Chem. Senses 27: 759–768, 2002

Influence of Test Duration on the Sensitivity of the Two-bottle Choice Test

Michael G. Tordoff and Alexander A. Bachmanov

Monell Chemical Senses Center, 3500 Market St., Philadelphia, PA 19104, USA

Correspondence to be sent to: Michael G. Tordoff, Monell Chemical Senses Center, 3500 Market St, Philadelphia, PA 19103-3308, USA. e-mail: tordoff@monell.org

Abstract

The long-term two-bottle choice test is commonly used as a simple screen to examine the acceptance of taste solutions by rodents. As part of an investigation of factors influencing the sensitivity of the two-bottle choice test, we determined the extent to which test duration influenced test sensitivity. C57BL6/J and 129X1/SvJ mice received four series of eight two-bottle tests, with each test lasting 1, 2, 4 or 6 days. Each series involved sequential tests with water, 2 mM saccharin, 5 and 50 mM citric acid, 30 and 300 μM quinine hydrochloride, 75 mM NaCl and 10% ethanol. There were significant differences between the strains in intake of saccharin, 5 and 50 mM citric acid, NaCl and ethanol in 4 and 6 day tests, but only saccharin and ethanol in 2 day tests, and 5 mM citric acid and ethanol in 1 day tests. To compare the sensitivity of the tests, we developed an analytical approach based on the comparison of deviations of individual 129X1/SvJ mice from the C57BL6/J strain mean. Our results suggest that to discriminate between strains or treatments when using 'standard' laboratory conditions and methods, 1 day tests are generally inadequate and 2 day tests are useful only if large effects are anticipated. Tests lasting 4 or 6 days are more sensitive, but conducting 6 day tests provides little additional benefit and sometimes is detrimental relative to conducting 4 day tests.

Introduction

The long-term, two-bottle choice test has been used for many years to assess the acceptability of various 'taste' solutions to animals. We use the word 'taste' advisedly here because the results of such tests are influenced by motivational, physiological and cognitive factors in addition to chemosensory ones. According to the most common method, individually housed rats are given access to two drinking spouts, one of which provides a taste solution and the other water. Generally, intakes of both the taste solution and water are monitored for at least 48 h. There are at least three reasons why 48 h tests are popular. First, intakes of rats over a 48 h period can be measured accurately, with little concern about spillage or evaporation as confounding factors. Secondly, the range of intakes typically observed in 48 h tests is sufficiently large to see differences between different groups of rats, treatments, or taste solutions. Thirdly, 48 h tests allow controls for the confounding effects of side preferences: some animals prefer one spout over the other [i.e. are left- or right- 'handed' (Biddle and Eales, 1996; Tordoff and Bachmanov, 2001a; Bachmanov *et al.*, 2002)] and so to counter this, the position of the drinking spouts is switched every 24 h. It is impractical to switch the spouts more often because this would fail to control for different intakes during different periods of the circadian cycle.

The recent impetus to understand the genetic basis of taste perception has led to the need to evaluate taste solution preferences of very large numbers of mice. For example, quantitative trait locus (QTL) analysis requires breeding and testing several hundred segregating hybrid (F_2) mice (Bachmanov *et al.*, 1997, 2001). Screening taste deficits induced by mutagenesis is expected to involve testing thousands of mice (The Jackson Laboratory, 2001). These studies face several problems in addition to the logistics of testing large numbers. One is that mice are used, and mice consume only a few milliliters of fluid per day, which results in only a small range of intakes and increases the relative contribution of spillage and evaporation to the results. Another is that the results of each mouse must stand alone. Unlike most earlier research, where inferences can be made based on the behavior of groups of animals, assigning the wrong phenotype to a mouse can lead to inaccurate QTL identification or wasted effort in breeding, genotyping and testing an uninteresting mouse's offspring. Thus, there is added importance to using very robust and reproducible tests. A third problem for investigators interested solely in the study of chemosensation is that interpretation of the results of long-term two-bottle choice tests is complicated by motivational, physiological and cognitive factors. However, the wide range of causes is an advantage if the purpose of the test is to screen the mice for any unusual behavior, which is often the case in genetic studies.

As part of an effort to increase the efficiency of the two-

bottle choice test for screening the offspring of mice with induced mutations (Tordoff and Bachmanov, 2001b), we attempted to examine the extent to which test duration influenced the sensitivity (i.e. statistical power) of two-bottle choice tests. The ultimate goal of this work is to provide efficient methods that can be used to identify mutant mice with outlying taste phenotypes relative to the normal, non-mutated population means. Under these circumstances, the identification of such outliers depends upon both the difference of the individual from the population mean and the variance of the population. Thus, tests that provide the greatest distinction between an outlier and its comparison group mean are considered to be the most powerful.

At present, no mice with induced mutations influencing taste have been identified. Consequently, as a model of 'outliers in relation to a population', we compared the results of two strains of mice, the C57BL/6J (B6) and 129X1/SvJ (129) strains, and for statistical purposes (see below) we treated the 129 strain as if they were outliers of the B6 strain. We chose to use these strains because our previous research has shown that they differ in preference for several taste solutions (Bachmanov *et al.*, 1996a,b, 1997, 2001). They are also used frequently in genetic studies, will be the first murine genomes to be sequenced and are likely candidates for mutagenesis experiments. The mice were given various taste solutions to consume during tests ranging in duration from 1 to 6 days. The solutions were representative sweet (2 mM saccharin), sour (5 and 50 mM citric acid), bitter (30 and 300 μM quinine hydrochloride; QHCl) and salty (75 mM NaCl) compounds, and 10% ethanol. They were chosen because they covered a wide range of acceptance, from avidly preferred (saccharin) to strongly disliked (QHCl). Moreover, according to previous research, there were very large differences between the B6 and 129 strains in saccharin and ethanol preference, moderate differences in citric acid and NaCl preference, and little if any difference in QHCl preference (Bachmanov *et al.*, 1996a,b, 1998).

Methods

Subjects were 16 male B6 and 16 male 129 mice, purchased from The Jackson Laboratory (Bar Harbor, ME). The mice were 12 weeks old at the start of testing. They were individually housed in plastic 'tub' cages (26.5 cm × 17 cm × 12 cm), with a stainless steel grid lid and wood shavings scattered on the floor. The vivarium was maintained at 23°C on a 12:12 h light/dark cycle, with lights off at 7 p.m. The mice were fed pelleted Teklad 8604 chow (Harlan, Madison, WI) and had deionized water to drink.

