From raw materials to validated system: the construction of a genomic library and microarray to interpret systemic perturbations in Northern bobwhite

Arun Rawat,1 Kurt A. Gust,2 Youping Deng,3 Natalia Garcia-Reyero,4 Michael J. Quinn Jr.,5 Mark S. Johnson,2 Karl J. Indest,2 Mohamed O. Elasri,1 and Edward J. Perkins2

1Department of Biological Sciences, University of Southern Mississippi, Hattiesburg; 2U.S. Army Engineer Research and Development Center, Environmental Laboratory, EP-P; 3SpecPro Incorporated, Environmental Laboratory, EP-P, Vicksburg; 4Jackson State University, Department of Chemistry, Jackson, Mississippi; 5U.S. Army Center for Health Promotion and Preventive Medicine, Aberdeen Proving Ground, Maryland

Submitted 28 January 2010; accepted in final form 16 April 2010

The utility of genomic tools for advancing virtually all disciplines within the life sciences has been broadly documented. Although genomic tools development for ecologically relevant nonmodel species has lagged relative to model species, advancements in sequencing technology, bioinformatics processing, and gene-expression platforms have led to an increasing number of nonmodel species having deep-coverage, well-annotated transcriptomes from which high-quality genomic tools have been produced (19, 22, 45, 64, 67, 68). These tools will prove indispensable as the focus of biological research and regulatory decision frameworks continue to shift toward systems-biology approaches (29, 38, 56).

The advent of ultra-high throughput DNA sequencing technologies has greatly expanded sequencing depth and coverage (44) providing the ability to achieve robust de novo transcriptome characterization for virtually any species at relatively low cost (45, 67). In concert, in silico methods for high-throughput sequence assembly and annotation including expressed sequence tag (EST) characterization, Gene Ontology (GO) classification and assignment to functional pathways have risen to the challenge of maximizing biological information connected to the transcriptome (30, 66, 68). A number of commercial sources (e.g., Agilent Technologies, Nimblegen, Affymetrix) provide on-demand synthesis of high-density oligonucleotide arrays directly from users’ annotated sequence databases yielding low-cost tools for investigation of highly integrated responses to various stimuli. Such expression-profiling tools are the engine for expanding the “universal language” with which to describe cellular responses (39).

Avian wildlife represents a diverse group of species that are critical to ecosystem function, yet few genomic tools are available to study them. Domestic chicken (Gallus gallus) represents the most robustly described avian species (11) and microarray tools developed for chicken have been shown to have diagnostic utility in avian cross-species investigations (13). However, the unique attributes of avian species extend beyond the repertoire of chicken to complex vocalizations and vocal learning which has led to a zebra finch (Taeniopygia guttata) genome sequencing project (http://www.genome.gov/11510730). The Northern bobwhite (Colinus virginianus) is important for numerous ecological (50, 52), methodological (34), and environmental management reasons (58), making it an excellent experimental avian wildlife model. The limited genomic description of Northern bobwhite consists mainly of 1,717 ESTs derived from a brain-tissue cDNA library (27) highlighting the need to greatly expand this information base.

In this work, we utilized pyrosequencing of a Northern bobwhite normalized cDNA library assembled from multiple tissues of both males and females followed by in silico annotation to robustly characterize the transcriptome. In addition to identifying 71,384 unique transcripts, comparison with the chicken genome provided 8,930 nonredundant transcript matches indicating that library provides ~48% coverage of all chicken (G. gallus) protein-coding genes. Oligonucleotide microarrays were developed from the transcripts to assess the transcriptional responses of liver from Northern bobwhite ex-
posed to the munitions compound 2,6-dinitrotoluene (2,6-DNT). 2,6-DNT is found as a soil and water contaminant on military facilities (49), posing potential risks to ground foraging birds including the Northern bobwhite. The Northern bobwhite microarray was used to characterize the systemic impacts of 2,6-DNT exposure and provide hypothetical mechanisms underlying the toxicological impacts of the environmental contaminant for use in ecological risk assessments.

MATERIALS AND METHODS

Tissue samples used to construct the normalized cDNA library for Northern bobwhite were collected in exposure studies conducted at the U.S. Army Center for Health Promotion and Preventative Medicine in Edgewood, MD. Exposure assays included two individual 14 day (d) exposures to the munitions constituent hexahydro-1,3,5-trinitro-1,3,5-triazine (35, 53), a 14d exposure to 2,6-DNT (35), and a 60d exposure to 2,6-DNT (52). In total, the RNA pool used to construct the Northern bobwhite cDNA library represents 179 tissue samples taken from 56 individual birds. Tissue types represented include brain, liver, testes, duodenum, colon, and feather pulp. All protocols were conducted consistent with good laboratory practices and were approved by the Institutional Animal Care and Use Committee at the U.S. Army Center for Health Promotion and Preventative Medicine.

Tissue Fixation and RNA Extraction

Immediately following euthanasia by CO2 asphyxiation, tissue samples were fixed in RNA Later (Ambion, Austin, TX) following the manufacturer’s recommendations. RNA extraction was conducted using RNaseasy Mini RNA extraction kits (Qiagen, Valencia, CA). RNA quality was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, Waldbronn, Germany) with RNA 6000 Nano LabChips (Agilent). Only samples with a 28s/18s ratio ≥ 1.4 and RNA integrity number ≥ 6 were used for downstream applications. The majority of RNA samples greatly exceeded these minimum requirements. The RNA compilation included 800 ng of total RNA from each of the 179 tissue samples, which was purified for poly(A) RNA using a NucleoTrap mRNA purification kit (Macherey-Nagel).

cDNA Library Construction and Normalization

The SMART PCR cDNA Synthesis Kit (Clontech Laboratories Mountain View, CA) was utilized to reverse-transcribe 0.5 μg of the poly(A) RNA into full-length cDNAs. The cDNA library was normalized prior to sequencing to capture both high and low abundance transcripts using the Trimmer cDNA Normalization Kit (Evrogen JSC, Moscow, Russia).

cDNA Sequencing and Processing

The normalized cDNA library was sequenced by 454 Life Sciences ( Branford, CT) using massively parallel pyrosequencing on a GS-FLX sequencer using a protocol to resolve 200 bp reads. Raw EST sequences and quality information are available at the Short Read Archive (SRA) division (http://www.ncbi.nlm.nih.gov/Traces/sra/) of GenBank under accession number SRA0094605. Sequences were trimmed for flanking SMART Adapters using Codon Code Aligner (Codon Code Dedham, MA) with the sequences shorter than 50 bp discarded. Poly (A/T) tails were removed from raw EST sequences with Codon Code Aligner, resulting in 467,708 high-quality sequences out of a total 478,142 reads. Sequence assembly was performed using CAP3 (32) with the minimum percent identity (the minimum percentage of identical bases in the aligned region) ≥80% and minimum alignment score (overlap match plus counting mismatches) ≥30 bp. Contiguous sequences (contigs) and singlets <200 bp were removed. The resulting 71,384 sequences (35,904 contigs and 35,480 singlets) were considered putative unique genes.

