MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2013) doi: 10.1111/1755-0998.12092

Transcriptome sequencing and microarray development for the woodrat (*Neotoma spp.*): custom genetic tools for exploring herbivore ecology

J. R. MALENKE,* B. MILASH,† A. W. MILLER* and M. D. DEARING*
*Department of Biology, University of Utah, 257 S. 1400 E., Salt Lake City, UT 84112, USA, †Bioinformatics Shared Resource, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT 84112, USA

Abstract

Massively parallel sequencing has enabled the creation of novel, in-depth genetic tools for nonmodel, ecologically important organisms. We present the *de novo* transcriptome sequencing, analysis and microarray development for a vertebrate herbivore, the woodrat (*Neotoma spp.*). This genus is of ecological and evolutionary interest, especially with respect to ingestion and hepatic metabolism of potentially toxic plant secondary compounds. We generated a liver transcriptome of the desert woodrat (*Neotoma lepida*) using the Roche 454 platform. The assembled contigs were well annotated using rodent references (99.7% annotation), and biotransformation function was reflected in the gene ontology. The transcriptome was used to develop a custom microarray (eArray, Agilent). We tested the microarray with three experiments: one across species with similar habitat (thus, dietary) niches, one across species with different habitat niches and one across populations within a species. The resulting one-colour arrays had high technical and biological quality. Probes designed from the woodrat transcriptome performed significantly better than functionally similar probes from the Norway rat (*Rattus norvegicus*). There were a multitude of expression differences across the woodrat treatments, many of which related to biotransformation processes and activities. The pattern and function of the differences indicate shared ecological pressures, and not merely phylogenetic distance, play an important role in shaping gene expression profiles of woodrat species and populations. The quality and functionality of the woodrat transcriptome and custom microarray suggest these tools will be valuable for expanding the scope of herbivore biology, as well as the exploration of conceptual topics in ecology.

Keywords: gene expression, herbivore, microarray, next-generation sequencing, transcriptome, vertebrate

Received 9 October 2012; revision received 31 January 2013; accepted 5 February 2013

Introduction

Gene expression patterns are an important component of phenotype and may be an adaptive element linking an organism to its variable environment (King & Wilson 1975; Ellegren & Sheldon 2008). Because many gene expression analyses require extensive genetic knowledge, in-depth analyses of gene expression patterns have been restricted historically to model systems and humans. Indeed, our understanding of gene expression as an adaptive component of an organism's biology has almost certainly been shaped by the biology of the limited number and taxonomy of model systems in which expression has been well studied (King & Wilson 1975; Ellegren & Sheldon 2008). On the other hand, ecologically important organisms are likely to have experienced more complex selective landscapes than laboratory models and to have responded over time with similarly complex adaptive strategies and phenotypic diversity. Exploring gene expression profiles in nonmodel, outbred organisms reacting to ecologically relevant challenges is the next step to understanding mechanisms of biodiversity as well as the coadaptive interactions that occur between species in natural communities.

Next-generation sequencing technologies allow ecologists and others with genetically under-studied organisms to begin to probe the genetic basis of ecological phenomena without requiring a huge amount of a priori genetic knowledge (Rokas & Abbot 2009; Wheat 2010). Next-generation sequencing has become more affordable, reliable and rapid. However, to be of greatest utility, this method requires a genomic (or transcriptomic) scaffold to which reads can be mapped and can be limited by transcript coverage, because many of the non-normalized reads will be for rRNA subunits.

Correspondence: Jael R. Malenke, Fax: (801) 581 4668; E-mail: malenke@biology.utah.edu

One way around these limitations is to use next-generation sequencing data to build a custom gene expression microarray (Garcia-Reyero *et al.* 2008; Vera *et al.* 2008; Bellin *et al.* 2009; Babik *et al.* 2010). Custom microarrays have an advantage in that they do not need an extensive genomic scaffold and can be designed to focus on transcripts or hypotheses of interest. Several instructive and interesting illustrations of this approach have been published recently (Vera *et al.* 2008; Babik *et al.* 2010; Wheat *et al.* 2011). For example, the heart tissue transcriptome of the bank vole (*Myodes glareolus*) was sequenced with next-generation technology (Babik *et al.* 2010). By taking a high-throughput sequencing approach, Babik *et al.* (2010) were able to describe 90–95% of heart-expressed genes in voles. The vole heart transcriptome microarray that resulted from this sequencing effort is a useful tool for investigating the relationship between heart function and aerobic metabolism.

Woodrats (*Neotoma* spp) are an excellent system for the implementation of functional genomics tools to explore ecological and evolutionary questions. Woodrats are herbivorous rodents distributed across North America, and their diets differ across habitats (Patton *et al.* 2008). Different species, and populations within a species, consume unique plant diets, some of which are heavily defended by plant secondary compounds. This diversity of diets has made woodrats an important ecological study system for examining how mammalian herbivores deal with plant defence (Lamb *et al.* 2004; Haley *et al.* 2007a; Magnanou *et al.* 2009; Skopec & Dearing 2011). In addition, recent efforts have begun to explore the molecular basis of detoxification in woodrats using tools such as heterologous microarrays. Woodrat mRNA has been hybridized to Norway rat chips (Skopec *et al.* 2007; Magnanou *et al.* 2009). In these cross-platform studies, the woodrat RNA hybridized to approximately 70% of the probes for Norway rat (*Rattus norvegicus*) and revealed patterns of differential expression across treatments. Finally, woodrats are a good candidate for a next-generation sequencing effort because there is a clear target tissue to assay for studies related to diet. Much of the metabolism of plant secondary compounds occurs through biotransformation pathways in the liver (Klaassen 2001), and other woodrat studies have revealed differences in expression and activity of specific liver biotransformation enzymes or enzyme classes with respect to diet (Haley *et al.* 2007a,b, 2008). By concentrating on the liver transcriptome, there is a limit to the number of transcripts expressed, ensuring better coverage of existing messages (Babik *et al.* 2010; Wheat 2010).

In this study, we describe the *de novo* sequencing, transcriptome analysis and subsequent creation of a gene expression microarray customized for the woodrat liver. We assessed the value of the custom array by comparing the performance of the homologous probes (from the *Neotoma* transcriptome) with that of the heterologous probes (from the Norway rat catalogue). The functionality of the woodrat array was tested by comparing gene expression patterns at two levels of taxonomic comparisons: interspecific comparisons between *Neotoma lepida* and two populations of *Neotoma bryanti*, and an intraspecific comparison of the two populations of *N. bryanti*.

Methods

Hepatic transcriptome sequencing and annotation

A single representative of *N. lepida* (female, mass: 103.4 g; captured near Barstow, CA, USA) was selected for 454 sequencing. Four days prior to dispatch, foliage from creosote bush (*Larrea tridentata*) was offered *ad lib* in addition to rabbit chow (Harlan Teklad 2031). Creosote foliage is a natural part of the diet of *N. lepida* near Barstow (Karasov 1989) and the animal readily consumed the offered foliage. One day after creosote foliage, the animal was given a single injection (IP) of Aroclor 1254 (1 mg/kg in 1 mL corn oil). Aroclor is a mixture of >200 polychlorinated biphenyls (Frame *et al.* 1996). It causes long-lasting induction of hepatic enzymes (Beebe *et al.* 1992) and is often used to induce hepatic enzymes in pharmacological studies (Richardson & Klaassen 2010). Dispatch of the animal was delayed for 3 days post-injection to permit adequate time for induction of detoxification enzymes (Pratt & Taylor 1990). The animal maintained mass over this time (final body mass: 106.8 g). The animal was dispatched, the liver removed and a 200 μg section was treated with RNA later.