During the experiment, the mice had continuous access for 104 days to two drinking tubes, one containing deionized water and the other containing a taste solution (which was sometimes deionized water also). The drinking tubes were placed to the (mouse's) right of the food hopper. Their tips were 15 mm apart and extended into the cage 25 mm [described in (Tordoff and Bachmanov, 2001a; Bachmanov *et al.*, 2002)]. Each spout had a 3.175 mm diameter hole from which the mice could lick fluids. Specifics of construction of the drinking tubes, details of the measurement procedures, and details of housing and maintenance conditions are available on the Monell Mouse Taste Phenotyping Project website (Tordoff and Bachmanov, 2001a).

Each of the 32 mice received four series of eight tests (a total of 32 tests). Each series was always conducted in the same order. The mouse first received a choice between two identical tubes of deionized water. It then received a choice between deionized water and the following compounds: 2 mM sodium saccharin, 5 mM citric acid, 50 mM citric acid, 0.03 mM QHCl, 0.3 mM QHCl, 75 mM NaCl and 10% ethanol. All compounds were obtained from Sigma Chemical Corp. (St Louis, MO), except for the ethanol, which was obtained from Pharmco Products, Inc. (Brookfield, CT). They were dissolved in deionized water and stored in Nalgene plastic bottles until required.

The four series differed in the duration of each test, which was 1, 2, 4 or 6 days. The order of the four series was counterbalanced using a Latin Square design, such that cohorts of four mice from each strain received the tests in each possible order. Once mice finished one series of tests they immediately began the next series. For tests lasting 2, 4 or 6 days, the taste solution was initially presented on the mouse's left but the positions of the two drinking tubes were switched every 24 h. For the 1 day tests, half the mice of each strain always received the taste solution on the right and the other half always received it on the left. Fluid intakes were measured to the nearest 0.1 ml every day in the middle of the light period. Body weights were measured at the beginning of each test series and at the end of the experiment.

Analyses and results

Unlike most studies, where the purpose is to determine whether differences exist between groups or treatments, here the purpose was to determine whether the tests differed in their statistical power to distinguish between the groups. This cannot be achieved using standard analytical methods. Below, we describe the methods used to collate the data, and then three approaches to analysis, along with their strengths and limitations.

Initial data collation and analysis

Fluid intakes from the two drinking tubes available each day for each mouse were collated according to the taste solution and test duration (8 taste solutions, including water × 4 test durations = 32 tests). Occasional data from a mouse were lost due to spillage or other technical errors (65 of 6756 measurements). When this occurred, average intakes from the remaining days of the test were used as a proxy for the

missing value. This allowed us to avoid problems caused by within-subject unequal group sizes.

For each mouse on each day, solution preference ratios were calculated based on the formula:

$$\text{preference (\%)} = \text{taste solution intake}/(\text{taste solution intake} + \text{water intake}) \times 100$$

For conditions in which water was presented in both drinking tubes, the tube presented on the mouse's left was considered as a taste solution. There were no significant differences between the strains in body weight. The B6 mice weighed 27 ± 0.5 g at the start and 31 ± 0.7 g at the end of the experiment. Corresponding weights for the 129 mice were 26 ± 0.5 g and 31 ± 0.6 g. This simplified subsequent analyses because there was no need to consider the effect of body weight on solution intake [see (Bachmanov *et al.*, 1998, 2002) for discussion].

Subsequent analyses were conducted in parallel, using taste solution intakes, water intakes and taste solution preferences as dependent variables. However, the analyses of intakes revealed little information of interest that was not provided by the analyses of preferences, so for brevity, only preferences are presented here. All hypothesis testing used a criterion of $P < 0.05$ for statistical significance, but exact probabilities are given below so that readers can use other criteria.

'Traditional' approach using analyses of variance (ANOVAs) to assess differences between strains and differences among tests of various durations

Thirty-two means were obtained from each mouse, one for each of the 32 tests, by averaging together its daily preference scores. The strain means calculated from these individual mouse means were compared using ANOVAs with factors of Strain (B6 or 129; between-subjects) and Test Duration (1, 2, 4 or 6 days; within-subjects). Separate analyses were conducted for each taste solution. When the effects of an ANOVA were significant, post-hoc *t*-tests were conducted to distinguish between individual pairs of means

Preferences for saccharin, 5 mM citric acid, NaCl and ethanol were significantly higher in B6 than 129 mice (Table 1). For saccharin, NaCl and ethanol, test duration significantly influenced mean preferences. Preferences in the 1 day test were significantly lower than those in the 4 day test (saccharin, ethanol) or 6 day test (saccharin, NaCl). However, there was no difference between the 2, 4 and 6 day tests in average preference scores for any taste solution. For saccharin preference, there was an interaction between strain and test duration, due to significantly lower preferences of B6 but not 129 mice during the 1 day test than during the 2, 4 and 6 day tests (Figure 2).

This 'traditional' approach to analysis has the advantage of using all the data collected to screen for strain differences or differences related to test duration. However, because the ANOVAs involve pooled error terms, they do not reflect differences in variance between individual pairs of tests. This is a critical issue if the statistical power of each test is to be evaluated. Also, because the ANOVAs are based on test means, they do not allow examination of differences related to particular days of a test.

Table 1 Results of ANOVA comparing mean strain values, test durations and strain × duration interactions for each test solution

Test	Effect (df)		
	Strain (1,30)	Test duration (3,90)	Strain × test duration (3,90)
Water	3.54	0.48	0.77
2 mM saccharin	49.35****	5.59**	3.64*
5 mM citric acid	14.63***	0.75	0.65
50 mM citric acid	3.19	1.47	0.13
30 μM QHCl	0.28	1.07	1.18
300 μM QHCl	0.37	0.71	1.00
75 mM NaCl	14.49***	3.66*	1.55
10% ethanol	88.92****	4.87**	2.16
Average	69.07****	1.78	1.17

The numbers in the body of the table are F values from 'traditional' strain × test duration ANOVAs. *$P < 0.05$, **$P < 0.005$, ***$P < 0.001$, ****$P < 0.00001$.

