Reliability of Threshold and Suprathreshold Methods for Taste Phenotyping: Characterization with PROP and Sodium Chloride

**Veronica Galindo-Cuspinera · Thierry Waeber ·
Nicolas Antille · Christoph Hartmann · Nicola Stead ·
Nathalie Martin**

Abstract The present study aimed to compare the accuracy and reliability of four standard methods used for classification of people as taster or non-tasters based on their sensitivity to PROP (6-*n*-propylthiouracil). A panel consisting of 21 subjects was tested for threshold and suprathreshold sensitivity of sodium chloride, PROP, and genotyped for *TAS2R38*. Two threshold methods, staircase and modified Harris–Kalmus, were used to obtain detection and recognition thresholds and compared for accuracy and repeatability. Similarly, two suprathreshold techniques, the just noticeable differences (JND) and the general labeled magnitude scale (gLMS), were used to determine Weber fractions and individual psychophysical functions and compared for accuracy and repeatability. Results show both threshold methods have been able to correctly separate people into two groups of tasters and non-tasters, with the staircase method having a lower variability among subjects. On the suprathreshold front, we found differences in sensitivity between tasters and non-tasters when comparing Weber fractions and psychophysical functions; however, our data suggest that clustering people without previous knowledge of their taster status is less accurate when using Weber fractions. Intensity ratings are more reliable to classify people into tasters and non-tasters. Results show that the staircase for threshold measurement and the gLMS methods are more reliable methods than Harris–Kalmus and JND for phenotyping people and can be used in large-scale studies in the quest to discover new genotype–phenotype associations.

V. Galindo-Cuspinera · T. Waeber · N. Antille · C. Hartmann ·
N. Stead · N. Martin
Food Consumer Interaction department, Nestlé Research Center,
Lausanne, Switzerland

Present Address:
V. Galindo-Cuspinera (✉)
NIZO food research,
Kernhemseweg 2,
6718 ZB Ede, The Netherlands
e-mail: veronica.galindo@fulbright.org

Introduction

Throughout the twentieth century different theories have been developed that could explain human responses to sensory stimuli. Threshold determination and psychophysical functions were established in order to determine levels of sensitivity to different sensory stimuli. However, advancement in the field is dependent on the availability of simple and reliable phenotyping methods, which can be applied in consumer and population studies to classify people based on their sensitivity to specific tastes. The present study aimed to compare the accuracy and reliability of four standard methods commonly used for phenotypic classification of people based on taste sensitivity or responsiveness. The methodologies tested were chosen based on their wide use in psychophysics studies and included two threshold methods: staircase methodology (SC) and modified Harris–Kalmus (m-HK), and two suprathreshold methods: just noticeable differences (JND)/ Weber fractions and general labeled magnitude scale (gLMS) intensity ratings.

For many years, people have been classified as non-tasters or tasters based on their ability to taste two compounds: phenylthiocarbamide (PTC) and the structurally related 6-*n*-propylthiouracil (PROP; Fox 1931; Blakeslee 1932;

Fischer and Griffin 1963). This characteristic has been proved to be a genetically controlled Mendelian recessive trait with heritability estimated at ~0.5 (Drayna et al. 2003). Perception of these compounds has been linked to a specific gene *TAS2R38* in chromosome 7 (Reed et al. 1999; Duffy et al. 2004), for which three specific genetic variations (single nucleotide polymorphisms or SNPs) make the final receptor unresponsive (Kim et al. 2003). Although the ability to taste PROP does not predict a person's ability to taste all bitters, good correlations have been found to other bitter compounds (Harris and Kalmus 1949a; Barnicot et al. 1951; Hall et al. 1975; Leach and Noble 1986; Sandell and Breslin 2006). However, correlations with PROP non-tasters cannot be extended to other qualities of taste (Drewnowski 1997, 1998; Ly and Drewnowski 2001), hence the need to look for people with different taste phenotypes that will help in elucidating other genes involved in taste perception. In the present study, we use PROP as a tool to compare accuracy of methods, given all the information available on this genetic trait and the strong sensory responses given by tasters and non-tasters to this compound.