RNA was extracted from liver tissue using RNeasy (Qiagen). Construction of the normalized cDNA library and sequencing was conducted at the Purdue University Genomics Core Facility using a protocol previously described (Meyer *et al.* 2009) modified to incorporate Titanium pA and pB priming sites in the final library molecules. Reads from the GS-FLX instrument were assembled into contigs using the Roche GS De Novo Assembler software. Contigs were annotated using TBLASTX against sequences from the REFSEQ database. We compiled a total of 74504 sequences in the reference database from the following species sources: *Rattus norvegicus*: 30785; *Mus musculus*: 42760; *Peromyscus* and other rodent genes: 959. We used the reciprocal best BLAST hit method to annotate the transcriptome, that is, each *Neotoma* contig was compared with the known rodent sequence database using TBLASTX. The best hit to the database was recorded (along with per cent identity, length of the region of similarity, *e*-value and number of bits), and that best hit was compared back to the *Neotoma*

contigs. A contig was annotated as the 'mutual best hit' to a rodent sequence if that contig was the top hit to the rodent sequence and *vice versa*. It was 'similar to' if the reference sequence was a best hit for the *Neotoma* contig, but not the reverse. To estimate number of unique genes and the coverage of the rat transcriptome by the woodrat contigs, the woodrat contigs were blasted against the 30 785 rat RNA sequences used in the annotation database and vice versa, and the number of matches tallied. Matches were those BLAST hits with *e*-values ≤ 0.1.

We used the GOSlim function in Blast2Go to generate Gene Ontology (GO) terms to describe the transcriptome. GOSlim uses a standardized set of terms from the Gene Ontology project and, therefore, permits comparisons across different data sets.

Custom microarray design

The custom *Neotoma* microarray was designed using eArray (Agilent). An 8 × 15 K array was implemented (15 744 total features), with the Agilent control grid consisting of 536 spots with positive and negative controls. The array was filled mostly with probes designed from woodrat transcriptome contigs using the 'GE Probe Design' tool with 'base composition methodology'. The entire transcriptome was uploaded to eArray to evaluate quality and cross-hybridization potential of the probes.

Four probe groups were designed from the targeted sets of woodrat contigs. The target contigs for the first probe group (Table 1) were contigs whose annotation matched a list of biotransformation genes ($n = 297$) extracted from heterologous woodrat-to-Norway rat arrays (Skopec *et al.* 2007; Magnanou *et al.* 2009). For this probe group alone, two probes per target sequence were designed using the 'best distribution methodology'. After these detoxification contigs were removed, the remaining contigs of the transcriptome were evaluated for their proximity to the 3′ end of the rodent message to which they were most homologous. Contigs that were not within 800 bp of the 3′ end of the annotated message were excluded. The target contigs for the second probe group were all contigs remaining that had $\geq 90\%$ sequence identity with their BLAST match in the region of highest homology (Table 1). The third probe group was created from those target contigs that had 80–89.99% sequence identity with their BLAST match in the region of highest homology and the fourth probe group from target contigs with 75–79.99% sequence identity with their BLAST match. One probe was designed for each contig in the second, third and fourth groups (Table 1). All probes with a base composition value =4 were excluded, which was <1% of all probes in each group.

Table 1 *Neotoma* custom expression microarray probe groups

Probe group names	Target contigs	Probes (final)	Replicated on Array
Woodrat biotransformation	480	(2×) 940	3
Woodrat 90–100% matches	2394	2389	2
Woodrat 80–89% matches	2085	2076	2
Woodrat 75–79% matches	881	881	1+*
Norway Rat detoxification metabolism	n.a.	374	3
Norway Rat oxidative stress	n.a.	288	2

*The 75–79% woodrat group used to fill in the remaining spots, and as a result, all probes in it were spotted once, and 649 of the 881 were spotted twice.

To compare heterologous to homologous binding efficiency, two additional probe groups were created from existing probes in the Agilent Norway rat catalogue. One consisted of probes related to liver detoxification pathways. It included the biotransformation features extracted from the Norway rat arrays used in previous cross-species hybridization arrays (Skopec *et al.* 2007; Magnanou *et al.* 2009), and probes returned from a GO search of Norway rat probes, using the GO terms 'drug metabolic process' (GO:0017144), 'xenobiotic metabolic process' (GO:0006805), 'superoxide metabolic process' (GO:0006801), 'nicotine metabolic process' (GO:0018933), 'quinolinate metabolic process' (GO:0046874). The second group consisted of Norway rat probes returned from a GO search using the term 'oxidative stress response' (GO:0006979).

A replicated noncontrol probe group was created from 23 probes from the 90% to 100% group and spotted on the array 10 times. Other probe groups were also spotted multiple times (Table 1). All probes were 60-m. Feature layout was randomized by eArray.

Woodrat trapping and feeding trial

To validate the sensitivity of the custom array, we designed experiments to compare expression across *Neotoma* species and populations. Desert woodrats (*N. lepida*) were trapped at Lytle Ranch Preserve in southern Utah (lat: 37.117514, long: −114.009661 Washington Co., UT, USA). Bryant's woodrats (*N. bryanti*) were trapped from two locations in California: Bryant's woodrats were trapped near Palm Springs, CA, USA (lat: 33.679616, long: −116.362018, Riverside Co., CA, USA) and from Ronald W. Caspers Wilderness Park (lat: 33.53367, long: −117.54965, Orange Co., CA, USA). Trapping locations

were determined using trapping records and species taxonomy from Patton *et al.* (2008). Lytle Ranch (LR) *N. lepida* and Palm Springs (PS) *N. bryanti* include creosote bush as a dietary component (Karasov 1989; Mangione *et al.* 2001). The Caspers Wilderness (CW) *N. bryanti* include oak (*Quercus agrifolio*) foliage in their diet and have no exposure to creosote bush. Animals were housed individually in the animal facilities at University of Utah and were given food and water *ad libitum* (Harlan Teklad formula 2031, rabbit chow) prior to the feeding trial. Experimental procedures were approved by the Institutional Animal Care and Use Committee (IACUC #07-02015).

Prior to the microarray trial, four animals from each of the three populations were fed Harlan Teklad 2031 for 8 days. The diet was prepared by mixing powered rabbit chow with acetone equal to 25% of the weight of the chow. To remove the acetone, the diet was dried under a fume hood for 24 h followed by vacuum for 1 h. Complete evaporation was confirmed gravimetrically. The addition of acetone served as a control reagent for other treatment diets that have plant secondary compounds added in acetone. Animals had *ad libitum* access to water and food. Leftovers were collected, additional diet was presented and animals were weighed, daily.

After the feeding trial, animals were dispatched using CO_2 asphyxiation; liver tissue and spleens were preserved in RNAlater (Sigma) following the manufacturer's protocol and frozen. Frozen tissues were kept at −80 °C. For the microarray analysis, RNA was extracted (RNAqueous) from the livers and sent to the University of Utah Microarray Core facility.

Microarray protocol

For each sample, the quality and concentration of the RNA was measured with an Agilent Bioanalyzer RNA Nanochip. The Agilent One-Color Quick Amp Labeling Kit was used to generate fluorescently labelled cRNA for one-colour microarray hybridizations. Agilent RNA spike-in controls were combined with input total RNA samples (50–1000 ng). The polyadenylated fraction of the total RNA sample was primed with oligo dT/T7 RNA polymerase promoter oligonucleotide sequences and cDNA synthesis accomplished by MMLV-RT. Following cDNA synthesis, T7 RNA polymerase and cyanine 3-CTP nucleotides were combined with the reaction mixture to simultaneously amplify the target material through the generation of cRNA and incorporate cyanine 3-CTP. Fluorescently labelled cRNA molecules were purified from the reaction mixture using the Qiagen RNeasy mini kit. The concentration of the purified samples was determined using a NanoDrop ND-1000 spectrophotometer.

Fluorescently labelled cRNA samples were fragmented and combined with Agilent Hi-RPM hybridization buffer. Microarray hybridizations were performed using Agilent SureHyb and incubated at 65 °C for 17 h, 10 r.p.m. Following incubation, microarray slides were washed for 1 min in each of the two wash buffers (6× SSPE, 0.005% *N*-lauroylsarcosine and 0.06× SSPE, 0.005% *N*-lauroylsarcosine), dipped in a solution of acetonitrile and dried.

Microarray slides were scanned in an Agilent Technologies G2505C Microarray Scanner at 5 μm resolution. Data captured from the scanned microarray image were loaded into Agilent Feature Extraction software version 10.5.1.1. The software automatically positions a grid and finds the centroid positions of each feature on the microarray. This information was used to calculate feature intensities, background measurements and spot statistics. Agilent's Feature Extraction software was used for spatial detrending.