Sequence Analysis, Open Reading Frame Prediction, and Domain Analysis

The homologs for the unique genes were interrogated with nonredundant database from the National Center for Biotechnology Information (NCBI) with Message Passing Interface (MPI) Blast (15) using high performance computing (http://cluster.vislab.usm.edu) on a 44-node cluster. Comparative genome analyses were performed using Refseq ID for protein, downloaded in February 2008 from http://www.ncbi.nlm.nih.gov. We derived the open reading frame (ORF) and frame direction of unique genes with the help of ESTScan (33, 48). Ortholog detection was performed with an in-house perl script to generate best basic local alignment tool (BLAST) E value and bit score pairs for chicken sequences (downloaded from NCBI, February 2008) and Northern bobwhite unique transcript dataset and further mapping between these two datasets performed in the database package MySQL (http://www.mysql.com). BLASTx orientations were preferentially used over frame directions from ESTScan in determining correct orientations for microarray probe design. For Interproscan domain analysis, the frame direction (and translated sequence) from ESTScan was used. Interproscan was used to search PROSITE, PRINTS, Pfam, ProDom, SMART, and Panther protein signatures for unique genes with significant blast hits (51).

GO and Kyoto Encyclopedia of Genes and Genomes Pathway Annotation

For the functional annotation of the Northern bobwhite orthologs, we performed the GO annotation against G. gallus annotation (downloaded from Ref. 20a, GOA Chicken 39.0, released December 15th 2008) using Web Gene Ontology Annotation Plot (WEGO) (72). The annotation derived was used to define second level GO function categories within each primary GO level (molecular function, cellular component, and biological process). The chicken and quail datasets were compared to find statistically significant relationships by Pearson χ2 test performed by WEGO on 2 × 2 matrices. Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway analysis was conducted using KEGG Orthology-Based Annotation System (KOBAS) web server (71) investigating raw sequences that had a significant blast hit. Sequences with significant KEGG Orthology (KO) were compared against G. gallus as a reference organism.

Northern Bobwhite Oligonucleotide Microarray

We created a 15,000-probe microarray using a ×15K custom oligonucleotide microarray platform (Agilent Technologies, Santa Clara, CA). The array was developed from a database consisting of sequences identifying putative transcript identities from BLASTx matches (E ≤ 10−2) to Northern bobwhite as well as from matches to closely related avian species including G. gallus, Numida meleagris, Coturnix japonica, Meleagris gallopavo, Callipepla gambelii, and Gallus varius to prioritize unique transcripts for inclusion on the array. A total of 8,454 nonredundant sequences with positive-frame orientations were identified and incorporated into the microarray design as described in the Supplemental Text.1 For the remaining positions available on the microarray, a total of 3,272 unique Refseq IDs from the remaining homologous sequences (nonavian species) in order of lowest E-value where E ≤ 10−3 were incorporated. Supplementary sequences were created for each due to lack of confidence in prediction yielding the 6,546 probe sequences needed to complete the microarray probe set (see below for anti-sense strand exclusion methods). The 15,000 target sequences were uploaded to eArray (Agilent Technologies) where 60-mer oligonucleotide probes

1 The online version of this article contains supplemental material.
were developed to represent each putative transcript on the microarray.

**Microarray Experimental Design, Hybridizations, and Data Extraction**

Male birds were selected for the investigation due to higher sensitivity to 2,6-DNT exposure relative to females (35). A completely randomized design was utilized to investigate differential gene expression in liver tissues among control birds and those exposed to 10 and 60 mg/kg/d doses of 2,6-DNT with each condition including four biological replicates. The Agilent One-Color Microarray Hybridization protocol (Agilent Technologies) was utilized for microarray hybridizations following manufacturer’s recommendations. We utilized 1 μg of total RNA for all hybridizations. An Axon GenePix 4000B Microarray Scanner (MDS Analytical Technologies, Toronto, Canada) was used to scan microarrays at 5 μm resolution. Data were extracted from microarray images using Agilent Feature Extraction software (Agilent Technologies). Analysis of internal control spikes indicated that signal data was within the linear range of detection.

**Microarray Data Analysis**

Background-subtracted adjusted median signal intensities were normalized to a per-chip basis (54) with R (53a) that transforms the signal intensity by dividing signal intensity for all the genes with the mean intensity in each array. Microarray analysis was performed using a HDArray library of functions available at http://www.r-project.org/, which utilizes a Bayesian probabilistic framework-based t-test to tests for differences in gene expression (9). The normalized data were imported into HDArray using Bioconductor (9a) to measure the confidence value associated with fold change for each gene (P < 0.01, unless stated otherwise). Results of Bayesian analysis were compared against a parametric, nonpaired t-test. The t-test was conducted using GeneSpring version 7.2 (Agilent Technologies) where data were first normalized with per-chip median scaling and cross-gene error model. The t-test (P < 0.05) assumed nonequal variance and incorporated a >1.8 fold-change requirement investigating the 9,711 probes that had present flags for all four replicate samples for all conditions.

The 3,272 sense-antisense probe pairs printed on the microarray were utilized to QC the microarray analysis for cross hybridization. One probe in each probe pair represents a nonsense sequence, and our expectation was that no target should have specific binding to it. When differentially expressed genes (DEGs) were examined, 4 and 19 probe pairs were coexpressed in liver (10 and 60 mg, respectively). These nonspecific DEGs were removed from our DEG list. The number of nonspecific DEGs were limited relative to the overall number of properly functioning probe pairs, which provided confidence in the representation of unique probes on the microarray, hybridization quality, and microarray analysis.

**Reverse-Transcription Quantitative Polymerase Chain Reaction**

The accuracy of the microarray results was assessed by reverse-transcription quantitative polymerase chain reaction (RT-qPCR) utilizing 35 primer sets (Supplemental Table S1) representing 34 unique genes of interest regarding the mechanistic toxicity of 2,6-DNT and one regulatory control (Northern bobwhite, 18S). Each condition included four biological replicates, identical to those used in the microarray experiment. See Supplemental Text for molecular methods. Applied Biosystems SDS 2.2 Software was utilized to resolve RT-qPCR data and the ∆ACT (Applied Biosystems, Foster City, CA) method used to quantify results. RT-qPCR data for primer sets that did not generate a single dissociation peak were excluded from analysis. The 18S regulatory control was unaffected by 2,6-DNT treatment (data not shown) and was used as the internal normalizer for relative quantification. The 95% confidence interval was calculated around the mean relative expression for each 2,6-DNT dose. Confidence intervals that did not include unity were considered differentially expressed relative to controls.

