Functional genomics of cactus host shifts in *Drosophila mojavensis*

LUCIANO M. MATZKIN,*THOMAS D. WATTS,BENJAMIN G. BITLER,CARLOS A. MACHADO and THERESE A. MARKOW

Department of Ecology and Evolutionary Biology, University of Arizona, PO Box 210088, Tucson, AZ 85721-0088, USA

Abstract

Understanding the genetic basis of adaptation to novel environments remains one of the major challenges confronting evolutionary biologists. While newly developed genomic approaches hold considerable promise for addressing this overall question, the relevant tools have not often been available in the most ecologically interesting organisms. Our study organism, *Drosophila mojavensis*, is a cactophilic Sonoran Desert endemic utilizing four different cactus hosts across its geographical range. Its well-known ecology makes it an attractive system in which to study the evolution of gene expression during adaptation. As a cactophile, *D. mojavensis* oviposits in the necrotic tissues of cacti, therefore exposing larvae and even adults to the varied and toxic compounds of rotting cacti. We have developed a cDNA microarray of *D. mojavensis* to examine gene expression associated with cactus host use. Using a population from the Baja California population we examined gene expression differences of third instar larvae when reared in two chemically distinct cactus hosts, agria (*Stenocereus gummosus*, native host) vs. organpipe (*Stenocereus thurberi*, alternative host). We have observed differential gene expression associated with cactus host use in genes involved in metabolism and detoxification.

Keywords: detoxification, *Drosophila mojavensis*, ecological genomics, Functional genomics, host shifts, microarray

Received 19 April 2006; accepted 10 July 2006

Introduction

Natural selection can shape patterns of variation at all levels of complexity, from gene to genomes to individuals and populations. Recently, technological advancements have made it possible to examine patterns of variation at the level of the transcriptome using microarrays (reviewed in Ranz & Machado 2006). Levels of interspecific transcriptional variation can differ dramatically across organisms and across studies (Jin *et al*. 2001; Oleksiak *et al*. 2002; Cheung *et al*. 2003; Ranz *et al*. 2003). Although the major control of this variation (*cis* vs. *trans*) is still debatable, there appears to be a large amount of epistasis (Brem *et al*. 2002; Rockman & Wray 2002; Yvert *et al*. 2003; Morley *et al*. 2004; Storey *et al*. 2005) and it is becoming apparent that natural selection has played a role in the transcriptional differences *between* species (Rifkin *et al*. 2003; Nuzhdin *et al*. 2004; Gilad *et al*. 2006). A question of

interest is how natural selection shapes the transcriptional variation *within* a species. Fortunately, microarray technology can be applied to ecologically interesting species, allowing for the incorporation of a species' ecology into its studies of genomic and transcriptome variation (Gibson 2002; Feder & Mitchell-Olds 2003).

Drosophila mojavensis with its well defined ecology and recently derived genomic tools (http://rana.lbl.gov/drosophila/) provides an excellent system to investigate the role of transcriptional evolution in an ecologically interesting organism. *Drosophila mojavensis* is endemic to the deserts of southwestern USA and northwestern Mexico (Heed 1978). Its cactophilic lifestyle implies that it oviposits, develops and feeds as adults in the necrotic tissues of specific cactus species (Heed 1982). There are four geographically and genetically isolated host races of *D. mojavensis*, each utilizing a different cactus host (Fellows & Heed 1972; Ruiz & Heed 1988; Reed *et al*. 2006). The Sonoran population utilizes the organpipe cactus (*Stenocereus thurberi*), Baja California the agria cactus (*Stenocereus gummosus*), Mojave Desert the barrel cactus (*Ferocactus*

Correspondence: Luciano Matzkin, Fax: (+1 520) 626 3522; E-mail: lmatzkin@email.arizona.edu

cylindraceus), and Catalina Island the prickly pear cactus (*Opuntia* spp.). The necroses of each different cactus species provide the different *D. mojavensis* populations with very distinct chemical environments (Heed 1978; Vacek 1979; Kircher 1982), which are both a function of the cactus tissues as well as the microfauna associated with the necrosis (Starmer 1982; Starmer *et al.* 1986). Depending on the chemical composition of each cactus host we can expect a different set of loci to be associated with the detoxification of the compounds, including orthologs to known detoxification genes such as P450s and Glutathione-S-transferase (*Gst*) (Feyereisen 1999; Enayati *et al.* 2005), and a set of loci associated with metabolic pathways depending on the alcohol/sugar composition of the cactus hosts.

The goal of this study was to identify loci and pathways that are differentially expressed in *D. mojavensis* third instar larvae during a cactus host shift. We built a microarray for *D. mojavensis* using 6520 random ESTs isolated from a cDNA library. We then exposed a Baja California population of *D. mojavensis* to its native host (agria) and to an alternative host (organpipe, the host in Sonora) and used the microarray to identify genes that were differentially regulated in each of the cactus treatments. We conclude that (i) cactus host usage has a significant effect on gene transcription; (ii) loci whose function involved detoxification are differentially regulated in response to a cactus host shift; and (iii) a subset of the differentially expressed loci may have arisen *de novo* in the *D. mojavensis* lineage. This represents the first study of transcriptional variation in an ecologically characterized *Drosophila* species.

Materials and methods

Development of the microarray

The *Drosophila mojavensis* stock used in the development of the microarray was originally collected in San Carlos, Sonora, Mexico in November of 2000. A laboratory culture was established using multiple females and maintained on standard banana/opuntia medium (Tucson *Drosophila* Species Stock Center, http://flyfood.arl.arizona.edu/opuntia.php3) in glass bottles. For the purpose of creating a comprehensive cDNA library, we collected tissue from several life stages. For collection of embryos, large numbers of adults were allowed to oviposit in a layer of yeast paste on banana/opuntia laboratory medium for 12 hours and then removed. The *D. mojavensis* embryos were then collected at 0, 12, and 24 h post oviposition. Embryos were collected by washing yeast paste containing oviposited eggs through a fine mesh screen, treating embryos to multiple washes with deionized water in a 50 mL conical tube, then snap frozen in liquid nitrogen and stored at −80 °C. For all other tissue collection, a large number of adult flies were allowed to oviposit on either banana/

opuntia laboratory medium or mashed organpipe cactus for 12 hours; samples were then collected accordingly during different times of development. From these two media, we collected *D. mojavensis* larvae every 12 hours from 36 h post oviposition to pupation (168 h) and pupae every 12 hours from pupation (180 h) to eclosion (288 h). Larvae and pupae were washed several times in deionized water, transferred to microcentrifuge tubes containing 50 μL of deionized water, and snap frozen and stored at −80 °C. The adults that emerged from the two media were collected at one and 10 days posteclosion. Females and males were collected for both age groups. Half of the 10 day old flies were allowed to mate, therefore obtaining a set of virgin and mated flies (females/males) for each of the two rearing media. Collected adults were placed in dry microcentrifuge tubes and frozen/stored as above.