Comparison of probability values using independent analyses of each test

A metric is needed in order to compare the statistical power of each test. If we had a single population, the metric would be a measure of sample variance, and the most powerful test would be the one that produced the least variable results (i.e. the smallest standard error). However, we have two populations (B6 and 129 mice), so the most powerful test is the one that produces the largest difference between the groups. The metric must account for both the difference between the group means and their variability. A test that produces large differences between the group means but much variability may not be as powerful as a test that produces small differences between the group means and little variability. The agreed upon metric for assessing the relative effect of mean difference and variance is *P*, the probability value assessed by a *t*-test or ANOVA. The smaller the *P* value, the greater the difference between the groups. Note that *P* is being used here not as a criterion for hypothesis testing (as in the 'traditional' approach), but as a measure of the statistical power of the test to distinguish between B6 and 129 mice.

To calculate test *P* values, the results of the 32 tests were treated as completely independent sets of data. The results of each 1 day test were analyzed by one-way ANOVAs with Strain as the only factor. The results of each 2, 4 and 6 day

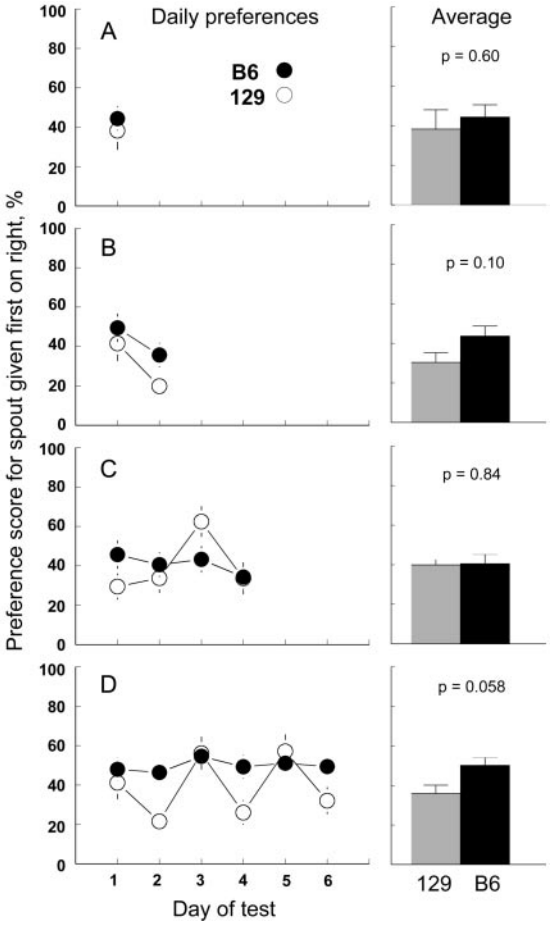

Figure 1 Daily and average preferences of B6 and 129 mice for water in the drinking spout first presented on the mouse's right (2, 4 and 6 day tests) or for half the mice of each strain on the right and the other half on the left (1 day tests). The alternating high and low preferences of the 129 strain are probably because of its strong side preference. **(A)** 1 day test; **(B)** 2 day test; **(C)** 4 day test; **(D)** 6 day test.

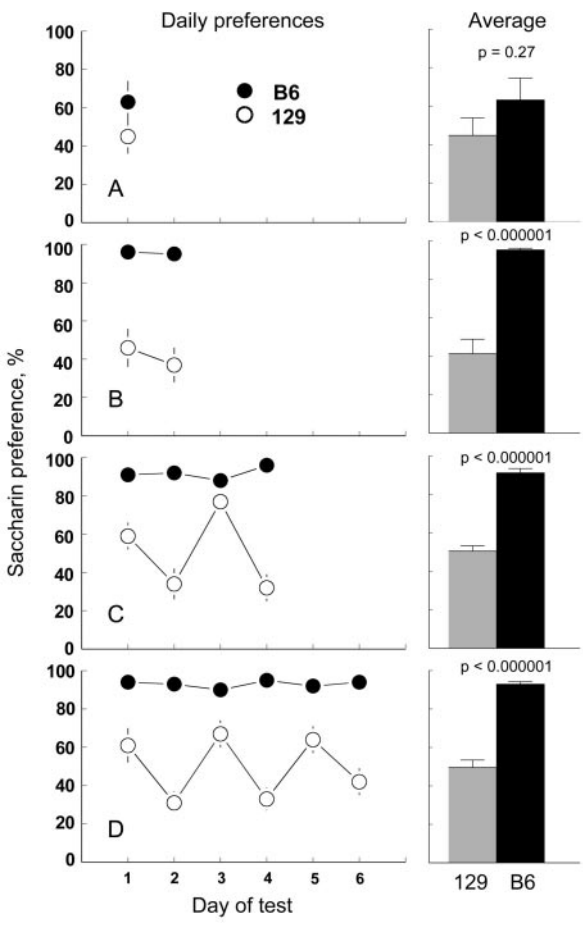

Figure 2 Daily and average preference ratios of B6 and 129 mice for 2 mM saccharin. **(A)** 1 day test; **(B)** 2 day test; **(C)** 4 day test; **(D)** 6 day test.

test were analyzed by two-way ANOVAs with factors of Strain and Day. Additional analyses were performed in which a third factor, Trial Pair, was introduced, to account for variance associated with the daily switching of the side the taste solution was presented.

The left-hand panels of Figures 1–8 show mean daily solution preferences of each strain, and the results of each ANOVA are presented in Table 2. Probabilities of differences between strains are presented over the bars in Figures 1–8 and are directly compared in Figure 9.

There were very few changes in preference as tests progressed. The effect of Day was significant in analyses involving preferences for water, 0.03 mM QHCl and NaCl over 2 days, water and saccharin over 4 days, and water, saccharin and ethanol over 6 days. However, these effects did not appear to be due to a consistent event, such as preferences gradually increasing or decreasing during a test. One exception was the daily 'switching' behavior shown by 129 mice given water or saccharin to drink, which led to Strain × Day interactions in 4 and 6 day water and saccharin preference. There were also Strain × Day interactions in the 6 day tests of 30 μM QHCl and ethanol preference. With 30 μM QHCl, this was due to diverging strain preferences on the last day of the test. With ethanol, this was due to a reduction in preference of the B6 but not 129 mice on the last 2 days of the test.

Analyzing the data as 2 day blocks rather than individual days removed the influence of the daily switching of tube positions, and the interaction observed on the last day of the 6 day test with 30 μM QHCl, but otherwise had little influence on the results.

These analyses had several advantages. First, the error term for each analysis was unaffected by the results of other taste tests. Secondly, comparisons between individual days or pairs of days could be made. Thirdly, the size of the probability value for each test's main effect of strain was a measure of the power of the test to distinguish between B6 and 129 mice.

