

A change of expression in the conserved signaling gene *MKK7* is associated with a selective sweep in the western house mouse *Mus musculus domesticus*

B. HARR,* C. VOOLSTRA,* T. J. A. J. HEINEN,* J. F. BAINES,* R. ROTTSCHIEDT,* S. IHLE,* W. MÜLLER,† F. BONHOMME‡ & D. TAUTZ*

*Institute for Genetics, Department of Evolutionary Genetics, Köln, Germany

†German Research Centre for Biotechnology, Braunschweig, Germany

‡Génome, Populations, Interactions, Adaptation, Université de Montpellier II, France

Keywords:

adaptation;
gene expression;
microarray;
mus musculus;
selective sweep.

Abstract

Changes in gene expression are known to occur between closely related species, but it is not yet clear how many of these are due to random fixation of allelic variants or due to adaptive events. In a microarray survey between subspecies of the *Mus musculus* complex, we identified the mitogen-activated protein-kinase-kinase *MKK7* as a candidate for change in gene expression. Quantitative PCR experiments with multiple individuals from each subspecies confirmed a specific and significant up-regulation in the testis of *M. m. domesticus*. Northern blot analysis shows that this is due to a new transcript that is not found in other tissues, nor in *M. m. musculus*. A *cis-trans* test via allele specific expression analysis of the *MKK7* gene in F1 hybrids between *domesticus* and *musculus* shows that the expression change is mainly caused by a mutation located in *cis*. Nucleotide diversity was found to be significantly reduced in a window of at least 20 kb around the *MKK7* locus in *domesticus*, indicative of a selective sweep. Because the *MKK7* gene is involved in modulating a kinase signalling cascade in a stress response pathway, it seems a plausible target for adaptive differences between subspecies, although the functional role of the new testis-specific transcripts will need to be further studied.

Introduction

It has long been speculated that changes in gene expression are a major contributor to evolution (King & Wilson, 1975; Davidson, 2001; Wray *et al.*, 2003). However, solid evidence for this inference has only recently accumulated. Across yeast, *Drosophila* and humans up to a quarter of all genes analysed show differences in expression among individuals of the same species (Jin *et al.*, 2001; Cowles *et al.*, 2002; Enard *et al.*, 2002; Fay *et al.*, 2004). The differences are heritable, i.e. are not

simply due to random fluctuations and it is possible to map the regions that affect gene expression (Schadt *et al.*, 2003; Stamatoyannopoulos, 2004). Thus, it is clear that gene expression could be a target of selection. However, global analysis of gene expression differences in primates and humans has suggested that most regulatory difference might accumulate in a neutral fashion (Khaitovich *et al.*, 2004). Other studies have shown that there is a strong stabilizing selection component (Denver *et al.*, 2005; Jordan *et al.*, 2005; Lemos *et al.*, 2005). This leaves open how often one should expect positive selection to affect transcription of genes. Here, we combine an analysis of gene expression differences between subspecies of the house mouse *Mus musculus* with an analysis of signatures of selective sweeps (Schlötterer, 2003) to differentiate between an adaptive and a neutral scenario.

Correspondence: Bettina Harr, Institute for Genetics, Department of Evolutionary Genetics, Köln, Germany.
Tel.: +49 221 470 2324; fax: +49 221 470 5975;
e-mail: harrb@uni-koeln.de

The close relationship among house mice, which naturally hybridize in many places where they come into contact (Boursot *et al.*, 1993; Guenet & Bonhomme, 2003) provides a powerful system to study the genetics of underlying recent adaptive divergence, including divergence in gene expression.

We focus on a tissue-specific expression difference at the 'Mitogen-Activated Protein Kinase Kinase 7' (*MKK7* or *Map2k7*) gene and confirm that this is associated with a *cis*-regulatory change, which most likely has been affected by positive selection.

Materials and methods

Animal materials

Microarray

For the pilot microarray study individuals from all four subspecies of *Mus musculus* were used: *M. m. musculus* (strain MGA from Georgia), *M. m. domesticus* (strain DEB, from Spain), *M. m. castaneus* (strain CIM from India) and *M. musculus* ssp (strain BID from Iran). Each animal was derived from wild caught animals that have been kept randomly bred, but in closed colonies for several generations in the genetic repository of GPIA lab in Montpellier. All animals were kept under standard laboratory conditions and killed at approximately 6 weeks of age. A single male was selected from each strain for RNA extraction.

Quantitative real time PCR

The qRT PCR gene was performed on mostly independent samples of all *musculus* subspecies. For subspecies *domesticus* and *musculus* we live trapped six individuals in Germany (around Cologne) and six individuals in the Czech Republic (around Prague), respectively, and transferred them to the lab where they were kept for 3–5 days under common conditions. All individuals were independently collected using a sampling strategy that eliminates the chance of collecting related animals by ensuring a minimal distance of 1 km between each collected mouse (Ihle *et al.*, 2006). The six animals from subspecies ssp and the three animals from subspecies *castaneus* were taken from the GPIA lab in Montpellier (see Table S1 for their sampling locations). All animals were killed in the same way as the animals for the microarray experiments.

Pyrosequencing

A wild-derived inbred strain of *musculus* (strain JPC 2821, from the Czech Republic) and *domesticus* (strain JPC 2705, from Germany) were used to generate F1 hybrids between these two subspecies. Both strains were kindly provided by J. Pialek. These two strains were also used for a Northern blot of total RNA (see below). For all experiments RNA was extracted from testis using Trizol (Invitrogen, USA) following the manufacturer's protocol.

Microarray analysis

Expression profiles were determined for 12 000 mouse transcripts using the MG_U74v2 Affymetrix® Genechip, MD, USA. The quality and integrity of the total RNA isolated from testis tissue of each mouse was controlled by running all samples on an Agilent Technologies 2100 Bioanalyser (Agilent Technologies; Waldbronn, Germany). For biotin-labelled target synthesis starting from 8 µg of total RNA, reactions were performed using standard protocols supplied by the manufacturer (Affymetrix; Santa Clara, CA). After hybridisation the GeneChips were washed, stained with SA-PE and read using an Affymetrix GeneChip fluidic station and scanner. For each of the four individuals (one per subspecies) two replicates were performed where the cDNA was labelled and hybridized twice independently.

