the signal in the data revealed by the biological replicates (solid lines). Thus, the difference between the dashed line and its corresponding solid line is the shot noise (see Anders and Huber, 2010). Fig. 3 shows that for genes with low read counts, the difference between mean counts is dominated by shot noise, but as coverage increased these two lines converge and the variance function beyond this point can be attributed to biological replication (Anders and Huber, 2010). This means that for genes with deep coverage, we have the power to detect biological differences in expression. At base means ≥1000 there are no genes with that many reads and the variance functions are no longer accurate.

3.4. GO terms and molecular markers present in candidates

Among the 9 *a priori* defined candidate genes expressed in kangaroo rat kidney, only two genes (*Slc12a1* and *Cyba*) contained SNPs. The total number of SNPs in *Slc12a1* was 7 and the ratio of transitions to transversions was 6:1. The only SNP in our *Cyba* reads was a transition. In addition to SNPs, we found a tri-nucleotide repeat (AGG) in *Aqp2* that consisted of five repeats. Within the set of 32 genes over expressed in kidney, we found 92 SNPs spread across 24 genes. The other eight genes did not contain any SNPs that passed our criteria. Additionally, 40 different microsatellites were located in 17 of these same 32 genes.

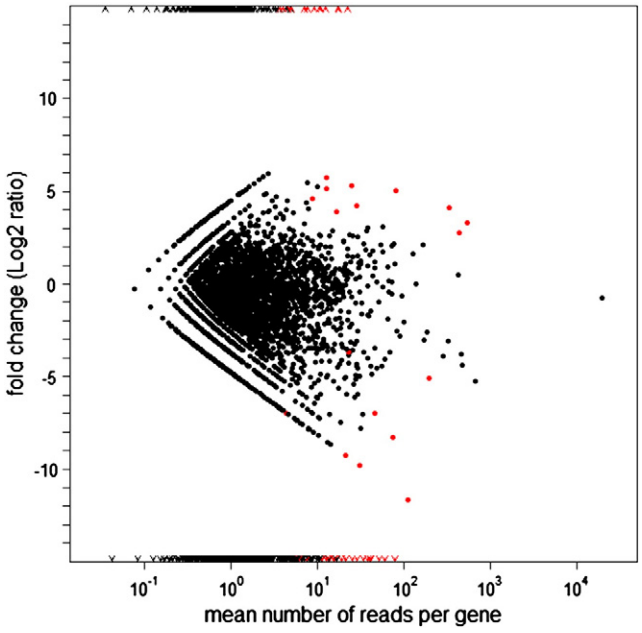

Fig. 1. Scatter plot of the fold change for each gene versus the mean number of reads across all libraries for that gene. Red dots represent genes that were significantly differentially expressed between kidney and spleen tissue at a 5% false discovery rate. Arrows at the top and bottom of the plot indicate genes where the fold change difference is > 15 or < −15, including values of ∞ or −∞ that result when there are 0 reads per gene in one of the two treatments. Genes with a positive fold change value are overexpresssed in spleen tissue while genes with a negative fold change are overexpressed in kidney tissue.

The Fisher's exact test revealed 88 GO terms which were overrepresented in annotations of our *a priori* candidates relative to the remainder of the kidney data set (Supplementary Table 2). Among these, 4 were terms in the cellular component (C) category, 15 from the molecular function (M) category, and 69 from biological process (P). Blast2GO (Götz et al., 2008) was used to filter these terms to include only GO terms at the most specific level of the gene ontology hierarchy. These filtered terms are displayed in Fig. 4, which shows the proportion of contigs from the *a priori* candidate genes that have been annotated with a given term along with the proportion of all other kidney contigs annotated with the same term.

There were 175 GO terms overrepresented among the annotations of the *a posteriori* DE genes relative to non-DE genes. Of these, 41 terms belong to the cellular component category, 58 to the molecular function category, and 76 to biological processes (Supplementary Table 3). As above, Blast2GO was used to produce a subset of most specific terms that are displayed in Figs. 5–7 (C, M, and P, respectively).

3.5. Testing for positive selection

We examined the *a priori* candidates as well as the DE genes for preliminary signs of positive selection through a comparison of *M. musculus, R. norvegicus*, and *D. spectabilis* sequences to estimate ω. The values of dn, ds, and ω for each alignment are presented in Supplementary Tables 4 and 5 for the *a priori* candidates and the *a posteriori* candidate genes, respectively. Each fragment represents a portion of the alignment before a stop codon was reached in the experimental *D. spectabilis* sequence. Values of $\omega > 1$ would have indicated positive selection and divergence in the amino acid composition of the *D. spectabilis* sequence from the *M. musculus* and *R. norvegicus* sequences. However, all of the alignments except for one had $\omega < 1$ and thus fail to show signs of positive selection over the alignment as a whole. The one alignment with $\omega > 1$ was *slc13a1* portion 1, and even in this case the model predicting positive selection was not significantly different from the null model 0, thus positive selection cannot be inferred. Additionally this was a small fragment and $\omega < 1$ for both a second portion of the gene and the concatenated sequence containing both fragments.

4. Discussion

We sought to identify genes that influence the ability of *D. spectabilis* to retain water during waste production with such efficiency that it is able to survive without drinking water. We employed a two pronged approach to achieve this goal. First, we defined a small set of candidate genes *a priori* from annotation of genes previously defined in distantly related model species. Second, we searched for differentially expressed