Most of the studies performed in the area of psychogenomics are based on detection threshold (Harris and Kalmus 1949c; Barnicot et al. 1951; Glanville and Kaplan 1965; Hall et al. 1975; Drewnowski et al. 1997; Wise et al. 2007), single tasting using only one concentration of the substance (Lawless 1980; Drewnowski et al. 2001), and concentration-intensity ratings (Delwiche et al. 2001; Tepper et al. 2001; Bufe et al. 2005; Dinehart et al. 2006). Threshold evaluation had been used as the gold standard methodology for classification (Blakeslee and Salmon 1935; Salmon and Blakeslee 1935; Harris and Kalmus 1949b, c; Fischer et al. 1961; Jian and Ennis 1998) until it was apparent that three distributions could be fitted to PROP threshold data (Bartoshuk 1991; Bartoshuk et al. 1994). One distribution is clearly associated with non-tasters but the other two taster distributions overlap substantially, which makes thresholds not the best measurement for classification of homozygous or heterozygous individuals (Reed et al. 1995). Controversy has developed on whether thresholds have a direct correlation to, or can predict, suprathreshold sensitivity. While in some studies, detection thresholds show good correlations with PTC sensitivity at suprathreshold levels (Drewnowski 1997; Bufe et al. 2005), having a low detection threshold does not ensure the same low sensitivity throughout the entire psychophysical function (Bartoshuk 1978; Pangborn and Pecore 1982; Bartoshuk 2000; Mojet et al. 2005). Relying on detection thresholds per se may cause misclassification of tasters (Bartoshuk 1978; Bartoshuk 2000; Keast and Roper 2007) and is considered by some scientists an unsatisfying way to study sensory experience because they refer to the dimmest sensations, outside of what people

experience in reality (Bartoshuk 2000). For these reasons, suprathreshold intensity ratings are often used in conjunction with thresholds (Drewnowski 1997; Bufe et al. 2005).

Other parameters commonly used to determine suprathreshold sensitivity in humans are the differential threshold and intensity ratings. The differential threshold or just noticeable differences was defined as the minimum amount by which stimulus intensity must be changed in order to produce a noticeable variation in sensory experience. Ernst Weber discovered that the differential threshold between two stimuli was not an absolute amount, but an amount in relation to the intensity of the first stimulus (Weber 1834). A parameter was defined from this theory known as Weber fraction, which is the ratio of JNDs to standard concentrations (Fechner 1987). This fraction is presumed to remain constant for a compound throughout the dynamic range; however, exceptions are seen at saturation levels and concentrations close to threshold. A systematic study on gustatory sensitivity was published in 1957, which shows average Weber fractions for four basic tastes (Schutz and Pilgrim 1957). Weber fractions have also been used to determine differences in sensitivity associated with aging and to evaluate sensitivity to different foods by PROP tasters or non-tasters (Prescott et al. 2004), but to our knowledge, no study has ever compared differences on PROP Weber fractions between tasters and non-tasters, which could make this method useful for phenotypic classification.

Another well-established methodology used to phenotype people consists of tasting five suprathreshold concentrations for PROP and sodium chloride (NaCl) and measuring the intensity using magnitude estimation (Bartoshuk et al. 1994; Lucchina et al. 1998). Variations on the number of solutions tested (Tepper et al. 2001) and on the scale have been used with this method. Instead of magnitude estimation, some researchers have used nine-point category scales and 15-cm line scales (Tepper and Nurse 1997), but these scales have shown severe ceiling effects (Lucchina et al. 1998; Bartoshuk 2000). The general labeled magnitude scale has given good results for this method and similar to magnitude estimation, avoids ceiling effects encountered with other scales. The gLMS is a semantic scale of perceptual intensity characterized by a quasi-logarithmic spacing of its verbal labels. The scale ranges from no sensation at the bottom of the scale to strongest imaginable sensation of any kind (top of the scale; Green et al. 1993). The gLMS had previously been shown to yield psychophysical functions and can be used to scale sensations of taste and smell when they are broadly distributed (Green et al. 1996). Although many methodologies are available to classify people based on their taste sensitivity, few studies have compared these methodologies in a systematic way.