**Northern Bobwhite Database Construction and Web Interface**

We developed a semiautomated pipeline for data extraction and parsing with PERL. BioPerl 1.6. We incorporated the annotated unique transcripts to construct the Northern Bobwhite database with the relational database management system MySQL (5.0.45) and Apache Tomcat Web server (2.2.10), which was used as container for HTML and PHP scripts. Users can access information such as UniGene sequence, number and name of ESTs contained in contig, predicted coding region, NCBI information like BLAST E-value, gene name, Uniprot accession, PDB, and EMBL ID at http://www.quailgenomics.info/searchmetadata.php.

**Public Availability of Gene Expression and Toxicology Data**


**RESULTS AND DISCUSSION**

**cDNA Sequencing, Sequence Assembly, and EST Processing**

Sequencing of the normalized cDNA library yielded 114.9 megabases of sequence in 478,142 reads with an average read length of 240 bp and ~95% of the total sequence read lengths distributed between 200 and 300 bp. We compared the unique gene transcripts derived from two unique sequence assembly approaches, CAP3 (32) and Newbler (454 Life Sciences) to determine the optimal assembly method for our sequence dataset. The results of the two sequence assembly algorithms were compared by searching each set against the NCBI nonredundant protein databases. The CAP3 assembly generated a greater number of significant matches than the Newbler assembly (21,912 vs. 16,175) at BLASTx, E ≤ 10−5 (Table 1) as well as a greater number of Refseq accession protein IDs (11,667 vs. 8,766). We also compared the gene and protein products in common among the two assemblies and found that 97% (6,158/6,349) of genes identified by Newbler were also present in the CAP3 assembly (Supplemental Table S2). The CAP3 assembly provided an additional 2,164 distinct genes compared with 191 identified by Newbler. Finally, in cross-species protein database comparisons, 80% (7,007/8,784) of Newbler protein hits were similar to CAP3 (Supplemental Table S3). The CAP3 assembly resulted in a higher number of significant matches, a greater number of distinct genes, and incorporated most of the protein database matches identified using Newbler. Given these optimal assembly results, the CAP3-assembled sequences were utilized for all the downstream analyses and annotations of ESTs summarized in Fig. 1.

A total of 82,064 unique sequences were identified comprising 37,685 contigs and 44,379 singlets after vector cleaning, adapter and polyA tract removal, and CAP3 sequence assembly. Nearly 64.4% and 19.8% of the contigs were composed of greater than two and nine ESTs, respectively (Supplemental
Fig. S1). Contigs and singlets having length >200 bp, totaling 71,384 unigenes, were selected for further analysis and annotation. The average length of the assembled contigs was 500 bp (min = 202 bp, max = 6,524 bp, and SD = 403). The number of transcripts having significant matches to known or postulated protein sequences, where BLAST (7) expectation scores (E) were $<10^{-5}$, was 39.2% of contigs (14,081 out of 35,904) and 22% (7,831 out of 35,480) of singlets (Table 1).

We recognized that the 71,384 unique putative transcripts identified were an overrepresentation of the expected overall Northern bobwhite transcriptome that likely resulted from nonoverlapping contigs and singletons leading to splits in transcript sequence coverage. Furthermore, many of the nonhomologous sequences might actually represent noncoding RNA, artifacts of sequencing and assembling errors or represent protein coding genes that still remain unknown. To examine the transcripts for the presence of noncoding RNAs, the unique transcripts were compared with integrated noncoding RNA sequences from the fly, human, and other eukaryote genomes (61). A total of 2,296 unique transcripts (BLASTn E $<10^{-5}$) showed similarity with noncoding RNA in eukaryotic genome databases as described in the Supplemental Text.

**Comparative Sequence Analysis**

We examined the degree to which genes present in Northern bobwhite are conserved across various species by comparison of the unique transcripts to genomes of several model species. Unique transcripts for Northern bobwhite were similar to 18,968 (26.57%), 16,579 (23.23%), 16,144 (22.62%), 6,100 (8.55%), 4,551 (6.38%), and 1,906 (2.67%) genes of chicken, human, mouse (Mus musculus), Drosophila melanogaster, Caenorhabditis elegans, and yeast (Saccharomyces cerevisiae), respectively at E $<10^{-10}$ (Table 2). Approximately 850 genes (1.19%) were similar to all six model organisms as well as chicken and 15,498 genes (21.71%) in mammalian model organisms (human and mouse), at E $<10^{-10}$ (Table 3). Not surprisingly, many genes with functions in central metabolism and protein synthesis such as eukaryotic translation initiation factor 1, glyceraldehyde-3-phosphate dehydrogenase, ribosomal protein, methylenetetrahydrofolate dehydrogenase, actin, lysyl/glycyl/threonyl tRNA synthetase, proteasome 26S ATPase, UDP-glucose pyrophosphorylase, heat shock protein, and fumarate hydratase were found to be conserved across all organisms at E $<10^{-10}$. Overall, Northern bobwhite had the greatest protein coding gene homology to chicken when compared against genomes of model organisms (Table 2); therefore, we selected the chicken genome to further investigate phylogeny.

**Assessment of Ortholog Detection**

Gene ortholog detection is extensively utilized to mine the increasing amount of sequence data generated by complete or partial genome sequencing projects. Unique transcripts were used to perform a BLAST search against the nr database (NCBI) to identify orthologs. The results depict the distribution of expressed sequence tag (EST) data searched against nr protein database by BLASTx.
Table 2. Genomic comparison of Northern bobwhite (Colinus virginianus) transcriptome against nr database for model organisms