Analysis of microarray performance

Prior to analysis, all control spots, nonuniform spots and population outlier spots were removed from the data set; intensity values were $\log_2$-transformed using Agilent Filter, software designed to simplify the processing of Agilent data (http://bioserver.hci.utah.edu/BioInfo/index.php/Software:AgilentFilter). Data from duplicate probes were combined, but different probes with the same annotation were maintained separately because it is possible that the original contigs were from different isoforms despite the shared match to a rodent gene.

In the resulting data set, we compared the performance of the microarray probes created from the *N. lepida* liver transcriptome to those probes mined from the Agilent Norway rat catalogue using the four LR *N. lepida*. Only the *N. lepida* samples were used because the original transcriptome was created from an *N. lepida* individual. We compared the intensities and qualities in several different ways. First, the intensities and qualities for each probe were averaged across the four *N. lepida* samples and then the averages for *Rattus* (n = 640) and *Neotoma* (n = 6286) probes were compared using a t-test. Because of the differences in probe number and the potentially important difference in function of those probes, a second analysis was performed limited to those probes with identical Genbank accession numbers in both the *Rattus* and *Neotoma* probe sets (a result of annotating the *Neotoma* transcriptome with rodent reference sequences). Where there were multiple probes for a given accession from either rat or woodrat, the intensities and quality values were averaged. The resulting 60 matched pairs of *Rattus* and *Neotoma* probes were compared using a paired t-test. For an interspecific

evaluation of the custom array, the LR *N. lepida* samples were compared with the PS *N. bryanti* samples using all *Neotoma* probes or the 60 woodrat-derived probes from the previous analysis. Intensities and quality values were analysed with a paired *t*-test.

To compare gene expression profiles across woodrat populations and species, the array data were batch uploaded to Genesifter 3.7. Due to significant differences in the probe performance, only the woodrat-derived probes were included. Prior to comparing expression profiles, the consistency of transcriptome response was evaluated by comparing overall gene expression profiles across individuals. In Genesifter, after normalizing the intensity data, individuals were clustered by gene expression using all woodrat-derived probes (n = 6286). Clustering parameters were distance/correlation, linkage/average, row centered/by genes.

Interspecific and intraspecific comparisons of woodrat gene expression

In Genesifter, three pairwise comparisons were created. The first was an interspecific comparison between LR *N. lepida* and PS *N. bryanti*. Both populations consume creosote and are from desert regions (Mojave and Sonoran, respectively), and thus might be expected to face similar ecological conditions. The second comparison was between LR *N. lepida* and CW *N. bryanti*, a population that inhabits ecologically distinct coastal woodland areas (Atsatt & Ingram 1983). The final comparison was between the PS and CW *N. bryanti* populations that consume different diets in different habitats (Atsatt & Ingram 1983). In all comparisons, the quality requirement was set to 1, and all individuals from both treatments were required to pass. The fold threshold for differences in gene expression was 2, and a *t*-test was performed (α = 0.05). The Benjamini & Hochberg (1995) false discovery rate was applied to the resulting significant gene list. Treatments were compared by examining the number and annotation of upregulated probes as well as the KEGG pathways and Gene Ontology terms represented by those probes. *Z*-scores were used to determine significant associations in these functional assays. Pathways or ontology terms with *z*-scores of >2 are considered to be significantly overrepresented in the results.

qPCR validation of microarray results

To validate the microarray results, qPCR was used to determine the relative expression of three genes: P450 (cytochrome) oxidoreductase (POR) and catechol-*O*-methyltransferase (COMT), which exhibited significant differences between specific pairs of populations, and superoxide dismutase (SOD), which exhibited no difference in expression. These genes were chosen because they are important in liver function and have been of significance in other studies examining woodrat detoxification strategies. β-actin was used as an endogenous control. For additional methods, please see the Supporting information.

Results

Transcriptome assembly and analysis

We obtained a total of 763 893 high-quality reads from the normalized *N. lepida* sample after 1.75 runs on a 454 platform. During assembly, 501 299 reads were fully incorporated into contigs, and an additional 72 341 reads were only partially incorporated into contigs. Average read length was 334 bp. The total number of contigs was 18 316 with an average size of 694 bp. Over 5000 contigs were between 1107 and 5601 bp in length (Fig. S1, Supporting information). Of the total number of contigs, 44.6% (8174) produced a mutual best hit from our reference database of rat, mouse, *Peromyscus* and other rodent sequences. Another 55% (10 076) were 'similar to' rodent reference sequences. Only 66 contigs did not blast to any known sequences.

By blasting the woodrat contigs against the rat RNA database, we found 9163 contigs had hits to 10 596 rat genes (34.4% coverage). The reverse BLAST search (rat genes against woodrat contigs), 10 603 rat genes (34.4%) had hits to 9173 contigs.

With respect to GO terms within biological process (Level 2), the woodrat liver transcriptome was characterized by terms related to cellular and metabolic processes and biological regulation (Fig. 1). Terms included in 'response to stimulus' were also abundant and comprised >10% of the terms in the transcriptome. This category includes terms related chemical stimulus such as drugs, toxins and other xenobiotics. With respect to molecular function (Level 2), the majority of GO terms were related to 'binding', which includes toxin binding (Fig. 2). 'Catalytic activity' represented more than 25% of all terms and includes activities such as oxidoreductase and hydrolase.

Analysis of microarray performance

Two of the 12 arrays failed to pass all 9 of Agilent's quality metrics. The two imperfect arrays passed seven and eight of nine quality metrics, respectively. On average, <0.2% of the features (spots) across all the arrays were flagged as nonuniform features.

The performance of probes designed from the *Neotoma* hepatic transcriptome was significantly better than that from the Agilent Norway Rat catalogue when

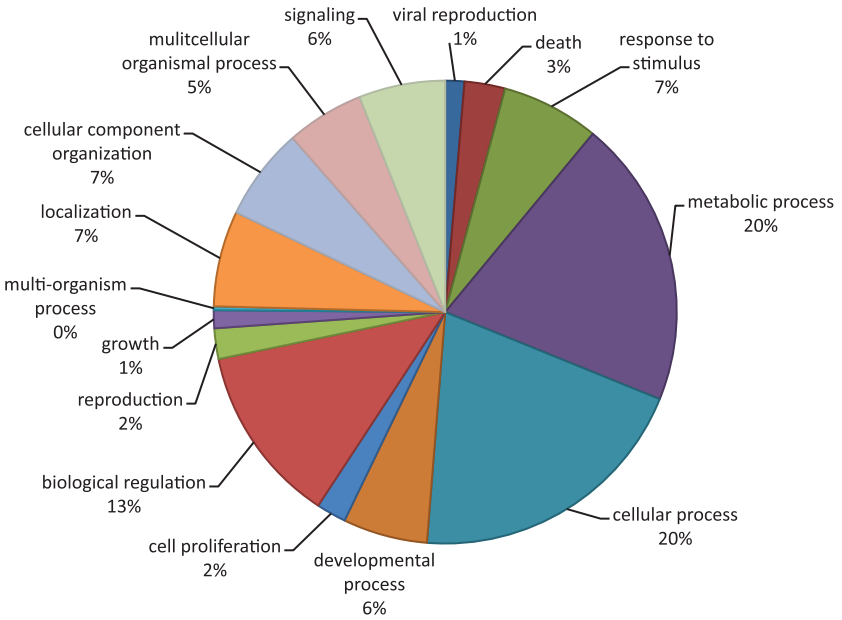

Fig. 1 Woodrat transcriptome gene ontology terms from level 2 of the biological process category. Percentage of transcriptome contigs comprising each term shown adjacent to term.

