PERGAMON

Neuroscience and Biobehavioral Reviews 26 (2002) 907–923

NEUROSCIENCE AND BIOBEHAVIORAL REVIEWS

www.elsevier.com/locate/neubiorev

Review

Identification and ranking of genetic and laboratory environment factors influencing a behavioral trait, thermal nociception, via computational analysis of a large data archive

Elissa J. Chesler[a], Sonya G. Wilson[a], William R. Lariviere[a], Sandra L. Rodriguez-Zas[b], Jeffrey S. Mogil[a,c,*]

[a]*Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA*
[b]*Department of Animal Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*
[c]*Department of Psychology, McGill University, Montreal, Que., Canada H3A 1B1*

Received 4 September 2002; revised 30 October 2002; accepted 30 October 2002

Abstract

Laboratory conditions in biobehavioral experiments are commonly assumed to be 'controlled', having little impact on the outcome. However, recent studies have illustrated that the laboratory environment has a robust effect on behavioral traits. Given that environmental factors can interact with trait-relevant genes, some have questioned the reliability and generalizability of behavior genetic research designed to identify those genes. This problem might be alleviated by the identification of the most relevant environmental factors, but the task is hindered by the large number of factors that typically vary between and within laboratories. We used a computational approach to retrospectively identify and rank sources of variability in nociceptive responses as they occurred in a typical research laboratory over several years. A machine-learning algorithm was applied to an archival data set of 8034 independent observations of baseline thermal nociceptive sensitivity. This analysis revealed that a factor even more important than mouse genotype was the experimenter performing the test, and that nociception can be affected by many additional laboratory factors including season/humidity, cage density, time of day, sex and within-cage order of testing. The results were confirmed by linear modeling in a subset of the data, and in confirmatory experiments, in which we were able to partition the variance of this complex trait among genetic (27%), environmental (42%) and genetic × environmental (18%) sources.

Keywords: Mice; Genetic; Environment; Nociception; Pain; CART; Data mining

Contents

* Corresponding author. Address: Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal, Que., Canada H3A 1B1. Tel.: +1-514-398-6085; fax: +1-514-398-4896.
E-mail address: jeffrey.mogil@mcgill.ca (J.S. Mogil).

0149-7634/02/$ - see front matter © 2003 Published by Elsevier Science Ltd.
PII: S0149-7634(02)00103-3

1. Introduction

1.1. Genetic and environmental factors in biobehavioral research

The ability to identify and manipulate genes influencing biomedical and behavioral traits is being hailed as a biological revolution in our time. An exclusive focus on genetic determinants will not fully succeed in explaining individual differences, however, because the proportion of the total variation due to genotype (i.e. broad-sense heritability) of most biobehavioral traits is less than 50% [100]. Early behavioral genetics results have been criticized for their lack of replicability [125]. Interactions of mouse genotype with the specific laboratory environment in which traits are examined [20,29] may have hampered the replicability of transgenic knockout studies [29]. Apparent disagreement in linkage mapping studies may also be explained by environmental factors [109]. The role of the environment on genetic studies can be striking. For example, of 17 quantitative trait loci (QTLs) identified in a *Drosophila melanogaster* life-span study, none were consistently present across five environments tested [123]. In a recent study using eight mouse strains, despite heroic attempts to standardize all facets of the laboratory environment in three different laboratories—including acquisition, husbandry and testing protocols—significantly different responses on various behavioral tasks were obtained at each site [29]. Although some have pessimistically concluded that the large site-specific effects (e.g. hyperlocomotion of mice tested in Edmonton relative to those in Albany or Portland) are detrimental to the generalizability and replicability of behavior genetic research, we would emphasize that reasonably high heritabilities and commonality of *relative* strain sensitivities across site were obtained.

The problem is more subtle when one considers 'natural' variation in the laboratory environment; that is, fluctuation of laboratory conditions within the range of control assumed by the researcher. To the extent that relevant environmental factors remain obscure, findings are dependent upon the particular set of conditions in which testing occurred. Some have thus argued for standardization of environmental conditions to improve replicability and comparison of results across laboratories [120,121,126]. However, standardization can lead to findings with low generalizability (i.e. low external validity), essentially rendering experiments uninformative, and argues instead for systematic variation of the testing environment [132,133]. In either case, although stringent control or systematic variation of all factors simultaneously may not be practical, control or variation of the more salient factors would help alleviate the problem. However, many actual sources of between-lab and especially within-lab variability remain unidentified, and even when known, such factors are rarely assessed simultaneously, and hence their relative impact is also unknown. We report that a computational approach can be used to identify such environmental factors in existing data sets. These data were recently published [24]. The reliability and robustness of this approach was confirmed by modeling and a controlled experiment, the results of which are reported here.

1.2. The tail-withdrawal test of thermal nociception

In the course of our ongoing study of the genetic mediation of nociception and analgesia over the last eight years, we have tested mice of varied genotypes on a common assay of thermal nociception, the 49°C hot water tail-immersion/withdrawal test [9,51,57]. This assay is a close cousin of the more common tail-flick test, which uses radiant heat from a high-wattage bulb as the noxious stimulus [33]. Although the hot water and radiant heat versions differ with respect to the stimulated surface area and heat transfer functions [73], they are generally considered equivalent. We have shown fairly high correlations between the sensitivity of inbred strains to the two assays ($r \geq 0.55$) [72], and are unaware of any published finding dissociating the assays qualitatively, and thus we are quite confident that our present results apply also to the

radiant heat tail-flick assay and thermal nociception more generally [72,84]. These two assays each possess both face validity and empirical validity in accurately predicting the clinical potency of opiate analgesics [30,117], although are not well representative of chronic clinical pain in humans [53,73,85,128]. The tail-flick/withdrawal reflex is a spinally mediated [55] flexion reflex under tonic inhibitory control by a descending antinociceptive system including a number of midbrain and brain stem loci (e.g. periaqueductal gray, rostral ventral medulla) [38,91].

Despite the relative stability of intraindividual tail-flick/withdrawal latencies compared to other assays [73, 85], a perusal of the large literature featuring these assays reveals widely varying 'baseline' values. This variability introduces a challenge to the generalizability of findings obtained from these studies, since both analgesia and hyperalgesia are, by definition, *changes* from the baseline. Especially in light of omnipresent 'ceiling effects' (i.e. the cut-off imposed for ethical reasons) and 'floor effects' (i.e. the minimal time taken for a response to be performed and recorded), the baseline values themselves impact greatly on estimates of analgesic and hyperalgesic potency and efficacy. This problem is magnified in those studies comparing more than one subject population (e.g. males vs. females, mutants vs. wildtypes, strain A vs. strain B). It would therefore be of practical advantage to more completely appreciate those factors impacting baseline latencies.

Of course, some of the variability derives from effects of the variation of physical properties of the stimulus. Stimulus intensity will impact observed response latencies due to the stimulus-response function inversely relating withdrawal latency to the rate of heat transfer to the tail [52,91], but also because the analgesic effects of restraint stress are minimized by the use of higher-intensity stimuli [31]. We have demonstrated that tail-withdrawal (TW) latencies are higher when mice are restrained in Plexiglas chambers for the entire experiment (even after habituation) compared to when they are removed from their home cage immediately prior to testing and restrained acutely [85]. Habituation to sham testing has been shown to reduce latencies in rats [76]; it is unclear whether this strategy is effective in mice. An additional stimulus-related factor specific to the tail-flick test is the precise locus of stimulation on the tail [92,134]; this factor is not highly relevant to the TW test since by convention investigators immerse the distal half of the tail.

Even when parametric factors are held constant, however, variability remains. Such interindividual variation is, of course, due to organismic and environmental factors, and their interaction. A large body of research has documented the modestly increased sensitivity to noxious stimuli of females of a number of species, including rodents and humans [10,23,39,65,106]. Increasing attention is also being paid to the role of inherited genetic variation in the mediation of pain sensitivity [79,82,86]. We and others have shown recently that these factors interact, such that sex differences in pain-related traits are observed in some rodent strains but not others [7,23,27,35,66]. Other organismic variables demonstrated to affect nociceptive sensitivity include ontogeny [26,40,113] and aging [67,90], weight [74,95, 107], and, in females, hormonal status [14,16,37,43,47].

Environmental factors affecting pain sensitivity have received even less attention, except for the clinically important finding that prior injury - and thus experience with pain - can produce longterm alterations in subsequent pain sensitivity [130,131]. A number of additional environmental factors able to affect nociception have been identified. Some are related to environmental conditions during development, such as prenatal stress to the mother [68,116], lead exposure [56,122], and neonatal handling [25,32,98]. Others are related to environmental conditions in adulthood, including stress exposure [21], housing density (see Section 4.3), diet [44,110], and magnetic fields [61]. Still others are related to the testing experience itself, including the time of day (and its relation to the photoperiod in which the subjects are maintained) (see Section 4.4), novelty stress associated with the testing room [1,6], and particularities of the experimenters themselves (a likely interpretation of the data of Ref. [29]).

Such environmental factors can interact with organismic factors. For example, isolation housing appears to be a stressor in female rats whereas crowding is a stressor in male rats [19]. There are strain differences in the conspecific aggression of group-housed males [77], and social defeat associated with that aggression is major stressor that can affect nociceptive and analgesic sensitivity [75]. The effects of housing density and photoperiod may be strain-dependent as well [15,22,34]. It should be noted also that recent findings have established that maternal behavior and stress reactivity can be transmitted non-genomically across generations [41].

To our knowledge, no attempt has ever been made to assess a large number of these modulatory factors simultaneously, either for pain or any other biobehavioral trait. Hence, the relative importance of the factors, and thus the factors that most require regulation in the laboratory, remains unknown. The last author has conducted or supervised a large number of experiments using the mouse 49°C TW test over the course of the last 10 years. These studies have focused on strain and sex differences in thermal nociception and antinociception, and a wide variety of genotypes have been employed. In all experiments, parametric factors have been kept constant, such that only certain organismic and environmental factors described above have been varied and recorded. Thus, an analysis of all these data together would be useful for examining the relative influence of genetic and laboratory environmental factors modulating this biobehavioral trait.

1.3. Study design

The archival data set consists of baseline TW latencies for each of 8034 naïve adult mice tested between August, 1993 and May, 2001, along with the following information,

where available, recorded on data sheets at the time of testing: genotype (i.e. strain, substrain and vendor; including 40 inbred, outbred, hybrid and mutant strains), sex, age, weight, testing facility, cage density, season, time of day, temperature, humidity, experimenter, and within-cage order of testing. Of these 12 factors, eight were amenable to analysis (see Section 2.3).

An analysis technique suitable for unbalanced data sets of high dimensionality was required for determination of the relative contribution of factors. Classification and regression tree (CART) analysis [18,112], an automated machine-learning technique, was thus used to characterize and obtain a preliminary ranking of the importance of the factors. In brief, this technique develops rules to partition the data based on the predicting factors, producing a decision tree that can be used to predict the value of the dependent measure from the factors considered in the model.