Microarray data analysis

Raw signal intensities were summarized according to two different algorithms, standard Affymetrix MA Suite 5.0 and *Rma* (Irizarry *et al.*, 2003), which consider only the perfect match oligos. Normalized (and log₁₀ transformed in the case of MA Suite 5.0) signal intensities as determined from both methods were used for statistical analyses. MA Suite 5.0 expression values were submitted to the Gene Expression Omnibus (accession number GSE2471). A one-way ANOVA was performed among subspecies with signal intensities as the dependent variable. We sought to isolate the subspecies contribution to a significant effect by means of *post hoc* Tukey's Honest Significant Difference (HSD) tests. Subspecies-specific differences in expression level were then extracted by conditioning on outcomes of the HSD tests to cases where the expression in one subspecies differed significantly from all others, but where these did not differ significantly among each other. The magnitude of change (i.e. 'fold change') for a subspecies-specifically expressed gene was calculated as the ratio of the signal intensity in this particular subspecies (averaged across the two replicates) to the average signal intensities across the remaining subspecies.

All data processing and statistical analyses were performed using *Perl* scripts (available from the authors upon request) and the statistical language *R*.

Quantitative real time PCR

A candidate gene from the microarray experiment was the *MKK7* gene, which showed striking expression differences in testis between *domesticus* and all other subspecies of house mice, including *musculus*. To verify this, we analysed an independent and larger sample using quantitative Real time PCR (qRT-PCR). The qRT-PCR was also performed on two genes (*Snape2* and *Cttn*) located directly downstream of *MKK7*.

Prior to the qRT-PCR, RNA was extracted from testis and reverse transcribed using the ThermoScript Reverse Transcriptase Kit (Invitrogen) according to the manufacturer's protocol. All Real Time PCR amplicons were obtained from ABI's 'Assay on Demand' selection and amplifications were performed according to the manufacturer's instructions using a 1 : 10 dilution of the cDNA per reaction. For each individual we performed a single cDNA synthesis, which was used in triplicate in the qRT-PCR. C_T values from each qRT-PCR reaction at the target gene were standardized relative to the house-keeping gene analysed in the same sample. Two different house-keeping genes were used (*Gapdh*, glyceraldehyd-3-phosphate dehydrogenase, Assay ID Mm99999915_g1 and *Tbp*, TATA box binding protein, Assay ID Mm00446973_m1). The target genes had the following Assay ID: *MKK7*: Mm00488759_m1, *Snapc2*: Mm00506001_m1 and *Cttn*: Mm01217030_g1. The same qRT-PCR set up was also used to assay expression levels at the *MKK7* gene in mouse strains JPC 2821 and JPC 2705, which were used in reciprocal crosses to generate F1 hybrids (see *Pyrosequencing of F1 hybrids*).

Quantitative real time PCR analysis

For each qRT-PCR reaction the C_T value for the house-keeping gene was subtracted from the C_T value of the target gene, yielding the ΔC_T value. These values were then averaged across the triplicate samples within an individual, yielding the average ΔC_T value. In analogy to the microarray data analysis, we performed a one way ANOVA using individuals within a subspecies as replicate, subspecies as fixed effect and average ΔC_T values as independent variable. The *post hoc* Tukey's HSD test was used again to identify significant pair-wise comparisons.

The following formula was used to calculate the difference in expression level (i.e. *fold change*) between *domesticus* and all other subspecies of house mouse from the ΔC_T values:

$$\text{Fold change} = 2^{(\overline{\Delta C_T - dom} - \overline{\Delta C_T - other})},$$

where $\overline{\Delta C_T - dom}$ is the average ΔC_T value across the *domesticus* individuals and $\overline{\Delta C_T - other}$ is the average ΔC_T value across the remaining subspecies (averaged first across individuals within subspecies and then across subspecies).

Pyrosequencing analysis of RNA from F1 hybrids

To characterize the underlying genetic basis of the expression change between *musculus* and *domesticus* at *MKK7*, we measured the allele-specific expression level of *MKK7* in F1 hybrids using pyrosequencing of cDNA, essentially following the method of Wittkopp *et al.* (2004). Pyrosequencing is a sequencing-by-synthesis method in which a cascade of enzymatic reactions yields detectable light proportional to incorporated nucleotides

(Ahmadian *et al.*, 2000,2006;). Thus, in an F1 hybrid, allele-specific expression can be measured by quantifying the amount of nucleotide incorporated across a single nucleotide difference distinguishing the two parental alleles (Wittkopp *et al.*, 2004). F1 hybrids were produced from a cross between a wild-derived inbred strain of *musculus* (strain JPC 2821) and *domesticus* (strain JPC 2705). To identify an SNP differentiating the parental alleles we amplified and sequenced the whole coding sequence of *MKK7* in the two parental strains. One SNP was identified and used to design a pyrosequencing assay using the program SNP PRIMER DESIGN FROM PYROSEQUENCING AB VERSION 1.0.1 (available at <http://primerdesign.pyrosequencing.com/jsp/TemplateInput.jsp>). The composition of the pyrosequencing assay, the position of the coding SNP and primer sequences are shown in Fig. S1.

Two reciprocal crosses were set up to generate the F1. From each cross we selected three male F1 mice, which were killed at the age of 8 weeks. Two male mice from the parental strain JPC 2821 and three male mice from the parental strain JPC 2705 were killed at the same age. RNA extraction from testis and cDNA synthesis was performed as before (see *Quantitative Real Time PCR*). Four independent cDNA syntheses were performed for each F1 individual (i.e. 4 cDNA reactions \times 6 F1 hybrid animals = 24 cDNA reactions in total). To correct for potential amplification biases between the subspecies-specific alleles in F1 hybrids we also simultaneously extracted DNA along with the RNA of each F1 individual (Wittkopp *et al.*, 2004), following the manufacturer's (Trizol, Invitrogen) protocol. PCR products from F1 hybrid DNA were pyrosequenced in duplicate, the PCR products from the cDNA of the 24 F1 hybrid animals were pyrosequenced once. In contrast to Wittkopp *et al.* (2004), we performed highly sensitive qRT-PCR to determine expression differences between the parental alleles instead of pyrosequencing on DNA/RNA extracted from mixtures of parental tissue. For the qRT-PCR the cDNA from each parental animal (two males of JPC 2821 and three males of JPC 2705) was used in triplicate.

The pyrosequencing reaction was set up according to the manufacturer's instructions.

If differential expression between the subspecies can be attributed to *trans* effects, we would expect both parental alleles to be expressed equally in the F1 hybrid (i.e. $\frac{\text{exp level}_{\text{Parent1}}}{\text{exp level}_{\text{Parent2}}} = 1$). Deviations from a ratio of unity are indicative of at least partial *cis* effects.

DNA sequencing

The sequence of the 5' upstream region of the *MKK7* (*NM_011944*) gene was downloaded from ENSEMBL (<http://www.ensembl.org>). Only ~2.7 kb of upstream sequence separate the coding region of the *MKK7* gene from the coding region of the neighbouring gene *NM_028175*, a transcript of unknown function. PCR

primers were designed to amplify 2.6 kb of the 5' upstream flanking sequence of the *MKK7* gene.