In the present study, we carried out a systematic comparison of four methodologies commonly used to assess threshold and suprathreshold sensitivity in humans. The methods tested were chosen mostly on popularity of use but also based on the specific parameter they measure. We chose two threshold methodologies to be compared, one that measures recognition threshold (modified Harris–Kalmus) and another that measures detection threshold (staircase). For suprathreshold evaluation, we chose to compare differential thresholds (Weber fractions) to psychophysical functions (gLMS) to assess which parameters more accurately separate the people as tasters or non-tasters. For evaluating the methodologies, three specific objectives were targeted in this research. The first consisted of assessing the accuracy of the method to classify people correctly based on their sensitivity to PROP. The second objective consisted of evaluating the repeatability of methods, for which we compared the reproducibility of individual results given for PROP and NaCl. The third objective consisted of evaluating the time required to perform each methodology, given that time is usually a strong constraint in large studies. The ultimate goal for this study was to identify two fast and reliable methods for measuring taste sensitivity that can be used for large cohort studies in the quest to discover new phenotype–genotype associations.

Methodology and Trials

Subjects

One hundred adults from the Nestlé Research Center were initially prescreened for taster status with a saturated PTC (phenylthiourea) strip (Carolina Genetics, USA). Based on their ability to detect the bitterness from PTC, 11 tasters and ten non-tasters were recruited for the study ($n=21$, 15 female, age range from 18 to 45). Research protocols were approved by the Ethics Commission of the University of Lausanne; all subjects agreed to participate by providing informed consent form. All tests were performed in a specialized sensory laboratory as individual sessions. A training session was performed with two concentrations of the five basic tastes to help the subjects recognize the taste quality and to familiarize them with the use of the gLMS. Subjects were asked to refrain from eating, drinking tea and coffee, smoking, or brushing their teeth 1 h prior to testing. Subjects were instructed to place 10 mL of solution in mouth for 5 s before spitting it out and making a rating. Subjects were required to rinse their mouth with deionized (DI) water four times between each tastant. Subjects were not allowed to participate to more than two half-hour sessions per day, having at least 1 h of rest in between. Subjects underwent at total of 18 half-hour individual

sessions in a period of 3 months. Four subjects failed to provide a sample for genetic evaluation given their early departure from the study, but as their psychophysical data was complete, we decided to include their data in the analysis.

Solutions

Sodium chloride, phenylthiocarbamide and, and n-6-propylthiouracil were purchased from Sigma Chemical as purum or European-Pharmacopoeia grade. All solutions were prepared with DI/MilliQ (Filter-Sterile) water to avoid differences in salt levels. Solutions were prepared 1 day in advance and kept under refrigeration conditions (4°C) for a maximum of 1 week. One hour before each session, solutions were placed in a water bath to equilibrate the temperature and kept at 21±1°C until the end of the tests. Filtered DI water was used as the blank stimulus and the rinsing solution. Solutions were presented in 20 mL plastic cups. All compound/method combinations were tested with each subject in a random order and in triplicate.

Methodologies

For purposes of this study we defined "subjective" methods as those procedures in which the subject needs to either give a rating on a scale (gLMS) and/or identify a taste quality (modified Harris–Kalmus). "Objective" methods refer to those procedures where the subject has only to identify the different sample (staircase and JND/Weber fractions). The four methodologies were tested in random order using NaCl or PROP as tastants until all repetitions were completed.

Threshold Methodologies

Concentration series for NaCl and PROP decreased by 1/6 log steps (factor equal to 1.468) from a moderate strong concentration (reference concentration). Concentrations ranged from 2.18e−3 to 3.2 mM for PROP and for NaCl 0.10 to 100 mM in 20 dilution steps.