RNA was individually extracted from all samples using the TRizol (Invitrogen) method. RNA quality was assessed using a spectrophotometer as well by running 1% formaldehyde agarose gels. For each stage (embryo, larva, pupa and adult), equal amounts of total RNA from each extraction were pooled and PolyA + mRNA was extracted from each pooled RNA using Oligotex mRNA minipreps (Qiagen Inc). We pooled 2 μg of mRNA from each life stage into a total mRNA sample, which was then used for library construction. The SMART cDNA Library Construction kit (Clontech Laboratories Inc.) was used to create the cDNA library. cDNA fragments were ligated to a λTriplEx2 vector and packaged into a phage (MaxPlax Lambda, Epicentre). The phage was used to transform BM25.8 *E. coli* cells. Approximately, 2.3 million transformants were created, assuring that most genes expressed are represented in the transformant pool. The library was amplified using the GENE-TRAPPER Protocol (Invitrogen). About 350 000 transformants were placed in each of six 500 mL bottles of LB/Carbenicillin (100 mg/mL)/SeaPrep agarose (FMC) and incubated overnight at 37 °C. Colonies were collected by centrifuging for 20 min at 8000 r.p.m. Transformants were plated and 40 000 clones were collected using a robotic colony picker (Q-bot, Genetix) housed at the Arizona Genomics Institute at the University of Arizona. Of the 40 000 clones, 6517 were PCR amplified, verified by running them in a 2% agarose gel, cleaned (using a 96-well Millipore PCR purification block on a Biomek FX robot) and rehydrated in a printing solution of 50% DMSO. In addition to the clones, we included 11 spots: one blank, five *D. mojavensis* alcohol dehydrogenase-1 (*Adh-1*) and five alcohol dehydrogenase-2 (*Adh-2*). All spots were printed in duplicate on GAPS II aminosilane coated slides (Corning) using a Virtek Chipwriter PRO microarray spotter.

Host shift experiment

A *Drosophila mojavensis* isofemale line originally collected from a necrotic agria cactus in La Paz, Baja California

(February, 2001) was utilized for the microarray experiments. This line is normally reared in standard banana/opuntia media. Since the purpose of this experiment was to observe the response of cactus host shifts, the cactus media must be as representative of the natural ecology of the system as possible. Given the significant effects of the microfauna of necrotic tissue on the life history and development of *D. mojavensis* (Fellows & Heed 1972; Starmer 1982; Starmer *et al.* 1986; Etges & Heed 1987; Etges 1989) we added a cocktail of the most common cactophilic microfauna. The cocktail included five yeast species (*Pichia cactophila*, *P. amethionina*, *Candida sonorensis*, *C. ingens*, and *Sporopachydermia cereana*, kindly provided by W. T. Starmer, Syracuse University) and a bacterial strain (*Pectobacterium cacticida*) commonly found in necrotic cacti (obtained from the American Type Culture Collection).

We followed a modified version of the Brazner *et al.* (1984) protocol for the preparation of the cactus necrosis. The cactus tissue used was first cut into cubes (~25 g), autoclaved (10lb-10min, 115 °C), placed into a sterile container and inoculated with the pectinolytic bacterium (*P. cacticida*). After two days, approximately 10 000 cells of each cactus yeast strain were applied onto the cactus cubes. Two days after the yeast inoculation, approximately 800 dechorionated, sterile embryos were collected, half of which were dispersed on the inoculated necrotic agria and half on necrotic organpipe (165 g of tissue each). Approximately 10-day old adults from the Baja California line were allowed to oviposit in embryo collection cages (Genesee Scientific) for 24 h, at which time the embryos were collected. Embryos were sterilized by first washing them with a saline solution, dechorionated with a Clorox solution, and re-washed with sterile saline (Starmer & Gilbert 1982). After 10 days, third instar larvae (~350 per treatment) were collected, snap frozen using liquid N2 and stored at –80 °C. The RNA was extracted from 10 groups of 20–30 larvae per treatment using RNeasy columns (Qiagen Inc.). The RNA was pooled per treatment, labelled (Cy3/Cy5) during the cDNA synthesis, and hybridized to the array. Five dye-flip replicates were performed (total of 10 slides). This level of technical replication allows for the most robust of signal to be evident. The labelling, PCR clean-ups, array printing, hybridizations and scanning were performed in the Genomic Analysis and Technology Core facility at the University of Arizona (protocols can be found at http://gatc.arl.arizona.edu/resources/protocols/index.php).

Microarray analysis

Hybridized slides were scanned using the Applied Precision arrayWoRxᵉ slide scanner and analysed using the softWoRxᵉ image analysis program. All of the raw intensity data was $\log_2$ transformed and then quantile-transformed (Bolstad *et al.* 2003). Our analyses were done using a two-step

mixed-model ANOVA of relative fluorescence intensities as suggested by Wolfinger *et al.* (2001). The first step is done globally using the complete dataset of transformed intensities (*Y*), which removes the variance associated with the dye and array design (random variables).

$$Y_{ij} = \mu + \text{ARRAY}_i + \text{DYE}_j + \text{ARRAY} \times \text{DYE}_{ij} + \text{Residual}_{ij}$$

The second step utilizes the residuals from the first step to do a gene-specific test. Therefore, per gene the model is:

$$\text{Residual}_{ijkl} = \mu + \text{ARRAY}_i + \text{DYE}_j + \text{CACTUS}_k + \text{ARRAY} \times \text{SPOT}_{il} + \text{Error}_{ijkl}$$

Where *i* is the number of different array designs (only two, organpipe derived RNA labelled with Cy3 and agria with Cy5, and vice versa), *j* the two dyes (Cy3/Cy5), *k* the two cactus treatments (agria/organpipe), and *l* the duplicate spotting on the array slides. Similar to the global step, ARRAY and its interaction with SPOT were treated as random variables, where DYE and CACTUS were fixed effects and CACTUS is the variable of interest. The CACTUS estimate and its *P*-value is what was used to create the volcano plot (see Fig. 1). The analysis was executed as described in Gibson & Wolfinger (2004) using the PROC MIXED model in SAS (ver. 9.0; SAS Institute Inc.). Given that multiple statistical tests are being executed a correction was needed. The Bonferroni method was utilized, but for microarray data sets this method is conservative (Jin *et al.* 2001). For example given an initial analysis of 6520 spots, a *P*-value must be lower than 7.67×10^{-6} to be able to call a test significant ($\alpha = 0.05/6520$). Therefore we also utilized the False Discovery Rate (FDR) Method of (Storey & Tibshirani 2003). The FDR method calculates the number of false positives within a set of significant values ($P < 0.05$) and then calculates a new significance probability, *q*. The test is considered to be significant, if $q < 0.05$.