In addition to the *MKK7* gene, we also sequenced four additional fragments located within a 47.5 kb region around the *MKK7* gene. Each fragment consisted of 700–1400 bp of noncoding sequence (either 5' flanking genes or intronic region). Table S2 gives the location of the sequenced fragments with respect to the genes in this region. All PCR products were directly sequenced in eight individuals (16 alleles) from a *musculus* population (from Kazakhstan), eight individuals (16 alleles) from a *domesticus* population (Germany) and one individual of the outgroup species *M. spicilegus*. As before, all individuals used for sequencing were collected from the wild using a sampling strategy that eliminates the chance of collecting related animals (Ihle *et al.*, 2006). The sampling area was similarly sized for both subspecies (40 × 70 km) and we adhered to a minimal distance of 1 km between each collected mouse.

For each locus a 50 µL PCR containing 100 ng of genomic DNA, 1.5 mM MgCl₂, 200 µM dNTPs, 1 µM of each primer and 0.5 U *Taq* polymerase (Eppendorf Master Taq) was performed. A typical cycling profile consisted of 30 cycles with 50 s at 94 °C, 50 s at 60 °C and 2 min at 72 °C. PCR products were purified using 96 well plates (Millipore Montag , IL, USA) according to the manufacturer's protocol. PCR products were sequenced in both directions using the BigDye sequencing chemistry on an ABI3700 automated sequencer. In the case of individuals heterozygous for an insertion/deletion, additional sequencing primers were designed to close the gaps. All PCR and sequencing primers are given in Table S3. Sequencing reactions were purified using Sephadex columns (Millipore, USA). Sequences have been submitted to GenBank with the accession numbers DQ078098-DQ078114 (5' upstream flanking sequence of *MKK7*) and DQ223578 - DQ223644.

DNA sequence data analysis

DNA sequences were aligned using *CLUSTALW* (Thompson *et al.*, 1994). We calculated four summary statistics of within population diversity levels: Watterson's θ_w (Watterson, 1975), based on the number of segregating (polymorphic) sites in the sample; π (Nei & Li, 1979), the average number of pair-wise differences in the sample; H (Fay & Wu, 2000), a statistic that contrasts intermediate frequency and high-frequency derived (with respect to the out group) variants, and D (Tajima, 1989), a statistic that contrasts intermediate and low frequency variants in a sample. Probabilities for Tajima's D and Fay and Wu's H were obtained using coalescent simulations under a standard neutral model and under the assumption of no recombination. All polymorphism indices and their significance were calculated using the program *COMPUTE* (Thornton, 2003).

Estimation of parameters of a selective sweep model

Tajima's D and Fay and Wu's H test of neutrality incorporate information of a single locus only. In contrast, the 'composite likelihood ratio' (CLR) test computes the likelihood of a selective sweep model vs. a neutral model for a dataset based on multiple linked loci along a recombining chromosome (Kim & Stephan, 2002). The test calculates the likelihood of observing a given number of derived variants at a polymorphic site under the selective sweep model (L_1 in Kim & Stephan, 2002) and compares this to that expected under a standard neutral model (L_0). The resulting maximum likelihood ratio from the observed data was compared with the distribution of ratios obtained from 1000 simulations of neutral data sets. Neutrality can be rejected if the LR value of the observed data is significantly larger than LR values generated under neutrality. The one-tailed P -value corresponds to the proportions of simulated neutral datasets having larger LR values than the observed dataset.

Since levels of heterozygosity were relatively low over the whole region in the *M. m. domesticus* population (see Results), we used a modified version of test A of (Kim & Stephan, 2002), where neutral data sets were generated conditioned on the observed number of segregating sites. For the *domesticus* case, where the CLR test rejected a neutral equilibrium model, we also analysed the observed data using a goodness of fit (GOF) test (Jensen *et al.*, 2005). This test compares the fit of a selective sweep model to the data against that of a generalized alternative model, with the goal of differentiating rejections due to selective sweeps from those potentially due to demographic history. GOF values obtained from the observed polymorphism data were compared to those estimated from 1000 data sets simulated under a selective sweep scenario.

Northern blot

Various alternative transcripts are described for the *MKK7* gene (Tournier *et al.*, 1997,1999). To identify the transcript that might be involved in the up-regulation at *MKK7* we performed a Northern blot. Ten microgram of total RNA isolated from several tissues (brain, liver, spleen, kidney, thymus, muscle, heart, testis, lung, blood, gut and skin) from a wild-derived strain of *M. m. musculus* (strain JPC 2821) and a wild-derived strain of *M. m. domesticus* (JPC 2705) was run on a denaturing agarose gel, blotted onto a nylon membrane (Pall) and hybridized against a 579 bp *MKK7*-specific probe. The probe was designed to cover the whole 'core'-region, i.e. all exons that are common to all known alternative transcripts of the gene. Since the gene is ubiquitously expressed (Su *et al.*, 2004) and only the conserved core of the transcript is targeted we selected one tissue (gut) to generate the probe by PCR amplification from *M. m. domesticus* using

the primers 5'-gagcatcgagattgaccaga-3' and 5'-gctcgatgtcatagtcagg-3'. The PCR fragment was radioactively labelled with [α - 32 P]-CTP (50 μ Ci, Hartmann) using the RediprimeII Random Labeling System (Amersham, USA). A Phosphoimager system (Fuji, USA) was used to detect the specific signals.

Results

Microarray

The microarray experiments were only performed for replicates of a single animal of each subspecies and we used them solely to identify candidate genes for further study. We found similar numbers of genes in each subspecies that were differentially expressed in one subspecies compared to all others. The numbers ranged from 15 in *domesticus* to 39 in *musculus* at a *P*-value cutoff of 0.005. When only genes with a minimum of twofold difference in expression level were included these numbers dropped to 10 (*domesticus*) and 17 (*musculus*) (Table 1). Similar results were obtained when the microarray raw data were analysed with the *RMA* algorithm (Irizarry *et al.*, 2003, data not shown). The list of subspecies-specific candidate genes from the microarray is available as Tables S4 (a)–(d).

Confirmation of the microarray result by real time PCR

Given that a large between-individual variance in gene expression has been observed in previous experiments (Oleksiak *et al.*, 2002), differential expression at a gene was conceivably specific to the particular individual rather than a subspecies-wide phenomenon. To examine this, we performed qRT-PCR using multiple and independently collected animals. Initially, four loci were selected for confirmation of the microarray results using quantitative real time PCR. Two of these were over-expressed specifically in *musculus* and two genes were over-expressed in *domesticus*. For both subspecies, we randomly selected two fully annotated genes with a large difference in expression. Out of the four genes, we could only confirm the expression difference at the *MKK7* gene (see below). The genes that were tested but could not be confirmed were *Agt*, *Igf2*, *Pdk3* (data not shown).