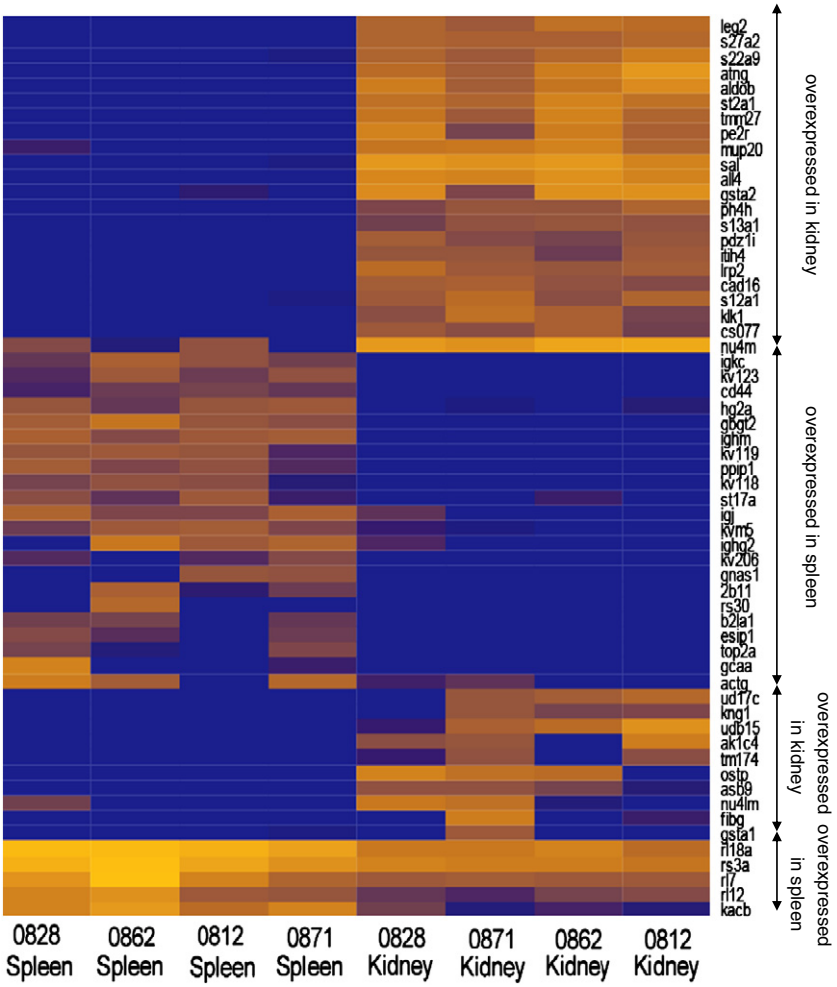

Fig. 2. Heat map illustrating the 59 most differentially expressed genes between kidney and spleen tissue. Sample names are listed along the bottom of the figure and gene names are listed along the right vertical axis. Fold-change values were transformed with a variance-stabilizing transformation to allow distance calculations for infinite fold-change differences that result when a gene is expressed in one tissue but has zero counts in the other. Orange is for hot or high expression, blue corresponds to cold or low expression.

genes and used those overexpressed in kidney to define another set of *a posteriori* candidate genes. The genes we identified in this study will form the basis for future comparative work examining in the role of natural selection on their sequence evolution and expression.

4.1. A priori candidate genes identified via model species

Out of our list of 38 *a priori* candidate genes annotated with Renal System Process related gene ontology terms, at least 9 *a priori* candidates were expressed in *D. spectabilis* kidney tissue. These nine genes include several involved in water reabsorption from kidney filtrate either through direct transport of water (*Aqp2* transports water back into the kidney from the collecting duct (Gomes et al., 2009; Ishibashi et al., 2009)) or through transport of solutes to create a concentration gradient for further transport of water. For instance, *Slc12a3* and *Slc12a1* encode solute carrier proteins that are involved in sodium reabsorption from the distal convoluted tubule and thick ascending limb, respectively, of the loop of Henle (Herbert, 1998; Herbert et al., 2004). As sodium is transported out of the filtrate by these channels, the surrounding kidney tissue becomes hypertonic relative to the filtrate. This in turn drives further transport of water out of the filtrate and back into the kidney through osmosis.

In addition to genes involved in water and solute transport, several *a priori* candidates were detected that control urine output and concentration through their interactions with other proteins in the kidney. Specifically, *Slc9a3r1* is not an actual transport protein, but it interacts with multiple proteins in the kidney, including sodium-hydrogen antiporter 3 (NHE3 which is encoded by *Slc9a3*) and Na-phosphate cotransporter 2a (Npt2a) (Weinman et al., 2006; Cunningham et al., 2007). Mice without functional copies of *Slc9a3r1* have been found to have increased phosphate and uric acid concentrations among other symptoms (Weinman et al., 2006; Cunningham et al., 2007). The ability of this gene's product to interact with multiple transport proteins and the changes in urine composition that occur in *Slc9a3r1* deficient mice highlight it as an appropriate candidate underlying osmoregulation.

Five other *a priori* candidate genes (*Agt*, *Agtr1a*, *Cyba*, *Hsd11b2*, and *Pcsk5*) were included as candidates due to annotation with the GO term "renal system process involved in regulation of systemic arterial blood pressure". Some of the mechanisms for control of blood pressure such as the renin-angiotensin system (RAS) can affect kidney function through reduced urine production and increased sodium absorption (Peart, 1965; Hall, 1986; Tamura et al., 1998; Paul et al., 2006). Angiotensinogen (product of *Agt*) is cleaved by renin in the kidney to produce Angiotensin I which is then converted to Angiotensin II by Angiotensin Converting Enzyme and binds to Angiotensin II receptor 1 (encoded by genes *Agtr1a* and *Agtr1b* in rodents, Mangrum et al., 2002) to exert its effects (Mangrum et al., 2002; Paul et al., 2006). Among the impacts of Angiotensin II are directly causing increased reabsorption of sodium in the proximal tubules, restricting blood flow

Table 4
32 genes significantly overexpressed in kangaroo rat kidney relative to spleen. Fold change values have been made positive to reflect higher expression relative to spleen, in cases where there are no reads in the spleen data set for a gene, a value of ∞ has been substituted. Read counts refer to the number of reads summed across all four libraries of each treatment.

Swiss-Prot symbol	Gene symbol	Gene name	Kidney read count	Spleen read count	Log2 fold change	P-value*
ak1c4	*Akr1C4*	Aldo-keto reductase family 1 member C4	184	0	∞	0.0230
aldob	*Aldob*	Aldolase B, fructose-bisphosphate	383	0	∞	0.00528
all4	*All4*	Allergen Fel d 4	747	0	∞	0.000403
asb9	*Asb9*	Ankyrin repeat and SOCS box protein 9	105	0	∞	0.0450
cs077	*C19orf77*	Transmembrane protein C19orf77 homolog	131	0	∞	0.0446
cad16	*Cdh16*	Cadherin-16	186	0	∞	0.00918
fibg	*Fgg*	Fibrinogen gamma chain	368	0	∞	6.37E-05
atng	*Fxyd2*	FXYD domain-containing ion transport regulator 2	427	0	∞	0.00499
gsta1	*Gsta1*	Glutathione S-transferase A1	104	1	6.97	0.0364
gsta2	*Gsta2*	Glutathione S-transferase A2	477	1	8.29	0.00975
itih4	*Itih4*	Inter-alpha-trypsin inhibitor heavy chain H4	139	0	∞	0.0289
klk1	*Klk1*	Kallikrein-1	279	0	∞	0.00148
kng1	*Kng1*	Kininogen-1	108	0	∞	0.0361
leg2	*Lgals2*	Galectin-2	240	0	∞	0.0119
lrp2	*Lrp2*	Low-density lipoprotein receptor-related protein 2	203	0	∞	0.0138
mup20	*Mup20*	Major urinary protein 20	498	2	6.97	0.00399
nu4m	*Mtnd4*	NADH-ubiquinone oxidoreductase chain 4	1536	27	5.07	0.0313
nu4lm	*Mtnd4L*	NADH-ubiquinone oxidoreductase chain 4 L	296	9	3.71	0.0446
ph4h	*Pah*	Phenylalanine-4-hydroxylase	156	0	∞	0.0289
pdz1i	*Pdzk1ip1*	PDZK1-interacting protein 1	120	0	∞	0.0477
pe2r	*Ptgr2*	Prostaglandin reductase 2	265	0	∞	0.0130
sal	*Sal1*	Salivary lipocalin	1137	1	11.64	0.000219
s12a1	*Slc12a1*	Solute carrier family 12 member 1	263	1	9.24	0.00888
s13a1	*Slc13a1*	Solute carrier family 13 member 1	123	0	∞	0.0361
s22a9	*Slc22a9*	Solute carrier family 22 member 9	250	1	9.80	0.0312
s27a2	*Slc27a2*	Very long-chain acyl-CoA synthetase	250	0	∞	0.00650
ostp	*Spp1*	Secreted phosphoprotein 1	382	0	∞	0.00168
st2a1	*Sult2a1*	Bile salt sulfotransferase 1	359	0	∞	0.00455
tm174	*Tmem174*	Transmembrane protein 174	91	0	∞	0.0450
tmm27	*Tmem27*	Collectrin	301	0	∞	0.00719
ud17c	*Ugt1a7c*	UDP-glucuronosyltransferase 1-7 C	150	0	∞	0.0280
udb15	*Ugt2b15*	UDP-glucuronosyltransferase 2B15	328	0	∞	0.00634