<table>
<thead>
<tr>
<th>Organisms</th>
<th>Human</th>
<th>Mouse</th>
<th>Fly</th>
<th>C. elegans</th>
<th>Yeast</th>
<th>Chicken</th>
<th>All Organisms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homology</td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
</tr>
<tr>
<td>E \leq 10^{-100}</td>
<td>735</td>
<td>1.03</td>
<td>714</td>
<td>1.00</td>
<td>155</td>
<td>0.22</td>
<td>101</td>
</tr>
<tr>
<td>E \leq 10^{-50}</td>
<td>3,160</td>
<td>4.43</td>
<td>3,050</td>
<td>4.27</td>
<td>791</td>
<td>1.11</td>
<td>549</td>
</tr>
<tr>
<td>E \leq 10^{-10}</td>
<td>11,938</td>
<td>16.72</td>
<td>11,506</td>
<td>16.12</td>
<td>3,260</td>
<td>4.57</td>
<td>2,299</td>
</tr>
<tr>
<td>E \leq 10^{-4}</td>
<td>16,579</td>
<td>23.23</td>
<td>16,144</td>
<td>22.62</td>
<td>6,100</td>
<td>8.55</td>
<td>4,551</td>
</tr>
<tr>
<td>E \leq 10^{-1}</td>
<td>18,386</td>
<td>25.76</td>
<td>17,964</td>
<td>25.17</td>
<td>7,912</td>
<td>11.08</td>
<td>6,180</td>
</tr>
<tr>
<td>E \leq 10^{-5}</td>
<td>52,998</td>
<td>74.24</td>
<td>53,420</td>
<td>74.83</td>
<td>63,472</td>
<td>88.92</td>
<td>65,204</td>
</tr>
<tr>
<td>No match</td>
<td>50,650</td>
<td>70.95</td>
<td>49,472</td>
<td>69.31</td>
<td>4,229</td>
<td>6.39</td>
<td>4,561</td>
</tr>
</tbody>
</table>

Table 3. Similar genes among model organisms represented in Northern bobwhite (C. virginianus) transcriptome

<table>
<thead>
<tr>
<th>Organism</th>
<th>E \leq 10^{-100}</th>
<th>E \leq 10^{-50}</th>
<th>E \leq 10^{-20}</th>
<th>E \leq 10^{-10}</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Human + mouse</td>
<td>697</td>
<td>0.98</td>
<td>2899</td>
<td>4.06</td>
</tr>
<tr>
<td>Human + fly + mouse</td>
<td>152</td>
<td>0.21</td>
<td>757</td>
<td>1.06</td>
</tr>
<tr>
<td>Human + fly + mouse + chicken</td>
<td>150</td>
<td>0.21</td>
<td>741</td>
<td>1.04</td>
</tr>
<tr>
<td>Human + fly + mouse + C. elegans</td>
<td>98</td>
<td>0.14</td>
<td>503</td>
<td>0.70</td>
</tr>
<tr>
<td>Human + fly + mouse + C. elegans + yeast</td>
<td>40</td>
<td>0.06</td>
<td>170</td>
<td>0.24</td>
</tr>
<tr>
<td>Human + fly + C. elegans + yeast + chicken</td>
<td>35</td>
<td>0.05</td>
<td>151</td>
<td>0.21</td>
</tr>
<tr>
<td>Human + fly + C. elegans + yeast + A. thaliana</td>
<td>35</td>
<td>0.05</td>
<td>151</td>
<td>0.21</td>
</tr>
<tr>
<td>Fly + C. elegans</td>
<td>99</td>
<td>0.14</td>
<td>504</td>
<td>0.71</td>
</tr>
<tr>
<td>Fly + C. elegans + yeast</td>
<td>40</td>
<td>0.06</td>
<td>170</td>
<td>0.24</td>
</tr>
<tr>
<td>Fly + C. elegans + yeast + A. thaliana</td>
<td>35</td>
<td>0.05</td>
<td>151</td>
<td>0.21</td>
</tr>
<tr>
<td>C. elegans + yeast</td>
<td>35</td>
<td>0.05</td>
<td>151</td>
<td>0.21</td>
</tr>
<tr>
<td>C. elegans + yeast + A. thaliana</td>
<td>37</td>
<td>0.05</td>
<td>159</td>
<td>0.22</td>
</tr>
</tbody>
</table>

Partial genome projects, which provides increasing accuracy in predicting ortholog and paralog relationships (12, 20). Prediction of protein-coding regions was established using ESTScan, which uses a hidden Markov model to identify coding regions, even if the quality is low and contains frame shifts, a common sequencing error (33). ORFs were found for 39,400 (22,432 contigs and 16,968 singlets) of the 71,384 unique transcripts with 20,660 located in the 5' end and 18,740 in the 3' end of the transcripts (see section “Protein domain prediction” below). The average read length for the ORFs was 352.3 bp (min = 60, max = 4,500; SD = 232.2). A total of 18,647 proteins from translated ORFs (47.3% of total) were found to have significant BLASTp matches at E \leq 10^{-5} and 18,499 (99.2%) of these mapped to the putative homologs. These predicted proteins were analyzed using a high-throughput automated annotation pipeline, yielding annotation for an additional 2.3% of proteins not recognized by BLASTp (see section “High-Throughput Automated Annotation”). The remaining predicted proteins (50.4%) detected by the ESTScan model may represent false positives or homologs that might be detected by less stringent blast cutoff values.

Phylogenetic trees provide accurate, however computationally intensive, detection of orthology for putative homologs. A less computationally demanding method is reciprocal BLAST hit (RBH) analysis, an alternative method frequently used as a shortcut for ortholog detection (46). The rationale is that two genes in different organisms are considered orthologous if they match as best hits when each full gene is queried against the other (10). We performed RBH to assess ortholog detection among Northern bobwhite and chicken (G. gallus, an intraorder phylogenetic relative of Northern bobwhite) across both the putative homologs and predicted ORFs that might lead to orthologous genes (Fig. 1). The blast hits (E \leq 10^{-5}) from chicken-bobwhite and bobwhite-chicken were sorted on minimum E value (and maximum bit score if more than one hit with same E value existed). We found 17,709 orthologs for the putative homologs and 15,441 orthologs for the predicted proteins from translated ORFs found by ESTScan. Out of orthologs predicted by ESTScan, 14,724 (95.3%) were also present in orthologs from the putative homologs, and the remaining in each dataset either were unique or had different ortholog matches. We observed that 17,800 putative homolog-ORF pairs (where matching protein identifiers were sorted on minimum E value for BLAST hits) were complementary in 85% of comparisons to the RBH pairs. To maximize the ortholog count for the Northern bobwhite, we conjoined ortholog sets derived from each method, resulting in 20,676 nonredundant orthologous unique transcripts that represent 8,825 unique gene products (Fig. 1). The number of orthologs corresponds to ~48% of the G. gallus transcriptome, suggesting that we have achieved coverage of approximately half of the Northern bobwhite transcriptome. The summation of these orthologs provides a broad representation of predicted molecular functions inherent in the Northern bobwhite transcriptome that we utilized to establish functional annotations as we describe in the following section.