Table 1 Number of candidate genes subspecies specifically expressed at significance threshold *P* < 0.005.

	<i>M. m. castaneus</i>	<i>M. m. musculus</i>	<i>M. m. domesticus</i>	<i>M. m. ssp.</i>
Number of genes	26	39	15	26
Number of genes > two-fold change	15	17	10	16

M. m. domesticus specific up-regulation of *MKK7*

Irrespective of whether the microarray raw data were summarized according to the MA Suite 5.0 or *RMA* algorithm, the microarray pilot experiment suggested that the *MKK7* gene is significantly differentially expressed in the *M. m. domesticus* individual relative to the individuals from the other subspecies ($F_{4,4} = 24.7$, $P = 0.0048$, MA Suite 5.0; $F_{4,4} = 99.8$, $P < 0.0001$, *RMA*). The results of qRT-PCR reveal that *MKK7* gene is significantly differentially expressed at the subspecies level rather than individual level ($F_{3,17} = 11.99$, $P < 0.0001$). The result is independent of the house-keeping gene used as a reference (*Tbp* or *Gapdh*, Fig. 1) and also holds true if the raw expression levels (un-normalized relative to a house-keeping gene) at *MKK7* are used. Pair-wise *post hoc* tests (Table 2) showed that only comparisons involving the *domesticus* subspecies are significant ($P < 0.05$). This also holds whether or not the data are normalized, as well as which house keeping gene is used for normalization (data not shown). The direction of the change was the same as in the microarray experiment and the magnitude of change was also very similar (11-fold up-regulation of *MKK7* in *domesticus* relative to other subspecies in the microarray experiment versus 5.3-fold (*Tbp*), 3.6-fold (*Gapdh*) and 4.95-fold (un-normalized) in the qRT-PCR).

Cis vs. *trans*-regulatory effects at *MKK7*

To characterize the underlying genetic basis of the change in expression we compared the relative abundance of subspecies-specific *MKK7* transcripts in F1

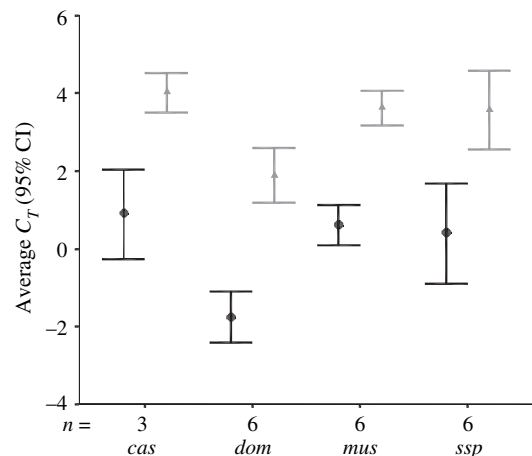


Fig. 1 Average AC_T values (and 95% confidence intervals) for the different subspecies of house mouse as measured in the qRT-PCR using Taq-Man probes. The number of individuals assayed per subspecies is given along the X-axis. Black bars: *MKK7* normalized relative to house-keeping gene *Tbp*. Grey bars: *MKK7* gene normalized relative to house-keeping gene *Gapdh*.

Table 2 *P*-values from pairwise comparisons between subspecies of *M. musculus* as obtained from one way ANOVA post hoc tests on the qRT-PCR data (average ΔC_T) from the *MKK7* gene.

Subspecies	Subspecies	<i>P</i> -value
<i>cas</i>	<i>dom</i>	0.027
	<i>mus</i>	0.85
	<i>ssp</i>	0.99
<i>dom</i>	<i>cas</i>	0.027
	<i>mus</i>	0.047
	<i>ssp</i>	0.012
<i>mus</i>	<i>cas</i>	0.85
	<i>dom</i>	0.047
	<i>ssp</i>	0.91
<i>ssp</i>	<i>cas</i>	0.99
	<i>dom</i>	0.012
	<i>mus</i>	0.91

hybrids generated from wild-derived inbred strains of *M. m. musculus* and *M. m. domesticus* employing a differential transcript quantification assay based on a discriminatory SNP via pyrosequencing. The parental strains (strains JPC 2821 and JPC 2705) show a 7.4-fold difference in expression level in the qRT-PCR, which is within the same order of magnitude as the expression difference observed in the qRT-PCR in natural populations and microarray experiments described above. The difference between the parental strains in *MKK7* expression is highly significant in a *t*-test ($t_3 = 26.6$, $P < 0.0001$).

We determined the ratios of the *domesticus/musculus* allele-specific expression in six F1's by pyrosequencing. Averaged across replicate cDNA syntheses within an individual, the ratio was 4.99 (range 3.8–5.96). An example is shown in Fig. 2, left panel. These ratios are significantly different from 1 (one sample *t*-test, $t_5 = 12.7$, $P < 0.0001$). When corrected for the slight amplification bias between the *M. m. domesticus*- and *M. m. musculus*-derived alleles, as estimated from pyrosequencing genomic DNA from F1 hybrid individuals (Fig. 2, right panel) the average ratio was 4.72 and still highly significantly different from the expected ratio of 1 under pure *trans* effects. If the change is caused by a pure *cis* effect the ratio of the *domesticus/musculus* allele-specific

expression level in the F1 hybrid should equal the ratio of the expression in pure species. This is not the case. The ratio in the F1 hybrid is approximately 70% of the ratio of the pure species as determined the qRT-PCR experiment. Both *cis* and *trans* effects could therefore play a role, although the ratios in the hybrids and the pure species were measured using different methods (pyrosequencing vs. direct qRT-PCR). It is probable that qRT-PCR is more sensitive in measuring expression differences than pyrosequencing, in which case the result would be biased more towards a *trans* effect. In any case, our results show that the expression change observed at *MKK7* between *domesticus* and *musculus* can be considered in large part due to *cis*-regulatory effects.

DNA sequence analysis

Compared to *musculus* the level of nucleotide polymorphism at *MKK7* was strongly reduced in *domesticus*. In the 2.6 kb region sequenced, only three polymorphic sites were recovered. As shown in Table 3, π was 22-fold lower at *MKK7* (fragment 2) in *domesticus* ($\pi_{\text{dom}} = 0.0002$) relative to *musculus* ($\pi_{\text{mus}} = 0.0044$). Similarly, θ_w was ~ 10 times lower in *domesticus*. The extremely low level of polymorphism at the *MKK7* locus in *domesticus* could be a signature of a recent selective sweep in this subspecies.