A total of 740 spots were sequenced for verification purposes. This included a large majority of the Bonferroni corrected significant spots, a portion of the FDR corrected significant spots plus a random set of spots. Sequences (i.e. spots on the array) that were identical were grouped and the analysis was then repeated. Identity of the spots was determined by implementing the BLASTX algorithm **1** against the complete set of *Drosophila melanogaster* proteins (Release 4.2). The threshold *e*-value was set at 1×10^{-10}, anything greater than this value was not considered a hit to *D. melanogaster*. In cases where there was no hit to the *D. melanogaster* database, the clone sequence was queried to the UC Santa Cruz Genome Browser (http://genome.ucsc.edu/) *Drosophila mojavensis* database (August **2** 2005 assembly) using the BLAST algorithm. If the query
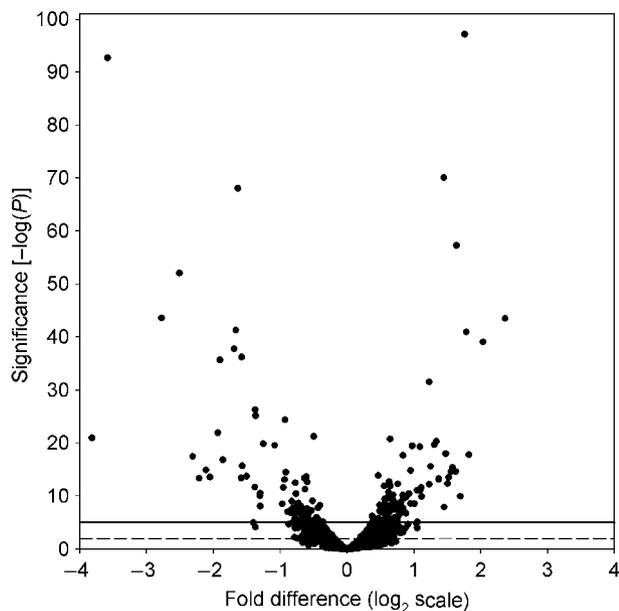
Fig. 1 Volcano plot of genes associated with host utilization in *Drosophila mojavensis*. Fold difference (organpipe minus agria expression) in $\log_2$ scale is on the *x*-axis. Positive values translate to a greater expression in the agria treatment and negative values a greater expression in the organpipe treatment. The *P*-value for each gene-specific ANOVA is on the *y*-axis. The solid lines indicate the cut-off value for significance using the Bonferroni correction (172 spots above the line) and the dashed line is the cut-off value for the FDR correction (1034 spots above the line).

result was within 200 bp upstream of a *D. mojavensis* predicted gene, the predicted gene sequence was queried against the *D. melanogaster* database and the identity of the best hit (*e*-value < 1×10^{-10}) recorded.

Real-Time PCR verification

For two differentially regulated loci, *Adh-2* and Glutathione-*S*-Transferase D1 (*GstD1*), we verified their change in expression using Real-Time PCR, as it tends to be more sensitive than microarray technology. Using a sample of the pooled RNA from the above microarray experiment, cDNA was created using an iScript cDNA Synthesis Kit (Bio-Rad Laboratories Inc.). This cDNA was used as the template for the Real-Time PCR which was done in a Bio-Rad iCycler IQ platform. The reaction mix was composed of 12.5 µL of IQ SYBR Green Supermix (Bio-Rad Laboratories Inc.), 0.25 µL of forward and reverse primers, 7 µL of dH$_2$O and 5 µL of cDNA template. For each locus and treatment 16 reactions were performed. For the purpose of standardizing the results, a control locus (Ribosomal subunit 16 s) was used. The forward/reverse primer sequences for the three loci were as follows: *Adh-2* (TTGAAGACAATCTTCGACAAGC/ACGCTCGATCTGGTAGTCGT); *GstD1* (GTCTACCTG-

GTGGAGAAGTACGGCAAGAC/TGGCGAACACCT-GAGGATAGTAGTAGTTGG); *16 s* (CTCGTCCAACCAT-TCATTCC/GAAATTTTAAATGGCCGCAGT). Efficiency for each locus was determined by running a dilution series (1000x, 100x, 10x, 1x) in triplicate. The results were standardized using the Pfaffl (2001) method and statistical significance was determined using a Pairwise Fixed Reallocation Randomization Test using the REST program (Pfaffl *et al*. 2002).

Results

From the original analysis of the 6520 spots (including *Adh-2* and *Adh-1*), 430 spots (225/205 in the organpipe/agria treatments) were significantly differentially expressed using the conservative Bonferroni correction. For the sake of simplicity we will refer to significantly expressed genes as those up-regulated in that particular treatment. Since all measurements are relative (organpipe treatment minus agria treatment) a gene that is significantly differentially expressed (i.e. up-regulated) in the agria treatment can be also viewed as being down-regulated in the organpipe treatment. Applying the FDR correction we observed 1514 significant spots (613/901 in the organpipe/agria treatments). We then sequenced a portion of the significant spots (Bonferroni and FDR corrected) as well as a set of nonsignificant spots for a total of 740. Clone sequences have been deposited to GenBank (accession numbers EC590007-EC590746). Identical sequences were grouped (total of 354 different sequences) and the above described statistical analysis was again performed on this smaller dataset. The BLASTX score, spot identity and expression ratio for all sequenced spots can be found in the supplementary data (including a FastA file containing the clones or contiged clones used for the BLAST analysis). A total of 6127 gene specific tests were performed, of those 173 (77/96 in the organpipe/agria treatments) were significant using the Bonferroni correction ($\alpha = 8.16 \times 10^{-6}$) and 1034 (378/656 in the organpipe/agria treatments) using the FDR correction. The volcano plot (Fig. 1) shows the range of expression differences observed, ranging from a 14.1-fold increase in organpipe to a 5.1-fold increase in agria. All expression data has been placed in the Gene Expression Omnibus under series entry #GSE5148.