Functional Annotation

Annotation provides functional context to genes/orthologs. To functionally annotate the 8,825 Northern bobwhite orthologs, we investigated and inferred putative GO (8) annota-
tions, finding matching annotations for 4,786 (54.2%) genes. The distribution of GO annotation within the first-level molecular function, biological process, and cellular component is shown in Fig. 2A. The number of second-level GO subcategories within each first-level GO category is represented for Northern bobwhite and chicken in Fig. 2, B–D. Statistically significant matches were observed among species for a variety of second-level GO categories including the top four most abundant categories for both for biological processes (“cellular process,” “metabolic process,” “biological regulation,” and “pigmentation”) and molecular functions (“binding,” “catalytic activity,” “molecular transducer activity,” and “transporter activity”). These results indicate similar enrichment in these ontology categories among Northern bobwhite and chicken. Annotations for these transcripts were inferred from the electronic annotation evidence level; however, with manual intervention, these orthologs are candidates that can be updated to Inferred from Sequence Orthology evidence level (http://www.geneontology.org/GO.evidence.shtml). The GO categories of these orthologs can be browsed through the GO browser implemented from amigo (28) that is available at Ref. 50a.

**High-Throughput Automated Annotation**

We utilized automated annotation of the Northern bobwhite transcript dataset so that we could efficiently identify protein
domains and incorporate proteins within functional molecular pathways (43). We applied different approaches for annotation, pathway assignment, and protein domain prediction to develop insight into the Northern bobwhite genome, which has limited information in public repositories.

Pathway assignment. Pathway information provides an integrated representation of biochemical reactions and functional interactions compared with the gene-centric GO analysis (63, 69). We collected pathway information from the KEGG database using KO to annotate unique gene sequences (37). KO is a classification of ortholog and paralog groups based on highly confident sequence similarity scores and is directed acyclic graph with hierarchy layout in four tiers (43). We assigned metabolic pathways based on sequence similarity (E-value ≤ 10^{-5}) to sequences with known KEGG pathways. Pathway identification of these unique transcripts were performed by statistical enrichment with Fisher exact test and Benjamini and Hochberg false discovery rate correction against the G. gallus pathways, resulting in 14,452 KO entries. In the second level of the KO (Fig. 3), endocrine and immune system in cellular processes, signal transduction in environmental information processing, translation in genetic information processing, cancer in human diseases, amino acid and carbohydrate metabolism in metabolism were found to be most abundant. Among the most frequently resolved identities at the third level of organization (Fig. 4) were receptors and channels, protein kinases, MAPK signaling pathway, calcium signaling pathway, and neuroactive ligand-receptor interaction.

Protein domain prediction. Interproscan is used to scan for protein signature (protein families, domains, and functional sites) from the databases PROSITE, PRINTS, Pfam, and ProDom (73). The output also allows for the visual inspection of the protein signature that was performed for unique transcripts and is freely available (see section “Northern Bobwhite Database Construction and Web Interface”). We compared the 20,954 putative coding regions predicted by ESTSCAN for 21,912 unigenes (E ≤ 10^{-5}) and found that 7,240 (33% of significant BLAST hits) were not assigned into any domain and family. The remaining 14,672 (66.95%) grouped into 2,453 domains and protein families. The most common domains were protein kinase (245), WD-40 repeats (122), ankyrin (96), zinc finger (91), and RNA recognition motif (84).

The 18,446 predicted coding regions for the remaining 49,475 nonsignificant unigenes (E > 10^{-5}) queried resulted in

---

**Fig. 3. Distribution of KEGG orthology (KO) for Northern bobwhite (C. virginianus). Represented is the 2nd level of hierarchical organization.**
385 (.78%) unique matches in Interpro database. The most common domains were TonB box N-terminal (52), phosphopantetheine attachment site (18), and PS00014 (16) involved in posttranslation modification and β-tubulin autoregulation binding site (12).

Microarray Analysis: Effects of 2,6-DNT Exposure on Transcript Expression

Several toxicological impacts have been observed in Northern bobwhite dosed with 2,6-DNT (52). The 10 and 60 mg/kg/d doses of 2,6-DNT were selected to represent the minimum dose at which significant impacts on toxicological endpoints were observed and a highly affective dose that elicited numerous significant impacts on toxicological endpoints approaching the threshold of mortality, respectively (52), hereafter referred to as L10 and L60. We compared the microarray results from Bayesian t-test with the results of parametric t-test and found a substantial overlap of DEG among the two analyses, including 182 and 396 common genes relative to controls for L10 and L60, respectively (Fig. 5, A and B). We proceeded by utilizing the results of Bayesian t-test for the purposes of investigating 2,6-DNT effects.

The number of DEG relative to unexposed controls (0 mg/kg/d) at $P \leq 0.01$ nearly doubled with the increase in 2,6-DNT dose from 372 to 750 in the L10 and L60 treatments, respectively (Supplemental Table S4) with 135 DEG in common among doses (Fig. 5C). KEGG pathway and GO associations were mined from the Northern bobwhite annotated library for all differentially expressed genes (Supplemental Tables S5–S7). Further enrichment of GO associations to establish biologically meaningful contexts connected to expression changes were performed with the help of the database for annotation, visualization, and integrated discovery (16, 31) and GenMAPP (14, 57).

Validation of Microarray Analysis

The correspondence among RT-qPCR and microarray results was generally good at 73.5% and 76.5% for the L10 and L60 treatments, respectively (Fig. 6). The correlation of fold-change results among the expression assays was highly signif-
observed toxicological phenotypes. Furthermore, genomic results detected 2,6-DNT impacts on gene expression (Fig. 6 and Supplemental Table S4) below the threshold at which adverse effects were manifested. We investigated pathways, GO terms and individual gene targets significantly impacted in 2,6-DNT treatments to determine potential mechanisms underlying observed gross-level effects and effects on blood chemistry and, further, explored genomics-directed observations to provide a systemic understanding of the general pharmacology of 2,6-DNT in Northern bobwhite (Table 4).

Gross-level effects. The high dose of 60 mg/kg/d, 2,6-DNT caused edema in the gastrointestinal (GI) tract of most birds with corresponding impacts on genes within the “prostaglandin synthesis and regulation” pathway (Table 4), which is key to the inflammatory response in mammalian models. Annexins (ANXA) are critical components of the inflammatory pathway (Supplemental Fig. S2), which are recognized to inhibit prostaglandin production in mammalian models (26). Anxa1,-2,-5,-13 were overexpressed (Fig. 6) at P value < 0.01 (Anxa7 and -8 were overexpressed at P value < 0.05) in the L60 treatment. Additionally, expression of the G protein-coupled receptor, endothelin A receptor (EDNRA), which regulates endothelin-1-mediated induction of prostaglandin E2 (40, 59), was reduced. Although prostaglandins are proinflammatory in most organs, they are recognized to have anti-inflammatory effects on GI mucosa (47). Increased expression of annexins as well as decreased expression of EDNRA can each contribute to reduced prostaglandin production, which is consistent with increased inflammation observed in the GI tract and liver of Northern bobwhite.