To confirm the accuracy of the results generated by CART, we performed fixed-effects modeling on a subset of the full data set, allowing a parametric estimate of the magnitude of factor effects through the computation of least-squares (LS) means based on a model of TW latency. We then performed a fully crossed and balanced experiment on several of the factors on a single day, allowing for partitioning of the variance and the determination of the precise proportion of trait variance accounted for by genetic and environmental variables. We were intrigued that a number of factors identified herein are not widely appreciated to affect nociception. For example, within-cage order of testing has never been reported to affect pain sensitivity measurements, and thus separate experiments were performed to confirm and characterize its effect.

2. Materials and methods

2.1. Subjects and husbandry

Subjects in all experiments were naïve, adult male and female mice of 40 different genotypes ($n = 15–928$/strain; see Table 1). All mice were housed in a 12:12 h light/dark cycle (lights on at 07:00 h) in a temperature-controlled (22 ± 2°C) vivarium, and given *ad lib* access to food (in Portland, OR: Purina Mouse Chow; in Champaign, IL: Harlan Teklad 8604) and tap water. The majority of mice were bred in-house. Others were purchased from The Jackson Laboratory (Bar Harbor, ME), Harlan Sprague Dawley (Indianapolis, IN), or Simonsen Laboratories (Gilroy, CA), or obtained from scientific collaborators. Unfortunately, in most cases no distinction was made on the original data sheets as to whether subjects were born in our vivaria or elsewhere. In the latter case, however, at least one week was allowed for mice to habituate to the new environment before any testing occurred.

In the Portland and Champaign sites, where the overwhelming majority of the data were obtained, similar husbandry protocols were followed. Virtually all mice were bred from single dam × single sire pairings in which the sire remained permanently. Mice were weaned at 18–21 d of age (runts were culled), and housed 1–7 per cage with their same-sex littermates. In some cases (and especially for males), subsets of same-sex siblings were housed in different cages to reduce housing density. All mice were housed in standard-sized ($11.5'' \times 7.5'' \times 5''$) polycarbonate cages with woodshaving or Bed-O-Cob $1/4''$ bedding, stainless steel wire bar cage tops, nestlets (for all mice in Champaign; for breeders and unweaned offspring only in Portland), and French square bottles with straight, open sipper tubes. Cages were kept on conventional 4- or 5-shelf stainless steel racks. Cages were changed at an average frequency of twice per week; unfortunately no record of precise changing days was available.

2.2. Nociceptive testing

Mice were brought on a rolling cart from a nearby vivarium to the testing room 30 min to 2 h before testing. At some point during this period they were handled by an experimenter, given an identifying mark on their tail with a permanent marker and, in most cases, weighed. Since latencies in this test are known to be affected by tail skin temperature [119], attention was paid to ensure that the ambient temperature in the testing room was 20–23°C, but no records were kept of actual temperatures during testing. For testing, mice were individually removed from their home cage and introduced to one of a small number of very similar cloth/cardboard 'pockets', into which they freely entered in almost every case. While lightly restrained in this manner, the distal half of the mouse tail was immersed in lightly-agitated water thermostatically controlled at 49 ± 0.2°C (VWR Model 1110 in Portland; Fisher Isotemp Model 2100 in Champaign), and the latency to a vigorous, reflexive withdrawal of the tail was measured to the nearest 0.1 s with a stopwatch held in the other hand.

To increase accuracy, two such measurements separated by 10–20 s were made and averaged. In a small percentage of cases, three measurements were made and the most similar two averaged. The mouse was then immediately returned to its home cage. The interval between testing one mouse and the next from the same cage ranged from 15 s to several minutes, and the cage top was left closed in this interval. It should be noted that in most of these experiments subjects were tested again at one or more time points after an analgesic manipulation (i.e. drug injection and/or stress exposure), but such data are not considered here.

Experimenters were extensively trained on the TW assay, and not permitted to acquire experimental data until they demonstrated the ability to obtain stable baselines and replicate known strain differences. JM taught the technique to SW and HH; SW in turn taught the technique to all other experimenters (see Table 1).

Table 1
Summary statistics of archival data set ($n = 8034$)

Factor type	Factor	Level	n	Comments
Organismic				
	Genotype (outbred)	CD-1	276	ICR stock from Harlan Sprague Dawley, Inc. (Indianapolis, IN)
		SW-ND4	105	Swiss-Webster stock from Harlan Sprague Dawley, Inc.
		SW-Sim	928	Swiss-Webster stock from Simonsen, Inc. (Gilroy, CA)
		SW-und.	65	Swiss-Webster stock from either Harlan or Simonsen (undetermined)
	Genotype (hybrid)	B6129F1	15	(C57BL/6J × 129P3/J)F_1
		B6AF2	15	(C57BL/6J × A/J)F_2
		B6D2F1	128	(C57BL/6J × DBA/2J)F_1
		B6D2F2	757	(C57BL/6J × DBA/2J)F_2
		C3HAF2	263	(C3H/HeJ × A/J)F_2
	Genotype (inbred)	129P3/J	211	Previously known as 129/J (The Jackson Laboratory, Bar Harbor, ME)
		A/J	368	
		AKR/J	250	
		BALB/cJ	276	
		C3H/HeJ	214	
		C3HeB/FeJ	133	
		C57BL/6J	744	
		C57BL/10J	278	
		C58/J	122	
		CBA/J	223	
		DBA/2J	563	
		LP/J	39	
		NOD/J	38	
		NON/J	28	
		RIIIS/J	122	
		SJL/J	27	
		SM/J	135	
		SWR/J	16	
	Genotype (mutant)	5HT1BKO	257	129-*Htr1b*^{tm1Hen} (maintained on a mixed 129 substrain background)
		CXBK	24	A recombinant inbred strain with an analgesia-relevant mutation
		DELTKO-1	217	129S6,C57BL/6-*Oprd1*^{tm1Pin}
		DELTKO-2	68	129S6-*Oprd1*^{tm1Pin}
		ENDKO	405	129S6, C57BL/6-*Pomc1*^{tm1Low}
		MUKO	60	129S6, C57BL/6-*Oprm*^{tm1Pin}
		OFQKO	62	129S6, C57BL/6-*Npnc1*^{tm1Pin}
		e/e	95	C57BL/6J-*Mc1r*^{e} (recessive yellow spontaneous mutants)
		Sombre	111	C3HeB/FeJ-*Mc1r*$^{E\text{-}so}$/*Mc1r*$^{E\text{-}so}$*Gli3*$^{Xt\text{-}J}$/ + (sombre spontaneous mutants)
	Genotype (selected)	HA	61	Mice selected for high stress-induced analgesia from outbred stock
		LA	57	Mice selected for low stress-induced analgesia from outbred stock
		HAR	147	Mice selected for high levorphanol analgesia from heterogeneous stock
		LAR	131	Mice selected for low levorphanol analgesia from heterogeneous stock
	Sex	Male	4109	
		Female	3766	
		Unknown	159	
	Age	<6 weeks	208	
		6–10 weeks	3052	
		>10 weeks	1209	
		Unknown	3565	The vast majority of these mice were likely 6–10 weeks old
	Weight	10.0–19.9 g	1666	(Male: 293; female: 1373)
		20.0–29.9 g	4612	(Male: 2812; female: 1800)
		$\geq$30.0 g	1037	(Male: 652; female: 385)
		Unknown	719	(Male: 401; female: 318)
Environmental—husbandry				
	Testing facility	Portland, OR	1787	The lab of Dr John K. Belknap, Oregon Health Sciences University
		Champaign, IL	5840	The PI's former lab at University of Illinois, Urbana-Champaign
		Milwaukee, WI	161	The lab of Dr Fred J. Helmstetter, University of Wisconsin-Milwaukee
		Piscataway, NJ	246	The lab of Dr John E. Pintar, UMDNJ
	Cage density	1	188	
		2	993	
		3	2396	

(*continued on next page*)

Table 1 (*continued*)

Factor type	Factor	Level	*n*	Comments
		4	2826	
		5	1019	
		6	349	Females only
		7	34	Females only
		Unknown	229	
Environmental—experiment-related				
	Year	1993	55	In Portland
		1994	97	In Portland
		1995	780	In Portland
		1996	843	In Champaign
		1997	583	In Champaign
		1998	846	In Champaign
		1999	2269	In Champaign and Milwaukee
		2000	1614	In Champaign
		2001	935	In Champaign and Piscataway
		Unknown	12	
	Season	Winter	2167	
		Spring	1690	
		Summer	1896	
		Fall	2269	
		Unknown	12	
	Temperature	<65.0 °F	12	Temperature measured in vivarium, not testing room
		65.0–69.9 °F	366	
		70.0–74.9 °F	5453	
		≥ 75.0 °F	8	
		Unknown	2195	
	Humidity	0–19.95%	788	Humidity measured in vivarium, not testing room
		20–39.95%	1750	
		40–59.95%	264	
		60–100%	423	
		Unknown	4809	
	Time of day	09:30–10:55 h	863	Refers to starting time of experiment; experiment duration was <2 h
		11:00–13:55 h	3746	
		14:00–17:00 h	3169	
		Unknown	256	
	Experimenter	AK	15	A female undergraduate
		AR	118	A male undergraduate
		BM	828	A male undergraduate
		CB	19	A male undergraduate
		EC	12	A female graduate student
		HH	259	A female graduate student
		JH	482	A male undergraduate
		JM	3376	The male principal investigator
		KM	190	A female undergraduate
		LN	12	A female undergraduate
		SW	2723	A female graduate student
	Order of testing	First	2649	
		Second	2386	
		Third	1744	
		Fourth	936	
		Fifth	249	
		Sixth	54	
		Seventh	4	
		Unknown	12	

2.3. CART analysis

CART is a machine learning technique ideal for large, unbalanced data sets with many predictors [18,112]. This method is often used in epidemiological research to empirically develop decision tress that can predict disease based on numerous predictors [17,46]. CART develops rules for partitioning data into subsets. A tree is grown from an initial parent node containing all of the data. This is done by exhaustively testing all possible splits by each predictor to identify the split that results in the most improvement, defined as the difference between the variance in the parent

node and the mean variance in the resulting two nodes. This is performed repeatedly until the data have been split completely. The resulting tree, which is now over-fit to the data, is then pruned using a cross-validation technique to select an optimal tree which is generalizable to subsets of the data, and of limited complexity. In addition to generating a tree which can be used to predict the phenotypic value of a subject tested under a particular set of conditions, this technique allows for the ranking of factors according to their role in reducing variance in the variety of contexts that are revealed in the process of splitting the data. Rankings are based on the relative variance reduction (improvement) attributed to each of the factors when used as a primary splitter or as one of the top five surrogates (factors which are highly correlated with the splitter, whose values can be used to generate similar results in the event of missing data on the primary splitter and whose importance may be masked by the splitter) at each node. The highest-ranked factor is arbitrarily assigned a score of 100 and the other scores are relative to that. Such relative rankings of variables, the major purpose of our simultaneous analysis of these factors, is not possible with parametric methods when sample sizes are not balanced across all cells in a design. Although CART is useful in the analysis of large data sets and shows a high agreement with linear modeling in such cases [18], the method is non-parametric and can not be used to estimate effect sizes in highly complex trees.