We performed several tests of the selective sweep hypothesis at *MKK7* in *domesticus*. We first tested the prediction that a locus that has recently been affected by a selective sweep should show an excess of rare alleles compared to a neutral model (Tajima, 1989; Braverman *et al.*, 1995, Fu, 1997) and an excess of high frequency derived alleles (Fay & Wu, 2000). Indeed, rare alleles (as measured by Tajima's *D*) and derived alleles at high frequency (as measured by Fay and Wu's *H*) are in excess at the *MKK7* gene in *domesticus*, though only Fay and Wu's *H* is significant as determined by simulations under a standard neutral model. The slight positive values of Tajima's *D* and Fay and Wu's *H* in *musculus* (Table 3) are inconsistent with a selective sweep in this subspecies. Thus, the predictions of the selective sweep are only fulfilled in *domesticus*.

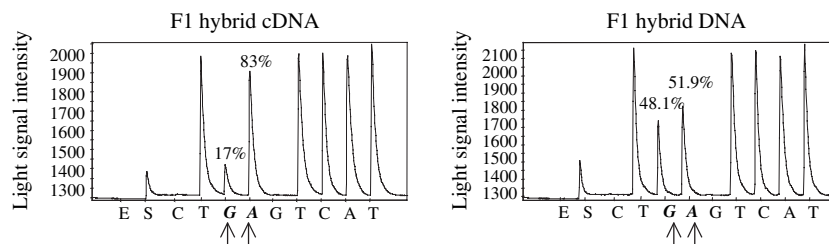


Fig. 2 Light signal intensity plot ('pyrogram') for the different nucleotides incorporated in the pyrosequencing reaction of F1 hybrid cDNA and DNA. The peak height in the pyrogram is directly proportional to the expression level of each allele. The arrows indicate the polymorphic site between the *domesticus* and the *musculus* allele (G: *musculus* allele-specific expression level, A: *domesticus* allele-specific expression level).

Fragment	1	2	3	4	5
Midpoint of fragment	3581150	3592462	3607209	3611897	3628673
Domesticus					
Number of alleles	16	16	16	16	16
Number of sites	660	2540	646	554	699
Number of segregating sites	2	3	0	0	2
π	0.0010	0.0002	0	0	0.0009
θ_W	0.0009	0.00036	0	0	0.0009
Tajima's D	0.2	-1.35	0	0	0.13
Probability Tajima's D \leq observed	0.62	0.12	1	1	0.49
FW-H	0.25	-3.36	0	0	0.2
Probability FW-H \leq observed	0.92	0.012*	1	1	0.77
Musculus					
Number of alleles	16	16	14	16	16
Number of sites	652	2614	638	553	670
Number of segregating sites	2	32	6	1	6
π	0.0007	0.0044	0.0033	0.0002	0.0035
θ_W	0.0009	0.0037	0.0030	0.0005	0.0027
Tajima's D	-0.58	0.76	0.42	-1.16	1.04
Probability Tajima's D \leq observed	0.35	0.84	0.72	0.33	0.85
FW-H	0.40	0.41	1.45	0.12	1.13
Probability FW-H \leq observed	0.85	0.54	0.99	0.57	0.98

FW-H: Fay and Wu's H.

*Probabilities are obtained from 10 000 simulations under a standard neutral model.

Another prediction of a selective sweep in a region is that variability levels are reduced locally, but recover distal and proximal to the selected site. We tested this prediction by sequencing four additional noncoding fragments upstream and downstream of the *MKK7* locus in the same population samples. The analysis of levels of polymorphism show a typical valley of reduced variability in a ~20 kb region around the *MKK7* gene in *domesticus* extending to two flanking genes (*Snapc2* and *Cttn*, Fig. 3). Upstream and downstream of this valley, nucleotide variability has substantially recovered (to ~5 times the variability observed at *MKK7*, Table 3). The same window of reduced polymorphism was found when θ_W values were used instead of π .

In contrast, no systematic pattern of reduced variability around the *MKK7* gene was found in *musculus* (Fig. 3). Only one locus showed low variability in *musculus* (fragment 4), but since flanking fragments on both sides show high polymorphism, we attribute this to random fluctuations. On average, variability was ~6 times higher in *musculus* compared to *domesticus* over the whole 50 kb region (Fig. 3 and Table 3).

To further confirm the selective sweep in *domesticus* we used the recently developed composite likelihood ratio test (Kim & Stephan, 2002), which has the advantage that it incorporates information from multiple loci along

Table 3 Summary of nucleotide polymorphism in 2.6 kb 5' upstream flanking region of the *MKK7* gene (fragment 2) and four *MKK7* flanking fragments in *M. m. domesticus* and *M. m. musculus*.

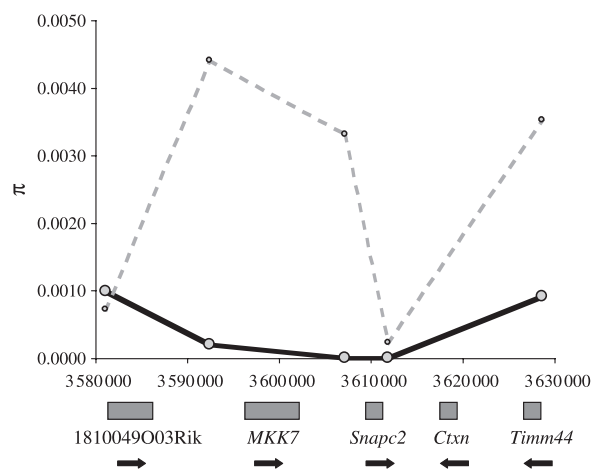


Fig. 3 Patterns of variability in the ~50 kb region around the *MKK7* region in *domesticus* (solid line). For comparison the variability in *musculus* in the same region is shown with a dashed line. Variability was estimated from 16 chromosomes for each subspecies.

the chromosome in the statistical test. The likelihood ratios of model A (ancestral and derived allele distinguished) of the test were calculated for two different θ_W values. One of the θ -values represents the value found

for the most variable fragment in the 50 kb region around *MKK7* in *domesticus* ($\theta_W = 0.001$, fragment 1) and the other is an average θ_W found across several unlinked loci in the *domesticus* genome ($\theta_W = 0.0023$, Harr, submitted). When compared to neutral simulations, the observed maximum likelihood ratio values for the whole 50 kb region is significantly higher than expected ($P = 0.005$ for $\theta_W = 0.001$, $P = 0.013$ for $\theta_W = 0.0023$), thus rejecting neutrality in favour of a selective sweep. The goodness of fit (GOF, Jensen *et al.*, 2005) statistic also suggests that a selective sweep model is more likely than a more generalized alternative model ($P = 0.24$ for $\theta_W = 0.001$ and $P = 0.28$ for $\theta_W = 0.0023$). The same likelihood ratio test applied to the polymorphism data in *musculus* showed no evidence for a selective sweep in this subspecies (data not shown).