2,6-DNT exposure resulted in liver lesions and oval cell/biliary hyperplasia, in 11 of 12 males and 10 of 12 females in the L60 treatment group, ultimately contributing to four deaths (52). Oval cell hyperplasia is believed to occur in mammalian species as a part of a hepatic-repair process when liver injury exceeds the proliferation capacity of normal hepatocytes, and the observed biliary hyperplasia indicates that the primary site of injury involved the bile duct epithelium (25). Investigation of GO terms including “response to wounding,” “wound healing,” and “hemostasis” indicated differential expression of a number of genes (Supplemental Table S7) including greatly increased expression (4.35-fold, log2, Fig. 6) of tissue factor (LOC429084) as well as increased expression of interleukin 10 receptor beta (IL10RB) in the L60 group. Increased expression of clotting factors such as (LOC429084) is consistent with initiation of clotting at the site of lesion injury (24), while the increased expression of interleukin 10 receptor, beta (IL10RB), is an indicator of enhanced immune response against wound infection.

Brown pigmentation was observed to accumulate in quail liver Kupffer cells in all high-dose (L60) birds accompanied by dark brown granular pigmentation in spleen, diarrhea, and weight loss in the majority of birds. 2,6-DNT and related chemicals such as 2,4,6-trinitrotoluene are known to cause anemia in mammalian species via erythropoiesis, which may in turn affect the primary blood conditioning organs including liver, kidney, and spleen (21, 62). Our results indicated that the gene coproporphyrinogen oxidase (CPOX) and a gene similar to uroporphyrinogen-III synthase (UROS, LOC426223), components of the porphyrin and chlorophyll metabolism pathway and an instrumental facilitator of heme biosynthesis, were
overexpressed in the L60 treatment (Fig. 6 and Table 4). Increased expression of CPOX and UROS genes (Supplemental Fig. S3), if uncoupled from heme synthesis, is consistent with accumulation of porphyrin byproducts (uroporphyrinogen III, protoporphyrinogen IX) causing porphyria (18). Porphyrin byproducts are indicators of impeded heme cycling and can also result in symptoms similar to those observed in 2,6-DNT exposed quail including diarrhea, loss of sensation, low red blood cell (RBC) count, and abnormal liver function (22a, 44a). While some gross-level effects, such as weight loss arise from complex impacts on physiology, evidence from transcriptomics provides additional insight into systemic-level responses of the organism. Our gene expression results provide plausible mechanisms underlying several gross-level effects of 2,6-DNT.

Effects on blood chemistry. Changes in blood chemistry parameters were some of the most sensitive indicators of 2,6-DNT exposure with impacts occurring at doses as low as 10 mg/kg/day. A significant decrease was seen in plasma-glucose levels in L10 and L60 birds in absence of 2,6-DNT-related changes in feed consumption (52). Consistent with reduced glucose levels and weight loss observed in Northern bobwhite, expression of genes represented in nine pathways and two GO categories related to carbohydrate and energy metabolism were affected (Table 4) with the majority of genes having decreased expression in 2,6-DNT exposures (Supplemental Tables S5 and S7). Investigation of the glycolysis and gluconeogenesis pathway (Fig. 7) for the L60 treatment indicated impacts on genes involved in establishing equilibrium between the glycolysis and gluconeogenesis processes. Phos-
phenoxypruvate carboxykinase 1 (PCK1), recognized as a major regulatory point for gluconeogenesis (23), had a 2.7-fold log_2 and a 3.6-fold log_2 decreased expression in L10 and L60 birds, respectively (Fig. 6). The concerted decreases in expression of (PCK1), aspartate aminotransferase (GOT1), and malate dehydrogenase 1 (MDH1) are consistent with a shift in equilibrium away from gluconeogenesis, therefore limiting glucose synthesis. In contrast, enolase 2 (ENO2), which utilizes the substrate formed in the PCK1-catalyzed reaction (phosphoenolpyruvate), was upregulated presumably in response to increased production of pyruvate from other metabolic pathways.

### Table 4. GO and KEGG pathway entries that best described the toxicological phenotypes observed in Northern bobwhite exposed to 2,6-DNT for 60d were used to gain toxicogenomic insights into the mechanisms underlying the toxicological effects

#### Gross-level Effects

**Edema in gastrointestinal tract**
- Liver lesions in 11 of 12 males and 10 of 12 females in the 60 mg/kg/d treatments. Oval cell/biliary hyperplasia
- GenMapp: prostaglandin synthesis and regulation (L60: 4 up, 1 down)
- GO:0006950: response to stress (L10: 5 up, 5 down)
- GO:0009611: response to wounding (L60: 2 up, 5 down)
- GO:0042060: wound healing (L60: 2 up, 3 down)

**Brown pigmentation accumulation in the Kupffer cells (liver) in all birds @ 60 mg/kg/d and dark brown granular pigmentation accumulation in spleen**
- KEGG:ggao0860: porphyrin and chlorophyll metabolism (L60: 2 up)
- GO:0007596: blood coagulation (L60: 2up, 3 down)
- GO:0007599: hemostasis (L60: 2 up, 3 down)

**Diarrhea & weight loss as well as death in 3 and 4 birds in the 40 and 60 mg/kg/d treatments respectively**

#### Blood Chemistry Results

**Reduction in plasma glucose levels**
- KEGG:ggao0010: glycolysis/gluconeogenesis (L60: 2 up, 5 down)
- KEGG:ggao0020: citrate cycle (TCA cycle) (L60: 1 up, 5 down; L10: 1 down. 1 common gene and response among doses)
- KEGG:ggao0030: pentose phosphate (L60: 1 down)
- KEGG:ggao0040: pentose and glucuronate interconversions (L60: 1 up, 1 down; L10: 1 up, 1 down). 2 common genes and responses among treatments)
- KEGG:ggao0051: fructose and mannose metabolism (L60: 1 up, 1 down; L10: 1 up, 1 down. 1 common gene and response among treatments)
- KEGG:ggao0052: galactose metabolism (L60: 1 up, 1 down; L10: 1 up, 1 common gene and response among treatments)
- KEGG:ggao00190: oxidative phosphorylation (L60: 3 up, 2 down, L10: 1 down. 1 common gene and response among treatments). 2 common gene responses and gene responses among treatments)
- KEGG:ggao00500: starch and sucrose metabolism (L60: 1 up, 5 down; L10: 1 down. 1 common gene and response among treatments)
- KEGG:ggao00620: pyruvate metabolism (L60: 2 up, 4 down; L10, 1 up, 2 down. 3 common gene responses and gene responses among treatments)
- GO:0006919: generation of precursor metabolites and energy (L60: 7 up, 15 down)
- GO:0016491: oxidoreductase activity (L60: 14 up, 25 down; L10: 9 up, 9 down)