Predictors entered into the model were genotype, sex, experimenter, time of day, season, humidity, order of testing, and cage density. Temperature, weight, and age were excluded because insufficient biologically relevant within-factor variability existed in the data set (see Table 1). Preliminary models indicated that testing facility may influence the trait; however, it was excluded from the final model because only one experimenter collected data in multiple facilities.

We were primarily interested in evaluating the relative rankings of these factors with regard to their association with TW latency, not in their predictive value. Thus, because this algorithm is biased towards the use of continuous factors (e.g. time of day, humidity) or high-level categorical factors (e.g. genotype, experimenter) as splitters, remedial measures were taken to increase the generalizability and validity of these rankings. For continuous factors a preliminary tree was grown to determine where splits tended to occur, and the data were then broken up into a moderate number of categories of equal range based on the rough locations of these splits. For all factors, a penalty was imposed on the improvement at each node equal to the number of levels of each factor divided by the total number of levels in the analysis.

2.4. Fixed-effects modeling

In an effort to estimate parametrically the magnitude of factor effects, a linear model fitting main effects and two-way interactions of the same eight factors modeled in CART was generated. This enabled us to estimate LS means for levels of these factors. There was insufficient information to test higher-order interactions, which are of questionable biological relevance anyway, and thus they were excluded from the model. Linear modeling was implemented using SAS v.6.12 PROC MIXED (SAS Institute, Cary, NC). This technique uses a likelihood-based approach to estimating model parameters, which is less sensitive to idiosyncrasies in the data structure such as empty cells or sample size imbalance than least squares methods. Data were log-transformed to satisfy model assumptions. A subset of the data was used for which no missing values were present ($n = 1772$). In addition, some factor levels were eliminated and others collapsed into fewer categories to facilitate estimability of the model (see Table 2); the most commonly represented experimenters and strains were retained. The model was reduced until no non-significant fixed effects remained. LS means were estimated based on this reduced model.

2.5. Balanced experiment

In order to test the hypotheses generated by modeling and CART data mining, and to determine the precise proportion of trait variance accounted for by genetic and environmental variables, a prospective experiment was performed. In this new experiment, a total of 192 naïve mice from three inbred strains (A/J, C57BL/6J and DBA/2J) were tested as described above on a single day (May 15, 2001), with all conditions of genotype × sex × within-cage order of testing represented. Furthermore, each mouse was tested in either the morning (10:00–11:00 h) or the afternoon (14:30–15:30 h), by one of two experimenters (JM and SW) whose data comprise the bulk of the archival data set. Factors held constant were age (42–45 d), weight (within 2 g of the mean for that strain and sex), and cage density (4 mice/cage). This experiment had a completely balanced design representing all of the easily manipulable factors identified using CART. As such, a 5-way fixed effects ANOVA model was used to analyze this experiment.

2.6. Order of testing experiments

We examined the influence of within-cage order of testing on TW latency using a strain identified as sensitive to the factor, Swiss-Webster (SW-Sim; Simonsen Labs). Mice of both sexes ($n = 8$/order/condition), housed in cages of four, were tested in one of two conditions: either returned to their home cage after baseline determination (our normal procedure), or transferred to a holding cage on the other side of the room (5 m away) after testing. If communication between previously tested and yet-to-be-tested mice is responsible for the phenomenon, the latter procedure may abolish the effect. This experiment was analyzed by

Table 2
Influence on thermal nociception of individual levels of genetic and environmental factors from archival data and from a controlled experiment

Factor Level[a]	LS means[b] (s)	*N*	Experiment[c] (s)	*N*
Experimenter				
BM	2.6 ± 0.18	166		
JH	2.0 ± 0.21	213		
JM	3.7 ± 0.36	505	3.4 ± 0.12	96
KM	3.0 ± 0.20	21		
SW	2.2 ± 0.22	867	2.1 ± 0.06*	96
Genotype				
129P3/J	2.8 ± 0.41	95		
A/J	2.8 ± 0.24	187	3.2 ± 0.15	64
AKR/J	2.2 ± 0.22	161		
BALB/cJ	3.8 ± 0.34	133		
C3H/HeJ	2.4 ± 0.16	108		
C57BL/6J	2.1 ± 0.11	408	1.9 ± 0.07*	64
C57BL/10J	2.1 ± 0.11	138		
C58/J	2.5 ± 0.30	88		
CBA/J	2.4 ± 0.34	129		
DBA/2J	2.6 ± 0.16	239	3.1 ± 0.14	64
RIIIS/J	3.0 ± 0.41	86		
Season				
See Fig. 3				
Cage density				
1–3	3.2 ± 0.35[d]	939		
4–6	2.0 ± 0.33	833		
Time of day				
08:00–10:55 h	3.1 ± 0.35	284	2.9 ± 0.13	96
11:00–13:55 h	2.2 ± 0.24	894		
14:00–17:00 h	1.8 ± 0.27	594	2.5 ± 0.10*	96
Sex				
Female	1.9 ± 0.30	888	2.7 ± 0.12	96
Male	2.1 ± 0.32	884	2.8 ± 0.12	96
Humidity				
See Fig. 3				
Order of testing				
First	2.3 ± 0.36	642	3.0 ± 0.19[e]	48
Second	2.0 ± 0.32	567	2.8 ± 0.18	48
Third	1.9 ± 0.30	359	2.6 ± 0.16	48
Fourth	2.1 ± 0.31	204	2.5 ± 0.15	48

Values represent mean ± SEM 49 °C TW latencies (in s). *Significantly different from all other levels of this factor, $p < 0.05$. No attempt was made to assess the significance of group differences from the raw data or LS means.

[a] Only factors and factor levels analyzed in the fixed effects model are presented.

[b] LS means from a subset of the archival data ($n = 1772$), with no missing values, analyzed by fixed-effects modeling. A comparison with analogous raw means of the full data set (not shown) reveals that this subset is remarkably representative of the full data.

[c] Means from a fully-crossed and -balanced experiment ($n = 192$) of May 15, 2001.

[d] LS means suggested that this factor may affect TW latencies in male mice only (not shown).

[e] A trend towards significance was obtained ($p = 0.14$); but see Fig. 6a.

a two-way order of testing × condition (home or holding cage) ANOVA.

In a second experiment, mice of the 11 inbred strains listed in Table 2 were tested for baseline sensitivity as per usual, administered one of a number of doses of morphine sulfate (5–40 mg/kg, i.p., in 10 ml/kg saline), and retested at 15, 30 and 60 min post-injection ($n = 8$–35/dose/order). Analgesia at each time point was expressed as percent analgesia: ([(post-injection latency − baseline latency)/(15-baseline latency) × 100]). Half-maximal analgesic doses (AD_{50}s) were calculated by the method of Tallarida and Murray [118].

2.7. Heritability

Estimates of broad-sense heritability in the archival data set and in the balanced experiment were calculated from variance component estimation by equating the expected mean squares to the observed mean squares in a one-way ANOVA fitting genotype as the only factor. This intraclass correlation approach to the estimation of heritability is valid in the presence of unequal numbers of mice in each strain.

3. Results

Raw summary statistics for this data set are shown in Table 1. The mean latency of all these observations is 3.1 s, with a standard deviation of 1.3 s. Although mean responses of the various strains tested differed profoundly (Fig. 1), the broad-sense heritability estimate obtained from these data (inbred strains only) was 24 ± 0.05%, leaving the majority of the variance to be accounted for by factors other than genotype.

3.1. CART analysis

Estimates of the ability of CART to explain variation in the TW data set are made by model fit statistics. The optimal tree selected by CART (can be viewed online; see Ref. [24]) accounted for 42% of the overall variance in TW latency (based on cross-validation) and had a resubstitution relative error of 49%, analogous to a multiple r^2 of 51%. This latter fit statistic is the correlation between actual observed latencies and those predicted for the factor levels of an individual observation by the tree. These model-fit statistics may represent underestimates, because of the remedial measures taken (see Section 2.3).

The factors, as ranked by CART, are shown in Fig. 2. As can be seen, experimenter and genotype were found to have the greatest association with TW latency, and the trait varies with other environmental factors including season, cage density, time of day (within the light phase), humidity and order of testing. An inspection of the optimal tree revealed some interesting properties of the factors. In agreement with our own work and that of others [10,80], sex differences can

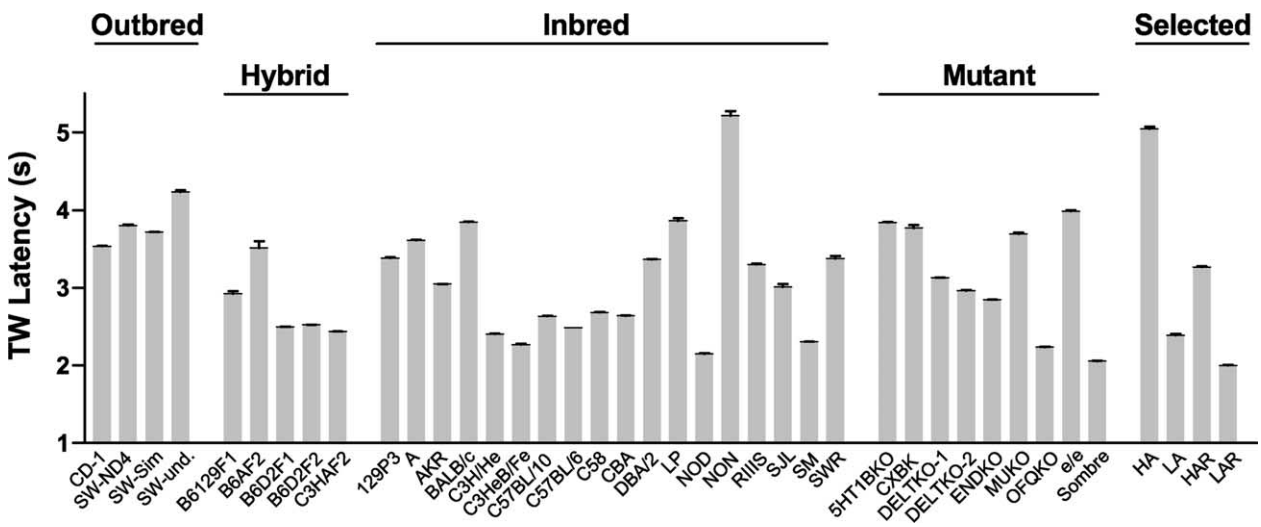

Fig. 1. Latency means (±SEM) on the 49 °C tail-withdrawal (TW) assay of 40 outbred, hybrid, inbred, mutant and artificially-selected populations (all genotypes having $n \geq 50$) tested from 1993 to 2001. Genotype nomenclature is fully described in Table 1.

be observed in mouse pain sensitivity. In every split by sex, females were found to be more sensitive than males to thermal nociception. In virtually every split by order of testing, the first mouse tested displayed a higher latency than all subsequently tested mice. In addition, late-day testing times, Spring testing dates and higher humidity were usually associated with lower latencies.