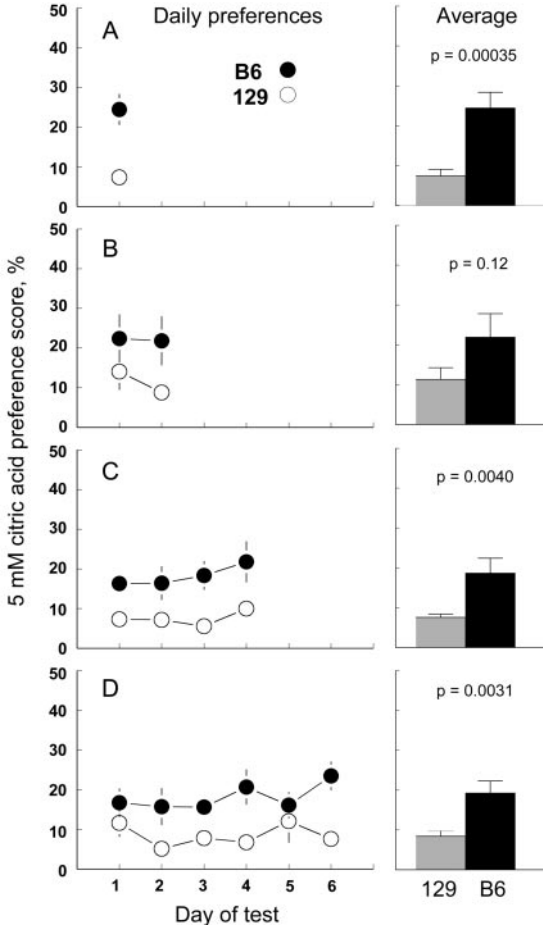

Figure 3 Daily and average preferences of B6 and 129 mice for 5 mM citric acid. **(A)** 1 day test; **(B)** 2 day test; **(C)** 4 day test; **(D)** 6 day test.

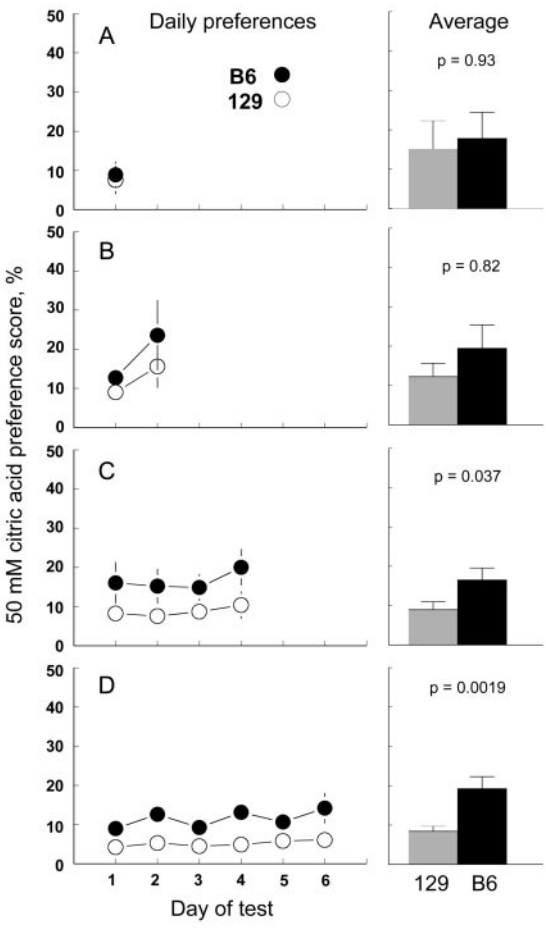

Figure 4 Daily and average preferences of B6 and 129 mice for 50 mM citric acid. **(A)** 1 day test; **(B)** 2 day test; **(C)** 4 day test; **(D)** 6 day test.

Analysis as if the 129 mice were outliers of the B6 strain

With the previous approach, the probability coefficient (P) provided a measure of the ability of the various tests to discriminate between B6 and 129 mice but did not evaluate whether P values differed significantly from one another. We could not determine, for example, whether the 4 day test with NaCl ($P = 0.0036$) was significantly better at discriminating B6 from 129 mice than was the 2 day test ($P = 0.037$). To provide measures that allowed the determination of significant differences between tests, we took the following approach, which we believe is novel.

The average preferences of each 129 mouse were normalized (converted to a z-score) based on the B6 strain mean and average of the B6 and 129 within-strain standard deviations. The z-scores determined for each of the four test durations could then be compared by ANOVA, such that differences between mean z-scores represented differences in the ability of the tests to discriminate the 129 from B6 strains.

An advantage of this approach is that it follows the methods for which the procedures are ultimately intended. Outliers are usually identified based on their deviation from the population mean by more than a preset number of standard deviations (typically 3). In effect, the approach used here was to consider the 129 mice as 'outliers' of the B6 group, calculate for each test how many standard deviations each 129 mouse differed from the B6 mean and then compare the values obtained from each test.

Table 3 presents means and standard errors of the 129 mice normalized to the B6 mean for each of the 32 tests conducted. There were significant differences related to test duration (i.e., 1, 2, 4 or 6 day tests) in z-scores obtained from the tests of saccharin, 5 mM citric acid, 50 mM citric acid, NaCl and ethanol (see Table 3 for F-values and differences between individual means). No differences related to test duration were found in analyses of water, 30 μM QHCl and 300 μM QHCl.

In a final analysis, z-scores from all the tests were compared using a two-way ANOVA with factors of Test Duration and Taste Solution. There was a highly significant interaction between these factors, $F(21,315) = 10.4$, $P < 0.000001$, and highly significant differences among the taste solutions, $F(7,105) = 44.1$, $P < 0.000001$. Many significant differences between pairs of means contributed to the sig-