The majority (85%) of the Bonferroni corrected significant genes were sequenced. Of these, 62 were up-regulated on the organpipe treatment and 84 on agria (104 and 129, respectively, for FDR corrected genes). Table 1 illustrates the function and numbers of significant genes observed in both treatments. Eighteen loci (25 using FDR) in the organpipe treatment and 39 (54 using FDR) in agria were of unknown function. These include loci which provided a BLASTX hit but with no known functional information as well as loci in which no hit to the *Drosophila melanogaster*

Table 1 Summary list of the biological process of the differentially regulated genes under the two different cactus host treatments. The numbers of Bonferroni and FDR (in parentheses) corrected significant genes are shown

Biological process	Agria	Organpipe
antimicrobial humoral response	4 (4)	0 (0)
axis specification	0 (0)	1 (1)
cell adhesion	0 (0)	0 (1)
cell motility	0 (0)	1 (2)
chromatin assembly	0 (0)	0 (4)
cytoskeleton	0 (1)	0 (2)
defence response	1 (1)	0 (1)
defence response — bacterial	1 (1)	2 (2)
defence response; protein folding	0 (0)	2 (6)
defence response; response to toxin	1 (1)	0 (2)
development	1 (1)	3 (3)
DNA repair/replication	1 (1)	0 (0)
electron transport	0 (0)	0 (1)
ethanol oxidation	0 (0)	1 (1)
heme oxidation	0 (0)	0 (1)
larval cuticle	2 (4)	11 (14)
lipid transport	0 (0)	1 (1)
metabolism	2 (2)	1 (2)
metabolism — carbohydrate	0 (1)	0 (1)
metabolism — carbohydrate; oxidoreductase activity	2 (2)	0 (0)
metabolism — chitin	3 (5)	0 (0)
metabolism — coenzyme	0 (0)	1 (1)
metabolism — lipid	4 (5)	3 (4)
metabolism — nucleotide	1 (1)	0 (0)
metabolism; oxidoreductase activity	0 (0)	1 (2)
mitochondrial electron transport	2 (2)	1 (1)
muscle contraction	0 (0)	1 (1)
nervous system	1 (1)	0 (1)
nitrogen compound metabolism	1 (1)	0 (0)
oogenesis	0 (1)	0 (0)
phosphorylation	0 (0)	1 (1)
protein biosynthesis	10 (28)	0 (3)
protein transport	0 (0)	1 (1)
proteolysis and peptidolysis	1 (4)	6 (10)
puparial adhesion	0 (0)	1 (1)
RNA binding	1 (1)	0 (0)
serine-type endopeptidase inhibitor	2 (3)	0 (0)
signal transduction	0 (0)	2 (3)
transcription	0 (0)	0 (1)
transport	4 (4)	3 (4)
unknown (found in *D. melanogaster*)	27 (38)	14 (18)
unknown (MSG + NMG)	12 (16)	4 (7)
Total	84 (129)	62 (104)

gene set was observed (see Table 1). Additionally, all genes with no BLASTX hit to *D. melanogaster* were queried to the complete RNA families database (Rfam, http://www.sanger.ac.uk/Software/Rfam/index.shtml) to search for noncoding RNA's. The only noncoding RNA (7SLRNA) found in our clone set did not significantly change its expression with host usage.
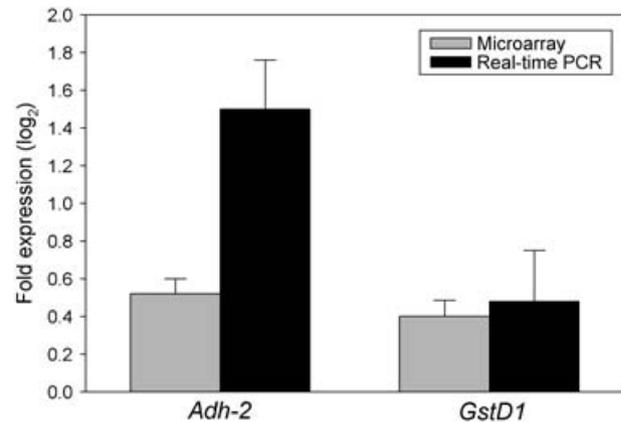
Fig. 2 Mean and standard error of the fold up-regulation (in $\log_2$ scale) of *Adh-2* and *GstD1* in the organpipe treatment calculated via the microarray or real-time PCR.

The 23 genes (16 from agria and seven from organpipe using FDR) that had no clear homology to any *D. melanogaster* gene were searched via BLAST to the assemblies of the other 10 sequenced *Drosophila* genomes (http://rana.lbl.gov/drosophila/). Eighteen of these 23 genes were exclusively found in *Drosophila mojavensis* and labelled MSG's (*mojavensis* specific genes), and the remaining five were labelled NMG's (not *melanogaster* genes). Furthermore, all significant and nonsignificant MSG's (32 genes total) were queried against the adult *D. mojavensis* EST library previously sequenced (ftp.ensembl.org). Seven out of the 32 MSG's were found in the adult EST library. Overall, there were seven significant MSG's in the organpipe treatment and 16 in agria.

Genes that perform a diverse set of functions appear to have been differentially regulated as a response to the experimental cactus host shift (Table 1). There are some differences in the types of genes that are affected by cactus host use. For example, a greater number of genes associated with protein biosynthesis were up-regulated in the agria treatment, but a greater number of genes associated with the larval cuticle were up-regulated in the organpipe treatment (see Table 1 and supplementary data). Overall, 17 genes involved with metabolism were observed in agria, while only 10 in organpipe. In the organpipe treatment, we observed the up-regulation of *GstD1*, a gene with known response to toxins (Tang & Tu 1994; Le Goff *et al.* 2001; Ranson *et al.* 2001; Chen *et al.* 2003).

To verify the microarray expression results, real-time PCR was performed on two significantly regulated genes, *Adh-2* and *GstD1* (Fig. 2). Microarray data shows a 1.5 fold increase in *Adh-2* expression when reared on organpipe. Using real-time PCR there was a 2.9 fold increase in *Adh-2* expression on organpipe ($P < 0.001$). A similar agreement between the microarray and real-time PCR data was observed for *GstD1*. From the microarray data, a 1.3 fold increase in

GstD1 was observed on organpipe, while a 1.4 fold increase ($P = 0.069$) was seen using real-time PCR technology. The efficiencies of the three amplicons (*Adh-2*, *GstD1* and *16 s*) were adequately high (97%, 98% and 92%, respectively).