Two additional genes (*Snape2* and *Cttn*) are located in the region of most strongly reduced variation around *MKK7*. We assayed both genes for differences in expression level between *domesticus* and the other subspecies of house mice using qRT-PCR, but found no difference (data not shown). Thus, only for the *MKK7* gene is there an overlap between expression change and being located in a region affected by a selective sweep.

MKK7 transcripts

Previous northern blot analysis using laboratory inbred strains of house mice showed that *MKK7* transcripts have a length of ~4 kb in all tissues analysed (Holland *et al.*, 1997; Tournier *et al.*, 1997). In testis tissue, however, an additional highly expressed shorter transcript of ~1.8 kb was found. Since laboratory strains of house mice are mixtures of genetic contributions from several subspecies (Wade *et al.*, 2002; Wade & Daly, 2005), we asked whether both transcripts are found in the testis of pure subspecies. Figure 4 shows a comparative northern blot analysis of *MKK7* expression in testis tissue from *musculus* and *domesticus*. The ~4 kb fragment is present in both subspecies but the shorter, ~1.8 kb fragment is expressed only in *domesticus*. We have also compared other tissues (see Methods) between our *musculus* and *domesticus* strains and found only the ~4 kb fragment (not shown). Thus, the ~1.8 kb variant of *MKK7* is specifically expressed in the testis of *domesticus* only.

Discussion

Evidence has recently accumulated in favour of rapid evolution of gene expression in association with new environmental challenges (e.g. Shapiro *et al.*, 2004; Kijimoto *et al.*, 2005). While the examples of Shapiro *et al.* (2004) and Kijimoto *et al.* (2005) are the evidence for fast evolution of gene expression in association with adaptation to novel conditions, they do not demonstrate the effect of positive selection in this process. Thus, gene expression divergence in itself does not imply that the

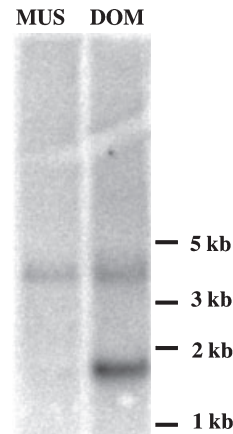


Fig. 4 Expression of *MKK7* in testis tissue of adult *M. m. musculus* (MUS) and *M. m. domesticus* (DOM). A northern blot of total RNA was hybridized against a *MKK7*-specific probe and exposed to a phospho-imaging screen. The positions of RNA size markers are illustrated.

difference is driven by selection. The reverse argument, however, can be applied to obtain evidence for a functional consequence of a gene expression change: if we can demonstrate the action of positive selection at a differentially expressed gene, it is likely that the observed change in expression has a functional consequence to the organism. Positive selection leaves a characteristic footprint in the genome. The spread of a positively selected mutation through a population (selective sweep) removes variability at the selected site and its flanking region, a process that has been referred to as 'hitchhiking' (Maynard Smith & Haigh, 1974). Thus, regions that recently experienced an episode of positive selection can be detected by a local reduction in variability (Schlötterer, 2003).

We identified the *MKK7* gene as specifically up-regulated in *domesticus* compared to the other subspecies of house mice and inferred positive selection as a likely explanation for the pattern of nucleotide polymorphism surrounding this gene. The genomic region of the *MKK7* gene shows signatures of a selective sweep, with substantially reduced nucleotide variation and an excess of derived polymorphic sites in high frequency in *domesticus* (Fay & Wu, 2000). The composite likelihood ratio test supports a selective sweep scenario and identifies the centre of a 50 kb region as the putative position of the selected site. In contrast to *domesticus*, no

evidence for a selective sweep at the *MKK7* region in *musculus* was found. Comparable signatures of local reduction in polymorphism level have previously been taken as evidence for positive selection in the respective regions (Harr *et al.*, 2002; Kingan *et al.*, 2003; Nair *et al.*, 2003; Catania *et al.*, 2004; Clark *et al.*, 2004; Schlenke & Begun, 2004; DuMont & Aquadro, 2005; Pool *et al.*, 2006).

The fact that expression divergence at *MKK7* is associated with a selective sweep in *domesticus* implies that the underlying nucleotide change(s) must be in *cis* rather than in *trans*. Using the pyrosequencing assay in F1 hybrids, we indeed find that a large component of the expression change in *M. m. domesticus* is caused by *cis* regulatory effects at *MKK7*. In *Drosophila* as well as in mice, *cis* effects have been shown to be a dominant source of inter-specific expression differences (Wittkopp *et al.*, 2004; Doss *et al.*, 2005). However, we cannot completely rule out that some additional *trans* effects contribute to the expression divergence observed at *MKK7* since the parental strains were assayed using a method other than the measurement of allele-specific expression as in the F1 hybrids. Additional *trans* effects could indeed explain the remaining variation that we see between the individuals of each subspecies. It is important to note that our observation of differential expression of the *musculus*-allele and the *domesticus*-derived allele in F1 hybrids also rules out the hypothesis that the expression change observed between the pure subspecies individuals might be due to different developmental stages assayed in the different subspecies.

It appears that the expression change in *domesticus* is not a simple up-regulation of a transcript that is also present in other subspecies, but rather involves the specific expression of a new, shorter transcript in *domesticus*. This transcript is only found in the testis (Holland *et al.*, 1997; Tournier *et al.*, 1997). Tournier *et al.* (1999) have shown that splice variants of the *MKK7* gene can lead to six different protein isoforms. However, the sizes of the known cDNAs from the region do not match the length of the two major transcripts seen on the Northern blots. The exact origin of the testis-specific transcript remains therefore unclear. Accordingly, we can so far not say whether a different promotor, different splicing or different poly-adenylation signal has led to the new transcript. However, given that the ~4 kb transcript is present in the same amounts in *domesticus* and *musculus*, it would seem most likely that the ~1.8 kb transcript is derived from a new promoter, rather than being a differentially processed form of the ~4 kb transcript.

The role of *MKK7*

The *MKK7* is part of the MAPK cascade, a highly conserved signalling pathway in all major groups of eukaryotes. On the amino acid level members of the MAPK cascades are characterized by a high level of

sequence conservation, which is indicative of strong purifying selection at these genes (Caenepeel *et al.*, 2004). Strong conservation of signalling genes is also observed on the gene expression level (Denver *et al.*, 2005).