* Significant p-values after a Benjamini–Hochberg correction.

to affect filtration rates in the kidney, and causing the release of aldosterone which further causes increased sodium absorption (Tamura et al., 1998; Mangrum et al., 2002). This increased sodium absorption as a result of the RAS has long been noted to result in decreased urine production (Peart, 1965; Hall, 1986).

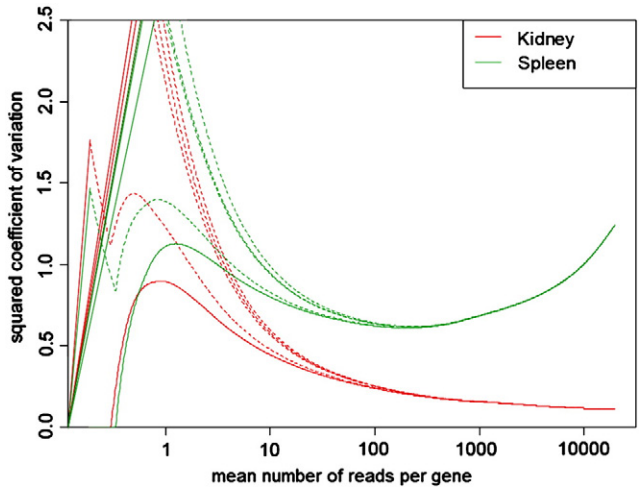

Fig. 3. Plot displaying the variance function for kidney (K, solid red line) and spleen (S, solid green line) treatments. The squared coefficient of variation is plotted versus the base mean values, which is related to the mean number of reads per gene across all libraries. The dashed lines represent the variance of all 8 individual libraries. The difference between the dashed and solid lines is the shot noise which is indicative of variance due to low read counts. When these converge, variation between samples is due to biological variation.

Knockout studies in mice show that the absence of two of our *a priori* candidates (*Agt* and *Agtr1a*) disrupts the ability of the RAS to increase sodium absorption. Under similar diets, mice without a functional *Agt* gene produce more urine and urine with a lower overall concentration when compared to wildtype mice that have a functional *Agt* (Tamura et al., 1998). Similarly, mice without functional Angiotensin II receptor 1a drink more water and produce more urine than wild type mice under identical salt intake (Mangrum et al., 2002). Our detection of these nine *a priori* candidate genes in *D. spectabilis* confirms that our methods have indeed identified heretofore uncharacterized genes involved in sodium reabsorption. Our data raise the possibility that salt reabsorption, a process that is common to general kidney function, is under strict control in *D. spectabilis* and is largely responsible for the increased ability of kangaroo rats to conserve water during urine production.

We failed to detect 29 of the *a priori* candidate genes, but of course the absence of evidence of genes in our dataset is not evidence of their absence from the *D. spectabilis* kidney transcriptome. The apparent absence of our other 29 *a priori* candidate genes may be due to several biological factors. The Heteromyidae and Muridae diverged 60–90 mya (Adkins et al., 2003; Huchon et al., 2007; Meredith et al., 2011) with the Timetree estimate at 78.9 mya (Honeycutt, 2009). Thus, the candidate genes we identified from murine rodents may be too divergent from the orthologous *D. spectabilis* genes to identify and annotate using a strict BLAST search. Furthermore, >50% of the contigs did not yield a significant BLAST hit from the non-redundant (nr) database. Many of these unidentified contigs could include orthologues of murine candidate genes, or could represent novel genes that perform similar functions. However, we chose to employ a conservative cut-off rather than risk erroneous annotation of contigs with paralogous sequences. Other candidate genes may have been expressed and sampled but were represented only as singletons. Singleton reads were included

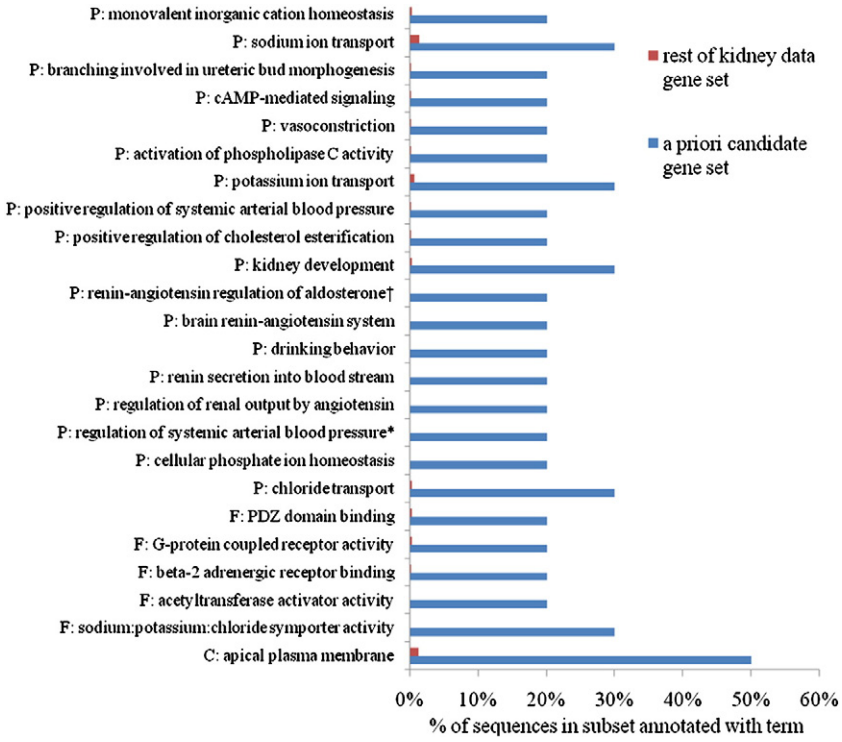

Fig. 4. GO terms that are significantly enriched in the *a priori* candidate data set relative to the total kidney data set after an FDR cutoff of <.05 and after Blast2GO filtered for the most specific terms. Bars represent the proportion of contigs in the subset that have been annotated with the GO term. *regulation of systemic arterial blood pressure, †renin-angiotensin regulation of aldosterone.

only in our expression test, and there at a conservative threshold of an e-value $\leq 1.00 \times 10^{-10}$. Some recent transcriptome studies have utilized singleton sequence data because many are biologically real (Meyer et al., 2009; Babik et al., 2010), but we have avoided them because many contain sequencing errors.