Discussion

Significant gene expression differences were observed as a result of cactus host utilization in *Drosophila mojavensis*. These transcriptional differences arise from the different chemical environments *D. mojavensis* larvae were exposed to. There are significant differences in cactus host chemistry between the four *D. mojavensis* populations (Heed 1978; Vacek 1979; Kircher 1982). It is these chemical differences that, among other factors, could be responsible for the specialization and genetic divergence that has occurred between the four *D. mojavensis* populations. Below we discuss our results based on the transcriptional differences observed in genes involved metabolic processes and detoxification pathways, two important types of processes that we predicted would be affected by cactus host shifts.

Metabolic response to cactus host shifts

A wide array of chemical compounds are found within cactus necrosis. The composition of these compound mixtures are both a function of the cactus and of the microfauna associated with the necrosis (Starmer 1982; Starmer *et al.* 1986). There are several differences in necrosis chemistry between the four cactus host of *Drosophila mojavensis*, involving the presence or absence of alkaloids, free sugars and triterpene glycosides (Kircher 1982). It is believed that the ancestral host of *D. mojavensis* is *Opuntia* (Heed 1982), therefore the adaptation to agria and organpipe have been most recent (< 1.5 million years, Matzkin 2004; Reed *et al.* 2006). Both of the hosts used in this study (agria and organpipe) are columnar cacti and therefore have some common features. While both agria and organpipe contain very low concentrations of free sugars and large amounts of lipids and triterpene glycosides (Kircher 1982), they differ considerably in which classes of lipids and triterpene glycoside they contain (Djerassi *et al.* 1954; Djerassi & Lippman 1955; Kircher 1982). For example, the agria cactus contains specific triterpenes (possible saponins) which were used by native people to poison fish (Bravo-Hollis 1978). Additionally, it is known that the sterol macdougallin, which is a major component of the agria cactus, inhibits larval development in insects (Cespedes *et al.* 2005). The low content of free sugars in cactus tissue provides a very unique habitat for *D. mojavensis*, forcing them to obtain energy from other sources. It is known that *D. mojavensis* are able to use alcohols, such as ethanol and 2-propanol, as an alternative source of energy (Starmer *et al.* 1977; Heed 1978; Brazner *et al.* 1984). In *Drosophila melanogaster*, a large portion of the carbon flux through alcohol metabolism pathway ends up as lipids stores (Freriksen *et al.* 1991; Heinstra & Geer 1991), and this is most likely the same pathway that *D. mojavensis* utilizes.

Our results show that host shifts clearly affect the expression of genes performing a wide variety of functions (Table 1). Unfortunately, for the large majority of these loci we know very little of their biology, yet general patterns of expression effects can be concluded from this dataset. There were a total of 27 genes (17 in agria, 10 in organpipe) with some role in metabolism in which its expression was affected by host use. Genes associated with lipid metabolism were differentially regulated under both treatments (5 in agria, 4 in organpipe), which could possibly be a response to the lipid content of the two cacti (Kircher 1982) as well as a response to a difference in alcohol environment (Vacek 1979; Freriksen *et al.* 1991; Heinstra & Geer 1991). Overall, the differential expression of metabolic genes reflects the different metabolic strategies (i.e. the turning on or off of different pathways) employed for the ability to survive, develop and reproduce in their respective necrotic cactus environment.

In addition to lipids, triterpenes and glycosides there are marked differences in the alcohol composition between the necrosis of agria and organpipe (Vacek 1979). This difference is reflected in some of the loci that were differentially regulated in this study. The gene formaldehyde dehydrogenase (*Fdh*, formerly known as octanol dehydrogenase or *Odh*) was up-regulated in the native agria cactus. In *D. melanogaster*, allozyme variation at *Fdh* explains a large portion of ethanol resistance, as well as explaining variation in life history characters of flies exposed to ethanol (Bokor & Pecsenye 1998; Bokor & Pecsenye 2000). Among the genes up-regulated in the organpipe treatment was *Adh*, a locus in which extensive work has been done on its role in alcohol detoxification (Chambers 1988). In *D. mojavensis* (and its sibling species *D. arizonae*) a duplication of the *Adh* locus occurred, creating a larval/ovarian (*Adh-1*) and a late-larval/adult (*Adh-2*) expressed paralog (Batterham *et al.* 1984; Atkinson *et al.* 1988). Since their duplication the two paralogs have functionally diverged and each of them has played a role in the host adaptation of *D. mojavensis* (Matzkin & Eanes 2003; Matzkin 2004; Matzkin 2005). In this study only the expression of *Adh-2* was affected (up-regulated in organpipe). Previous studies have shown two allozyme alleles for *Adh-2* (termed Slow and Fast). The Slow allele is found at a high frequency in mainland populations, where *D. mojavensis* uses organpipe cactus, and the Fast allele is found in Baja California populations, where *D. mojavensis* uses agria cactus (Heed 1978). The two alleles have different substrate specificities, the Fast allele has a significantly greater activity on 2-propanol (Matzkin 2005), which is the alcohol found in large quantities in

agria (i.e. Baja California) (Vacek 1979). The expression differences observed at *Fdh*, *Adh-2* as well as those in lipid metabolism suggest that the environmental alcohol differences between the necrosis of agria and organpipe can elicit a substantial response.

Detoxification pathways

Another important class of genes affected by the host shift were those with a role in chemical (detoxification) and immune defence. Eleven such loci were over expressed in the alternative host, organpipe, relative to the native host, agria (3 loci). Some of these loci are known to respond to toxins such as CG33178, a Microsomal glutathione-S-transferase-like gene (*Mgstl*; up-regulated in agria), and metallothionein A (*MtnA*) and Glutathione-S-transferase D1 (*GstD1*), both up-regulated in organpipe. The *MtnA* gene has been extensively studied for its involvement in heavy metal detoxification (Posthuma & Van Straalen 1993). Although to our knowledge heavy metal concentrations of either cactus host have never been determined, heavy metal accumulation from soils has been reported in the saguaro cactus (*Carnegiea gigantea*) (Kolberg & Lajtha 1997). Therefore, the differential expression of *MtnA* could present differences in heavy metal concentrations between the two cacti. Of the two classical detoxification gene families (P450 and *Gst*) only *Gst*'s were found to be differentially regulated. In all four Sonoran desert cactophiles (*Drosophila mojavensis*, *Drosophila mettleri*, *Drosophila pachea* and *Drosophila nigrospiracula*) P450s can be induced and used in a detoxification response to alkaloid exposure (Danielson *et al*. 1998; Fogleman & Danielson 2000). Unlike the other cactus hosts, agria and organpipe do not contain alkaloids (Kircher 1982), and therefore it is possible that P450s were not induced in this study. Further, another possible reason for the lack of P450 overexpression in our study is that the array was created from a random sample of clones, consequently it is possible that they are not present in our microarray.