The large difference in expression level at the *MKK7* gene between *domesticus* and all other subspecies is striking, given the short evolutionary time that separates the different house mouse subspecies and the overall constraint on the evolution of these genes.

The expression change at *MKK7* is manifested in the testis of individuals from the subspecies *domesticus*. Previous studies, both in mouse and in *Drosophila* showed that changes in signalling genes result in a meiotic drive phenotype with a competitive advantage of sperm carrying the mutant allele (Merrill *et al.*, 1999; Bauer *et al.*, 2005). In both cases transmission distortion is caused by an up-regulation in the respective gene in the testis of the organism. Whether or not the up-regulation of *MKK7* is connected with competitive advantages of sperm remains to be seen. In *C. elegans*, *MKK7* was shown to be involved in a regulatory cascade of innate immunity (Kim & Ausubel, 2005), but an analogous function in the mouse would not explain the specific up-regulation in testis.

Our data suggest that although we deal with a highly conserved gene in a central cellular signalling cascade, changes in the expression of this gene appear to be directly connected to short-term adaptations. Studies using natural hybrid zones (Payseur *et al.*, 2004) may show whether the change in expression level at *MKK7* in testis contributes to reproductive isolation between house mouse subspecies.

Acknowledgments

We would like to thank B. Schmitz for technical assistance and J. Pialek and Annie Orth for maintaining wild mice stocks. B. Bimova, E. Bozikova introduced us to handling and breeding of wild mice. We are thankful to T. Price and H. Tang for statistical advice and T. Price for helpful discussion and comments on the manuscript. The work has been funded by an Emmy Noether fellowship by the DFG to B. Harr and a Volkswagen Stiftung grant to D. Tautz.

References

- Ahmadian, A., Gharizadeh, B., Gustafsson, A.C., Sterky, F., Nyren, P., Uhlen, M. & Lundeberg, J. 2000. Single-nucleotide polymorphism analysis by pyrosequencing. *Anal. Biochem.* **280**: 103–110.
- Ahmadian, A., Ehn, M. & Hober, S. 2006. Pyrosequencing: history, biochemistry and future. *Clin. Chim. Acta.* **363**: 83–94.
- Bauer, H., Willert, J., Koschorz, B. & Herrmann, B.G. 2005. The t complex-encoded GTPase-activating protein Tagap1 acts as a transmission ratio distorter in mice. *Nat. Genet.* **37**: 969–973.

- Boursot, P., Auffray, J.C., Brittondavidian, J. & Bonhomme, F. 1993. The evolution of house mice. *Annu. Rev. Ecol. Syst.* **24**: 119–152.
- Braverman, J.M., Hudson, R.R., Kaplan, N.L., Langley, C.H., Stephan, W. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* **140**: 783–96.
- Caenepeel, S., Charydzak, G., Sudarsanam, S., Hunter, T. & Manning, G. 2004. The mouse kinome: discovery and comparative genomics of all mouse protein kinases. *Proc. Natl. Acad. Sci. USA* **101**: 11707–11712.
- Catania, F., Kauer, M.O., Daborn, P.J., Yen, J.L., Ffrench-Constant, R.H. & Schlotterer, C. 2004. World-wide survey of an accord insertion and its association with DDT resistance in *Drosophila melanogaster*. *Mol. Ecol.* **13**: 2491–2504.
- Clark, R.M., Linton, E., Messing, J. & Doebley, J.F. 2004. Pattern of diversity in the genomic region near the maize domestication gene *tb1*. *Proc. Natl. Acad. Sci. USA* **101**: 700–707.
- Cowles, C.R., Hirschhorn, J.N., Altshuler, D. & Lander, E.S. 2002. Detection of regulatory variation in mouse genes. *Nat. Genet.* **32**: 432–437.
- Davidson, E.H. 2001. *Genomic Regulatory Systems: Development and Evolution*. Academic Press, San Diego.
- Denver, D.R., Morris, K., Strelman, J.T., Kim, S.K., Lynch, M. & Thomas, W.K. 2005. The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nat. Genet.* **37**: 544–548.
- Doss, S., Schadt, E.E., Drake, T.A. & Lusis, A.J. 2005. Cis-acting expression quantitative trait loci in mice. *Genome. Res.* **15**: 681–691.
- DuMont, V.B., & Aquadro, C.F. 2005. Multiple signatures of positive selection downstream of notch on the X chromosome in *Drosophila melanogaster*. *Genetics* **171**: 639–653.
- Enard, W., Khaitovich, P., Klose, J., Zollner, S., Heissig, F., Giavalisco, P., Nieselt-Struwe, K., Muchmore, E., Varki, A., Ravid, R., Doxiadis, G.M., Bontrop, R.E. & Paabo, S. 2002. Intra- and interspecific variation in primate gene expression patterns. *Science* **296**: 340–343.
- Fay, J.C., McCullough, H.L., Sniegowski, P.D. & Eisen, M.B. 2004. Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. *Genome. Biol.* **5**: R26.
- Fay, J.C. & Wu, C.I. 2000. Hitchhiking under positive darwinian selection. *Genetics* **155**: 1405–1413.
- Fu, Y.X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**: 915–25.
- Guenet, J.L. & Bonhomme, F. 2003. Wild mice: an ever-increasing contribution to a popular mammalian model. *Trends Genet.* **19**: 24–31.
- Harr, B., Kauer, M. & Schlotterer, C. 2002. Hitchhiking mapping: a population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **99**: 12949–12954.
- Holland, P.M., Suzanne, M., Campbell, J.S., Noselli, S. & Cooper, J.A. 1997. *MKK7* is a stress-activated mitogen-activated protein kinase functionally related to hemipterous. *J. Biol. Chem.* **272**: 24994–24998.
- Ihle, S., Ravaoarimanana, I., Thomas, M. & Tautz, D. 2006. An analysis of signatures of selective sweeps in natural populations of the house mouse. *Mol. Biol. Evol.* **23**: 915–925.
- Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B. & Speed, T.P. 2003. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31**: e15.
- Jensen, J.D., Kim, Y., DuMont, V.B., Aquadro, C.F. & Bustamante, C.D. 2005. Distinguishing between selective sweeps and demography using DNA polymorphism data. *Genetics* **170**: 1401–1410.
- Jin, W., Riley, R.M., Wolfinger, R.D., White, K.P., Passador-Gurgel, G. & Gibson, G. 2001. The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nat. Genet.* **29**: 389–395.
- Jordan, I.K., Marino-Ramirez, L. & Koonin, E.V. 2005. Evolutionary significance of gene expression divergence. *Gene* **345**: 119–126.
- Khaitovich, P., Weiss, G., Lachmann, M., Hellmann, I., Enard, W., Muetzel, B., Wirkner, U., Ansoerge, W. & Paabo, S. 2004. A neutral model of transcriptome evolution. *PLoS Biol* **2**: E132.
- Kijimoto, T., Watanabe, M., Fujimura, K., Nakazawa, M., Murakami, Y., Kuratani, S., Kohara, Y., Gojobori, T. & Okada, N. 2005. *Cimp1*, a novel astacin family metalloproteinase gene from East African cichlids, is differentially expressed between species during growth. *Mol. Biol. Evol.* **22**: 1649–1660.
- Kim, D.H. & Ausubel, F.M. 2005. Evolutionary perspectives on innate immunity from the study of *Caenorhabditis elegans*. *Curr. Opin. Immunol.* **17**: 4–10.
- Kim, Y. & Stephan, W. 2002. Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* **160**: 765–777.
- King, M.C. & Wilson, A.C. 1975. Evolution at two levels in humans and chimpanzees. *Science* **188**: 107–116.
- Kingan, S.B., Tatar, M. & Rand, D.M. 2003. Reduced polymorphism in the chimpanzee semen coagulating protein, semenogelin I. *J. Mol. Evol.* **57**: 159–169.
- Lemos, B., Meiklejohn, C.D., Caceres, M. & Hartl, D.L. 2005. Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories. *Evolution Int. J. Org. Evolution* **59**: 126–137.
- Maynard Smith, J. & Haigh, J. 1974. The hitch-hiking effect of a favorable gene. *Genet. Res.* **23**: 23–35.
- Merrill, C., Bayraktaroglu, L., Kusano, A. & Ganetzky, B. 1999. Truncated *RanGAP* encoded by the segregation distorter locus of *Drosophila*. *Science* **283**: 1742–1745.
- Nair, S., Williams, J.T., Brockman, A., Paiphun, L., Mayxay, M., Newton, P.N., Guthmann, J.P., Smithuis, F.M., Hien, T.T., White, N.J., Nosten, F. & Anderson, T.J. 2003. A selective sweep driven by pyrimethamine treatment in southeast asian malaria parasites. *Mol. Biol. Evol.* **20**: 1526–1536.
- Nei, M. & Li, W.-H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA* **76**: 5269–5273.
- Oleksiak, M.F., Churchill, G.A. & Crawford, D.L. 2002. Variation in gene expression within and among natural populations. *Nat. Genet.* **32**: 261–266.
- Payseur, B.A., Krenz, J.G. & Nachman, M.W. 2004. Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. *Evolution Int. J. Org. Evolution* **58**: 2064–2078.
- Pool, J.E., Bauer Dumont, V., Mueller, J.L. & Aquadro, C.F. 2006. A scan of molecular variation leads to the narrow localization of a selective sweep affecting both Afrotropical