4.2. A posteriori candidate genes identified via differential expression

We identified 32 genes that were overexpressed in kidney tissue relative to spleen tissue; of these, 24 were expressed only in kidney. We used sequence read counts to estimate relative gene expression and to identify potential *a posteriori* candidate genes. Some of these genes are of interest due to their functions in the kidney (including *Slc12a1*) or the function of genes with which they interact or regulate. For instance, *fxyd2* encodes the sodium/potassium-transporting ATPase subunit gamma, and although it is not required for the sodium/potassium-transporting ATPase's function of sodium transport, *fxyd2* can modify the activity of sodium/potassium-transporting ATPase in the kidney (Therien et al., 1997). The *a posteriori* candidates also include multiple genes involved in reabsorption of sodium in the kidney. This is of interest due to the fact that this function was seen in many of the detected *a priori* genes and is the function of the most highly expressed gene in our total kidney dataset (*Slc12a1*) as well. Such genes were missed in the screening for *a priori* candidates either due to the lack of annotation with a renal system process GO term or annotation with GO terms that were too specific to be included in the general categories we used. Their presence highlights the importance of a strong osmotic gradient to retention of water in the kidney of *D. spectabilis*.

Fig. 3 shows a great deal of shot noise at low and medium base mean values (which correspond to the mean number of reads from a gene across the eight libraries). At these read count levels we could increase our signal to noise ratio through increasing sequencing depth and thus would have greater power to detect fine scale expression differences. However, Fig. 3 shows that at high base mean values there is little shot noise (area where the dashed and solid lines converge) and sufficient power to detect significant expression differences for 59 genes from $^1/_2$ plate of 454 sequencing. The genes overexpressed in kidney tissue provide defined targets for more extensive study and the sequences obtained in our study provide the genomic infrastructure needed for future q-PCR or further RNA-seq experiments to test for differential gene expression at a finer scale in this and other Heteromyid species.

4.3. GO terms and molecular markers present in candidates

The GO terms overrepresented in the sequences of the *a priori* and *a posteriori* candidate genes provide a set of annotations that are congruent with the notion that these genes play a significant role in limiting urine output. For instance, these include terms such as “drinking behavior” and “cellular ion homeostasis” in the overrepresented terms for the *a priori* candidates. Overrepresented terms for the *a posteriori* data set included several transport related terms which were absent from the *a priori* candidate gene enriched terms such as “bile acid and bile salt transport”, “lipoprotein transport”, and “sodium:potassium-exchanging ATPase complex”. Other GO terms were overrepresented in both *a priori* and *a posteriori* candidates, including “sodium ion transport”, “potassium ion transport”, and “sodium:potassium:chloride symporter activity”. The presence of these terms in both sets of candidates provides a framework for classifying other genes present in our kidney data set that are annotated with these terms as additional candidates and targets for further study.

The molecular markers identified in this study provide useful markers for population genetic studies in this species. SNP and microsatellite identification has been successfully conducted as part of several transcriptome sequencing projects using next generation sequencing (see examples in reviews by Ekblom and Galindo, 2011 and Garvin et al., 2010). As in the studies described in those reviews,

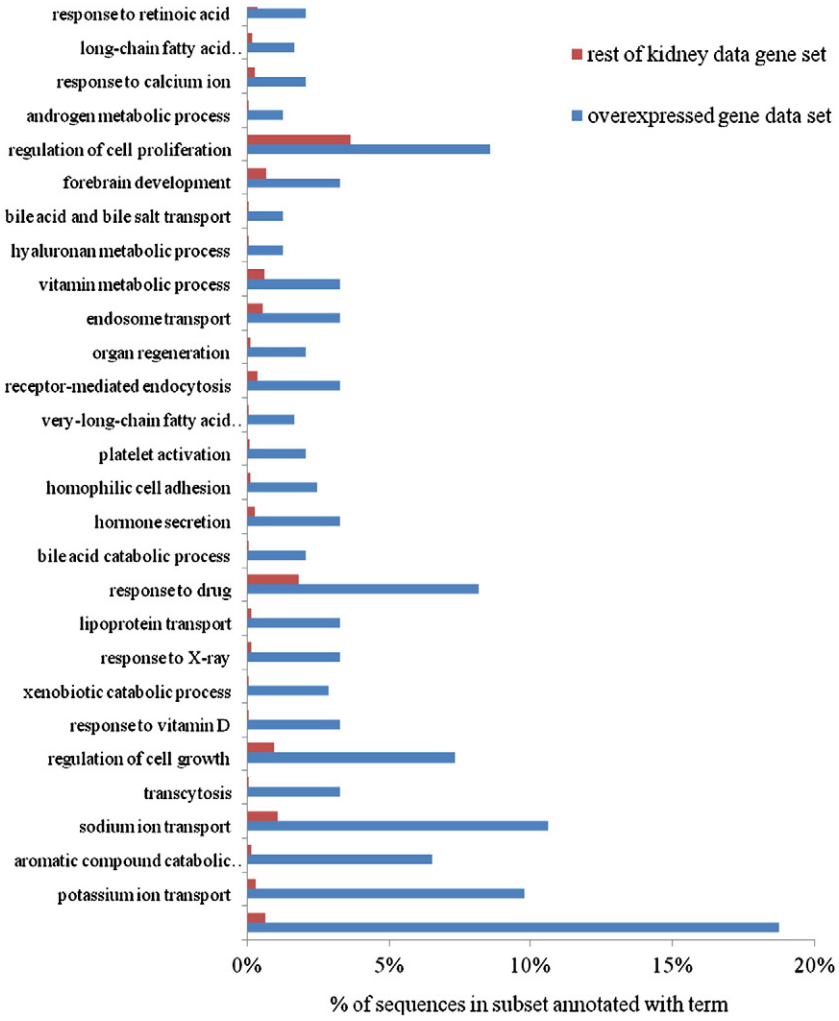

Fig. 5. Most specific biological process GO terms that are significantly enriched in the overexpressed data set relative to the total kidney data set after an FDR cutoff of <.05 and after Blast2GO filtered for the most specific terms. Bars represent the proportion of contigs in the subset that have been annotated with the GO term. *mitochondrial electron transport, NADH to ubiquinone.

the markers identified here are predominantly from coding sequence. Existing 'neutral' microsatellite markers in *D. spectabilis* (Davis et al., 2000; Waser et al., 2006) provide a suitable backdrop for population genetic comparisons among marker sets (e.g. Busch et al., 2009).