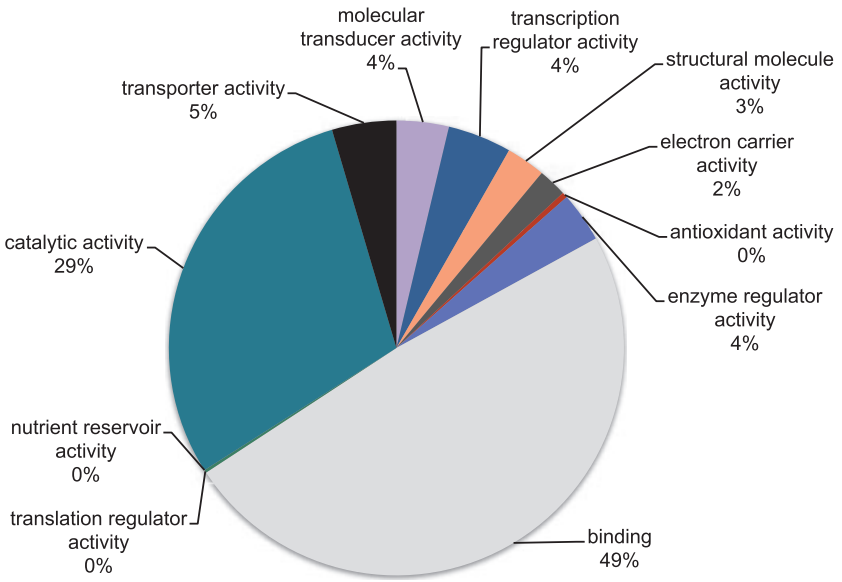

Fig. 2 Woodrat transcriptome gene ontology terms from level 2 of the molecular function category. Percentage of transcriptome contigs comprising each term shown adjacent to term.

exposed to *Neotoma* samples. Over the entire array, the probes from the Agilent Norway Rat catalogue had average intensity values 38% lower than the *Neotoma* probes and quality values 40% lower (Table 2A). A similar pattern was found in the paired comparison between *Rattus* and *Neotoma* probes annotated to the same Genbank reference sequence. The per cent reductions were 37% and 35% for intensity and quality, respectively (Table 2A).

There were also significant differences in array performance between samples from *N. lepida*, the species from which the original transcriptome was sequenced, and those from *N. bryanti*, a closely related woodrat species. The *N. bryanti* samples had significantly higher intensity and quality values than the *N. lepida* samples (Table 2B). However, the differences were much smaller in this comparison relative to the *Neotoma* versus *Rattus* comparisons (Table 2A); the intensity of *N. lepida* samples were 5% lower across 60 matched probes than *N. bryanti*, and the quality values were, on average, 6% lower.

Using all the custom probes, woodrat samples were clustered according to overall gene expression profile, including probes that were not significantly different in any analyses. Individuals grouped by treatment, indicating shared responses within a treatment relative to between treatments (Fig. 3).

Table 2 Probe performance on the custom woodrat microarray

(A) Comparison between probes derived from the *Neotoma lepida* hepatic transcriptome and the Agilent Norway Rat catalogue				
Probe origin comparison	*Neotoma* (±SD)	Rattus (±SD)	% Reduction	
All probes				
Intensity	10.00 ± 3.81	6.14 ± 3.86	38.63	d.f. = 771.6, $t = -24.19$, $P < 0.0001$
Quality	0.864 ± 0.295	0.514 ± 0.441	40.51	d.f. = 698.2, $t = -19.62$, $P < 0.0001$
Matched probes				
Intensity	11.59 ± 3.77	7.31 ± 4.20	36.91	d.f. = 59, $t = -7.27$, $P < 0.0001$
Quality	0.899 ± 0.247	0.584 ± 0.429	35.04	d.f. = 59, $t = -5.41$, $P < 0.0001$
(B) Comparison between *N. lepida* and *Neotoma bryanti* samples on a small subset of detoxification-related probes of *N. lepida* origin				
Interspecific comparison	LR *N. lepida*	PS *N. bryanti*	% Reduction	
All probes				
Intensity	10.00 ± 3.81	10.48 ± 3.80	4.58	d.f. = 6279, $t = 52.37$, $P < 0.0001$
Quality	0.864 ± 0.295	0.945 ± 0.203	8.57	d.f. = 6279, $t = 31.96$, $P < 0.0001$
Matched probes				
Intensity	11.59 ± 3.77	12.20 ± 3.71	5.03	d.f. = 59, $t = 5.27$, $P < 0.0001$
Quality	0.899 ± 0.247	0.955 ± 0.159	5.86	d.f. = 59, $t = 2.75$, $P < 0.01$

LR, Lytle Ranch; PS, Palm Springs.

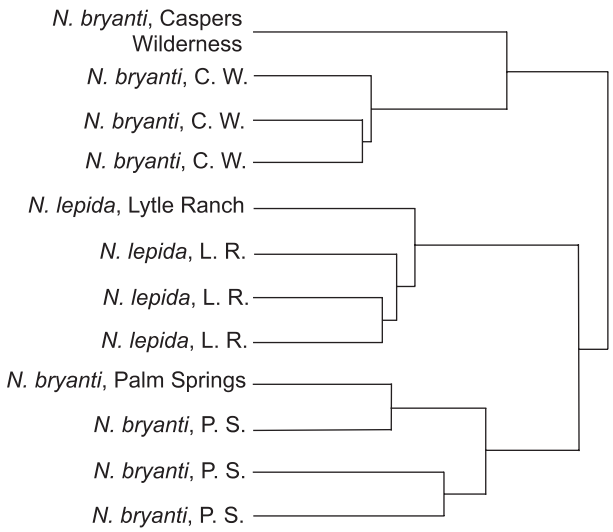

Fig. 3 Cluster analysis of individual woodrat samples calculated from overall gene expression profiles. Individuals are specified by species and population designations.

Interspecific and intraspecific comparisons of woodrat gene expression

In the comparison of ecologically similar LR *N. lepida* and PS *N. bryanti*, there were 191 probes with significantly different expression. Ninety-three probes were upregulated in the LR animals (i.e. downregulated in PS), and 98 probes were upregulated in the PS animals (Table S1, Supporting information). The difference in expression for the most upregulated probes in *N. lepida* was 10-fold; the top probe in *N. bryanti* was upregulated 35-fold.

When the upregulated genes (not probes) in each treatment were categorized by KEGG pathway, there were fewer significantly represented KEGG pathways in *N. lepida* than in *N. bryanti* (Table 3). The top two KEGG pathways (by gene numbers) upregulated in *N. bryanti* were 'drug metabolism—cytochrome P450' and 'metabolism of xenobiotics by cytochrome P450', possibly indicating overlapping detoxification genes. KEGG pathways may have only one characteristic gene upregulated and still be significantly better represented than expected by chance. Likewise, a single upregulated gene may co-occur in multiple pathways. The best-represented pathways in *N. bryanti* were different than *N. lepida* ('chemokine signalling pathway' and 'osteoclast differentiation').

Significant GO terms were assembled from each of the three main categories: biological processes, cellular component and molecular function. Within a category, terms across the tiers were first filtered to include only those with significant, positive *z*-scores (>2) and ordered by the number of genes upregulated in each category. GO terms were kept only if they had ≥ 10 genes contributing to their overrepresentation in a category. This method revealed more significant GO terms, as well as biotransformation functionality, associated with PS woodrats than LR woodrats (Table S2, Supporting information).