- and cosmopolitan populations of *Drosophila melanogaster*. *Genetics* **172**:1093–1105.
- Schadt, E.E., Monks, S.A., Drake, T.A., Lusk, A.J., Che, N., Colinayo, V., Ruff, T.G., Milligan, S.B., Lamb, J.R., Cavet, G., Linsley, P.S., Mao, M., Stoughton, R.B. & Friend, S.H. 2003. Genetics of gene expression surveyed in maize, mouse and man. *Nature* **422**: 297–302.
- Schlenke, T.A. & Begun, D.J. 2004. Strong selective sweep associated with a transposon insertion in *Drosophila simulans*. *Proc. Natl. Acad. Sci. USA* **101**: 1626–1631.
- Schlötterer, C. 2003. Hitchhiking mapping–functional genomics from the population genetics perspective. *Trends Genet.* **19**: 32–38.
- Shapiro, M.D., Marks, M.E., Peichel, C.L., Blackman, B.K., Nereng, K.S., Jonsson, B., Schluter, D. & Kingsley, D.M. 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**: 717–723.
- Stamatoyannopoulos, J.A. 2004. The genomics of gene expression. *Genomics* **84**: 449–457.
- Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., Cooke, M.P., Walker, J.R. & Hogenesch, J.B. 2004. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. USA* **101**: 6062–6067.
- Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- Thompson, J.D., Higgins, D.G. & Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680.
- Thornton, K. 2003. Libsequence: a C++ class library for evolutionary genetic analysis. *Bioinformatics* **19**: 2325–2327.
- Tournier, C., Whitmarsh, A.J., Cavanagh, J., Barrett, T. & Davis, R.J. 1997. Mitogen-activated protein kinase kinase 7 is an activator of the c-Jun NH2-terminal kinase. *Proc. Natl. Acad. Sci. USA* **94**: 7337–7342.
- Tournier, C., Whitmarsh, A.J., Cavanagh, J., Barrett, T. & Davis, R.J. 1999. The MKK7 gene encodes a group of c-Jun NH2-terminal kinase kinases. *Mol. Cell. Biol.* **19**: 1569–1581.
- Wade, C.M. & Daly, M.J. 2005. Genetic variation in laboratory mice. *Nat. Genet.* **37**: 1175–1180.
- Wade, C.M., Kulbokas, III, E.J., Kirby, A.W., Zody, M.C., Mullikin, J.C., Lander, E.S., Lindblad-Toh, K. & Daly, M.J. 2002. The mosaic structure of variation in the laboratory mouse genome. *Nature* **420**: 574–578.
- Watterson, G.A. 1975. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**: 256–276.
- Wittkopp, P.J., Haerum, B.K. & Clark, A.G. 2004. Evolutionary changes in cis and trans gene regulation. *Nature* **430**: 85–88.
- Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P., Pizer, M., Rockman, M.V. & Romano, L.A. 2003. The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* **20**: 1377–1419.

Supplementary Material

The following supplementary materials are available for this article online:

Table S1. Sample locations and number of generations mouse strains that were kept in the lab.

Table S2. Location and genes associated with sequenced fragments in the *MKK7* region on mouse chromosome 8.

Table S3. Primer sequences and amplification conditions for sequencing of 2.6 kb of 5' upstream sequence of the *MKK7* gene and other flanking sequence fragments.

Table S4A. Genes differentially expressed between *M. m. castaneus* and all other subspecies.

Table S4B. Genes differentially expressed between *M. m. domesticus* and all other subspecies.

Table S4C. Genes differentially expressed between *M. m. musculus* and all other subspecies.

Table S4D. Genes differentially expressed between *M. m. ssp.* and all other subspecies.

Fig. S1. Pyrosequencing assay for the *MKK7* gene.

This material is available as part of the online article from <http://www.blackwell-synergy.com>

Received 2 January 2006; revised 3 March 2006; accepted 7 March 2006