4.4. Testing for positive selection

We have performed initial tests for selection using transcriptome data, but these estimates of ω failed to demonstrate any evidence of positive selection on the portions of the genes tested in this study. Instead ω values were <1 (except for one fragment of *slc13a1*), which is indicative of a low proportion of non-synonymous changes and purifying selection. The failure to detect positive selection is somewhat unsurprising as it has been argued that pairwise and whole gene estimates of dn/ds ratios are conservative (Nielsen, 2005), and more often selection will act on individual sites within a gene (Yang et al., 2000). For this reason many studies using dn/ds ratios and similar divergence based tests for selection have taken advantage of codon based models (Nielsen and Yang, 1998; Yang et al., 2000) which utilize a phylogenetic framework to compare multiple models in a likelihood ratio test to detect selection (Yang, 2007). These include studies that have utilized partial or full CDS derived from next generation sequencing data (Künster et al., 2010; Renaut et al., 2010; Brieuc and Naish, 2011). In the Brieuc and Nash study, 15 of 152 surveyed genes showed signs of positive selection and of these only three sequences had a ω >1, with 9 of the genes displaying sequence wide values of ω <.05. Therefore the possibility still exists that positive selection is acting on some of the candidate genes identified in this study. For confident inference of selection from such divergence tests, robust phylogenetic relationships and further taxonomic sampling are needed (and are under way in our lab).

5. Conclusions and future directions

This project has generated more than 400,000 sequence reads that have been used to identify the presence of expected candidate genes as well as to identify novel target genes that may underlie the physiological adaptation of water conservation in desert rodents. Efficient osmoregulation in *D. spectabilis* is no doubt influenced by behavior and other physiological mechanisms (e.g., hormonal control). However, the genes identified in this study – particular those which show high levels of DE and have known functions in water or solute transport – are likely to be important drivers of evolutionary adaptations to arid environments.

Of particular interest is *Slc12a1*, one of the murine candidate genes that was identified in *D. spectabilis* and overexpressed in the kidney transcriptome. Mutations in *Slc12a1* cause Bartter's syndrome in humans (Simon et al., 1996a; Herbert, 1998; Adachi et al., 2007).

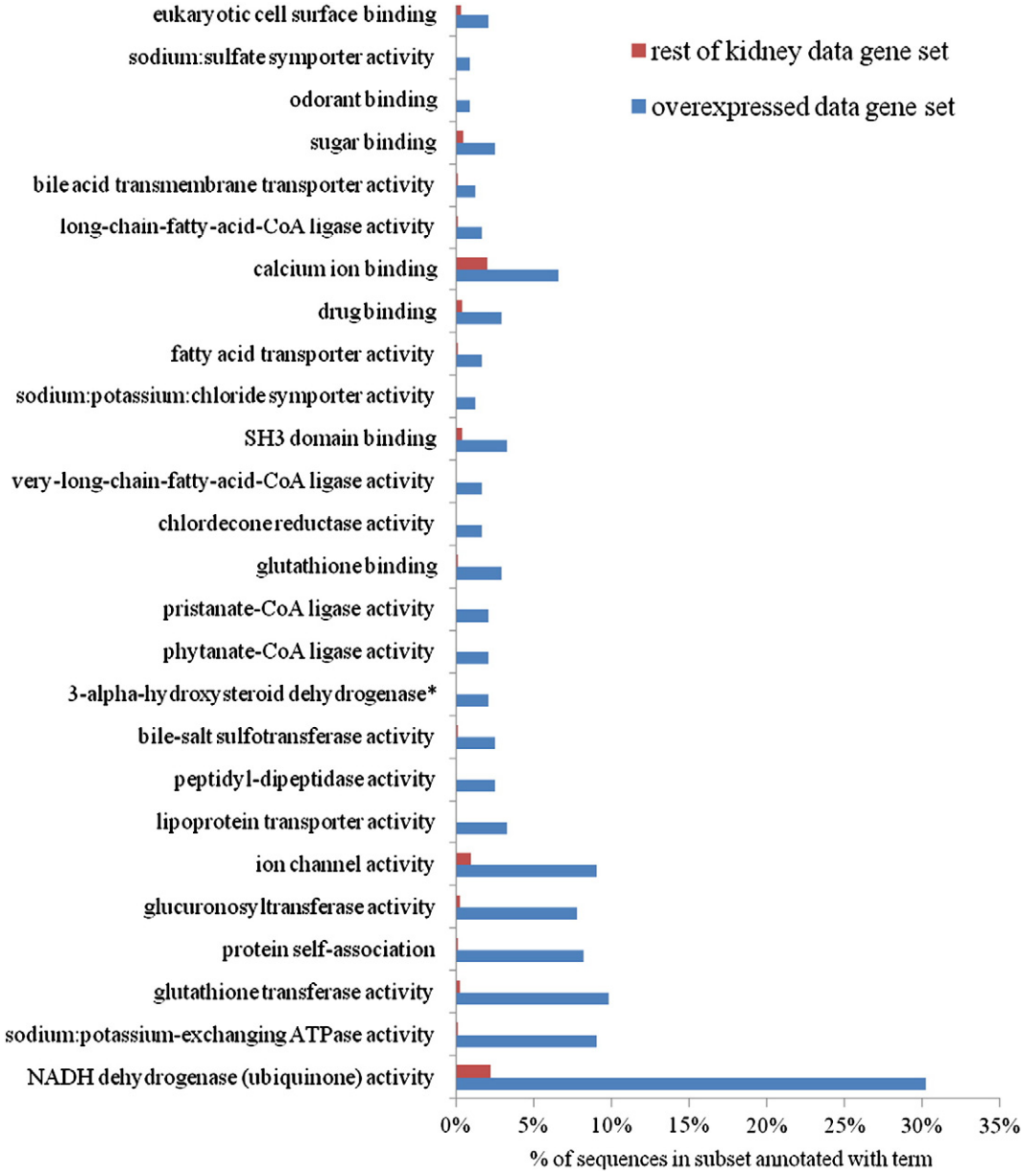

Fig. 6. Most specific molecular function GO terms that are significantly enriched in the overexpressed data set relative to the total kidney data set after an FDR cutoff of <.05 and after Blast2GO filtered for the most specific terms. Bars represent the proportion of contigs in the subset that have been annotated with the GO term. *3-alpha-hydroxysteroid dehydrogenase (B-specific) activity.

Among the symptoms of this disorder is a compromised ability to concentrate urine (Herbert, 1998). Meanwhile, mutations in *Slc12a3* are responsible for another human kidney defect, Gitelman syndrome (Simon et al., 1996b), and additional SNP variants of *Slc12a3* have been associated with susceptibility to diabetic nephropathy (Tanaka et al., 2003). These lines of evidence underscore the likelihood that these genes and others responsible for renal sodium and ion transport are integral to the enhanced urine concentrating ability of *D. spectabilis*.