There were 354 probes differently expressed in the ecologically distinct, interspecific comparison between LR *N. lepida* and CW *N. bryanti*. About 240 were upregulated in *N. lepida*; 154 were upregulated in *N. bryanti*

Table 3 Overrepresented KEGG pathways: (A) Lytle Ranch (LR) *Neotoma lepida* (B) Palm Springs (PS) *Neotoma bryanti*

KEGG pathway	No. genes	Total on Array	*Z*-score
(A) Upregulated in LR *N. lepida*			
Chemokine signalling pathway	3	35	2.77
Osteoclast differentiation	3	28	3.29
Bladder cancer	2	16	2.99
Chagas disease (American trypanosomiasis)	2	27	2
Cysteine and methionine metabolism	2	23	2.28
Toll-like receptor signalling pathway	2	21	2.45
Galactose metabolism	1	7	2.31
Maturity onset diabetes of the young	1	5	2.86
Olfactory transduction	1	7	2.31
Pantothenate and CoA biosynthesis	1	7	2.31
Primary immunodeficiency	1	7	2.31
(B) Upregulated in PS *N. bryanti*			
Drug metabolism—cytochrome P450	8	66	6.56
Metabolism of xenobiotics by cytochrome P450	5	54	4.28
Protein processing in endoplasmic reticulum	5	69	3.56
Steroid hormone biosynthesis	4	26	5.35
Circadian rhythm—mammal	3	9	7.25
Starch and sucrose metabolism	3	23	4.16
Tryptophan metabolism	3	28	3.65
Ascorbate and aldarate metabolism	2	19	2.94
Butanoate metabolism	2	17	3.17
Fat digestion and absorption	2	10	4.42
Glyoxylate and dicarboxylate metabolism	2	13	3.77
Lysine degradation	2	30	2.08
Other types of O-glycan biosynthesis	2	15	3.44
Pentose and glucuronate interconversions	2	21	2.74
Porphyrin and chlorophyll metabolism	2	26	2.33
Propanoate metabolism	2	29	2.14
Pyruvate metabolism	2	27	2.26
Synthesis and degradation of ketone bodies	2	5	6.54
Terpenoid backbone biosynthesis	2	7	5.43
Tyrosine metabolism	2	22	2.65

KEGG pathways assembled from lists of upregulated genes ordered by the number of gene representatives in each pathway with a significant *z*-score (>2). To limit comparisons, the number of pathways was capped at 20.

(Table S3, Supporting information). The top differentially expressed probe in *N. lepida* was upregulated by 1663-fold; the top probe in *N. bryanti* was upregulated by 131-fold.

There were relatively few KEGG pathways overrepresented in the *N. lepida*, and those reflected a number of metabolic processes, including 'drug metabolism—other enzymes' (Table 4A). There were more KEGG pathways for *N. bryanti*, a number of which had biotransformation functions ('drug metabolism—cytochrome P450', 'metabolism of xenobiotics' and 'drug metabolism—other enzymes'; Table 4B). In *N. lepida,* metabolic processes played a large role in the biological process GO terms, whereas transferase activities played a dominant role in the molecular function terms. In *N. bryanti*, response to stimulus terms were overrepresented in biological process, and terms related to membranes were common in the cellular component category (Table S4, Supporting information).

There were 100 probes differently expressed in the intraspecific comparison between PS and CW *N. bryanti*. Seventy-two were upregulated in the PS animals; 28 upregulated in the CW animals (Table S5, Supporting information). The most upregulated probe in the PS animals had 62-fold higher expression than the CW equivalent. The top probe for CW was 37-fold upregulated. The best-represented KEGG pathways for PS *N. bryanti* were 'drug metabolism—cytochrome P450' and 'metabolism of xenobiotics by cytochrome P450', similar to the results in the interspecific comparison (Table 5). The paucity of genes in the KEGG pathways for the CW animals likely reflects the smaller number of genes that responded positively in these animals. This was also the case for GO terms; no terms passed the filtering requirements for CW woodrats (Table S6, Supporting information).

qPCR validation of microarray

The qPCR results for COMT, POR and SOD corroborated the microarray data (Fig. S2, Supporting information). For each gene, we chose a specific pair of populations to compare. Consistent with the microarray results (Table S1 and Table S3 respectively, Supporting information), the COMT gene was expressed higher by PS *N. bryanti* than by LR *N. lepida* (2.95-fold), and the POR gene was expressed higher by CW *N. bryanti* than by LR *N. lepida* (3.04-fold). SOD exhibited no differences between the CW and PS *N. bryanti*. Data were submitted to Dryad database (Malenke *et al*. 2013).

Discussion

Woodrats (*Neotoma spp*) are ecologically interesting animals that have been used to inform topics from paleoclimate

Table 4 Overrepresented KEGG pathways: (A) Lytle Ranch (LR) *Neotoma lepida* and (B) Casper's Wilderness (CW) *Neotoma bryanti*

KEGG pathway	No. genes	Total on Array	*Z*-score
(A) Upregulated on LR *N. lepida*			
Arachidonic acid metabolism	6	29	4.62
Drug metabolism—other enzymes	5	39	2.84
Glycine, serine and threonine metabolism	5	29	3.67
Pyrimidine metabolism	4	34	2.33
Nitrogen metabolism	2	13	2.1
Lysine biosynthesis	1	3	2.6
(B) Upregulated on CW *N. bryanti*			
Drug metabolism—cytochrome P450	8	66	3.73
Tryptophan metabolism	7	28	6.06
Protein processing in endoplasmic reticulum	6	69	2.27
Fatty acid metabolism	5	37	3.22
Metabolism of xenobiotics by cytochrome P450	5	54	2.22
Drug metabolism—other enzymes	4	39	2.22
Retinol metabolism	4	42	2.05
Steroid hormone biosynthesis	4	26	3.21
Bile secretion	3	28	2
Butanoate metabolism	3	17	3.09
Circadian rhythm—mammal	3	9	4.75
Pyruvate metabolism	3	27	2.08
Starch and sucrose metabolism	3	23	2.41
Fat digestion and absorption	2	10	2.76
Glyoxylate and dicarboxylate metabolism	2	13	2.26
Nonhomologous end-joining	2	6	3.88
Other types of O-glycan biosynthesis	2	15	2
Synthesis and degradation of ketone bodies	2	5	4.33
Terpenoid backbone biosynthesis	2	7	3.52
Cyanoamino acid metabolism	1	2	3.49

KEGG pathways assembled from lists of upregulated genes ordered by the number of gene representatives in each pathway with a significant *z*-score (>2). To limit comparisons, the number of pathways was capped at 20.

reconstruction (Nowak *et al.* 1994; Jackson *et al.* 2005), population genetic structure (Matocq 2002; Mendez-Harclerode *et al.* 2007) and behavioural strategies for dealing with plant secondary compounds (Torregrossa *et al.* 2012). In this study, we contribute transcriptomic and gene expression tools. The individual reads of the woodrat transcriptome were relatively long and

Table 5 Overrepresented KEGG pathways for (A) Palm Springs (PS) and (B) Casper's Wilderness (CW) *Neotoma bryanti*

KEGG pathway	No. genes	Total on Array	*Z*-score
(A) Upregulated in PS *N. bryanti*			
Drug metabolism—cytochrome P450	4	66	3.07
Metabolism of xenobiotics by cytochrome P450	3	54	2.46
Arachidonic acid metabolism	2	29	2.39
Ascorbate and aldarate metabolism	2	19	3.22
Lysine degradation	2	30	2.32
Other types of *O*-glycan biosynthesis	2	15	3.75
Pentose and glucuronate interconversions	2	21	3.01
Porphyrin and chlorophyll metabolism	2	26	2.59
Starch and sucrose metabolism	2	23	2.83
Steroid hormone biosynthesis	2	26	2.59
Fat digestion and absorption	1	10	2.2
Fatty acid elongation in mitochondria	1	8	2.54
Intestinal immune network for IgA production	1	11	2.06
Lysine biosynthesis	1	3	4.51
Renin-angiotensin system	1	6	3.04
Synthesis and degradation of ketone bodies	1	5	3.38
Terpenoid backbone biosynthesis	1	7	2.76
(B) Upregulated in CW *N. bryanti*			
Regulation of actin cytoskeleton	2	54	2.76
Amoebiasis	1	20	2.37
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	1	16	2.73
Cell adhesion molecules (CAMs)	1	16	2.73
Cytosolic DNA-sensing pathway	1	21	2.3
Dilated cardiomyopathy	1	11	3.42
DNA replication	1	11	3.42
ECM–receptor interaction	1	6	4.79
Glycerophospholipid metabolism	1	22	2.23
Glycosaminoglycan biosynthesis—heparan sulphate	1	12	3.25
Hypertrophic cardiomyopathy (HCM)	1	13	3.1
Intestinal immune network for IgA production	1	11	3.42
Jak-STAT signalling pathway	1	25	2.04
Melanoma	1	21	2.3
Mismatch repair	1	8	4.09
Nucleotide excision repair	1	19	2.45

KEGG pathways assembled from lists of upregulated genes ordered by the number of gene representatives in each pathway with a significant *z*-score (>2). The number of pathways was capped at 20.

combined into large, well-annotated contigs. The subsequently designed custom woodrat array had high technical quality scores, and the biological replicates for a given treatment were similar in their overall gene expression patterns, indicating repeatability. In addition, the woodrat samples had higher quality and intensity values when hybridized to woodrat-specific probes compared with the cross-hybridized Norway rat probes. Finally, comparisons of the gene expression profiles between woodrat species and within woodrat species suggest that the activities and processes captured by the arrays reflect important ecological and evolutionary patterns, with specific focus on the biotransformation functions of woodrat livers.