**Reduction in RBCs and hemoglobin concentrations**
- KEGG:ggao00860: porphyrin and chlorophyll metabolism (L60: 2 up)

**Reduction in albumin (>50%), globulin, and total protein**
- Albumin (ALB, NP_990592) (L60: up; L10 down)
- KEGG:ggao04120: ubiquitin-mediated proteolysis (L60: 4 down; L10: 3 up, 4 down)
- GO:0006519: amino acid and derivative metabolic process (L60: 4 up, 8 down)
- GO:0006530: amino acid metabolic process (L60: 4 up, 7 down)
- GO:0009063: amino acid catabolic process (L60: 1 up, 3 down)
- GO:0009308: amino catabolic process (L60: 1 up, 3 down)
- GO:0016504: protease activator activity (L10: 1 up, 1 down)

**Reduction in aspartate amino transferase concentrations**
- Aspartate aminotransferase (also known as glutamic-oxaloacetic transaminase, GOT) (L60: 1 down)

**Increase in uric acid**
- Aspartate aminotransferase (also known as glutamic-oxaloacetic transaminase, GOT) (L60: 1 down)
- KEGG:ggao0860: porphyrin and chlorophyll metabolism (L60: 2 up)
- KEGG:ggao0220: urea cycle and metabolism of amino groups (L60: 2 up, 1 down)

**Reduction in Na⁺ and K⁺ ion concentration**
- KEGG:ggao0860: porphyrin and chlorophyll metabolism (L60: 2 up)
- KEGG:ggao0910: nitrogen metabolism (L 60: 2 down; L10, 1 down)
- GO:0044270: nitrogen compound catabolic process (L60: 1 up, 3 down)
- GO:0030001: metal ion transport (L60: 11 up, 3 down)
- GO:0006812: cation transport (L60: 13 up, 3 down)
- GO:0015672: monovalent inorganic cation transport (L60: 12 up, 0 down)

Continued

Physiol Genomics • VOL 42 • www.physiolgenomics.org
response to low substrate concentrations and in response to decreased plasma-glucose levels.

Significant reductions in RBC concentrations, plasma albumin levels (>50% reduction), aspartate aminotransferase concentrations, total proteins, globulin concentration, and Na⁺ and K⁺ ion concentrations were also seen in L60 birds, whereas uric acid concentrations nearly doubled in L60 birds (52). As described above, the KEGG pathway “porphyrin and chlorophyll metabolism” involved in the “heme biosynthesis pathway” (Supplemental Fig. S3) was affected in the L60 treatments including over-expression of CPOX and UROS (Fig. 6, Supplemental Table S5). Increased expression of these genes is consistent with a compensatory effort to return hemoglobin and RBCs to normal levels.

A significant increase in albumin transcript copy number (ALB, 3.58-fold, log₂, Fig. 6) is consistent with a response to compensate for marked reductions in plasma-albumin levels observed by Quinn et al. (52). Reductions in plasma-globulin and total proteins correspond with decreased expression of the majority of genes associated with the “ubiquitin-mediated proteolysis” pathway (Supplemental Table S5), as well as the GO categories related to amine and amino acid catabolism (Table 4, Supplemental Tables S6 and S7), indicating a potential inhibition of protein cycling initiated by 2,6-DNT exposure.

Reduced blood concentrations of aspartate aminotransferase, also known as glutamic-oxaloacetic transaminase (GOT1), at 60 mg/kg/day 2,6-DNT, corresponded with reduced GOT1 transcript expression in L60 (Fig. 6). In addition to the role of GOT1 in glycolysis and gluconeogenesis described above, it and number of additional genes are components of the metabolic pathways “nitrogen cycle,” “urea cycle and metabolism of amino acid groups,” and “porphyrin and chlorophyll metabolism,” as well as members of the GO category “nitrogen compound catabolic process,” which were impacted in the L60 treatment (Table 4, Supplemental Table S5 and S6). Additionally, genes including agmatine ureohydrolase (AGMAT) and ornithine decarboxylase 1 (ODC1), which are involved in the “urea cycle and metabolism of amino acid groups” pathway (Supplemental Table S5), had increased expression in L60 birds (Fig. 6). These overall impacts, and most specifically the increased expression of components of the “urea cycle and metabolism of amino acid groups” pathway, are consistent with elevated blood-urea concentrations observed in birds exposed to high doses of 2,6-DNT (52).

Finally, observed reductions in Na⁺ and K⁺ ion concentrations were accompanied by a predominant increase in transcription expression within the GO categories: “metal ion transport,” “cation transport,” and “monovalent inorganic cation transport” (Table 4 and Supplemental Table S7). Cation transport is fundamental to a variety of physiological processes and finding the ultimate cause of depletion is a difficult task. A potential result of depletion is manifested in the increased expression of the Na⁺/glucose cotransporter 1 (SLC5A1, Fig. 6) a probable response to increase absorption of glucose from dietary sources (65) to supplement already marginalized glucose supply to cells.

Genomics-directed observations. In addition to having toxicological phenotypes guide our investigation of mechanisms of action, we utilized genomic results to identify impacts beyond the results provided in the toxicological bioassays to improve assessment of 2,6-DNT pharmacology and provide connectivity among systemic effects. Regarding 2,6-DNT pharmacology, pathways involved in phase I and phase II xenobiotic metabolism (“metabolism of xenobiotics by cytochrome P450” and “glutathione metabolism,” respectively) had increased expression in the L60 treatment (Table 4 and Supplemental Table S5). Specifically, glutathione-S transferase
Glycolysis and Gluconeogenesis

**Gene Database**
- **Gene:** hg_hsa_20070517.gtf
- **Expression Dataset:**
  - **Name:** liver6.gene
  - **Color:** Red
- **Legend:** Fold Change
  - Upregulated
  - Downregulated
  - No criteria met
  - Not found

**Author:** Nick Fidelman, Kari Dahlquist
**Maintained by:** GenMAPP.org
**E-mail:** genmapp@gladstone.ucsf.edu
**Last modified:** This pathway was converted from human by homology
**Copyright © 2009, Gladstone Institutes**
**License:** Creative Commons Attribution 3.0 Unported License

**Physiol Genomics** • VOL 42 • www.physiolgenomics.org
(GSTA) and cytochrome P450 (CYP7A1) had increased expression within these pathways (Fig. 6). TNT and related compounds (including 2,6-DNT) can induce these mechanisms (17, 36, 55), which may ultimately influence toxicity (60). Additionally, the numerous results indicating that pathways involved in nitrogen and urea metabolism (described above) are being impacted by 2,6-DNT suggest the potential for denitrification of 2,6-DNT and processing of resultant nitrogenous metabolites.