In terms of future directions, we aim to expand our analysis of osmoregulation to include other heteromyid species from the two remaining subfamilies within this group. These include species from wet (rainforest; *Heteromys desmarestianus*) and dry (desert; *Chaetodipus baileyi*) environments to permit a comparative evaluation of sequence evolution and gene expression in the kidney. The candidate and DE genes identified in the current study will serve as a necessary starting point for our broader phylogenetic and environmental comparison. Additionally, the spleen data (used as a reference in this study) will allow for evolutionary comparisons of the immune response genes that are expressed in heteromyid rodents.

Acknowledgments

This work was primarily funded through a supplement to a National Science Foundation award (DEB-0816925). Additional funding to N.J.M comes from an NSF doctoral dissertation improvement grant (DEB-1110421). J.A.D's lab is also supported by University Faculty Scholar funds provided through Purdue's Office of the Provost. We thank Phillip San Miguel, Sam Martin, and Purdue University's Genomics Core Facility for help with cDNA sequencing. We also thank two anonymous reviewers for suggestions and comments that helped to improve the manuscript. Additionally, we are grateful for comments and feedback on this manuscript from the DeWoody Lab. Reads and contigs from PCAP assembly from the 454 sequencing has been deposited in datadryad.org (doi:10.5061/dryad.sf654).

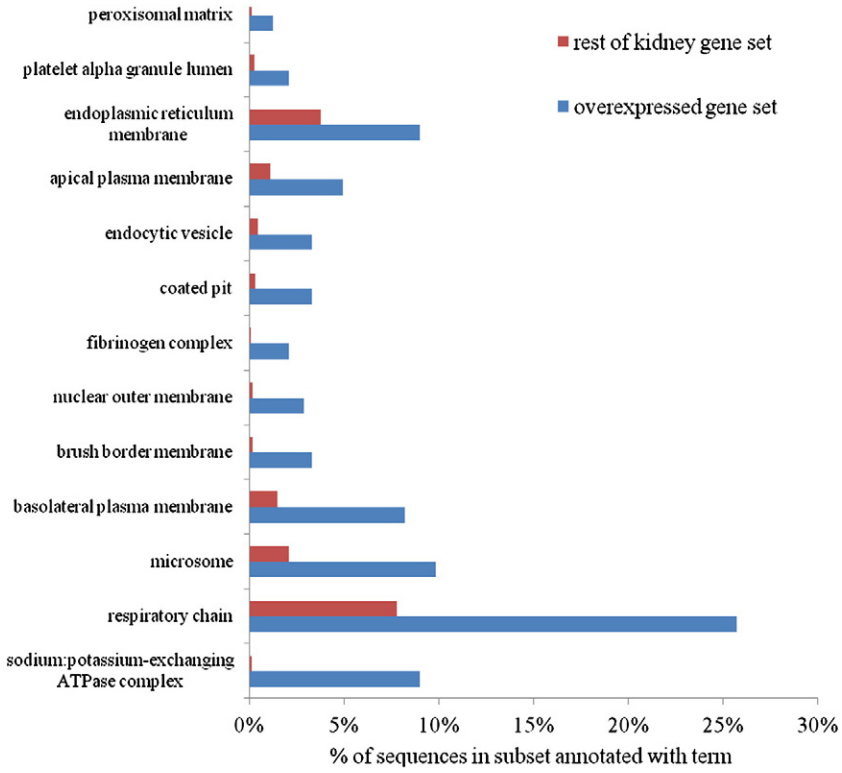

Fig. 7. Most specific cellular component GO terms that are significantly enriched in the overexpressed data set relative to the total kidney data set after an FDR cutoff of <.05 and after Blast2GO filtered for the most specific terms. Bars represent the proportion of contigs in the subset that have been annotated with the GO term.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.cbd.2012.07.001.

References

't Hoen, P.A.C., Ariyurek, Y., Thygesen, H.H., Vreugdenhil, E., Vossen, R.H.A.M., de Menezes, R.X., Boer, J.M., van Ommen, G.J.B., den Dunnen, J.T., 2008. Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. Nucleic Acids Res. 36, 1–11.

Adachi, M., Asakura, Y., Sato, Y., Tajima, T., Nakajima, T., Yamamoto, T., Fujieda, K., 2007. Novel *SLC12A1* (NKCC2) mutations in two families with Bartter Syndrome Type 1. Endocr. J. 54, 1003–1007.

Adkins, R.M., Walton, A.H., Honeycutt, R.L., 2003. Higher-level systematic of rodents and divergence time estimates based on two congruent nuclear genes. Mol. Phylogenet. Evol. 26, 409–420.

Alexander, L.F., Riddle, B.R., 2005. Phylogenetics of the new world rodent family Heteromyidae. J. Mammal. 86, 366–379.

Al-kahtani, M.A., Zuleta, C., Caviedes-Vidal, E., Garland, T., 2004. Kidney mass and relative medullary thickness of rodents in relation to habitat, body size, and phylogeny. Physiol. Biochem. Zool. 77, 346–365.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Anders, S., Huber, W., 2010. Differential expression analysis for sequence count data. Genome Biol. 11, R106.

Babik, W., Stuglik, M., Qi, W., Kuenzli, M., Kuduk, K., Koteja, P., Radwan, J., 2010. Heart transcriptome of the bank vole (*Myodes glareolus*): towards understanding the evolutionary variation in metabolic rate. BMC Genomics 11, 390.

Best, T.L., 1988. *Dipodomys spectabilis*. Mamm. Species 311, 1–10.

Beuchat, C.A., 1996. Structure and concentrating ability of the mammalian kidney: correlations with habitat. Am. J. Physiol. 271, R157–R179.

Blüthgen, N., Brand, K., Čajavec, B., Swat, M., Herzel, H., Beule, D., 2005. Biological profiling of gene groups utilizing gene ontology. Genome Inform. 16, 106–115.

Brieuc, M.S.O., Naish, K.A., 2011. Detecting signatures of positive selection in partial sequences generated on a large scale: pitfalls, procedures and resources. Mol. Ecol. Resour. 11, 172–183.

Busch, J.D., Waser, P.M., DeWoody, J.A., 2009. The influence of density and sex on patterns of fine-scale genetic structure. Evolution 63, 2302–2314.

Cain III, J.W., Krausman, S., Rosenstock, S., Turner, J.C., 2006. Mechanisms of thermoregulation and water balance in desert ungulates. Wildl. Soc. Bull. 34, 570–581.

Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., Lewis, S., 2009. AmiGO Hub, Web Presence Working Group. AmiGO: online access to ontology and annotation data. Bioinformatics 25, 288–289.

Clark, A.G., Glanowski, S., Nielsen, R., Thomas, P.D., Kejariwal, A., Todd, M.A., Tanenbaum, D.M., Civello, D., Lu, F., Murphy, B., Ferriera, S., Wang, G., Zheng, X., White, T.J., Sninsky, J.J., Adams, M.D., Cargill, M., 2003. Inferring nonneutral evolution from human–chimp–mouse orthologous gene trios. Science 302, 1960–1963.