3.2. Fixed-effects modeling

LS means and standard errors for levels of these factors are shown in Table 2 and Fig. 3. These represent estimates of the means assuming balanced sample sizes in higher-order interactions, and are thus less biased than raw means which are weighted by sample size. That is, LS means give parametric estimates of the within-factor variation in latencies. This analysis also supports the relative importance of experimenter and genotype as the most influential factors in that the largest effects are observed, with a 1.7 s difference between the means of the highest and lowest ranked experimenters and genotypes. Other factors had smaller ranges, consistent with their ranking by CART. Although this analysis is a corroborative parametric analysis of the archival data, the results are not causal evidence of the influence of particular factors. In the model developed to estimate these LS means, the interaction of virtually every factor with genotype was significant. The implications of such interactions are exemplified in Fig. 4, showing the LS means of data from three mouse strains tested by two experimenters.

3.3. Balanced experiment

This experiment was analyzed by a 5-way experimenter × strain × sex × time × order of testing (3 × 2 × 2 × 2 × 2 × 4) ANOVA. For experimenter, genotype and time of day factors, the influence of levels suggested by the raw data and LS means were confirmed as significant (all $p < 0.01$; see Table 2). For sex and order of testing, trends in the same direction as the raw data and LS means were seen, but significant differences were not obtained in these strains and with this sample size ($p = 0.38$ and 0.14, respectively), attesting to the relatively low impact of these factors. Sex exerted more influence as a factor that interacts with environment. Collectively, Fig. 5 shows that 87% of the overall trait variance in this experiment could be explained by genotype (27%), environmental factors (42%) and genotype × environment interactions (18%).

3.4. Order of testing

A two-way ANOVA revealed a significant interaction of experimental condition with order of testing ($F_{3,55} = 3.1$, $p < 0.05$). Thus, as Fig. 6a shows, even the lowest ranking factor from the CART analysis, order

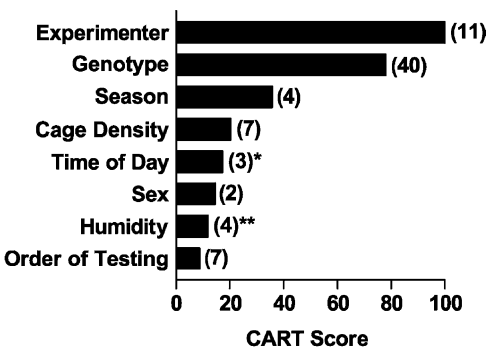

Fig. 2. Factor importance rankings computed by CART. Bars represent scores relative to the highest-ranked factor (100). Numbers in parentheses represent number of factor levels. This figure is adapted from a previously published table [24]. *Time of day levels were: early (09:30–10:55 h), midday (11:00–13:55 h), and late (14:00–17:00 h). **Humidity levels were: high (≥60%), medium-high (40–59%), medium–low (20–39%), and low (<20%).

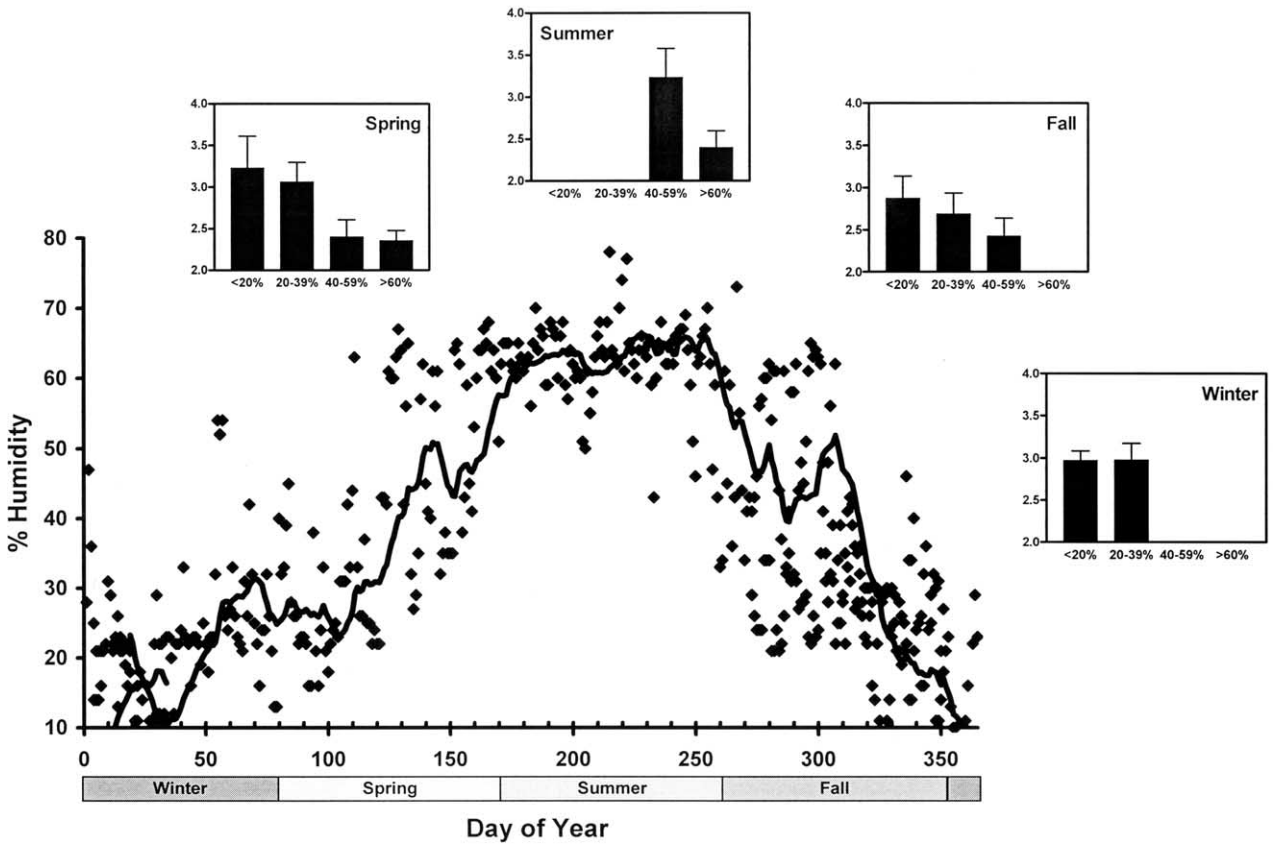

Fig. 3. Influence of humidity and season on 49 °C TW latencies in 1772 inbred mice. Main graph shows vivarium humidity values (midpoint of daily range) for each day. The trend line represents a weekly moving average of the values. Insets show humidity × season interaction LS means (TW latency in seconds) calculated from these data. Only humidity classes per season with $n > 30$ are shown. As can be seen, TW latencies tend to decrease with increases in humidity, except perhaps in the Winter.

of testing, can exert a significant effect in a controlled experiment using a sensitive strain. This effect was eliminated by preventing the exposure of naïve mice to previously tested mice (Fig. 6a).

Fig. 6b shows that order of testing was also found to affect morphine analgesia. Based on non-overlapping 95% confidence intervals, AD_{50} estimates derived from the fourth mice tested in each cage were significantly higher than those derived from the first mice tested. In addition to this potency shift, a clear reduction in morphine efficacy was seen in later-tested mice.

4. Discussion

The results of the present analysis and experiment confirm the important role of genotype in the mediation of thermal nociceptive sensitivity in the mouse, in accordance with previous findings from our laboratory (from subsets of the present data) [83] and those of others (see Refs. [8,79, 82] for reviews). Work is proceeding apace towards the identification of QTLs underlying these strain differences, and the responsible genes within these QTLs [81]. The results also confirm the fact that despite the easily demonstrated role of inherited genetic factors, environmental factors and gene × environment interactions play an even larger role in the production of variability in this trait. For the first time, an attempt has been made to consider many of the relevant factors together.

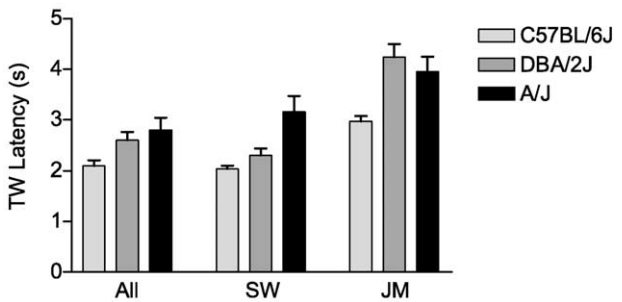

Fig. 4. Genotype × experimenter interaction revealed by LS means. Bars represent LS mean (± SEM) 49 °C TW latencies of three inbred strains (C57BL/6J, DBA/2J and A/J) tested by all experimenters in the data subset described in Table 2 combined (All), or by individual experimenters SW or JM. Although in each case C57BL/6 mice were more sensitive to thermal nociception than A/J, the relative position of the DBA/2 strain was dependent on the experimenter: DBA/2J ≈ C57BL/6J for SW, but DBA/2J ≈ A/J for JM. This pattern was observed again in the balanced experiment (not shown).