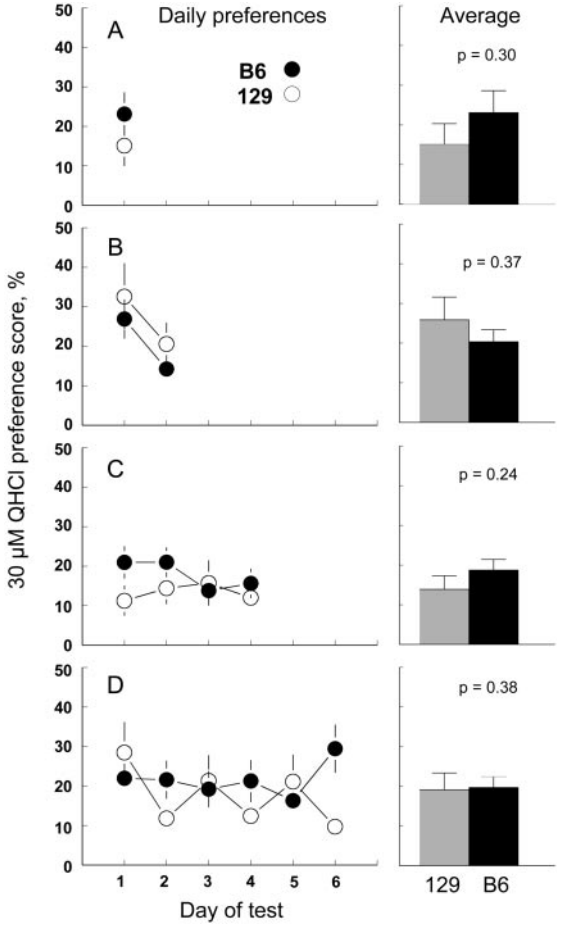

Figure 5 Daily and average preferences of B6 and 129 mice for 30 µM QHCl. **(A)** 1 day test; **(B)** 2 day test; **(C)** 4 day test; **(D)** 6 day test.

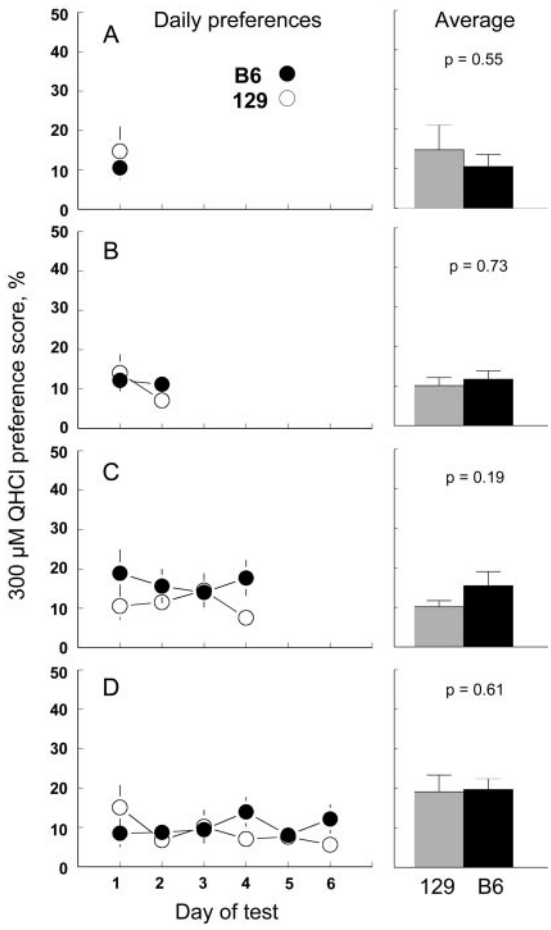

Figure 6 Daily and average preferences of B6 and 129 mice for 300 µM QHCl. **(A)** 1 day test; **(B)** 2 day test; **(C)** 4 day test; **(D)** 6 day test.

nificant interaction (Table 3). In general, the differences between the B6 and 129 strains were significantly greater in tests involving saccharin and ethanol than in tests involving 5 mM citric acid and NaCl. They were significantly greater in tests involving 5 mM citric acid and NaCl than in tests involving 30 or 300 µM QHCl. The differences between B6 and 129 strains observed in tests involving water and 50 mM citric acid fell in between those observed in tests involving 5 mM citric acid or NaCl and QHCl.

Discussion

The results confirm and extend earlier findings that B6 mice have stronger preferences for saccharin, ethanol and citric acid than do 129 mice (Bachmanov *et al.*, 1996a,b, 1998). The B6 mice also had stronger preferences for NaCl than did 129 mice. This appears to contradict the results of our earlier work (Bachmanov *et al.*, 1996a, 1998), but it is due to a substrain difference. Whereas our earlier studies used C57BL/6ByJ and 129P3/J strains, here we used C57BL/6J and 129X1/SvJ strains, which are more commonly used in genetic research. A direct comparison found that 75 mM NaCl preference is higher in NaCl-experienced C57BL/6J than 129X1/SvJ mice but lower in NaCl-experienced C57BL/6ByJ than 129P3/J mice (unpublished results).

Using the conventional level of $P < 0.05$ as the criterion for significance, the 4 and 6 day tests distinguished between the B6 and 129 strains' preferences for saccharin, 5 mM citric acid, 50 mM citric acid, NaCl and ethanol; the 2 day tests distinguished preferences for saccharin, NaCl and ethanol only, and the 1 day tests distinguished preferences for 5 mM citric acid and ethanol only. The large difference between B6 and 129 mice in 5 mM citric acid preference found in the 1 day test was probably spurious given its lack of consistency with the small differences in citric acid preference observed on the first day of longer tests. Moreover, preferences of both strains for all solutions were lower in 1 day tests than in longer tests. We conclude that the 1 day tests have little power to discriminate between B6 and 129 mice.

There appeared to be a substantial advantage conferred by testing mice for 4 or 6 days rather than 2 days. This is most clearly illustrated in Figure 9. Using conventional levels of significance, B6 and 129 mice differed in 2 day tests of only two compounds but in 4 and 6 day tests of five