The other class of detoxification genes, *Gst*'s, are involved in the detoxification of a wide array of compounds and generally function on hydrophobic organic compounds (Atkins *et al*. 1993). In mammals *Gst*'s are induced during exposure to carcinogens and are involved in drug resistance (Hayes & Pulford 1995), while in insects it has been largely examined in the context of insecticide resistance (Enayati *et al*. 2005). In this study, one member of this gene family with high homology to the *Drosophila melanogaster GstD1* locus was differentially regulated. In both *D. melanogaster* and in *Anopheles gambiae GstD1 has* been implicated in the resistance of these species to the insecticide DDT [1,1,1,-trichloro-2,2-bis-(*p*-chlorophenyl)ethane] (Tang & Tu 1994; Le Goff *et al*. 2001; Ranson *et al*. 2001; Chen *et al*. 2003). In addition to being differentially expressed in organpipe, a

nucleotide sequence survey strongly suggests that *GstD1 has* gone through a period of adaptive protein evolution (Matzkin unpublished). Null alleles of the *D. melanogaster* ortholog of the locus up-regulated in agria (*Mgstl*) significantly reduce lifespan of adults, lowering the efficacy of the cellular detoxification system (Toba & Aigaki 2000).

Overall, the data suggest that several detoxification pathways are utilized when exposed to necrotic cacti. What remains unclear are the identities of the specific chemical signals that *D. mojavensis* utilizes to modify gene expression. One possible strategy that the flies could employ is to have a general detoxification response, especially for detoxification gene families that might share common biochemical properties such as *Gst*'s. For example, in vertebrates, pre-exposure to the triterpene lupeol (a component of organpipe cactus) induces high *Gst* and *catalase* expression when exposed to cytotoxins (Sudharsan *et al*. 2005). In this study both *catalase* and a *Gst* gene were up-regulated when exposed to organpipe. This might be a possible mechanism for increasing the expression of several detoxification genes when exposed to a novel chemical environment.

Novel genes associated with the cactophilic lifestyle

New genes and gene functions are created through a process of duplication, mutation and selection, and several models have been proposed to describe the mechanism for the creation, maintenance and diversification of new genes (Ohta 1988; Hughes 1994; Force *et al*. 1999; Francino 2005). One commonality between these models is that the molecular and functional divergence between paralogs is eventually driven by natural selection. In this study we found a total of 18 genes (32 genes including nonsignificant spots) that appear to be unique to *Drosophila mojavensis* (MSG's) and possibly to other members of the repleta group. These loci do not appear to be noncoding RNA and furthermore a few of these loci have been previously sequenced from an adult EST library of *D. mojavensis*. We would not expect all of our MSG's to be present in the adult EST library given that our microarray contain cDNA for all life stages and our experiment involved only third instar larvae. The molecular evolution of these loci still needs to be determined, but these data suggest that some of these loci might have arisen *de novo* in *D. mojavensis* (or its recent ancestor), and could be members of pathways that have been recruited as a response to the cactophilic lifestyle of *D. mojavensis*.

Conclusions

The central aim of this study was to examine the pattern of transcriptional regulation associated with cactus host utilization in *Drosophila mojavensis*. As predicted from the different chemical composition of the cactus host of *D. mojavensis*, a wide array of detoxification and metabolic

genes were differentially expressed as a response to a host shift. In addition, the observation of genes specific to the lineage leading to *D. mojavensis* suggest that novel gene functions have been recruited possibly as a response to the cactophilic ecology of this species. The present study focused on one population of *D. mojavensis*, future work will assess the intra — and interpopulation transcriptional variation to examine the different (or similar) detoxification strategies utilized across the entire range of this species.

Acknowledgements

Supplementary material

The supplementary material (Appendix S1) is available from http://www.blackwellpublishing.com/products/journals/suppmat/MEC/MEC3102/MEC3102sm.htm

References

Atkins WM, Wang RW, Bird AW, Newton DJ, Lu AYH (1993) The catalytic mechanism of glutathione-S-Transferase (gst): Spectroscopic determination of the Pk (A) of Tyr-9 in rat alpha-1–1 gst. *Journal of Biological Chemistry*, **268**, 19188–19191.

Atkinson PW, Mills LE, Starmer WT, Sullivan DL (1988) Structure and evolution of the *Adh* genes of *Drosophila mojavensis*. *Genetics*, **120**, 713–723.

Batterham P, Chambers GK, Starmer WT, Sullivan DT (1984) Origin and expression of an alcohol dehydrogenase gene duplication in the genus Drosophila. *Evolution*, **38**, 644–657.

Bokor K, Pecsenye K (1998) Comparative influence of *Odh* and *Adh* loci on alcohol tolerance in *Drosophila melanogaster*. *Genetics Selection Evolution*, **30**, 503–516.

Bokor K, Pecsenye K (2000) Differences in the effect of ethanol on fertility and viability components among laboratory strains of *Drosophila melanogaster*. *Hereditas*, **132**, 215–227.

Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **19**, 185–193.

Bravo-Hollis H (1978) *Las cactaceas de Mexico*. Universidad Nacional Autonoma de Mexico, Mexico D.F.

Brazner J, Aberdeen V, Starmer WT (1984) Host-plant shifts and adult survival in the cactus breeding *Drosophila mojavensis*. *Ecological Entomology*, **9**, 375–381.

Brem RB, Yvert G, Clinton R, Kruglyak L (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science*, **296**, 752–755.

Cespedes CL, Salazar JR, Martinez M, Aranda E (2005) Insect growth regulatory effects of some extracts and sterols from *Myrtillocactus geometrizans* (Cactaceae) against *Spodoptera frugiperda* and *Tenebrio molitor*. *Phytochemistry*, **66**, 2481–2493.

Chambers GK (1988) The Drosophila alcohol dehydrogenase gene enzyme system. *Advances in Genetics Incorporating Molecular Genetic Medicine*, **25**, 39–107.

Chen L, Hall PR, Zhou XE *et al.* (2003) Structure of an insect [delta]-class glutathione S-transferase from a DDT-resistant strain of the malaria vector *Anopheles gambiae*. *Acta Crystallographica Section D*, **59**, 2211–2217.