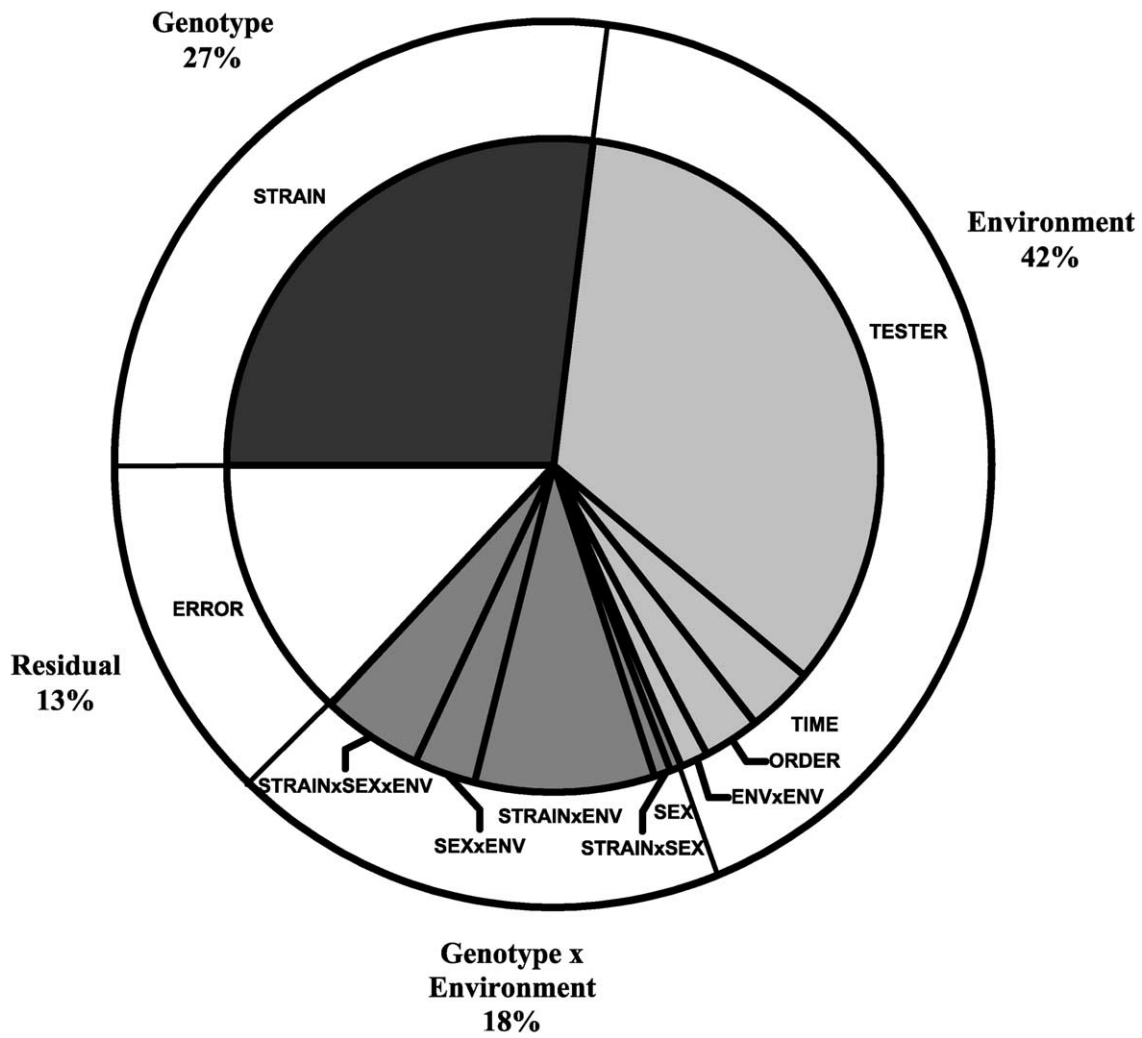

Fig. 5. Partitioning the variance of 49°C TW test variability. Shown are percentages of the corrected total variance in a fully balanced and crossed study performed on A/J, C57BL/6J and DBA/2J mice on a single day. The overall ANOVA multiple $r^2 = 0.87$, indicating that the model accounted for a large portion of the variance in this data set. For simplicity of representation, interactions with experimenter (TESTER), time of day (TIME) and within-cage order of testing (ORDER) are jointly termed ENV. Sex is represented as a genotype × environment factor, although this status is debatable. Regardless of whether sex is considered a purely environmental factor, a purely genetic factor, or an interaction, in this case the influence of sex by itself is miniscule (0.4%); it is the sex × environment interactions that account for 7.9% of the total variance.

4.1. Influence of experimenter

As previously discussed [24], the high importance of experimenter is likely in agreement with the recent findings of Crabbe and colleagues [29], who simultaneously tested a common set of mouse strains on a number of behavioral assays at three different sites using identical equipment, husbandry and testing protocols. Although the relative ranking of the strains in their study was similar at each site, the absolute performance on some behaviors differed greatly from site to site. This variability can only be accounted for by factors not explicitly controlled, notably including the particular experimenters in each laboratory. Precisely what differentiated the experimenters in the present study remains unknown. Neither experimenter age, sex or experience level seemed to correlate with the observed differences. Regarding the former it is essential to note that the experimenter effect is not of the same magnitude in all strains, as would be expected if the difference in latencies obtained by different experimenters was due to a simple additive effect of reaction time. Regarding the effect of experience, we note that baseline latencies of mice tested by SW decreased rather consistently during the course of graduate school (1997 mean: 3.6 ± 0.08 s; 2001 mean: 2.3 ± 0.03 s), whereas JM's latencies remained constant during this period. Although it is tempting to suggest that SW may have been becoming progressively more expert at the technique, it still begs the question as to why JM, with more experience still, consistently collected data with *higher* latencies. JM was, of course, the oldest member of the lab. However, JM did not appear to display longer reaction times in standard reaction time tests, although this was not evaluated rigorously.

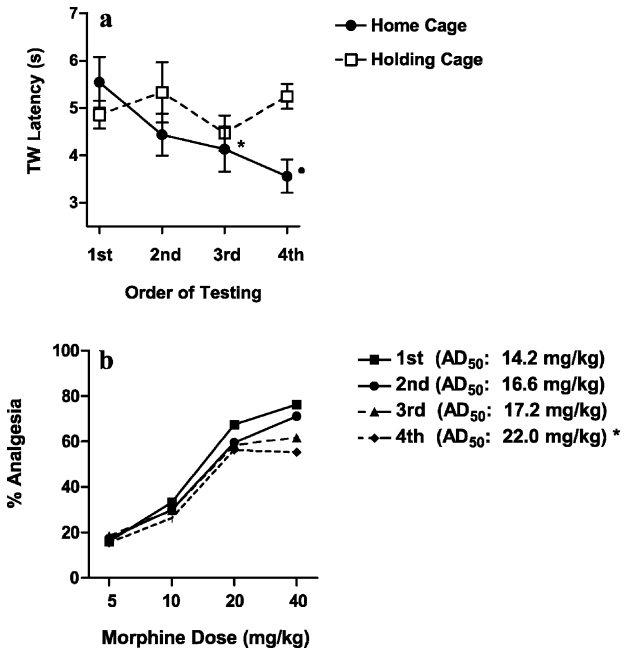

Fig. 6. Influence of within-cage order of testing on nociceptive sensitivity (a) and morphine analgesia (b). Symbols in (a) represent mean (± SEM) 49 °C TW latencies of mice tested and then immediately returned to their home cages or transferred to a holding cage after testing. In the home cage group only, testing affects the latencies of subsequently tested mice. Symbols in (b) represent maximal mean percent analgesia scores (see text) of mice (home cage) tested for morphine inhibition (5–40 mg/kg) of 49 °C TW nociception. Error bars are omitted for clarity. Half-maximal analgesic doses (AD_{50}s) presented in parentheses, were compared statistically on the basis of their 95% confidence intervals (not shown). *Significantly different from first mice, $p < 0.05$. • Significantly different from first mice and Holding Cage (fourth) mice, $p < 0.05$.

Overall, it seems likely that differential animal handling, perhaps inducing different levels of stress in the subjects, is responsible for the 'experimenter effect'. The phenomenon studied presently is perhaps especially highly influenced by experimenter because the mice are handled *as they are being tested*, in contrast to many other nociceptive and behavioral assays. It is also certainly possible that different experimenters affect latencies on this test by a more generalized mechanism, for example their particular odor.

It is important to note that although the grand mean of all strains varied greatly depending on who was conducting the testing, the ranked sensitivities of each strain were generally preserved (not shown). However, the fact that the experimenter doing the testing is more highly associated with TW latencies than the subject's genotype emphasizes the importance of environmental variables.

4.2. Influence of season and humidity

The present data appear to suggest that sensitivity to a noxious thermal stimulus is higher in conditions of high humidity (Fig. 3), which in natural circumstances in temperate climates would correspond to late spring, summer and early fall. Circannual (i.e. seasonal) and weather effects on pain-related traits are not entirely unknown, having been studied as early as 1887 [78]. Circannual variation in concentrations of endorphins in human cerebrospinal fluid has been reported, with lower values in the summer and higher values in the winter [124]. Human arthritic sufferers report more pain in conditions of high humidity and low temperature [5,114], but this may be more a matter of perception than reality [54]. Seasonal influences have been shown to affect analgesic magnitude and even neurochemistry in rodents [59,101,129], but we are unaware of any systematic study of the influence of season or humidity on experimental pain sensitivity in either humans or laboratory animals.

Although our vivaria were maintained on a constant 12:12 h light/dark cycle, the obviously uncontrolled humidity (Fig. 3) could serve as a seasonal cue to the animals. Although *Mus musculus domesticus* is fecund year-round, the percentage of females producing litters and the size of those litters are lower in the winter, in both wild and laboratory stocks [36]. It is tempting to suggest that the season/humidity effect on nociception seen here may be related to reproduction, given the considerable anatomical and neurochemical overlap of central nervous system circuitry devoted to pain modulation and reproductive control [13].

More generally, we would suggest that the season/humidity effect may have measurable influences on any number of behavioral traits [89]. It is virtually unheard of for such parameters to be reported in the biobehavioral literature; these data suggest a need for their reporting and/or control.

4.3. Influence of cage density

Housing density in the laboratory setting is known to affect any number of morphological and biochemical variables [50,87], notably including plasma corticosterone levels [96]. There exists a large literature on the effects of social isolation on thermal nociception and analgesia [4,28,34,45,70,88,94,104,111,115], with virtually all groups showing a hypoalgesic effect of isolation (during development or in adulthood) on baseline sensitivity of adult rodents. The present data are in agreement with this literature. In contrast, only one published study has looked at the modulatory effects of crowding on nociception, comparing the responses of rats housed 10 per cage (with 95 cm^2 of floor space per rat) to those housed 3 per cage (with 710 cm^2) [99]. Crowding affected nociception in the tailshock and paw pressure tests, but not the 50°C TW test.

Of interest is the fact that our fixed-effects modeling suggested that cage density affected nociception largely in males (see Table 2). It has been shown in rats that, based on plasma corticosterone levels, crowding stresses males (likely related to inter-male aggression and the maintenance of dominance hierarchies) but calms females [19].