One unique advantage of working with woodrats is that they are in the same rodent superfamily (Rodentia: Muroidea) as rats and mice. This phylogenetic relatedness allows some molecular and genetic tools that have been developed for rats and mice to be used successfully on woodrat samples (Haley *et al.* 2007a; Skopec & Dearing 2011). This relatedness is likely also responsible for the thoroughness of contig annotation in the woodrat transcriptome. Using a custom BLAST database made exclusively of rodent reference sequences (41% *Rattus*, 57% *Mus*, 1.3% *Peromyscus* and other rodents), we were able to annotate all but 66 contigs (0.3% of contigs) as either a 'mutual best hit' or 'similar to' in a reciprocal best BLAST search. This is a higher rate of reliable annotation than reported in similar *de novo* transcriptome sequencing efforts, where the per cent of annotations with equivalent *e*-values ranges from 24% to 60% (Babik *et al.* 2010; Bruder *et al.* 2010; Pauchet *et al.* 2010; Castoe *et al.* 2011; Milan *et al.* 2011).

In terms of functionality, the woodrat transcriptome revealed both similarities to other transcriptomes as well as unique markers that reflect the nature of the tissue sample (liver) and the animal's ecology. The most heavily represented GO terms in the molecular function ontology were binding and catalytic activity. These umbrella terms include activities and processes (such as toxin binding and oxidoreductase activity, respectively) that relate to hepatic function and, thus, might be expected because liver tissue was sequenced and woodrats are of particular interest due to their ability to metabolize xenobiotics. However, binding and catalytic activity are also the two most common molecular function GO terms in recent transcriptomes built from a variety of tissues and taxa, including caterpillars (*Manduca sexta*, Pauchet *et al.* 2010), clams (*Ruditapes philippainarum*, Milan *et al.* 2011) and pythons (*Python molurus bivittatus*, Castoe *et al.* 2011) indicating a universal expression pattern. On the other hand, within the biological process ontology, there is less overlap between well-represented woodrat terms (cellular and metabolic processes and biological regulation) and the top terms in other *de novo* transcriptomes. In addition, the term, 'response to stimulus' that includes many of the enzyme pathways related to xenobiotic metabolism, encompassed >10% of the transcriptome. In another nonmodel mammal, the domestic ferret (*Mustela putorius furo*), the 'response to stimulus' term composes 6% of the transcriptome (Bruder *et al.* 2010), while in *Manduca sexta*, an insect in which xenobiotic metabolism has been studied extensively, the 'response to stimulus' term makes up only approximately 3% of the transcriptome (Pauchet *et al.* 2010). This reflects the functional importance of the woodrat liver for metabolizing plant secondary compounds.

The success of the woodrat transcriptome annotation using other rodent reference sequences begs the question of whether it is worthwhile to develop species-specific genetic tools when cross-species tools are available. Previous work on woodrats has indicated that cross-species molecular tools are useful but limited. For example, Magnanou *et al.* (2009) used Agilent's Norway rat microarray to compare gene expression across populations of woodrats. Although there was a multitude of expression differences across experimental treatments, overall, only 72% of the genes had quality scores high (>0.75) enough to be included in the analyses suggesting a loss of a quarter of the data as a result of cross-species hybridization. We directly tested the value of species-specific hybridization by including both Norway rat probes and woodrat probes on the woodrat microarray. Our results underscore the importance of species-specific probes. The quality values of sample hybridized to Norway rat probes were 35% lower than for woodrat probes matched for function. Of greater concern were the significantly lower-intensity values of Norway rat probes compared with the woodrat probes. Low-quality spots should be excluded from analyses by quality thresholds, but low-intensity spots might be maintained in the data set and cause pattern misinterpretation (Bar-Or *et al.* 2007). The shape of the frequency distribution of intensities was also different for Norway rat and woodrat probes (not show here), which makes correcting for differences in intensity not simply a matter of examining relative fold change effects. In the light of these differences, the data from the Norway rat probe were excluded from the remainder of analyses.

Interestingly, when the performance of *N. lepida* and *N. bryanti* samples were compared, the differences in quality and intensity were much smaller (*c.* 5%) and in an unexpected direction. *N. bryanti* samples performed better on the array, even though the probes were designed from the *N. lepida* transcriptome. Obviously, the congeneric woodrat species are much more closely related than woodrats and Norway rats, but the reason

for the increase in hybridization quality and intensity is unclear. Perhaps this result is due to population-level differences between the LR (UT) experimental animals and the Barstow (CA) individual used to develop the transcriptome. Differences in cross-strain hybridization have been noted in other systems (Bar-Or *et al.* 2007). Unfortunately, building additional microarrays for every species of *Neotoma*, not to mention every population, is not feasible at this time.

Microarrays are broad scale assays, designed to give an overall profile of gene expression across treatments. In ecological studies, microarrays have been used to examine how taxa share gene expression profiles or differ in their gene expression profiles under certain conditions, with the goal of understanding how selection has shaped gene expression responses (Whitehead 2012). However, in interspecific comparisons and even in interpopulation comparisons, the effect of selection on gene expression profiles is confounded by the effect of genetic drift. As phylogenetic distance increases, differences in gene expression between taxa accumulate by random chance (Whitehead & Crawford 2006). The solution to this problem is to perform multiple comparisons, holding either phylogenetic distance constant or ecological similarity constant. For example, in a study on fish adaptation to water temperature, Oleksiak *et al.* (2002) compared cold and warm water populations of one killifish species (*Fundulus heteroclitus*) to a warm water population of a sister species (*Fundulus grandis*), thus varying either the ecological conditions or phylogenetic distance. Using these comparisons, the researchers found a suite of genes that behaved more similarly in the interspecific, ecologically similar comparison than in the conspecific, ecologically distinct comparison, suggesting the action of nonrandom evolutionary forces (Oleksiak *et al.* 2002; Whitehead & Crawford 2006).

As a test of the custom array, we conducted a similar set of comparisons on separate species that share a similar ecological niche in the desert southwest and feed on creosote foliage (LR *N. lepida* and PS *N. bryanti*; Karasov 1989; Patton *et al.* 2008). Conversely, the CW *N. bryanti* animals are conspecific of the PS animals, but live in oak habitat, where they feed on oak foliage and other local plants (Atsatt & Ingram 1983). Thus, they are ecologically distinct from both other populations. If selection for ecological adaptation was affecting the gene expression profiles of these populations, one hypothesis is that the interspecific comparisons between the two desert populations should have fewer differentially expressed genes than the interspecific comparison between the desert population and the coastal population. This could result either from stabilizing selection acting on the ecologically similar populations or from directional selection differentiating the ecologically distinct populations. Differences accumulated via genetic drift should be the same in both comparisons because the phylogenetic distance is the same. Indeed, this is the pattern we see in these comparisons. There were 85% more differentially expressed genes across woodrat species with different ecological niches than across species with similar ecological niches. The fewest differences occurred in the conspecific comparison between PS and CW animals, as would be predicted for two closely related populations of the same species.