Cheung VG, Conlin LK, Weber TM *et al.* (2003) Natural variation in human gene expression assessed in lymphoblastoid cells. *Nature Genetics*, **33**, 422–425.

Danielson PB, Foster JLM, McMahill MM, Smith MK, Fogleman JC (1998) Induction by alkaloids and phenobarbital of family 4 cytochrome P450s in *Drosophila*: Evidence for involvement in host plant utilization. *Molecular and General Genetics*, **259**, 54–59.

Djerassi C, Lippman AE (1955) Terpenoids. XIII. The structures of the cactus triterpenes machaeric acid and machaerinic acid. *Journal of the American Chemical Society*, **77**, 1825–1828.

Djerassi C, Geller LE, Lemin AJ (1954) Terpenoids. VIII. The structures of the cactus triterpenes gummosogenin and longispinogenin. *Journal of the American Chemical Society*, **76**, 4089–4091.

Enayati AA, Ranson H, Hemingway J (2005) Insect glutathione transferases and insecticide resistance. *Insect Molecular Biology*, **14**, 3–8.

Etges WJ (1989) Evolution of developmental homeostasis in *Drosophila mojavensis*. *Evolutionary Ecology*, **3**, 189–201.

Etges WJ, Heed WB (1987) Sensitivity to larval density in populations of *Drosophila mojavensis*: Influences of host plant variation on components fitness. *Oecologia*, **71**, 375–381.

Feder ME, Mitchell-Olds T (2003) Evolutionary and ecological functional genomics. *Nature Review Genetics*, **4**, 649–655.

Fellows DF, Heed WB (1972) Factors affecting host plant selection in desert-adapted cactiphilic Drosophila. *Ecology*, **53**, 850–858.

Feyereisen R (1999) Insect P450 enzymes. *Annual Review of Entomology*, **44**, 507–533.

Fogleman JC, Danielson PB (2000) Analysis of fragment homology among DNA sequences from cytochrome P450 families 4 and 6. *Genetica*, **110**, 257–265.

Force A, Lynch M, Pickett FB *et al.* (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, **151**, 1531–1545.

Francino MP (2005) An adaptive radiation model for the origin of new gene functions (vol 37, pg 573, 2005). *Nature Genetics*, **37**, 911–911.

Freriksen A, Seykens D, Scharloo W, Heinstra PWH (1991) Alcohol dehydrogenase controls the flux from ethanol into lipids in Drosophila larvae: a C-13 NMR-study. *Journal of Biological Chemistry*, **266**, 21399–21403.

Gibson G (2002) Microarrays in ecology and evolution: a preview. *Molecular Ecology*, **11**, 17–24.

Gibson G, Wolfinger R (2004) Gene expression profiling using mixed models. In: *Genetic Analysis of Complex Traits Using SAS* (ed. Saxton AM), pp. 251–278. SAS Users Press, Cary.

Gilad Y, Oshlack A, Smyth GK, Speed TP, White KP (2006) Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature*, **440**, 242–245.

Hayes JD, Pulford DJ (1995) The glutathione S-Transferase supergene family: Regulation of gst and the contribution of the isoenzymes to cancer chemoprotection and drug resistance. *Critical Reviews in Biochemistry and Molecular Biology*, **30**, 445–600.

Heed WB (1978) Ecology and genetics of sonoran desert Drosophila. In: *Ecological Genetics: the Interface* (ed. Brussard PF), pp. 109–126. Springer-Verlag.

Heed WB (1982) The origin of Drosophila in the sonoran desert. In: *Ecological Genetics and Evolution: the Cactus-Yeast-Drosophila Model System* (ed. Barker JSF, Starmer WT), pp. 65–80. Academic Press, New York.

Heinstra PWH, Geer BW (1991) Metabolic control analysis and enzyme variation—nutritional manipulation of the flux from ethanol to lipids in Drosophila. *Molecular Biology and Evolution*, **8**, 703–708.

Hughes AL (1994) The evolution of functionally novel proteins after gene duplication. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **256**, 119–124.

Jin W, Riley RM, Wolfinger RD *et al.* (2001) The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nature Genetics*, **29**, 389–395.

Kircher HW (1982) Chemical composition of cacti and its relationship to sonoran desert Drosophila. In: *Ecological Genetics and Evolution: the Cactus-Yeast-Drosophila Model System* (ed. Barker JSF, Starmer WT), pp. 143–158. Academic Press, New York.

Kolberg K, Lajtha K (1997) Population dynamics and trace metal biogeochemistry of the saguaro cactus (*Carnegiea gigantea*) in the sonoran desert. *Ecoscience*, **4**, 223–231.

Le Goff G, Bride JM, Cuany A, Berge JB, Amichot M (2001) The sibling species *Drosophila melanogaster* and *Drosophila simulans* differ in the expression profile of glutathione S-transferases. *Comparative Biochemistry and Physiology B-Biochemistry and Molecular Biology*, **129**, 837–841.

Matzkin LM, Eanes WF (2003) Sequence variation of alcohol dehydrogenase (*Adh*) paralogs in cactophilic Drosophila. *Genetics*, **163**, 181–194.

Matzkin LM (2004) Population genetics and geographic variation of alcohol dehydrogenase (*Adh*) paralogs and glucose-6-phosphate dehydrogenase (*G6pd*) in *Drosophila mojavensis*. *Molecular Biology and Evolution*, **21**, 276–285.

Matzkin LM (2005) Activity variation in alcohol dehydrogenase paralogs is associated with adaptation to cactus host use in cactophilic *Drosophila*. *Molecular Ecology*, **14**, 2223–2231.

Morley M, Molony CM, Weber TM *et al.* (2004) Genetic analysis of genome-wide variation in human gene expression. *Nature*, **430**, 743.

Nuzhdin SV, Wayne ML, Harmon KL, McIntyre LM (2004) Common pattern of evolution of gene expression level and protein sequence in *Drosophila*. *Molecular Biology and Evolution*, **21**, 1308–1317.

Ohta T (1988) Evolution by gene duplication and compensatory advantageous mutations. *Genetics*, **120**, 841–847.

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

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time rt-PCR. *Nucleic Acids Research*, **29**, 2002–2007.

Pfaffl MW, Horgan GW, Dempfle L (2002) Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic Acids Research*, **30**, e36.

Posthuma L, Van Straalen NM (1993) Heavy-metal adaptation in terrestrial invertebrates: a review of occurrence, genetics, physiology and ecological consequences. *Comparative Biochemistry and Physiology C-Pharmacology Toxicology and Endocrinology*, **106**, 11–38.