4.4. Influence of time of day

A considerable literature, dating back to 1912 [49], documents the influence of circadian rhythms on pain sensitivity and analgesia in both humans and laboratory animals (see Refs. [71,97] for reviews). This literature is quite contradictory in humans, with the chronobiology of pain depending to a large degree on the particular clinical syndrome or experimental model studied [71]. In rodents and for thermal nociception, the situation is somewhat clearer, with decreased sensitivity (on the tail-flick and hotplate tests) in the dark (active) phase compared to the light (resting) phase reported in 11 of 14 applicable studies reviewed by Perissin and colleagues [97]. It is much more difficult, however, to discern a clear pattern in the literature *within* a photophase, and many studies simply compared day to night. Of the mouse studies with multiple diurnal data points, our findings appear to be in contrast with some [42,60,93], but in accordance with the work of Golombek and colleagues [48]. In all cases, though, the hotplate test was used, and unlike the TW test the former assay is affected by (rhythmic) activity levels. It should also be noted that circadian rhythms of nociception are clearly strain-dependent [22,103]. It is interesting that β-endorphin concentrations in pain-relevant loci in the rat (thalamus and periaqueductal gray) peak at 08:00 h [64], which may confer the 'morning analgesia' effect that we see here. In any case, researchers should be aware that circadian rhythms may exert effects within photoperiods in addition to between them.

4.5. Influence of sex

Sex exerted more influence as a factor that interacts with environment (Fig. 5), implying that biological sex differences may manifest themselves in differential sensitivity to environmental factors. Our data thus suggest that sex differences in this trait may largely be an epiphenomenon of the sex/testing environment interaction. Studies of human pain sensitivity may corroborate this in that sex differences seem highly related to anxiety [108] and attention [63], factors that are highly variable in experimental as well as clinical situations.

4.6. Influence of order of testing

We are unaware of any prior demonstration that within-cage order of testing may affect nociceptive or analgesic sensitivity. The results of the experiment shown in Fig. 6a strongly suggest that this phenomenon is mediated by inter-animal communication. It is unclear at the present time whether this communication is achieved via chemosensation [102] (the release of a pheromonal 'alarm substance' [2]) or ultrasonic vocalization (the emission of an 'alarm cry' [12]). We suspect that the former possibility is more likely, since mice apparently do not emit fear-related ultrasonic vocalizations as do rats [11]. Both the release of alarm substances and ultrasonic vocalization behavior are known to be genotype-dependent in rats [3,69], which may explain why the phenomenon was more robust in SW:Sim mice than in the three inbred strains tested on May 15, 2001.

In an elegant recent study, Raber and Devor [105] demonstrated that the phenotype of rats artificially selected for low autotomy (LA) behavior (self-mutilation of a denervated limb; a model of neuropathic pain), despite its high heritability could be overridden by exposure to the soiled bedding of high-autotomy (HA) rats. The effect was even stronger when LA rats were actually housed with HA rats, whether or not the cagemates were familiar and whether or not the HA rats were actually denervated and performing autotomy. The precise relevance of this finding to the present order-of-testing phenomenon is unclear given the vast difference in time scales, but Raber and Devor's [105] study certainly illustrates in dramatic fashion how social variables, including chemical communication among cagemates, can 'trump' genetic susceptibility.

Regardless of the method of communication, it is interesting to speculate on the nature of the message being sent in the present situation. Presumably, the mouse is communicating distress at the handling and/or the exposure to a noxious stimulus. Without further experimentation, it remains unclear whether later-tested mice are rendered more or less fearful (i.e. stressed) than earlier-tested mice. It is also the case that stress can increase or decrease sensitivity to pain (i.e. produce stress-induced hyperalgesia or stress-induced analgesia, respectively) depending on any number of situational parameters [58,62].

Although order of testing was the least influential factor in the present data set, it is one that may be broadly relevant to any number of common behavioral assays. Also, the relevance of this factor to pain research is magnified by our observation that measurements of the potency and/or efficacy of analgesic drugs like morphine are even more greatly affected by order of testing, with the first mouse tested from a cage $>50\%$ more sensitive to the drug than the fourth mouse (Fig. 6b). This effect does not appear to be specific to morphine analgesia, being evident in analogous data sets using different (non-opioid) centrally acting analgesics (data not shown).

4.7. Genetic/environmental correlations and interactions

The evaluation of strain effects, the estimation of heritability of traits, and even the testing of novel mutations assumes an ability to dissociate environmental sources of phenotypic variability from genetic sources of variability. When these become confounded in the course of running experiments, a rather frequent occurrence, the environmental effects become correlated with and may mask or exaggerate genetic effects. That is, differential effects of the environment could result in over- or underestimation

of genetic effects, particularly in the event that a phenotype is conditionally observed in particular environments.

It is clear from these data that genetic and environmental factors are interacting to importantly affect thermal nociceptive sensitivity. We suspect that stress, and genetically determined responsivity to stress, may represent crucial mediators of this interaction. Although current biases in the pain research field would tend to attribute differences in basal TW latencies to afferent fiber activity or spinal processing, it is equally possible that differential activation of descending modulatory circuitry—either analgesic [62] or hyperalgesic [58]—is responsible. Given the fact that the behavior measured here is indeed a spinal reflex, one might expect other pain behaviors (and other behaviors more generally) to be even more importantly influenced by environmental factors.

4.8. Test condition variation or standardization?

The present work may have implications for the current debate regarding standardization of testing conditions in behavior genetics [120,121,126]. Though it may be feasible to standardize testing conditions across testing situations for some factors, it may never be practical to fully control laboratory parameters. In other words, true 'standardization' is not possible. As such, the research is always assuming construct validity and reliability of a measure in the presence of fluctuating environmental variables. The best protection against erroneous interpretation of results through false attribution of effects to the intended manipulation rather than to systematic nuisance variables is to repeatedly test a single construct in a diverse set of conditions (as we have done here), and through a diverse set of measures [72,83,84]. It is dangerously limiting to generalizability to restrict the testing conditions for behavioral assessment of a single trait known to vary across assays or assay parameters. However, practical issues, including cost of performing the additional assays required and effective database construction for large-scale projects, do require that standardization rather than systematic variation of parameters be attempted. Recent efforts address these issues through empirical optimization of standards as a means of choosing robust standards that are both informative and reliably measured [127]. This alternative to the use of parameters derived by historical accident may be a necessary compromise to systematic variation of laboratory environment over the course of an experiment.

4.9. Conclusions and relevance

The study of neurobehavioral traits, particularly using a genetic approach, is an ambitious undertaking performed primarily because of the societal importance of understanding such traits, and despite sub-optimal conditions of low heritability and high environmental influence [100]. This enterprise is being refined by our increasing understanding of the role of the laboratory environment and a revisiting of methodological issues, particularly as assays developed in other species are being tailored for use in the mouse. Assessing the adequacy of our assumption that the laboratory is sufficiently regulated for the study of behavioral traits across large numbers of studies will be essential for the completion of large-scale projects such as those currently being proposed in this post-genomic era. The ability to extract knowledge of the influence of the context in which experiments are performed from an aggregate body of experiments may facilitate more intelligent approaches to addressing environmental influence, other than the requisite assumption that laboratory inconsistency is a necessary evil in the enterprise of behavioral research.

The consequences and the applicability of this approach need not be limited to behavioral genetics research. It is possible to use data mining as performed here to study the influence of any large number of factors on a measure, as a means of reducing the pool of variables studied in controlled experiments. Virtually every laboratory that has repeatedly performed a single assay of neuroscientific interest has extant resources for such analysis accumulated over the years, and the role of laboratory and organismic factors on a vast array of traits can be evaluated quantitatively in these archives for hypothesis generation prior to further experimentation.

The present study demonstrates that for a biobehavioral trait like thermal nociceptive sensitivity, it is possible to use an analytic approach to identify and rank both genetic and environmental factors associated with trait variance. Ultimately, the operation of all the factors we considered - and many others we were unable to consider here-needs to be further explicated with mechanistic studies in mice and humans. Given that an overwhelming proportion of variability in nociceptive sensitivity is accounted for by environmental factors and their interaction with genes, the mere elucidation of pain genes will not be sufficient to explain individual differences. Once the relevant genes are found, systematic investigation of interactions between these genes and environmental factors like those identified herein may yield clinically important information that may facilitate individualized pharmacologically and behaviorally based treatment strategies.

Acknowledgements

We thank Drs John Belknap, John Crabbe, Marshall Devor, Ron Melzack and the anonymous reviewers for their helpful comments on the manuscript, Dr Larry Hubert for suggesting CART analysis, and Ms Brenda Edwards and her staff for excellent animal care and record-keeping. This work was supported by US Public Health Service grants DA11394, DE12735 and DA15191 to J.S.M.

References

[1] Abbott FV, Franklin KBJ, Connell B. The stress of a novel environment reduces formalin pain: possible role of serotonin. Eur J Pharmacol 1986;126:141–4.

[2] Abel EL. The pituitary mediates production or release of an alarm chemosignal in rats. Horm Behav 1994;28:139–45.

[3] Abel EL. Response to alarm substance in different rat strains. Physiol Behav 1992;51:345–7.

[4] Adler MW, Mauron C, Samanin R, Valzelli L. Morphine analgesia in grouped and isolated rats. Psychopharmacologia (Berl.) 1975;41: 11–14.

[5] Aikman H. The association between arthritis and the weather. Int J Biometeorol 1997;40:192–9.

[6] Aloisi AM, Albonetti ME, Carli G. Behavioral effects of different intensities of formalin pain in rats. Physiol Behav 1995;58: 603–10.

[7] Barrett AC, Cook CD, Terner JM, Roach EL, Syvanthong C, Picker MJ. Sex and rat strain determine sensitivity to κ opioid-induced antinociception. Psychopharmacology 2002;160:170–81.

[8] Belknap JK, O'Toole LA. Studies of genetic differences in response to opioid drugs. In: Harris RA, Crabbe JC, editors. The genetic basis of alcohol and drug actions. New York: Plenum Press; 1991. p. 225–52.

[9] Ben-Bassat J, Peretz E, Sulman FG. Analgesimetry and ranking of analgesic drugs by the receptacle method. Arch Int Pharmacodyn Ther 1959;122:434–47.

[10] Berkley KJ. Sex differences in pain. Behav Brain Sci 1997;20: 371–80.

[11] Blanchard DC, Griebel G, Blanchard RJ. Mouse defensive behaviors: pharmacological and behavioral assays for anxiety and panic. Neurosci Biobehav Rev 2001;25:205–18.

[12] Blanchard RJ, Blanchard DC, Agullana R, Weiss SM. Twenty-two kHz alarm cries to presentation of a predator, by laboratory rats living in visible burrow systems. Physiol Behav 1991;50:967–72.

[13] Bodnar RJ, Commons K, Pfaff DW. Central neural states relating sex and pain. Baltimore: Johns Hopkins University Press; 2002.

[14] Bon K, Lanteri-Minet M, Menetrey D, Berkley KJ. Sex, time-of-day and estrous variations in behavioral and bladder histological consequences of cyclophosphamide-induced cystitis in rats. Pain 1997;73:423–9.