An additional observation provides a systemic link to the impacts on energy metabolism discussed above, in this case, involving perturbations caused by 2,6-DNT on transcript expression of genes and pathways involved in lipid metabolism. An isomer of 2,6-DNT, 2,4-dinitrotoluene (2,4-DNT), has been observed to impact lipid metabolism in the fish model, fathead minnow (Pimephales promelas), resulting in decreased expression of a number of genes that regulate fatty acid synthesis and causing phospholipid accumulation in the liver (70). Similarly, 2,6-DNT caused decreased expression of several genes involved in lipid metabolism in Northern bobwhite liver (L60) including fatty acid synthase (FASN), a gene similar to long-chain acyl-CoA synthetase 3 (LOC424810), 3-hydroxy-3-methylglutaryl-CoA reductase (HMGCR), a gene similar to delta7-sterol reductase (Loc422982), a gene similar to acyl-CoA synthetase long-chain family member (ACSL1), acetyl-CoA carboxylase alpha (ACACA), sterol CoA desaturase (SCD) apolipoprotein B (APOB), and acyl-CoA dehydrogenase family 11 (ACAD11) (Supplemental Table S4). In the fathead minnow model, impacts of 2,4-DNT appeared to be modulated via the peroxisome proliferative activated receptor α (PPARα) and peroxisome proliferative activated receptor γ (PPARγ) pathways. No significant effect was observed on PPARα in Northern bobwhite liver; however, PPARγ was downregulated ($P = 0.03, -1.0$ fold log$_2$ in L10) in addition to peroxisome proliferative activated receptor γ coactivator 1 (PPARGC1A, $-1.1$-fold log$_2$ in L10, $-1.1$-fold, log$_2$ in L60). These impacts on components of the lipid metabolism pathway are an additional indicator of 2,6-DNT-induced perturbation of lipids.

![Image of a pathway diagram](Fig. 8. Effects of 2,6-DNT exposure on the peroxisome proliferative activated receptor (PPAR) pathway in liver tissue of Northern bobwhite dosed with 60 mg/kg/d, 2,6-DNT in a 60d exposure. Significant impacts on transcript expression relative to controls are represented by red (increased expression) and green (decreased expression) highlights.)
energy metabolism in Northern bobwhite, which ultimately may have forced the stress leading to many of the observed toxicological impacts (Fig. 8).

Impacts on Energy Metabolism is Central to 2,6-DNT Toxicity

Consumption of 2,6-DNT impacted expression of genes that facilitate energy metabolism including both gluconeogenesis (Fig. 7) and lipid metabolism (Fig. 8). Gluconeogenesis is the process by which noncarbohydrate, carbon substrates are converted to glucose, the paramount short-term energy storage molecule. Marked reductions in PCK1 expression, in addition to reduced expression of GOT1 and MDH1, indicate impaired potential for glucose generation (23), which corresponds with reduced glucose levels in blood in the absence of feeding rate changes in Northern bobwhite. From the perspective of energy metabolism, lipids represent medium- to long-term storage molecules. Expression of PPARG1 and PPARGC1A, genes that represent key control points for initiation of fatty acid metabolism (Fig. 8) and mitochondrial biogenesis (41), was reduced, as was the case for a variety of additional genes involved in lipid metabolic pathways. Reduced expression of these genes is indicative of impairment of lipid catabolism, which can ultimately reduce availability of lipid as a substrate for production of cellular energy in times of energetic debt. A study investigating PPARA1 knockout mice indicated increased concentrations of long-chain fatty acids in plasma in addition to impairments in citric acid cycle flux, enhanced urea cycle activity, and increased amino acid catabolism (42). If PPAR impacts are conserved in bird species, these results indicate that reduced blood-protein levels and increased blood-urea levels may have resulted from enhanced catabolism of amino acids. Furthermore, these results suggest that 2,6-DNT-induced impairment of glucose and lipid metabolism in Northern bobwhite may have caused reduced availability of carbohydrate- and lipid-based substrates for energy generation, elevating amino acid-based biomolecules as a predominant substrate for this process. This would additionally contribute to a principle gross-level effect observed in 2,6-DNT exposure, weight loss/wasting.

Conclusions

In this paper, we have presented the entire life cycle of development of an annotated microarray from sequencing, annotation, and application to understanding sublethal impacts of 2,6-DNT dosing in Northern bobwhite. We performed comparative genome analyses with model organisms and GO annotation for 8,825 unique orthologous Northern bobwhite genes. Furthermore, we annotated this species with high-throughput annotation and transitioned this knowledge to develop a custom microarray for Northern bobwhite. Microarray analysis of liver tissue from Northern bobwhite exposed to 2,6-DNT indicated a variety of molecular impacts resulting from 2,6-DNT exposures that were consistent with overt toxicological symptoms. Key molecular-to-phenotype impacts included: prostaglandin pathway-mediated inflammation, increased expression of heme synthesis pathway in response to anemia, a shift in energy metabolism toward protein catabolism via inhibition of control points for glucose and lipid metabolic pathways, PCK1 and PPARGC1, respectively, as well as a variety of additional insights. The conservation of genomic responses among species including parallel impacts of a 2,6-DNT isomer on gene expression observed in fathead minnow (70) indicates the potential for commonality of mechanisms of action across distantly related phylogenetic relatives. Overall, our observations illustrate not only the utility of the genomic infrastructure developed for Northern bobwhite for assessment of systemic perturbations, but also the potential for generalization of genomic responses among phyla.

ACKNOWLEDGMENTS

We thank Glover George for assisting in establishing MPI-BLAST and use of http://cluster.vislab.usm.edu and Dr. Mehdi Pirooznia for providing the application BatchBlastExtractor available at http://mbc.b.usm.edu. Permission was granted by the Chief of Engineers to publish this information.

GRANTS

This work was supported by the U.S. Army Environmental Quality and Installations Basic Research Program and Toxigenomics Focus Area. We also gratefully acknowledge the support of the Mississippi Functional Genomics Network (National Center for Research Resources Grant P20 RR-016476) and Engineer Research and Development Center Grant W912HZ-08-C-0032.

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

REFERENCES