Ranson H, Rossiter L, Ortelli F *et al.* (2001) Identification of a novel class of insect Glutathione S-transferases involved in resistance to DDT in the malaria vector *Anopheles gambiae*. *Biochemical Journal*, **359**, 295–304.

Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL (2003) Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science*, **300**, 1742–1745.

Ranz JM, Machado CA (2006) Uncovering evolutionary patterns of gene expression using microarrays. *Trends in Ecology and Evolution*, **21**, 29–37.

Reed LK, Nyboer M, Markow TA (2006) Evolutionary relationships of *Drosophila mojavensis* geographic host races and their sister species *Drosophila arizonae*. *Molecular Ecology*.

Rifkin SA, Kim J, White KP (2003) Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nature Genetics*, **33**, 138–144.

Rockman MV, Wray GA (2002) Abundant raw material for cis-regulatory evolution in humans. *Molecular Biology and Evolution*, **19**, 1991–2004.

Ruiz A, Heed WB (1988) Host-plant specificity in the cactophilic *Drosophila mulleri* species complex. *Journal of Animal Ecology*, **57**, 237–249.

Starmer WT, Heed WB, Rockwood-Sluss ES (1977) Extension of longevity in *Drosophila mojavensis* by environmental ethanol: Differences between subraces. *Proceedings of the National Academy of Sciences of the USA*, **74**, 387–391.

Starmer WT (1982) Analysis of the community structure of yeasts associated with the decaying stems of cactus.1. *Stenocereus gummosus*. *Microbial Ecology*, **8**, 71–81.

Starmer WT, Gilbert DG (1982) A quick and reliable method for sterilizing eggs. *DIS*, **58**, 170–171.

Starmer WT, Barker JSF, Phaff HJ, Fogleman JC (1986) Adaptations of Drosophila and yeasts — their interactions with the volatile 2-propanol in the cactus microorganism Drosophila model system. *Australian Journal of Biological Sciences*, **39**, 69–77.

Storey JD, Akey JM, Kruglyak L (2005) Multiple locus linkage analysis of genomewide expression in yeast. *PLoS Biology*, **3**.

Storey J, Tibshirani R (2003) Statistical significance for genome-wide studies. *Proccedings of the National Academy of Sciences, USA*, **100**, 9440–9445.

Sudharsan PT, Mythili Y, Selvakumar E, Varalakshmi P (2005) Cardioprotective effect of pentacyclic triterpene, lupeol and its ester on cyclophosphamide-induced oxidative stress. *Human and Experimental Toxicology*, **24**, 313–318.

Tang AH, Tu CPD (1994) Biochemical characterization of *Drosophila* Glutathione S-Transferases D1 and D21. *Journal of Biological Chemistry*, **269**, 27876–27884.

Toba G, Aigaki T (2000) Disruption of the Microsomal glutathione S-transferase-like gene reduces life span of *Drosophila melanogaster*. *Gene*, **253**, 179.

Vacek DC (1979) The microbial ecology of the host plants of *Drosophila mojavensis*. PhD, University of Arizona.

Wolfinger RD, Gibson G, Wolfinger ED *et al.* (2001) Assessing gene significance from cdna microarray expression data via mixed models. *Journal of Computational Biology*, **8**, 625–637.

Yvert G, Brem RB, Whittle J *et al.* (2003) Trans-acting regulatory variation in saccharomyces cerevisiae and the role of transcription factors. *Nature Genetics*, **35**, 57–64.

Luciano Matzkin is a PERT Postdoctoral Fellow interested in the genetical basis of adaptation focusing on the transcriptional and molecular variation of ecologically interesting organisms (cactophilic *Drosophila*). Thomas Watts is a Senior Research Specialist and Kristin Sweetser is Research Assistant in the Markow lab. Benjamin Bitler was an undergraduate Research Assistant in the Markow lab, currently is a Cancer Biology doctoral student at the University of Arizona. Carlos Machado is an Assistant Professor interested in population genetics, phylogenetics, plant-insect coevolution, and the genetic basis of speciation in insects. Therese Markow is a Regents Professor who has extensively worked on the ecology and evolution of cactophilic *Drosophila*.

Author Query Form

Dear Author,

    During the copy-editing of your paper, the following queries arose. Please respond to these by marking up your proofs with the necessary changes/additions. Please write your answers on the query sheet if there is insufficient space on the page proofs. Please write clearly and follow the conventions shown on the attached corrections sheet. If returning the proof by fax do not write too close to the paper's edge. Please remember that illegible mark-ups may delay publication.

Many thanks for your assistance.

No.	Query	Remarks
1	Is BLASTX a computer program? Then it should be formatted with small capitals here and elsewhere. Please check.	
2	This is BLAT in the original text. I've changed it to BLAST. Please check.	

MARKED PROOF

Please correct and return this set

Please use the proof correction marks shown below for all alterations and corrections. If you wish to return your proof by fax you should ensure that all amendments are written clearly in dark ink and are made well within the page margins.

Instruction to printer	Textual mark	Marginal mark
Leave unchanged	··· under matter to remain	Stet
Insert in text the matter indicated in the margin	ʎ	New matter followed by ʎ
Delete	⊢ through matter to be deleted	ꝰ
Delete and close up	⊖ through matter to be deleted	ꝰ
Substitute character or substitute part of one or more word(s)	/ through letter or ⊢ through word	New letter or new word
Change to italics	— under matter to be changed	⊔⊔⌐
Change to capitals	≡ under matter to be changed	≡
Change to small capitals	= under matter to be changed	=
Change to bold type	∿ under matter to be changed	∿
Change to bold italic	≋ under matter to be changed	⊔⊔⌐
Change to lower case	Encircle matter to be changed	≢
Change italic to upright type	(As above)	⊔⌐
Insert 'superior' character	/ through character or ʎ where required	ˠ under character e.g. ˠ
Insert 'inferior' character	(As above)	ʎ over character e.g. ʎ̃
Insert full stop	(As above)	⊙
Insert comma	(As above)	,
Insert single quotation marks	(As above)	ˠ and/or ˠ
Insert double quotation marks	(As above)	ˠ and/or ˠ
Insert hyphen	(As above)	⊞
Start new paragraph	⌐	⌐
No new paragraph	⌐	⌐
Transpose	⊔⊓	⊔⊓
Close up	linking ⊂ letters	⊂
Insert space between letters	ʎ between letters affected	#
Insert space between words	ʎ between words affected	#
Reduce space between letters	⊤ between letters affected	⊤
Reduce space between words	⊤ between words affected	⊤