[15] Bonnet KA, Hiller JM, Simon EJ. The effects of chronic opiate treatment and social isolation on opiate receptors in the rodent brain. Opiates and endogenous opioids. Amsterdam: Elsevier/North Holland Press; 1976. p. 335–44.

[16] Bradshaw H, Miller J, Ling Q, Malsnee K, Ruda MA. Sex differences and phases of the estrous cycle alter the response of spinal cord dynorphin neurons to peripheral inflammation and hyperalgesia. Pain 2000;85:93–9.

[17] Breault JL, Goodall CR, Fos PJ. Data mining a diabetic data warehouse. Artif Intell Med 2002;26:37–54.

[18] Breiman L. Classification and regression trees. Pacific Grove, CA: Wadsworth; 1984.

[19] Brown KJ, Grunberg NE. Effects of housing on male and female rats: crowding stresses males but calms females. Physiol Behav 1995;58: 1085–9.

[20] Cabib S, Orsini C, Le Moal M, Piazza PV. Abolition and reversal of strain differences in behavioral responses to drugs of abuse after a brief experience. Science 2000;289:463–5.

[21] Caggiula AR, Antelman SM, Aul E, Knopf S, Edwards DJ. Prior stress attenuates the analgesic response but sensitizes the corticosterone and cortical dopamine responses to stress 10 days later. Psychopharmacology 1989;99:233–7.

[22] Castellano C, Puglisi-Allegra S, Renzi P, Oliverio A. Genetic differences in daily rhythms of pain sensitivity in mice. Pharmacol Biochem Behav 1985;23:91–2.

[23] Chen Z, Inoue M, Hopkins E, Pellegrino MJ, Nowakowki R, Mogil JS, Kest B, Allen RG, Ueda H, Pintar JE. Nociceptin/OFQ gene knockout reveals an antinociceptive role for the nociceptin/OFQ in the mouse. Soc Neurosci Abstr 1999;25:1472.

[24] Chesler EJ, Wilson SG, Lariviere WR, Rodriguez-Zas SL, Mogil JS. Influences of laboratory environment on behavior. Nature Neurosci 2002;5:1101–2.

[25] Clausing P, Mothes HK, Opitz B, Kormann S. Differential effects of communal rearing and preweaning handling on open-field behavior and hot-plate latencies in mice. Behav Brain Res 1997;82:179–84.

[26] Conway CM, Martinez J, Lytle LD. Maturational changes in the thermal nociceptive responses of developing rats. Dev Psychobiol 1998;33:47–60.

[27] Cook CD, Barrett AC, Roach EL, Bowman JR, Picker MJ. Sex-related differences in the antinociceptive effects of opioids: importance of rat genotype, nociceptive stimulus intensity, and efficacy at the m opioid receptor. Psychopharmacology 2000;150: 430–42.

[28] Coudereau J-P, Monier C, Bourre J-M, Frances H. Effect of isolation on pain threshold and on different effects of morphine. Prog Neuro-Psychopharmacol Biol Psychiatry 1997;21:997–1018.

[29] Crabbe JC, Wahlsten D, Dudek BC. Genetics of mouse behavior: interactions with laboratory environment. Science 1999;284: 1670–2.

[30] Cunningham DJ, Benson WM, Hardy JD. Modification of the thermal radiation method for assessing antinociceptive activity in the rat. J Appl Physiol 1957;11:459–64.

[31] d'Amore A, Chiarotti F, Renzi O. High-intensity nociceptive stimuli minimize behavioral effects induced by restraining stress during the tail-flick test. J Pharmacol Toxicol Meth 1992;27:197–201.

[32] d'Amore A, Mazzucchelli A, Loizzo A. Long-term changes induced by neonatal handling in the nociceptive threshold and body weight in mice. Physiol Behav 1995;17:1195–7.

[33] D'Amour FE, Smith DL. A method for determining loss of pain sensation. J Pharmacol Exp Ther 1941;72:74–9.

[34] Defeudis FV, Defeudis PA, Somoza E. Altered analgesic responses to morphine in differentially housed mice. Psychopharmacology 1976;49:117–8.

[35] DeLeo JA, Rutkowski MD. Gender differences in rat neuropathic pain sensitivity is dependent on strain. Neurosci Lett 2000;282: 197–9.

[36] Drickamer LC. Seasonal variation in fertility, fecundity and litter sex ratio in laboratory and wild stocks of house mice (*Mus domesticus*). Lab Anim Sci 1990;40:284–8.

[37] Drury RA, Gold RM. Differential effects of ovarian hormones on reactivity to electric footshock in the rat. Physiol Behav 1978;20: 187–91.

[38] Fields HL, Heinricher MM. Brainstem modulation of nociceptor-driven withdrawal reflexes. Ann NY Acad Sci 1989;563:34–44.

[39] Fillingim RB, Maixner W. Gender differences in the responses to noxious stimuli. Pain Forum 1995;4:209–21.

[40] Fitzgerald M. Developmental biology of inflammatory pain. Br J Anaesth 1995;75:177–85.

[41] Francis D, Diorio J, Liu D, Meaney MJ. Nongenomic transmission across generations of maternal behavior and stress responses in the rat. Science 1999;286:1155–8.

[42] Frederickson RCA, Burgis V, Edwards JD. Hyperalgesia induced by naloxone follows diurnal rhythm in responsivity to painful stimuli. Science 1977;198:756–8.

[43] Frye CA, Bock BC, Kanarek RB. Hormonal milieu affects tailflick latency in female rats and may be attenuated by access to sucrose. Physiol Behav 1992;52:699–706.

[44] Frye CA, Cuevas CA, Kanarek RB. Diet and estrous cycle influence pain sensitivity in rats. Pharmacol Biochem Behav 1993;45:255–60.

[45] Gentch C, Lichtsteiner M, Frischknecht H-R, Feer H, Siegfried B. Isolation-induced locomotor hyperactivity and hypoalgesia in rats

are prevented by handling and reversed by resocialization. Physiol Behav 1988;43:13–16.
[46] Gerald LB, Tang S, Bruce F, Redden D, Kimerling ME, Brook N, Dunlap N, Bailey WC. A decision tree for tuberculosis contact investigation. Am J Respir Crit Care Med 2002;15:1122–7.
[47] Gintzler AR. Endorphin-mediated increases in pain threshold during pregnancy. Science 1980;210:193–5.
[48] Golombek DA, Escolar E, Burin LJ, De Brito Sanchez MG, Cardinali DP. Time-dependent melatonin analgesia in mice: inhibition by opiate or benzodiazepine antagonism. Eur J Pharmacol 1991;194:25–30.
[49] Grabfield GP, Martin EG. Variations in the sensory threshold for faradic stimulation in normal human subjects. I. The diurnal rhythm. Am J Physiol 1912;31:300–8.
[50] Grewal IS, Miller A, Sercarz EE. The influence of mouse housing density on autoimmune reactivity. Autoimmunity 1997;26:209–14.
[51] Grotto M, Sulman FG. Modified receptacle method for animal analgesimetry. Arch Int Pharmacodyn 1967;165:152–9.
[52] Hamann SR, Martin WR. Thermally evoked tail avoidance reflex: input-output relationships and their modulation. Brain Res Bull 1992;29:507–9.
[53] Hammond DL. Inference of pain and its modulation from simple behaviors. In: Chapman CR, Loeser JD, editors. Advances in pain research and therapy: issues in pain management. New York: Raven Press; 1989. p. 69–91.
[54] Hawley DJ, Wolfe F, Lue FA, Moldofsky H. Seasonal symptom severity in patients with rheumatic diseases: a study of 1424 patients. J Rheumatol 2001;28:1900–9.
[55] Irwin S, Houde RW, Bennett DR, Hendershot LC, Seevers MH. The effects of morphine, methadone and meperidine on some reflex responses of spinal animals to nociceptive stimulation. J Pharmacol Exp Ther 1951;101:132–43.
[56] Jackson HC, Kitchen I. Perinatal lead exposure impairs opioid but not non-opioid stress-induced antinociception in developing rats. Br J Pharmacol 1989;97:1338–42.
[57] Janssen PAJ, Niemegeers CJE, Dony JGH. The inhibitory effect of fentanyl and other morphine-like analgesics on the warm water induced tail withdrawal reflex. Arzneim-Forschung (Drug Res) 1963;13:502–7.
[58] Jorum E. Analgesia or hyperalgesia following stress correlates with emotional behavior in rats. Pain 1988;32:341–8.
[59] Kavaliers M, Galea LAM. Sex differences in the expression and antagonism of swim stress-induced analgesia in deer mice vary with the breeding season. Pain 1995;63:327–34.
[60] Kavaliers M, Hirst M. Daily rhythms of analgesia in mice: effects of age and photoperiod. Brain Res 1983;279:387–93.
[61] Kavaliers M, Ossenkopp K-P. Exposure to rotating magnetic fields alters morphine-induced behavioural responses in two strains of mice. Neuropharmacology 1984;89:440–3.
[62] Kelly DD, editor. Ann NY Acad Sci 1986;467.
[63] Keogh E, Hatton K, Ellery D. Avoidance versus focused attention and the perception of pain: differential effects for men and women. Pain 2000;85:225–30.
[64] Kerdelhue B, Karteszi M, Pasqualini C, Reinberg A, Mezey E, Palkovits M. Circadian variations in beta-endorphin concentrations in pituitary and in some brain nuclei of the adult male rat. Brain Res 1983;261:243–8.
[65] Kest B, Sarton E, Dahan A. Gender differences in opioid-mediated analgesia: animal and human studies. Anesthesiology 2000;93: 539–47.
[66] Kest B, Wilson SG, Mogil JS. Sex differences in supraspinal morphine analgesia are dependent on genotype. J Pharmacol Exp Ther 1999;289:1370–5.
[67] Kim YI, Na HS, Yoon YW, Nahm SH, Ko KH, Hong SK. Mechanical allodynia is more strongly manifested in older rats in an experimental model of peripheral neuropathic. Neurosci Lett 1995; 199:158–60.
[68] Kinsley CH, Mann PE, Bridges RS. Prenatal stress alters morphine- and stress-induced analgesia in male and female rats. Pharmacol Biochem Behav 1988;30:123–8.
[69] Knapp DJ, Kampov-Polevoy AB, Overstreet DH, Breese GR, Rezvani AH. Ultrasonic vocalization behavior differs between lines of ethanol-preferring and nonpreferring rats. Alcohol: Clin Exp Res 1997;21:1232–40.
[70] Kostowski W, Czlonkowski A, Rewerski W, Piechocki T. Morphine action in grouped and isolated rats and mice. Psychopharmacology 1977;53:191–3.
[71] Labrecque G, Vanier M-C. Biological rhythms in pain and in the effects of opioid analgesics. Pharmacol Ther 1995;68:129–47.
[72] Lariviere WR, Wilson SG, Laughlin TM, Kokayeff A, West EE, Adhikari SM, Wan Y, Mogil JS. Heritability of nociception. III. Genetic relationships among commonly used assays of nociception and hypersensitivity. Pain 2002;97:75–86.
[73] Le Bars D, Gozariu M, Cadden SW. Animal models of nociception. Pharmacol Rev 2001;53:597–652.
[74] Marks HE, Fargason BD, Hobbs SH. Reactivity to aversive stimuli as a function of alterations in body weight in normal and gonadectomized female rats. Physiol Behav 1972;9:539–44.
[75] Miczek KA, Thompson ML, Shuster L. Opioid-like analgesia in defeated mice. Science 1982;215:1520–2.
[76] Milne RJ, Gamble GD. Habituation to sham testing procedures modifies tail-flick latencies: effects on nociception rather than vasomotor tone. Pain 1989;39:103–7.
[77] Miner LL, Elmer GI, Pieper JO, Marley RJ. Aggression modulates genetic influences on morphine analgesia as assessed using a classical mendelian cross analysis. Psychopharmacology 1993;111: 17–22.
[78] Mitchell S. The relations of pain to weather. Am J Med Sci 1887;73: 305–29.
[79] Mogil JS. The genetic mediation of individual differences in sensitivity to pain and its inhibition. Proc Natl Acad Sci USA 1999;96:7744–51.
[80] Mogil JS, Chesler EJ, Wilson SG, Juraska JM, Sternberg WF. Sex differences in thermal nociception and morphine antinociception in rodents depend on genotype. Neurosci Biobehav Rev 2000;24:375–89.
[81] Mogil JS, Richards SP, O'Toole LA, Helms ML, Mitchell SR, Belknap JK. Genetic sensitivity to hot-plate nociception in DBA/2J and C57BL/6J inbred mouse strains: possible sex-specific mediation by δ2-opioid receptors. Pain 1997;70:267–77.
[82] Mogil JS, Sternberg WF, Marek P, Sadowski B, Belknap JK, Liebeskind JC. The genetics of pain and pain inhibition. Proc Natl Acad Sci USA 1996;93:3048–55.
[83] Mogil JS, Wilson SG, Bon K, Lee SE, Chung K, Raber P, Pieper JO, Hain HS, Belknap JK, Hubert L, Elmer GI, Chung JM, Devor M. Heritability of nociception. I. Responses of eleven inbred mouse strains on twelve measures of nociception. Pain 1999;80:67–82.
[84] Mogil JS, Wilson SG, Bon K, Lee SE, Chung K, Raber P, Pieper JO, Hain HS, Belknap JK, Hubert L, Elmer GI, Chung JM, Devor M. Heritability of nociception. II. Types of nociception revealed by genetic correlation analysis. Pain 1999;80:83–93.
[85] Mogil JS, Wilson SG, Wan Y. Assessing nociception in murine subjects. In: Kruger L, editor. Methods in pain research. Boca Raton, FL: CRC Press; 2001. p. 11–39.
[86] Mogil JS, Yu L, Basbaum AI. Pain genes? natural variation and transgenic mutant. Annu Rev Neurosci 2000;23:777–811.
[87] Nagy TR, Krzywanski D, Li J, Meleth S, Desmond R. Effect of group vs. single housing on phenotypic variance in C57BL/6J mice. Obes Res 2002;10:412–5.
[88] Naranjo JR, Fuentes JA. Association between hypoalgesia and hypertension in rats after short-term isolation. Neuropharmacology 1985;24:167–71.
[89] Nelson RJ, Badura LL, Goldman BD. Mechanisms of seasonal cycles of behavior. Annu Rev Psychol 1990;41:81–108.