Detection Threshold: Staircase Method

This test permits assessment of the detection threshold, (the lowest concentration at which a person first detects the presence of a sensation, in this case a taste sensation). The assay was based on a two alternative forced choice test (2AFC) where the assessor was presented with two solutions starting at dilution step 12 (2.15 mM salt and 0.047 mM PROP), one solution containing the compound to be tested and the other DI water; each time, the assessor had to identify the solution containing the taste. A wrong

answer moved up (increased) the concentration to be tested while correct answers moved down the concentration following a staircase procedure (four down, one up with five reversals; Bartoshuk 1978). Thresholds were calculated by averaging the log concentration of the last four reversals (change of direction) and expressed as geometric means.

Recognition Thresholds: Modified Harris–Kalmus Method

This test assesses the recognition threshold, (the lowest concentration at which a person is able to clearly characterize a taste). Subjects received a series of cups containing 20 solutions in ascending order of concentration for a single compound. Subjects were asked to taste the solutions one by one and to identify which one had a clear taste. When subjects identified a cup eliciting a clear taste they were asked to identify the quality of that taste. If the subject identified a quality different to that of the target taste compound or could not identify the quality, they proceeded with the next higher concentration in the series. When subjects identified the correct quality, they were presented with a sorting test made of six cups, three containing DI water and three containing the target solution previously identified as having the taste. The subjects' task was to group the solutions in two sets and identify those that contained the taste. If they succeeded in sorting the solutions, that concentration was labeled as their recognition threshold. If subjects failed, they were presented with another six-cup sorting task containing the next highest concentration. Failure to sort the solutions correctly into two groups led to an increase of concentration until the recognition threshold was obtained or subjects reached the highest concentration in the series.

Suprathreshold Methodologies

Just Noticeable Differences and Weber Fractions

This test permits assessment of the just noticeable difference 'JND' or 'ΔI' (the lowest difference in concentration that is clearly perceptible to the assessor) between concentrations of a compound, at a suprathreshold concentration range. The assay is based on a 2AFC. Solutions were prepared following 1/24 log steps (factor equal to 1.1007) from a predefined baseline concentration up (total of 15 dilutions per test). Subjects were presented with two cups, one cup containing the solution at the baseline concentration and the other cup a solution at a higher concentration (starting at dilution step 5) and presented in random order. Subjects had to identify the solution with the stronger taste. A wrong answer moved up (increased) the difference of concentration while correct answers moved down this difference, following

a staircase procedure (three down, one up with five reversals). JND were calculated by averaging the concentrations of the four last reversals (change of direction) minus the reference concentration. JND were determined at Iso-Molar (for NaCl and PROP) and isointense levels (for PROP) see Fig. 1:

- Iso-Molar levels: two baseline levels were chosen for NaCl (100 and 32 mM). One iso-Molar baseline concentration was chosen for PROP equal in concentration to step 4 in the gLMS test (0.32 mM) in order to compare the results from both methods.
- Isointense levels: given that for some subjects the iso-Molar concentration could be too close to threshold or saturation levels, where the Weber's law no longer applies, two baseline levels were chosen in order to assess Weber fractions at isointense levels (the intensity perceived of the baseline concentration is matched for every subject). The levels were chosen after running the test with gLMS on PROP and based on the psychophysical functions for tasters and non-tasters (Figs. 1 and 4a). Concentrations correspond to a weak bitter level on the gLMS (0.1 and 2 mM PROP).

In order to compare JND values at different levels, the data was transformed into Weber fractions. Weber fractions were calculated from the JND using the following formula:

$$\frac{\Delta I}{I} = K$$

Where ΔI is represents the just noticeable difference, I represents the reference stimulus, and K is the Weber fraction.
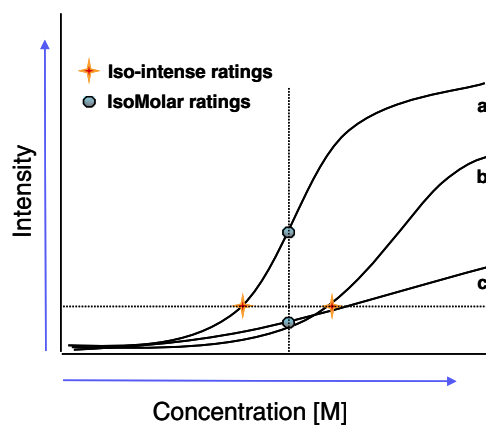