Cresko, W.A., Amores, A., Wilson, C., Murphy, J., Currey, M., Phillips, P., Bell, M.A., Kimmel, C.B., Postlethwait, J.H., 2004. Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. Proc. Natl. Acad. Sci. U. S. A. 101, 6050–6055.

Cunningham, R., Brazie, M., Kanumuru, S., Xiaofei, E., Biswas, R., Wang, F., Steplock, D., Wade, J.B., Anzai, N., Endou, H., Shenolikar, S., Weinman, J., 2007. Sodium–hydrogen exchanger regulatory factor-1 interacts with mouse urate transporter 1 to regulate renal proximal tubule uric acid transport. J. Am. Soc. Nephrol. 18, 1419–1425.

Davis, C., Keane, B., Swanson, B., Loew, S., Waser, P.M., Strobeck, C., Fleischer, R.C., 2000. Characterization of microsatellite loci in banner-tailed and giant kangaroo rats. *Dipodomys spectabilis and D. ingens*. Mol. Ecol. 9, 642–644.

Eisenberg, J., 1963. The Behavior of Heteromyid Rodents. University of California Publications in Zoology.

Ekblom, R., Galindo, J., 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity 107, 1–15.

Eo, S.H., Doyle, J.M., Hale, M.C., Marra, N.J., Ruhl, J.D., DeWoody, J.A., 2012. Comparative transcriptomics and gene expression in larval tiger salamander (*Ambystoma tigrinum*) gill and lung tissues as revealed by pyrosequencing. Gene 492, 329–338.

Faircloth, B.C., 2008. Msatcommander: detection of microsatellite repeat arrays and automated, locus-specific primer design. Mol. Ecol. Resour. 8, 92–94.

Feder, M.E., Mitchell-Olds, T., 2003. Evolutionary and ecological functional genomics. Nat. Rev. Genet. 4, 651–657.

Garvin, M.R., Saitoh, K., Gharrett, A.J., 2010. Application of single nucleotide polymorphisms to non-model species: a technical review. Mol. Ecol. Resour. 10, 915–934.

Gomes, D., Agasse, A., Thiébaud, P., Delrot, S., Gerós, H., Chaumont, F., 2009. Aquaporins are multifunctional water and solute transporters highly divergent in living organisms. Biochim. Biophys. Acta 1788, 1213–1228.

Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nueda, M.J., Robles, M., Talón, M., Dopazo, J., Conesa, A., 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36, 3420–3435.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.

Hale, M.C., McCormick, C.R., Jackson, J.R., DeWoody, J.A., 2009. Next-generation pyrosequencing of gonad transcriptomes in the polyploidy lake sturgeon (*Acipenser fulvescens*): the relative merits of normalization and rarefaction in gene discovery. BMC Genomics 10, 203.