[90] Neri M, Agazzani E. Aging and right-left asymmetry in experimental pain measurement. Pain 1984;19:43–8.
[91] Ness TJ, Gebhart GF. Centrifugal modulation of the rat tail flick reflex evoked by graded noxious heating of the tail. Brain Res 1986; 386:41–52.
[92] Ness TJ, Jones SL, Gebhart GF. Contribution of the site of heating to variability in the latency of the rat tail flick reflex. Brain Res 1987; 426:169–72.
[93] Ohdo S, Ogawa N, Song JG. Chronopharmacological study of acetylsalicylic acid in mice. Eur J Pharmacol 1995;293:151–7.
[94] Panskepp J. Brief social isolation, pain responsivity, and morphine analgesia in young rats. Psychopharmacology 1980;72:111–2.
[95] Pare WP. Age, sex and strain differences in the aversive threshold to grid shock in the rat. J Comp Physiol Psychol 1969;69:214–8.
[96] Peng X, Lang CM, Drozdowicz CK, Ohlsson-Wilhelm BM. Effect of cage population density on plasma corticosterone and peripheral lymphocyte populations of laboratory mice. Lab Anim 1989;23: 302–6.
[97] Perissin L, Facchin P, Porro CA. Diurnal variations in tonic pain reactions in mice. Life Sci 2000;67:1477–88.
[98] Pieretti S, d'Amore A, Loizzo A. Long-term changes induced by developmental handling on pain threshold: effects of morphine and naloxone. Behav Neurosci 1991;105:215–8.
[99] Pilcher CWT, Browne J. Environmental crowding modifies responding to noxious stimuli and the effects of μ- and κ-agonists. Life Sci 1982;31:1213–6.
[100] Plomin R. The role of inheritance in behavior. Science 1990;248: 183–8.
[101] Popova NK, Koriakina LA. Genetic and seasonal characteristics of the effect of stress on pain sensitivity in mice. Biull Eksp Biol Med 1986;191:11–13.
[102] Prasad BC, Reed RR. Chemosensation: molecular mechanisms in worms and mammals. Trends Genet 1999;15:150–3.
[103] Puglisi-Allegra S, Castellano C, Oliverio A. Circadian variations in stress-induced analgesia. Brain Res 1982;252:373–6.
[104] Puglisi-Allegra S, Oliverio A. Social isolation: effects on pain threshold and stress-induced analgesia. Pharmacol Biochem Behav 1983;19:6479–681.
[105] Raber P, Devor M. Social variables affect phenotype in the neuroma model of neuropathic pain. Pain 2002;97:139–50.
[106] Riley III JL, Robinson ME, Wise EA, Myers CD, Fillingim RB. Sex differences in the perception of noxious experimental stimuli: a meta-analysis. Pain 1998;74:181–7.
[107] Roane DS, Porter JR. Nociception and opioid-induced analgesia in lean (Fa/–) and obese (fa/fa) Zucker rats. Physiol Behav 1986;38: 215–8.
[108] Rollman GB. Gender differences in pain: role of anxiety. Pain Forum 1995;4:231–4.
[109] Roubertoux PL, Le Roy-Duflos I. Quantitative trait locus mapping: fishing strategy or replicable results? Behav Genet 2001;31:141–8.
[110] Shir Y, Ratner A, Raja SN, Campbell JN, Seltzer Z. Neuropathic pain following partial nerve injury in rats is suppressed by dietary soy. Neurosci Lett 1998;240:73–6.
[111] Siegfried B, Frischknecht HR. Naltrexone-reversible pain suppression in the isolated attacking mouse. Behav Neural Biol 1988;50: 354–60.
[112] Steinberg D, Colla P. CART: non-structured non-parametric data analysis. San Diego, CA: Salford Systems; 1995.
[113] Stolberg A, Frenk H. Development of water-immersion produced analgesia. Dev Psychobiol 1995;28:247–55.
[114] Strusberg I, Mendelberg RC, Serra HA, Strusberg AM. Influence of weather conditions on rheumatic pain. J Rheumatol 2002;29:335–8.
[115] Szikszay M, Benedek G. Tolerance to morphine induced by chronic mild environmental stressors. Acta Physiol Hung 1989;73:447–53.
[116] Szuran T, Zimmerman E, Pliska V, Pfister HP, Welzl H. Prenatal stress effects on exploratory activity and stress-induced analgesia in rats. Dev Psychobiol 1991;24:361–72.
[117] Taber RI. Predictive value of analgesic assays in mice and rats. Adv Biochem Psychopharmacol 1974;8:191–211.
[118] Tallarida RJ, Murray RB. Manual of pharmacologic calculation. New York: Springer; 1981.
[119] Tjolsen A, Hole K. The tail-flick latency is influenced by skin temperature. APS J 1993;2:107–11.
[120] van der Staay FJ, Steckler T. Behavioural phenotyping of mouse mutants. Behav Brain Res 2001;125:3–12.
[121] van der Staay FJ, Steckler T. The fallacy of behavioral phenotyping without standardisation. Genes Brain Behav 2002;1:9–13.
[122] Vickers C, Paterson AT. Two types of chronic lead treatment in C57BL/6 mice: interaction with behavioural determinants of pain. Life Sci 1986;39:47–53.
[123] Vieira C, Pasyukova EG, Zeng ZB, Hackett JB, Lyman RF, Mackay TF. Genotype-environment interaction for quantitative trait loci affecting life span in *Drosophila melanogaster*. Genetics 2000;154: 213–27.
[124] von Knorring L, Almay BGL, Johansson F, Terenius L, Wahlstrom A. Circannual variation in concentrations of endorphins in cerebrospinal fluid. Pain 1982;12:265–72.
[125] Wahlsten D. Single-gene influences on brain and behavior. Annu Rev Psychol 1999;50:599–624.
[126] Wahlsten D. Standardizing tests of mouse behavior: reasons, recommendations, and reality. Physiol Behav 2001;73:695–704.
[127] Wahlsten D, Crabbe JC, Metten P, Rustay N, Cooper SF. Assessing the reliability of tests of behavior in standard inbred mouse strains. Soc Neurosci Abstr 2001;27:#312.1.
[128] Wilson SG, Mogil JS. Measuring pain in the (knockout) mouse: big challenges in a small mammal. Behav Brain Res 2001;125: 65–73.
[129] Winters WD, Hance AJ, Cadd GC, Lakin ML. Seasonal and sex influences on ketamine-induced analgesia and catalepsy in the rat: a possible role for melatonin. Neuropharmacology 1986;25: 1095–101.
[130] Woolf CJ. Evidence for a central component of post-injury pain hypersensitivity. Nature 1983;308:686–8.
[131] Woolf CJ, Salter MW. Neuronal plasticity: increasing the gain in pain. Science 2000;288:1765–8.
[132] Wurbel H. Behavioral phenotyping enhanced-beyond (environmental) standardization. Genes Brain Behav 2002;1:3–8.
[133] Wurbel H. Behaviour and the standardization fallacy. Nature Genet 2000;26:263.
[134] Yoburn BC, Morales R, Kelly DD, Inturrisi CE. Constraints on the tailflick assay: morphine analgesia and tolerance are dependent upon locus of tail stimulation. Life Sci 1984;34:1755–62.