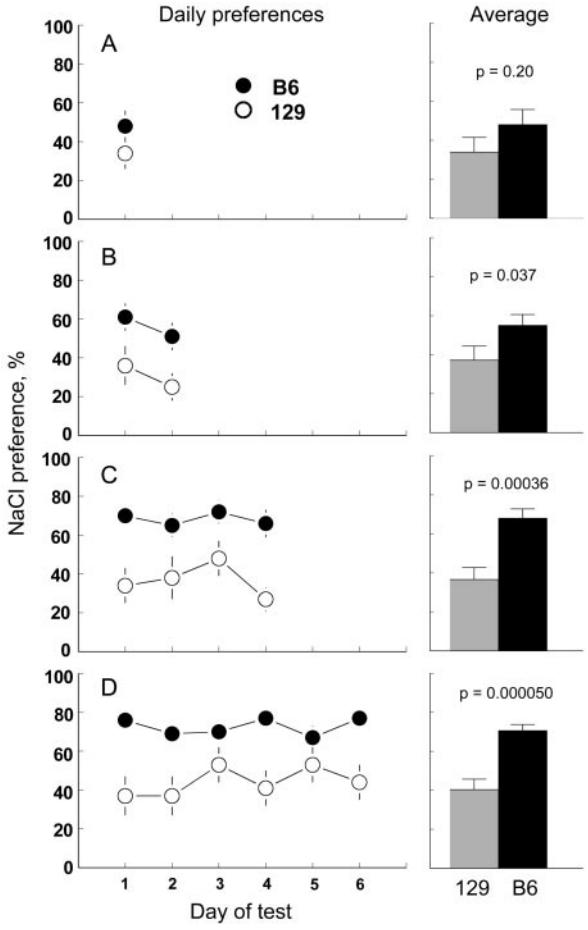

Figure 7 Daily and average preference ratios of B6 and 129 mice for 75 mM NaCl. **(A)** 1 day test; **(B)** 2 day test; **(C)** 4 day test; **(D)** 6 day test.

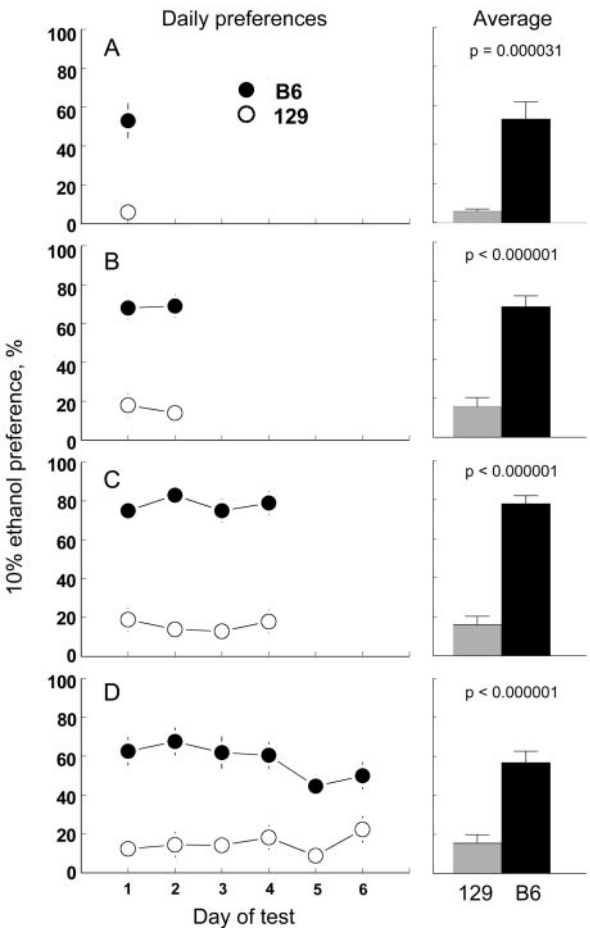

Figure 8 Daily and average preference ratios of B6 and 129 mice for 10% ethanol. **(A)** 1 day test; **(B)** 2 day test; **(C)** 4 day test; **(D)** 6 day test.

compounds. Analysis of *z*-scores revealed that overall, 4 day tests had significantly more statistical power to resolve between the B6 and 129 strains than did the 2 day tests. However, 4 day tests also had significantly more sensitivity than did 6 day tests overall, and 2 and 6 day tests were of equivalent sensitivity (Table 3, bottom row). This was because the B6–129 difference was significantly less pronounced during 6 day tests of saccharin and ethanol than during 4 day tests. For the tests involving citric acid and NaCl, there was a tendency for 6 day tests to produce higher levels of significance than 4 day tests (Figure 9), and this was significant according to a *z*-score analysis of 50 mM citric acid. Whereas there were extremely large differences between B6 and 129 mice in saccharin and ethanol intake, the differences for citric acid and NaCl were only moderate. We conclude from this that the greatest resolution is produced by 4 day tests when large differences exist between the strains, but 6 day tests may confer a slight additional advantage when only subtle strain differences are present.

The greater sensitivity of 4 or 6 day compared with 1 or 2 day tests appears to be due simply to the increased length of the tests. There was no evidence that preferences progressively diminished or increased as the mice gained experience with a particular taste solution (with the possible exception of the reduction in preference shown by B6 mice on the last 2 days of their 6 day test with ethanol). We expected that longer tests would produce less within-strain variability because any momentary aberration of intake by an individual mouse would be 'averaged out' over time. However, this was not the case. Long and short tests had similar within-strain variances. Instead, the greater effectiveness of longer tests appeared to be due almost entirely to greater separation of the strain means (note, for example, the standard errors on the bars of Figures 1–8, the standard errors showing the within-strain variance of 129 mice in Table 3 and the strain distributions shown in Figure 10). We have no ready explanation for this. The possibility that prior experience (e.g. carry-over effects) obscures differences between the strain means appears unlikely because in most cases, preferences on the first pair of days was very similar to preferences on later pairs of days in the 4 and 6 day tests. Alternatively, perhaps frequent (i.e. every 2 days or less) but not infrequent switching of the taste solution confuses the mice in a manner that influences their

Table 2 *F*-values for each ANOVA comparing daily preferences of each strain for each test duration

Test	Test duration			
	1 day	2 days	4 days	6 days
Water				
Strain	0.28	2.85	0.04	3.87
Day	–	7.13*	3.05	5.17
Strain × day	–	0.35	2.38	2.88
2 mM saccharin				
Strain	1.25	50.38***	160.57***	114.10***
Day	–	0.62	6.31*	5.17*
Strain × day	–	1.03	10.78**	6.54*
5 mM citric acid				
Strain	16.22**	2.62	9.70**	10.32**
Day	–	1.03	2.04	5.14*
Strain × day	–	0.67	0.43	1.38
50 mM citric acid				
Strain	0.01	0.05	4.74*	11.61**
Day	–	3.66	0.57	0.84
Strain × day	–	0.22	0.10	0.80
30 μM QHCl				
Strain	1.12	0.82	1.45	0.78
Day	–	7.32*	0.56	1.95
Strain × day	–	0.00	1.22	3.48
300 μM QHCl				
Strain	0.37	0.12	1.82	0.26
Day	–	2.03	0.18	0.87
Strain × day	–	1.14	1.18	2.16
75 mM NaCl				
Strain	1.71	4.08	16.13**	14.71**
Day	–	6.17*	1.58	0.98
Strain × day	–	1.90	0.71	2.18
10% ethanol				
Strain	24.01**	53.61***	108.17***	37.25***
Day	–	0.38	0.66	3.10
Strain × day	–	0.07	1.15	2.82

The numbers in the body of the table are *F*-values from ANOVAs of each if the 32 tests conducted. Degrees of freedom: all strain comparisons = (1,30); effect of day and interactions, 2 day = (1,30); 4 day = (3,90); 6 day = (5,150). *$P < 0.05$, **$P < 0.01$, ***$P < 0.00001$.