Fig. 1 Determination of iso-Molar and isointense bitterness level. Expected psychophysical functions for PROP tasters (*a*) and non-tasters (*b* and *c*); a *horizontal line* indicates the concentrations at which both groups experience the same bitter intensity (isointense level). The *vertical line* indicates the intensity perceived by each group when exposed to the same PROP concentration (Iso-Molar level)

Intensity Ratings

Subjects used a vertical general labeled magnitude scale to rate the perceived taste intensity of PROP, PTC, and NaCl. The scale was divided by verbal adjectives logarithmically spaced corresponding to barely detectable (corresponding value of 1.5), weak (6), moderate (17), strong (35), very strong (52), and strongest imaginable (100). Subjects were presented with six cups containing different concentrations of NaCl, PROP, or PTC (one compound per session). Subjects were initially trained to identify each basic taste quality by presenting them with exemplars of weak and moderate intensity. Subjects were trained on the use of the scale with the previous solutions and with verbal descriptors referring to previous experiences with food and other sensations (Green et al. 1993; Hansen et al. 2006). All subjects were able to differentiate the two concentration levels and identify the quality attribute. The rating took place in triplicate on different days. Subjects were asked to taste each sample for 5 s before spitting it out and rating the perceived intensity on a gLMS paper scale. Concentrations used for gLMS rating were prepared based on a 1/2 log steps (factor equal to 3.162) and they comprise: PROP 0, 0.032, 0.10, 0.32, 1.0, and 3.2 mM; PTC 0, 0.056, 0.18, 0.56, 1.8, and 5.6 mM; NaCl 0, 10, 31.6, 100, and 316 mM and 1 M.

Taster Status

Subjects were prescreened as tasters or non-tasters based on their response to a saturated PTC paper strip. After recruitment, intensity ratings given to a 0.56 mM PTC solution were obtained for all subjects and used for clustering of groups. This data was then compared to the individual genetic haplotype to corroborate the initial classification and further confirmed by comparing results from individual psychophysical functions for PTC and PROP.

Genetic Analysis

DNA Sampling

Buccal cell samples were collected from 17 subjects at the end of the study (four subjects were no longer available for cell collection). The Gentra Puregene Buccal Cell Kit (400) and DNeasy-Blood and Tissue Kit (Qiagen, Switzerland) were used for both collecting the cells and purification of DNA. Samples were obtained by brushing a cytobrush inside of both cheeks for isolation of genomic DNA and purified following the manufacturer instructions.

Real-Time PCR Genotyping

SNP genotyping was performed using ready-made assays (Applied Biosystems, Switzerland). The assays used were: C_8876467_10–SNP 1 P49A; C_9506827_10–SNP 2 A262V. The polymerase chain reaction (PCR) reaction was run on the ABI 7900HT Fast-Real Time PCR machine (Applied Biosystems, Switzerland) following manufacturer's instructions. Cycling conditions were: 10 min at 95°C followed by 50 cycles of 15 s at 92°C and 1 min at 60°C.

PCR and Sequencing

To characterize SNP2 (A262V), primers (OTL594 and 595) were designed to amplify a 600 bp fragment of hTAS2R38 containing SNP2. To confirm the real-time PCR amplification data, primers (OTL592 and 593) were designed to amplify a 513 bp fragment containing the SNP 1 (P49A; AmpliGoldTaq: Applied Biosystems, Switzerland). Cycling conditions were: 5 min at 94°C, followed by 30 cycles of 30 s at 94°C, 30 s at 50°C and 1 min at 72°C, followed by 7 min at 72°C and ∞ 4°C. Gene-specific primers used were: (OTL 592 Forward Primer) 5' ACC AGG TCT TTT AGA TTA GCC AAC T 3'; (OTL 593 Reverse Primer) 5' CAA AGA ATA ATA CCC AGG AGC ATC 3'; (OTL 594 Forward Primer) 5' GCT TGG CAA GCT GGG TCT CG 3'; (OTL 595 Reverse Primer) 5' TCC GGG AAT CTG CCT TGT GGT 3'.