What are the potential functional differences that distinguish these populations from each other? In the comparison between the ecologically distinct LR *N. lepida* and CW *N. bryanti*, there are many GO terms overrepresented by upregulated genes in both populations. In the LR animals, the biological process ontology is dominated by terms relating to metabolic processes, and the molecular function ontology has a number of terms relating to transferase activity. Transferase activity is one of the primary functions of conjugation enzymes (e.g. glutathione-*S*-transferase, GST), which comprise one of two major classes of biotransformation enzymes in the liver. In the CW animals, biotransformation function is also well represented in the biological process ontology through terms relating to 'response to stimulus' and in transferase activity and oxidoreductase activity in the molecular function ontology. As expected, biotransformation is also well represented among upregulated KEGG pathways. In LR animals, 'drug metabolism' is among the top KEGG pathways, whereas in CW animals, 'drug metabolism—cytochrome P450', 'metabolism of xenobiotics by cytochrome P450' and 'drug metabolism—and other enzymes' are all significantly overrepresented. This preponderance of biotransformation-related functional terms might appear to be an artefact of the array design, which specifically included gene sets focused on biotransformation. However, all of the terms have significant *z*-scores, meaning that biotransformation function is more commonly upregulated than expected by chance, even given the large number of biotransformation features on the array.

Conversely, in the comparison between the ecologically similar LR *N. lepida* and PS *N. bryanti*, the biotransformation signal is much weaker. Overall, there were fewer significantly different GO terms in this comparison, perhaps as a result of the smaller number of probe differences. This patterns fit the hypothesis that the two ecologically similar populations would not have as many differentially expressed genes in the suite of genes relating to plant secondary compound metabolism because they have experienced similar selection for gene expression via a shared diet. Indeed, the differences in expression revealed by this comparison may give some clues as to how these two populations differ in their dietary

biotransformation. For instance, there are members of the cytochrome P450 (P450s) enzyme superfamily upregulated in both populations (Table S1, Supporting information). Many P450s are hepatic enzymes that oxidize xenobiotic substrates and are thought to be involved in 75% of phase I (functionalization) metabolism (Guengerich 2007). In woodrats, P450 enzyme activity and expression levels have been correlated with multiple plant components, including the phenolic resin produced by creosote bush and terpene-rich juniper foliage (Lamb *et al.* 2004; Haley *et al.* 2007b, 2008; Skopec *et al.* 2007; Magnanou *et al.* 2009), suggesting they are commonly used biotransformation tools. Conversely, members of GST and UDP glucuronosyltransferase (UGT) families appear as relatively highly expressed only in the PS *N. bryanti* animals (Table S1, Supporting information). Both GST and UGT are important phase II (conjugation) enzymes, which add conjugate groups to xenobiotic compounds to make them more excretable in the bile or urine (Klaassen 2001). Their upregulation in PS *N. bryanti* relative to the LR *N. lepida* suggests that these two populations may have evolved divergent strategies to cope with similar dietary challenges.

Species in the genus *Neotoma* have been the subject of many ecological and physiological studies exploring how mammalian herbivores deal with the variety and abundance of plant secondary compounds in their diets (Mangione *et al.* 2001; Lamb *et al.* 2004; Haley *et al.* 2007a,b, 2008; Skopec *et al.* 2008; Skopec & Dearing 2011). Much of this body of work compares woodrat individuals fed different diets—with components of their native plant toxins in contrast with novel plant components or control diets. In this data set, all the animals were fed what has been considered a 'control' diet (alfalfa-based rabbit chow), yet there are many metabolic and biotransformation differences in gene expression across the treatments. It is possible that these differences are constitutive; that is, they represent evolutionary (or developmental) differences in gene expression that exist regardless of actual diet. Constitutive differences might be more likely between the populations that are both ecologically distinct and phylogenetically distinct, thus differing as a result of both selection and drift. Another possibility is that the differences in metabolism-related expression are induced by the rabbit chow diet and that different populations of woodrats consuming rabbit chow respond by differentially regulating suites of hepatic genes. In other work, woodrats on rabbit chow have not experienced weight loss or intake reduction, as have animals on many of the treatment diets augmented with plants secondary compounds (Mangione *et al.* 2001; Skopec *et al.* 2008; Torregrossa *et al.* 2012). However, alfalfa produces minor quantities of secondary metabolites [e.g. phytoalexins (Baldridge *et al.* 1998) and flavones (Maxwell *et al.* 1989)], so the expression differences observed may be a response to the control diet. These results suggest the woodrat-specific microarray is sensitive enough to document expression differences ensuing from even relatively benign diets.

Conclusion

The introduction of next-generation sequencing technology has enabled a new burst of genetic and genomic tools for organisms with little likelihood of attracting the kind of intensive time and money required to create those resources via older methods. The woodrat transcriptome and microarray represent prime examples of functioning and informative tools where little prior genomic information existed until recently. The normalized woodrat transcriptome appears to have captured a biologically relevant snapshot of liver function. Using this, the custom woodrat array outperforms previous cross-species hybridizations and has already resulted in informative comparisons of ecologically and phylogenetically distinct woodrat populations with respect to selection and genetic drift. Now, the opportunity exists to combine our ecological understanding of the diverse *Neotoma* genus with the genetic underpinnings of behaviour and physiology. This combination of ecology and genetics may facilitate an understanding of the strategies employed by wild, outbred mammals to adapt to their environment.

Acknowledgements

We thank K. Kohl, M. Skopec, A. Fitzgerald and J. Varner for help with the woodrats, and S. Laverty for help organizing and reorganizing sequencing data. Funding was provided by an NSF grant (IOS 0817527) to M.D.D.

References

Atsatt PR, Ingram T (1983) Adaptation to oak and other fibrous, phenolic-rich foliage by a small mammal, *Neotoma fuscipes*. *Oecologia*, **60**, 135–142.

Babik W, Stuglik M, Qi W *et al.* (2010) Heart transcriptome of the bank vole (*Myodes glareolus*): towards understanding the evolutionary variation in metabolic rate. *BMC Genomics*, **11**, 390.

Baldridge GD, O'Neill NR, Samac DA (1998) Alfalfa (*Medicago sativa* L.) resistance to the root-lesion nematode, pratylenchus penetrans: defense-response gene mRNA and isoflavonoid phytoalexin levels in roots. *Plant Molecular Biology*, **38**, 999–1010.

Bar-Or C, Czosnek H, Koltai H (2007) Cross-species microarray hybridizations: a developing tool for studying species diversity. *Trends in Genetics*, **23**, 200–207.

Beebe L, Fox SD, Riggs CW *et al.* (1992) Persistent effects of a single dose of aroclor 1254 on cytochromes P450IA1 and IIb1 in mouse lung. *Toxicology and Applied Pharmacology*, **114**, 16–24.

Bellin D, Ferrarini A, Chimento A *et al.* (2009) Combining next-generation pyrosequencing with microarray for large scale expression analysis in non-model species. *BMC Genomics*, **10**, 555.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*, **57**, 289–300.

Bruder CE, Yao S, Larson F *et al.* (2010) Transcriptome sequencing and development of an expression microarray platform for the domestic ferret. *BMC Genomics*, **11**, 251.

Castoe TA, Fox SE, De Koning AJ *et al.* (2011) A multi-organ transcriptome resource for the burmese python (*Python molurus bivittatus*). *BMC Research Notes*, **4**, 310.

Ellegren H, Sheldon BC (2008) Genetic basis of fitness differences in natural populations. *Nature*, **452**, 169–175.

Frame GM, Wagner RM, Carnahan JC *et al.* (1996) Comprehensive, quantitative congener-specific analyses of eight aroclors and complete pcb congener assignments on DB-1 capillary GC columns. *Chemosphere*, **33**, 603–623.

Garcia-Reyero N, Griffitt RJ, Liu L *et al.* (2008) Construction of a robust microarray from a non-model species (largemouth bass) using pyrosequencing technology. *Journal of Fish Biology*, **72**, 2354–2376.

Guengerich FP (2007) Cytochrome P450 and chemical toxicology. *Chemical Research in Toxicology*, **21**, 70–83.

Haley SL, Lamb JG, Franklin MR *et al.* (2007a) Xenobiotic metabolism of plant secondary compounds in oak (*Quercus agrifolia*) by specialist and generalist woodrat herbivores, genus neotoma. *Journal of Chemical Ecology*, **33**, 2111–2122.

Haley SL, Lamb JG, Franklin MR *et al.* (2007b) Xenobiotic metabolism of plant secondary compounds in juniper (*Juniperus monosperma*) by specialist and generalist woodrat herbivores, genus *Neotoma*. *Comparative Biochemistry and Physiology Part C: Toxicology and Pharmacology*, **146**, 552–560.