Hale, M.C., Jackson, J.R., DeWoody, J.A., 2010. Discovery and evaluation of candidate sex- determining genes and xenobiotics in the gonads of lake sturgeon (*Acipenser fulvescens*). Genetica 138, 745–756.
Hall, J., 1986. Control of sodium excretion by angiotensin II: intrarenal mechanisms and blood pressure regulation. Am. J. Physiol. Regul. Integr. Comp. Physiol. 250, R960–R972.
Herbert, S.C., 1998. Roles of Na–K–2Cl and Na–Cl cotransporters and ROMK potassium channels in urinary concentrating mechanism. Am. J. Physiol. Renal Physiol. 275, F325–F327.
Herbert, S.C., Mount, D.B., Gamba, G., 2004. Molecular physiology of cation-coupled Cl^- cotransport: the SLC12 family. Pflügers Arch. 447, 580–593.
Holdenried, R., 1957. Natural history of the bannertail kangaroo rat in New Mexico. J. Mammal. 38, 330–350.
Honeycutt, R.L., 2009. Rodents (Rodentia). In: Hedges, S.B., Kumar, S. (Eds.), The Timetree of Life. Oxford Univ. Press, Oxford, UK, pp. 490–494.
Howell, A.B., Gersh, I., 1935. Conservation of water by the rodent *Dipodomys*. J. Mammal. 16, 1–9.
Huang, X.Q., Wang, J.M., Aluru, S., Yang, S.P., Hillier, L., 2003. PCAP: a whole-genome assembly program. Genome Res. 13, 2164–2170.
Huchon, D., Chevret, P., Jordan, U., Kilpatrick, C.W., Ranwez, V., Jenkins, P.D., Brosius, J., Schmitz, J., 2007. Multiple molecular evidences for a living mammalian fossil. Proc. Natl. Acad. Sci. U. S. A. 104, 7495–7499.
Hudson, M.E., 2008. Sequencing breakthroughs for genomic ecology and evolutionary biology. Mol. Ecol. Resour. 8, 3–17.
Ishibashi, K., Hara, S., Kondo, S., 2009. Aquaporin water channels in mammals. Clin. Exp. Nephrol. 13, 107–117.
Jones, W.T., 1984. Natal philopatry in bannertailed kangaroo rats. Behav. Ecol. Sociobiol. 15, 151–155.
Künster, A., Wolf, J.B., Backström, N., Whitney, O., Balakrishnan, C.N., Day, L., Edwards, S.V., Janes, D.E., Schlinger, B.A., Wilson, R.K., Jarvis, E.D., Warren, W.C., Ellegren, H., 2010. Comparative genomics based on massive parallel transcriptome sequencing reveals patterns of substitution and selection across 10 bird species. Mol. Ecol. 19, 266–276.
Mangrum, A.J., Gomez, R.A., Norwood, V.F., 2002. Effects of AT_{1A} receptor deletion on blood pressure and sodium excretion during altered dietary salt intake. Am. J. Physiol. Renal Physiol. 283, F447–F453.
Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L.I., Jarvie, T.P., Jirage, K.B., Kim, J., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., Rothberg, J.M., 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376–380.
Meredith, R.W., Janečka, J.E., Gatesy, J., Ryder, O.A., Fisher, C.A., Teeling, E.C., Goodbla, A., Eduardo, E., Simao, T.L.L., Stadler, T., Rabosky, D.L., Honeycutt, R.L., Flynn, J.J., Ingram, C.M., Steiner, C., Williams, T.L., Robinson, T.J., Burk-Herrick, A., Westerman, M., Ayoub, N.A., Springer, M.S., Murphy, W.J., 2011. Impacts of the cretaceous terrestrial revolution and KPg extinction on mammal diversification. Science 334, 521–524.
Meyer, E., Aglyamova, G.V., Wang, S., Buchanan-Carter, J., Abrego, D., Colbourne, J.K., Willis, B.L., Matz, M.V., 2009. Sequencing and *de novo* analysis of a coral larval transcriptome using 454 GSFlx. BMC Genomics 10, 219.
Murray, D., Doran, P., MacMathuna, P., Moss, A., 2007. *In silico* gene expression analysis —an overview. Mol. Cancer 6, 50.
Nachman, M.W., 2005. The genetic basis of adaptation: lessons from concealing coloration in pocket mice. Genetica 123, 125–136.
Nachman, M.W., Hoekstra, H.E., D'Agostino, S.L., 2003. The genetic basis of adaptive melanism in pocket mice. Proc. Natl. Acad. Sci. U. S. A. 100, 5268–5273.
Nielsen, R., 2005. Molecular signatures of natural selection. Annu. Rev. Genet. 39, 197–218.
Nielsen, R., Yang, Z., 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148, 929–936.
Nylander, J.A.A., 2004. MrAIC.pl. Program Distributed by the Author. Evolutionary Biology Centre, Uppsala University.
Orr, H.A., Coyne, J.A., 1992. The genetics of adaptation: a reassessment. Am. Nat. 140, 725–742.
Paul, M., Mehr, A.P., Kreutz, R., 2006. Physiology of local renin-angiotensin systems. Physiol. Rev. 86, 747–803.
Peart, W.S., 1965. The renin-angiotensin system. Pharmacol. Rev. 17, 143–182.
Renaut, S., Nolte, A.W., Bernatchez, L., 2010. Mining transcriptome sequences towards identifying adaptive single nucleotide polymorphisms in lake whitefish species pairs (*Coregonus* spp. Salmonidae). Mol. Ecol. 19, 115–131.
Schmidt-Nielsen, B., 1948. Rodent life in desert areas. Anat. Rec. 101, 696.
Schmidt-Nielsen, B., 1952. Renal tubular excretion of urea in kangaroo rats. Am. J. Physiol. 170, 45–56.
Schmidt-Nielsen, K., 1964. Desert Animals Physiological Problems of Heat and Water. Oxford Univ. Press, Oxford, UK.
Schmidt-Nielsen, B., O'Dell, R., 1961. Structure and concentrating mechanism in the mammalian kidney. Am. J. Physiol. 200, 1119–1124.
Schmidt-Nielsen, B., Schmidt-Nielsen, K., 1950. Evaporative water loss in desert rodents in their natural habitat. Ecology 31, 75–85.
Schmidt-Nielsen, K., Schmidt-Nielsen, B., 1952. Water metabolism of desert mammals. Physiol. Rev. 32, 135–166.
Schmidt-Nielsen, B., Schmidt-Nielsen, K., Brokaw, A., Schneiderman, H., 1948a. Water conservation in desert rodents. J. Cell. Comp. Physiol. 32, 331–360.
Schmidt-Nielsen, K., Schmidt-Nielsen, B., Brokaw, A., 1948b. Urea excretion in desert rodents exposed to high protein diets. J. Cell. Comp. Physiol. 32, 361–379.
Schmidt-Nielsen, K., Schmidt-Nielsen, B., Schneiderman, H., 1948c. Salt excretion in desert mammals. Am. J. Physiol. 154, 163–166.
Shapiro, M.D., Marks, M.E., Peichel, C.L., Blackman, B.K., Nereng, K.S., Jónsson, B., Schluter, D., Kingsley, D.M., 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. Nature 428, 717–723.
Simon, D.B., Karet, F.E., Hamdan, J.M., Di Pietro, A., Sanjad, S.A., Lifton, R.P., 1996a. Bartter's syndrome, hypokalaemic alkalosis with hypercalciuria, is caused by mutations in the Na-K-2Cl cotransporter NKCC2. Nat. Genet. 13, 183–188.
Simon, D.B., Nelson-Williams, C., Bia, M.J., Ellison, D., Karet, F.E., Molina, A.M., Vaara, I., Iwata, F., Cushner, H.M., Koolen, M., Gainza, F.J., Gitelman, H.J., Lifton, R.P., 1996b. Gitelman's variant of Bartter's syndrome, inherited hypokalaemic alkalosis, is caused by mutations in the thiazide-sensitive Na–Cl co-transporter. Nat. Genet. 12, 24–30.
Stinchcombe, J.R., Hoekstra, H.E., 2007. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity 100, 158–170.
Subramanian, A.R., Kaufmann, M., Morgenstern, B., 2008. DIALIGN-TX: greedy and progressive approaches for segment-based multiple sequence alignment. Algorithms Mol. Biol. 3, 6.
Tamura, K., Umemura, S., Sumida, Y., Nyui, N., Kobayashi, S., Ishigami, T., Kihara, M., Sugaya, T., Fukamizu, A., Miyazaki, H., Murakami, K., Ishii, M., 1998. Effect of genetic deficiency of angiotensinogen on the renin-angiotensin system. Hypertension 32, 223–227.
Tanaka, N., Babazono, T., Saito, S., Sekine, A., Tsunoda, T., Haneda, M., Tanaka, Y., Fujioka, T., Kaku, K., Kawamori, R., Kikkawa, R., Iwamoto, Y., Nakamura, Y., Maeda, S., 2003. Association of solute carrier family 12 (sodium/chloride) member 3 with diabetic nephropathy, identified by genome-wide analyses of single nucleotide polymorphisms. Diabetes 52, 2848–2853.
Therien, A.G., Goldshleger, R., Karlish, S.J.D., Blostein, R., 1997. Tissue-specific distribution and modulatory role of the γ subunit of the Na,K-ATPase. J. Biol. Chem. 272, 32628–32634.
Vimtrup, B., Schmidt-Nielsen, B., 1952. The histology of the kidney of kangaroo rats. Anat. Rec. 114, 515–528.
Vorhies, C.T., Taylor, W.P., 1922. Life history of the kangaroo rat, *Dipodomys spectabilis spectabilis* Merriam. U.S.D.A. Tech. Bullet. No. 1091, pp. 1–39.
Waser, P.M., Busch, J.D., McCormick, C.R., DeWoody, J.A., 2006. Parentage analysis detects cryptic pre-capture dispersal in a philopatric rodent. Mol. Ecol. 15, 1929–1937.
Weinman, E.J., Mohanlal, V., Stoycheff, N., Wang, F., Steplock, D., Shenolikar, S., Cunningham, R., 2006. Longitudinal study of urinary excretion of phosphate, calcium, and uric acid in mutant NHERF-1 null mice. Am. J. Physiol. Renal Physiol. 290, F838–F843.
Wheat, C.W., 2010. Rapidly developing functional genomics in ecological model systems via 454 transcriptome sequencing. Genetica 138, 433–451.
Yang, Z., 1997. Paml: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13, 555–556.
Yang, Z., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.
Yang, Z., Nielsen, R., Goldman, N., Pedersen, A.K., 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155, 431–449.
Zhu, Y.Y., Machleder, E.M., Chenchik, A., Li, R., Siebert, P.D., 2001. Reverse transcriptase template switching: a SMART (TM) approach for full-length cDNA library construction. Biotechniques 30, 892–897.