PCR reactions were purified using the Qiaquick PCR Purification Kit (Qiagen, Switzerland) and sequenced using a sequencing Service (Microsynth, Switzerland) using primers OTL601 for SNP1 and OTL600 for SNP2. The sequences of the primers used were: (OTL 600 Forward Primer) 5' AGC AGA CCT CAC TTC ACA GT 3'; (OTL 601 Forward Primer) 5' TAG CCA ATT AGA GAA GTG AC 3'.

Sequences of the PCR products were aligned using the ClustalW2 Tool (www.ebi.ac.uk) against the sequence of hTAS2R38 Accession Number NM_176817.2. The nucleotide(s) at the position of the SNP was identified visually from the chromatograms.

Data Analysis

All results are expressed as geometric means and geometric standard error. The statistical analyses consisted of comparing repeatability of methods and assessing the ability of the methods to separate between tasters and non-tasters. A 95% confidence level was applied for all tests.

To compare repeatability of the methods, coefficients of variation (CV = standard deviation/mean) of log-transformed data were computed. The CV was preferred over the standard

deviation because it allows correcting for the differences in means between methods (heteroscedasticity) and between tasters and non-tasters. CVs were calculated individually and compared within subjects across methods, using a two-tailed Wilcoxon signed rank test for paired data. The use of standard parametric tests was not possible because basic assumptions on normality and equality of variances were in most cases not met, hence the choice of using a non-parametric approach.

To assess the ability of the methods to separate between tasters and non-tasters, two different approaches were used:

With A Priori Knowledge of Taster Status

The first approach consisted in assessing the differences in sensitivity between the two predefined groups for thresholds and suprathreshold ratings using a one-tailed Wilcoxon signed rank test. The basic assumptions are that detection and recognition thresholds as well as Weber fractions (iso-Molar) are expected to be higher for non-tasters than for tasters, while gLMS ratings are expected to be higher for tasters than for non-tasters. Only in the case of Weber fractions at isointense levels, a two-tailed Wilcoxon signed rank test for two sample data was used.

Without Previous Knowledge of Taster Status

The second approach consisted of applying the k-means algorithm for each method in order to get two clusters of subjects from the raw data without any preliminary classification. The k-means was applied on log-transformed data because this algorithm relies on the Euclidean distance. The resulting clusters were then compared to the taster/non-taster classification obtained from the PTC strip and genetic haplotype using the Hubert–Arabie index (Hubert and Arabie 1985). This index ranges between −1 and 1 where high values indicate that the two clusters compared are very similar. Values of 0.64 would indicate around 10% misclassification.

Results

Taster Status

Twenty-one subjects were clustered based on their response to a one solution test of 0.56 mM PTC. This classification was corroborated by their individual haplotype for *TAS2R38* (Table 1) and further confirmed by comparing the individual psychophysical functions for PTC. The classification as tasters and non-tasters was initially done in order to evaluate the accuracy of the methods to correctly identify the different phenotypes when testing the same group of subjects.

Threshold Methodologies

Detection and recognition thresholds were determined for PROP and NaCl in triplicate using the staircase or modified Harris–Kalmus methodologies. Figure 2a and b show the average thresholds per subject for each method and for both test compounds. Average group thresholds for NaCl were 1.77 mM with SC and 5.12 mM with m-HK method, showing a larger variability among individuals with recognition thresholds as compared to detection thresholds (Fig. 2a). PROP thresholds varied from 0.031 mM to 0.56 mM with SC and 0.028 mM to 0.55 mM with m-HK for tasters and non-tasters (Table 2). Figure 2b clearly shows differences in threshold between tasters and non-tasters using both threshold methodologies. As expected, PROP average thresholds for tasters are lower than for non-tasters. Although values vary slightly between methods (Table 2), both methodologies are capable to distinguish tasters from non-tasters.