Haley SL, Lamb JG, Franklin MR *et al.* (2008) 'Pharm-ecology' of diet shifting: biotransformation of plant secondary compounds in creosote (*Larrea tridentata*) by a woodrat herbivore, *Neotoma lepida*. *Physiological and Biochemical Zoology*, **81**, 584–593.

Jackson ST, Betancourt JL, Lyford ME *et al.* (2005) A 40,000-year woodrat-midden record of vegetational and biogeographical dynamics in north-eastern Utah, USA. *Journal of Biogeography*, **32**, 1085–1106.

Karasov WH (1989) Nutritional bottleneck in a herbivore, the desert woodrat (*Neotoma lepida*). *Physiological Zoology*, **62**, 1351–1382.

King MC, Wilson AC (1975) Evolution at two levels in humans and chimpanzees. *Science*, **188**, 107–116.

Klaassen C (2001) *Casarett and Doull's Toxicology: The Basic Science of Poisons*. McGraw Hill, New York.

Lamb JG, Marick P, Sorensen J *et al.* (2004) Liver biotransforming enzymes in woodrats neotoma stephensi (muridae). *Comparative Biochemistry and Physiology Part C: Toxicology and Pharmacology*, **138**, 195–201.

Magnanou E, Malenke JR, Dearing MD (2009) Expression of biotransformation genes in woodrat (*Neotoma*) herbivores on novel and ancestral diets: identification of candidate genes responsible for dietary shifts. *Molecular Ecology*, **18**, 2401–2414.

Malenke J, Milash B, Miller A, Dearing D (2013) Data from: transcriptome sequencing and microarray development for the woodrat (*Neotoma spp.*): custom genetic tools for exploring herbivore ecology. *Dryad Digital Repository*. doi:10.5061/dryad.25696.

Mangione AM, Dearing D, Karasov W (2001) Detoxification in relation to toxin tolerance in desert woodrats eating creosote bush. *Journal of Chemical Ecology*, **27**, 2559–2578.

Matocq MD (2002) Phylogeographical structure and regional history of the dusky-footed woodrat, *Neotoma fuscipes*. *Molecular Ecology*, **11**, 229–242.

Maxwell CA, Hartwig UA, Joseph CM, Phillips DA (1989) A chalcone and two related favonoids released from alfalfa roots induce node genes of rhizobium meliloti. *Plant Physiology*, **91**, 842–847.

Mendez-Harclerode FM, Strauss RE, Fulhorst CF *et al.* (2007) Molecular evidence for high levels of intrapopulation genetic diversity in woodrats (*Neotoma micropus*). *Journal of Mammalogy*, **88**, 360–370.

Meyer E, Aglyamova GV, Wang S *et al.* (2009) Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFLx. *BMC Genomics*, **10**, 219.

Milan M, Coppe A, Reinhardt R *et al.* (2011) Transcriptome sequencing and microarray development for the manila clam, *Ruditapes philippinarum*: genomic tools for environmental monitoring. *BMC Genomics*, **12**, 234.

Nowak CL, Nowak RS, Tausch RJ, Wigand PE (1994) A 30000 year record of vegetation dynamics at a semi-arid locale in the great basin. *Journal of Vegetation Science*, **5**, 579–590.

Oleksiak MF, Churchill GA, Crawford DL (2002) Variation in gene expression within and among natural populations. *Nature Genetics*, **32**, 261–266.

Patton JL, Huckaby DG, Alvarez-Castañeda ST (2008) *The Evolutionary History and a Systematic Revision of Woodrats of the Neotoma lepida Group*. UC Publications in Zoology, University of California Press, Berkeley, California.

Pauchet Y, Wilkinson P, Vogel H *et al.* (2010) Pyrosequencing the manduca sexta larval midgut transcriptome: messages for digestion, detoxification and defence. *Insect Molecular Biology*, **19**, 61–75.

Pratt WB, Taylor P (1990) *Principles of Drug Action: The Basis of Pharmacology*, 3rd edn. Churchill Livingstone, London.

Richardson TA, Klaassen CD (2010) Role of UDP-glucuronosyltransferase (UGT) 2B2 in metabolism of triiodothyronine: effect of microsomal enzyme inducers in sprague dawley and UGT2B2-deficient Fischer 344 rats. *Toxicological Sciences*, **116**, 413–421.

Rokas A, Abbot P (2009) Harnessing genomics for evolutionary insights. *Trends in Ecology and Evolution*, **24**, 192–200.

Skopec MM, Dearing MD (2011) Differential expression and activity of catechol-o-methyl transferase (COMT) in a generalist *Neotoma albigula*) and juniper specialist (*Neotoma stephensi*) woodrat. *Comparative Biochemistry and Physiology Part C: Toxicology and Pharmacology*, **154**, 383–390.

Skopec MM, Haley S, Dearing MD (2007) Differential hepatic gene expression of a dietary specialist (*Neotoma stephensi*) and generalist (*Neotoma albigula*) in response to juniper (*Juniperus monosperma*) ingestion. *Comparative Biochemistry Physiology, Part D Genomics and Proteomics*, **2**, 34–43.

Skopec MM, Haley S, Torregrossa AM, Dearing MD (2008) An oak (*Quercus agrifolia*) specialist (*Neotoma macrotis*) and a sympatric generalist (*Neotoma lepida*) show similar intakes and digestibilities of oak. *Physiological and Biochemical Zoology*, **81**, 426–433.

Torregrossa AM, Azzara AV, Dearing MD (2012) Testing the diet-breadth trade-off hypothesis: differential regulation of novel plant secondary compounds by a specialist and a generalist herbivore. *Oecologia*, **168**, 711–718.

Vera JC, Wheat CW, Fescemyer HW *et al.* (2008) Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Molecular Ecology*, **17**, 1636–1647.

Wheat CW (2010) Rapidly developing functional genomics in ecological model systems via 454 transcriptome sequencing. *Genetica*, **138**, 433–451.

Wheat CW, Fescemyer HW, Kvist J *et al.* (2011) Functional genomics of life history variation in a butterfly metapopulation. *Molecular Ecology*, **20**, 1813–1828.

Whitehead A (2012) Comparative genomics in ecological physiology: toward a more nuanced understanding of acclimation and adaptation. *Journal of Experimental Biology*, **215**, 884–891.

Whitehead A, Crawford DL (2006) Variation within and among species in gene expression: raw material for evolution. *Molecular Ecology*, **15**, 1197–1211.

J.R.M. contributed to experimental and microarray design and analysis, and paper writing. B.M. contributed to transcriptome analysis, array design, and data accessioning. A.W.M. contributed to the microarray validation. M.D.D. contributed to experimental design, transcriptomic analysis and paper writing.

Data Accessibility

Microarray design and data: GEO data series #GSE43941.

Short read library and assembled transcriptome: NCBI TSA BioProject #PRJNA188250, BioSample #SAMN01915240.

Microarray validation (intensity and quality data) and qPCR validation: Dryad entry doi:10.5061/dryad.25696.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1 Gene lists from interspecific comparison of Lytle Ranch *Neotoma lepida* and Palm Springs *Neotoma bryanti*.

Table S2 Significantly overrepresented Gene Ontology terms for: (A) Lytle Ranch *Neotoma lepida* (B) Palm Springs *Neotoma bryanti*.

Table S3 Gene lists from interspecific comparison between Lytle Ranch *Neotoma lepida* and Casper's Wilderness *Neotoma bryanti*.

Table S4 Overrepresented Gene Ontology terms: (A) Lytle Ranch *Neotoma lepida* and (B) Casper's Wilderness *Neotoma bryanti*.

Table S5 Gene lists from intraspecific comparison between the Palm Springs and Casper's Wilderness *Neotoma bryanti* populations.

Table S6 Overrepresented Gene Ontology terms upregulated on (A) Palm Springs and (B) Casper's Wilderness *Neotoma bryanti*.

Fig. S1 Frequency distribution of contig lengths (bp) from woodrat transcriptome.

Fig. S2 Relative expression of superoxide dismutase, P450 (cytochrome) oxidoreductase and catechol-*O*-methyltransferase determined with qPCR.