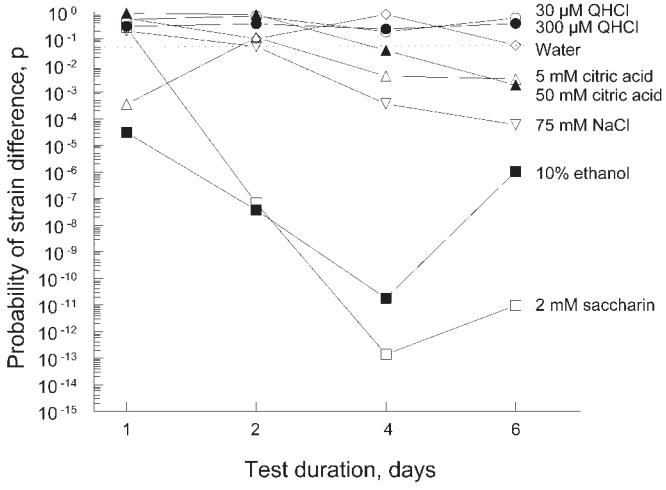

Figure 9 Probabilities associated with the difference between B6 and 129 mice for each of 32 tests. Dotted line shows $P < 0.05$. Note that the smaller the *P* value, the greater the power of the statistical test to discriminate B6 from 129 mice.

solution preference. With long tests, they may learn to respond more precisely to a taste solution even though the perception of the solution is unchanged (i.e. they may learn to perform appropriately). Whatever the reason, the results show that it can be misguided to evaluate a test based solely on variability of response.

One result worthy of discussion was the 'switching' shown by 129 mice given water or saccharin to drink. The 129 mice have the strongest predisposition to drink from the left hand spout when given two spouts to chose from, among 28

Table 3 *z*-score means and standard errors for each test duration of 129 mice relative to the B6 strain means

Test	*F*(3,45)	Test duration			
		1 day	2 days	4 days	6 days
Water	2.49	0.19 ± 0.31	0.67 ± 0.25*	0.04 ± 0.18	0.88 ± 0.26*
2 mM saccharin	27.66	0.44 ± 0.23[a]	3.31 ± 0.46[b]*	4.06 ± 0.28[b]*	1.07 ± 0.32[a]*
5 mM citric acid	8.19	1.54 ± 0.15[a]*	0.60 ± 0.17[b]*	1.23 ± 0.09[a]*	1.24 ± 0.15[a]*
50 mM citric acid	7.19	0.10 ± 0.26[a]	0.39 ± 0.18[a]*	0.75 ± 0.20[b]*	1.35 ± 0.08[c]*
30 μM QHCl	1.71	0.37 ± 0.24	–0.32 ± 0.33	0.40 ± 0.28	0.04 ± 0.31
300 μM QHCl	2.41	–0.23 ± 0.34	0.20 ± 0.25	0.52 ± 0.15*	0.52 ± 0.21*
75 mM NaCl	7.17	0.46 ± 0.25[a]	0.74 ± 0.29[a]*	1.46 ± 0.29[b]*	1.67 ± 0.32[b]*
10% ethanol	17.63	1.86 ± 0.15[a]*	2.62 ± 0.23[b]*	3.60 ± 0.25[c]*	2.11 ± 0.21[a]*
Average	13.06	0.59 ± 0.14[a]*	1.03 ± 0.11[b]*	1.51 ± 0.10[c]*	1.11 ± 0.12[b]*

The values in the right four columns of the table show the number of standard deviations the 129 mice differed from the B6 mean (i.e. *z*-scores). Values in the column labeled '*F*(3,45)' are the results of ANOVAs based on the *z*-scores of each mouse (based on the B6 strain mean).

*$P < 0.05$, 129 strain different from B6 strain, according to paired *t*-test of *z*-scores. For each taste solution, means with different subscripts differ significantly from each other, according to separate ANOVAs for each taste solution. According to an omnibus ANOVA of all data, differences between means of >0.62 are significant at the $P < 0.05$ level (post-hoc *t*-test).