In order to compare the repeatability of methods, we computed individual coefficients of variation obtained with both threshold methodologies using log-transformed data. CVs were compared within subject with a two-tailed paired Wilcoxon ranked test. No significant difference on the CV between SC and m-HK methods was observed at 95% confidence level for NaCl thresholds ($p=0.92$) nor PROP thresholds ($p=0.11$). However, the trend is for PROP detection thresholds to have smaller coefficients of variation (average CV 0.046) as compared to PROP recognition thresholds (average CV 0.071). In general, data dispersion shows higher variability for m-HK method as compared to SC (Fig. 2a and b error bars).

Suprathreshold Methodologies

JNDs/Weber Fractions

Differences in sensitivity measured as JND and transformed to Weber fractions were obtained using two baseline levels for NaCl and three baseline levels for PROP in triplicate. JNDs for PROP were measured using iso-molar and isointense concentrations. Only the iso-molar ratings were used to evaluate the method against gLMS intensity ratings. The test with isointense concentrations was performed separately in order to evaluate a new hypothesis. We wanted to know if at isointense levels of perception, Weber fractions remain constant.

Iso-Molar Baseline Concentrations Figure 3 shows the individual Weber fractions for each compound/baseline concentration. Group average Weber fractions for NaCl are similar for both baseline concentrations: 0.166 (at 100 mM baseline) and 0.160 (at 30 mM baseline). These

Table 1 Classification as taster or non-taster based on the individual rating for 0.56 mM PTC solution, cutting point gLMS=17 (moderate). Comparison of phenotype to the genotype analysis for TAS2R38 gene obtained by PCR and gene sequencing, n=21 (11 tasters) to corroborate taster status. Last column indicate screening results with PTC strip test N non-taster, Y taster

Subject	gLMS ratings for 0.56mM PTC	Phenotype	SNP1 P/A 49	SNP2 A/V 262	Haplotype	Phenotype (PTC rating)/genotype match	PTC strip test
1	1.1	Non-taster	–	–	–		N
3	1.0	Non-taster	AA	VV	AV-AV	√	N
4	2.7	Non-taster	AA	VV	AV-AV	√	N
5	4.0	Non-taster	AA	AV	AV-AA	√	N
15	6.9	Non-taster	–	–	–		N
19	1.3	Non-taster	AA	AV	AV-AA	√	N
22	1.4	Non-taster	–	–	–		N
23	3.9	Non-taster	AA	VV	AV-AV	√	N
26	4.8	Non-taster	AA	VV	AV-AV	√	N
12	8.1	Non-taster	AA	AV	AV-AA	√	N
6	27.8	Taster	PA	AV	PA-AV	√	Y
8	43.9	Taster	PP	AA	PA-PA	√	Y
10	57.4	Taster	PA	AV	PA-AV	√	Y
11	30.9	Taster	PA	AV	PA-AV	√	Y
13	32.4	Taster	PP	AA	PA-PA	√	Y
16	32.2	Taster	PP	AV	PA-PV	√	Y
18	59.4	Taster	PA	AV	PA-AV	√	Y
20	33.7	Taster	PA	AV	PA-AV	√	Y
21	34.6	Taster	PA	AV	PA-AV	√	Y
24	19.2	Taster	–	–	–		Y
25	41.9	Taster	PA	AV	PA-AV	√	Y