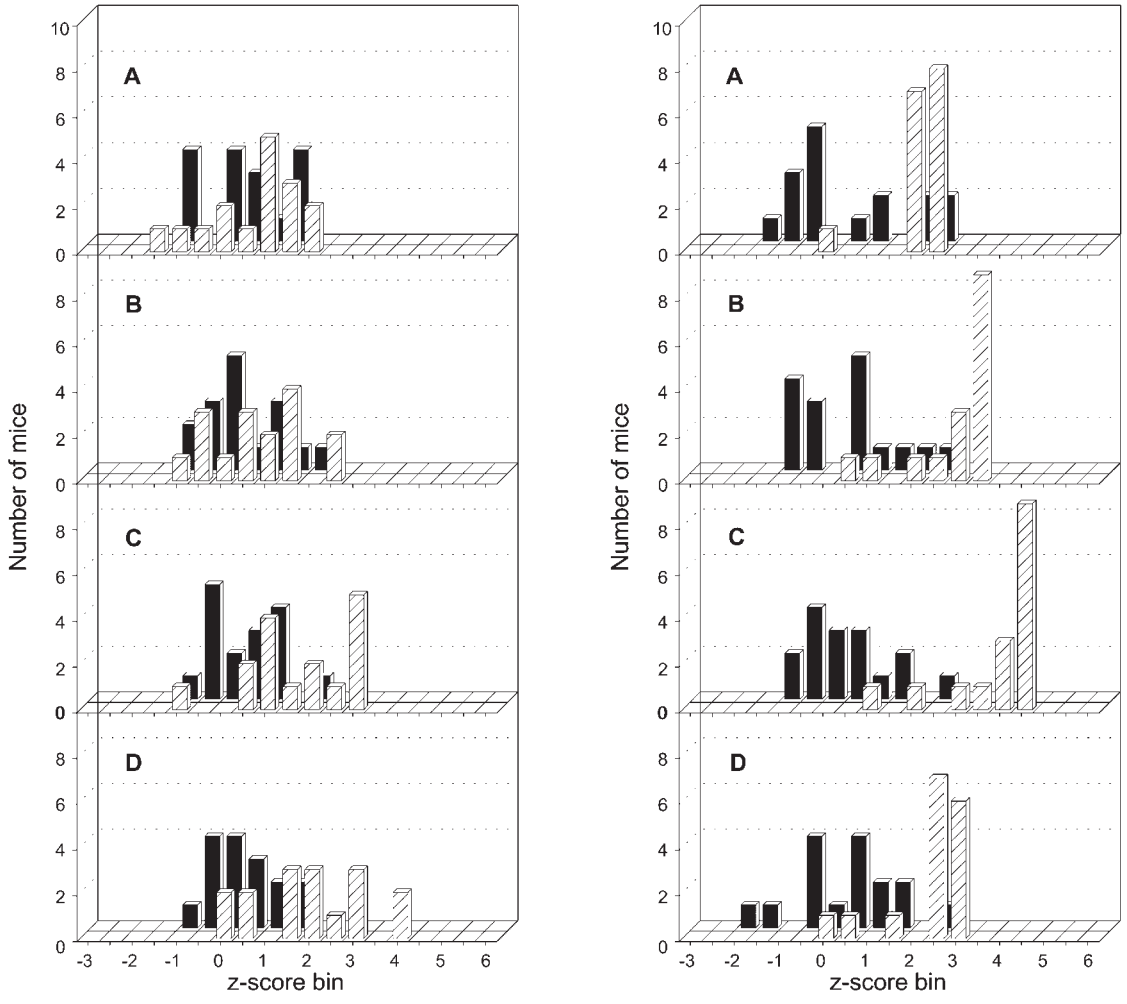

Figure 10 Distribution of preferences for NaCl (left) or ethanol (right) by 16 B6 (black) and 16 129 (shaded) mice. **(A)** 1 day test; **(B)** 2 day test; **(C)** 4 day test; **(D)** 6 day test. Note that the larger differences between strains observed with longer-duration tests are due to differences in strain means and not reduced within-strain variation.

strains tested (Tordoff and Bachmanov, 2001a; Bachmanov *et al.*, 2002). Thus the 'switching' can be explained by the mice drinking predominantly from the left hand tube irrespective of the taste solution. This effect occurred only when the solution was neutral but not when there was a strong taste response. The implication is that the 129 mice either cannot taste 2 mM saccharin solution or find it insufficiently pleasant to override their side preference.

We conclude that, at least for this study, 4 or 6 day two-bottle choice tests provide significantly more sensitivity than do 1 or 2 day tests. However, it would be wrong to generalize this conclusion to all other test situations. The 2 day tests failed only for subtle comparisons. They competently distinguished between the saccharin and ethanol preferences of B6 and 129 mice (Ps < 0.0000001). Also, in many studies, several concentrations of a taste compound are tested in ascending series. Under these circumstances, several short but similar tests of different concentrations may contribute statistical power to reveal subtle strain differences as a main effect of combined analyses.

As well as test duration, several factors, including the magnitude of differences expected and the variability of the animals to be tested, must be taken into account when estimating the expected statistical power of two-bottle choice tests. The best test duration to use for a given experiment is ultimately a compromise between the requirement to produce sensitive and reliable results on the one hand and the cost and effort involved on the other. The conclusions made here do not provide a recipe for every experiment involving preferences. But, all things being equal, under the 'standard' conditions used in our laboratory and many others, we suggest that the extra sensitivity provided by 4 or 6 day tests may offset the extra time and effort involved in conducting them. Perhaps more importantly, the present paper describes an approach to optimize tests depending on a particular stimulus or research goal.

Acknowledgements

This work was supported by National Institutes of Health grant AA-12715. We thank Diane Pilchak, Anthony Acquaviva and Julie Williams for technical support.

References

Bachmanov, A.A., Reed, D.R., Tordoff, M.G., Price, R.A. and **Beauchamp, G.K.** (1996a) *Intake of ethanol, sodium chloride, sucrose, citric acid, and quinine hydrochloride solutions by mice: a genetic analysis*. Behav. Genet., 26, 563–573.

Bachmanov, A.A., Tordoff, M.G. and **Beauchamp, G.K.** (1996b) *Ethanol consumption and taste preferences in C57BL/6ByJ and 129/J mice*. Alcohol Clin. Exp. Res., 20, 201–206.

Bachmanov, A.A., Reed, D.R., Ninomiya, Y., Inoue, M., Tordoff, M.G., Price, R.A. and **Beauchamp, G.K.** (1997) *Sucrose consumption in mice: major influence of two genetic loci affecting peripheral sensory responses*. Mammal. Genome, 8, 545–548.

Bachmanov, A.A., Tordoff, M.G. and **Beauchamp, G.K.** (1998) *Voluntary sodium chloride consumption by mice: differences among five inbred strains*. Behav. Genet., 28, 117–124.

Bachmanov, A.A., Li, X., Reed, D.R., Ohmen, J.D., Li, S., Chen, Z., Tordoff, M.G., de Jong, P.J., Wu, C., West, D.B., Chatterjee, A., Ross, D.A. and **Beauchamp, G.K.** (2001) *Positional cloning of the mouse saccharin preference (Sac) locus*. Chem. Senses, 26, 925–933.

Bachmanov, A.A., Reed, D.R., Beauchamp, G.K. and **Tordoff, M.G.** (2002) *Body weight, food intake, water intake, and drinking spout side preference of 28 mouse strains*. Behav. Gen., 32, 435–443.

Biddle, F.G. and **Eales, B.A.** (1996) *The degree of lateralization of paw usage (handedness) in the mouse is defined by three major phenotypes*. Behav. Genet., 26, 391–406.

The Jackson Laboratory (2001) *Neuroscience mutagenesis facility*. http://www.jax.org/nmf/documents/overview.html

Tordoff, M.G. and **Bachmanov, A.A.** (2001a) *Food intakes, water intakes, and spout side preferences*. http://aretha.jax.org/pub-cgi/phenome/mpdcgi?rtn=studies/details&id=63

Tordoff, M.G. and **Bachmanov, A.A.** (2001b) *Monell mouse taste phenotyping project*. http://www.monell.org/MMTPP

Accepted December 3, 2001