results are in line with previous reports showing an average Weber fraction of 0.153 (Schutz and Pilgrim 1957), although we found large differences among individual responses for NaCl. Weber fractions for PROP were determined at an iso-Molar baseline concentration of 0.32 mM. Average Weber fractions show significant differences between tasters and non-tasters when using iso-Molar concentrations (Fig. 3c, p=0.04). Tasters show in average
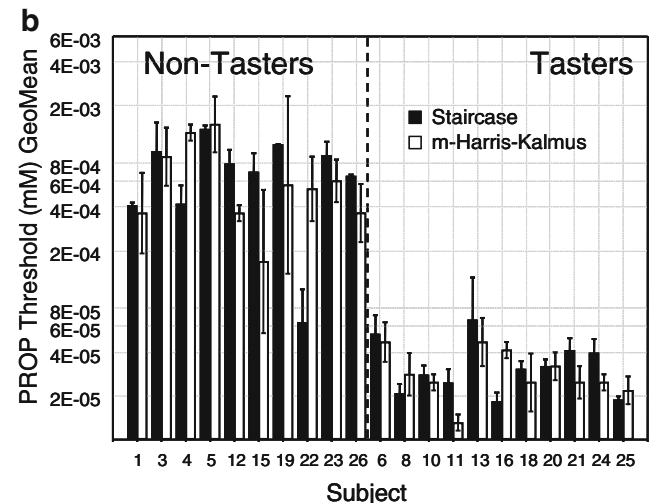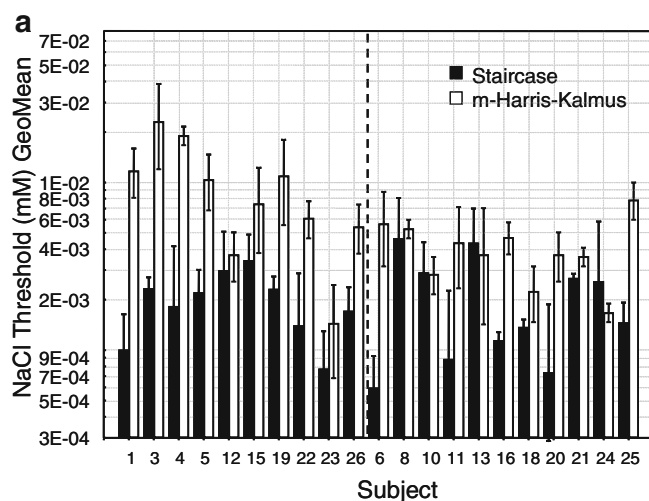

Fig. 2 Individual thresholds for NaCl and PROP as determined by two methodologies: staircase method (detection threshold) vs. modified Harris–Kalmus method (recognition thresholds). **a** Average detection (black bars) and recognition (white bars) thresholds per subject for NaCl and **b** PROP, n=3 repetitions. Subjects are ordered according to PROP taster status and results are plotted in semi-logarithmic space

Table 2 Average ratings and confidence limits (lower LCL and upper UCL) per group/method and comparison of methods' accuracy for classifying people as tasters or non-tasters based on PROP sensitivity

Threshold methods	Concentration PROP (M)	Taster threshold GeoMean (mM)	95% LCL-UCL mM	Non-taster threshold GeoMean (mM)	95% LCL-UCL mM	Wilcoxon (p) taster vs. non-taster	Hubert–Arabie index
Staircase	N/A	0.031	0.025–0.038	0.56	0.39–0.82	<0.001[a]	0.81
Harris Kalmus	N/A	0.028	0.023–0.033	0.55	0.36–0.84	<0.001[a]	1

[a] Indicates significant differences at $\alpha = 0.05$. The Huber–Arabie index (−1 to 1) indicates the level of misclassification obtained when comparing clusters with and without a priori knowledge of taster status. An index of 0.64 indicates about 10% misclassification

higher sensitivity to PROP reflected in lower Weber fraction when measured at iso-Molar concentrations; however, some people can be easily misclassified based solely on this parameter (Fig. 3c, subjects 4, 6, 8, and 15). Further research is needed to explain the overlap on Weber fractions between tasters and non-tasters. To the best of our knowledge, this constitutes the first report on the differences in PROP Weber fractions as they relate to genetic status.

Isointense Baseline Concentration The isointense baseline concentrations for PROP were derived from the average psychophysical functions for tasters and non-tasters. As it is known that Weber's law does not apply when the perceived intensity is close to either saturation or threshold, this procedure was done to ensure all subjects were able to clearly perceive the taste of PROP. Average Weber fraction did not show significant differences at isointense levels (Fig. 3d and e,
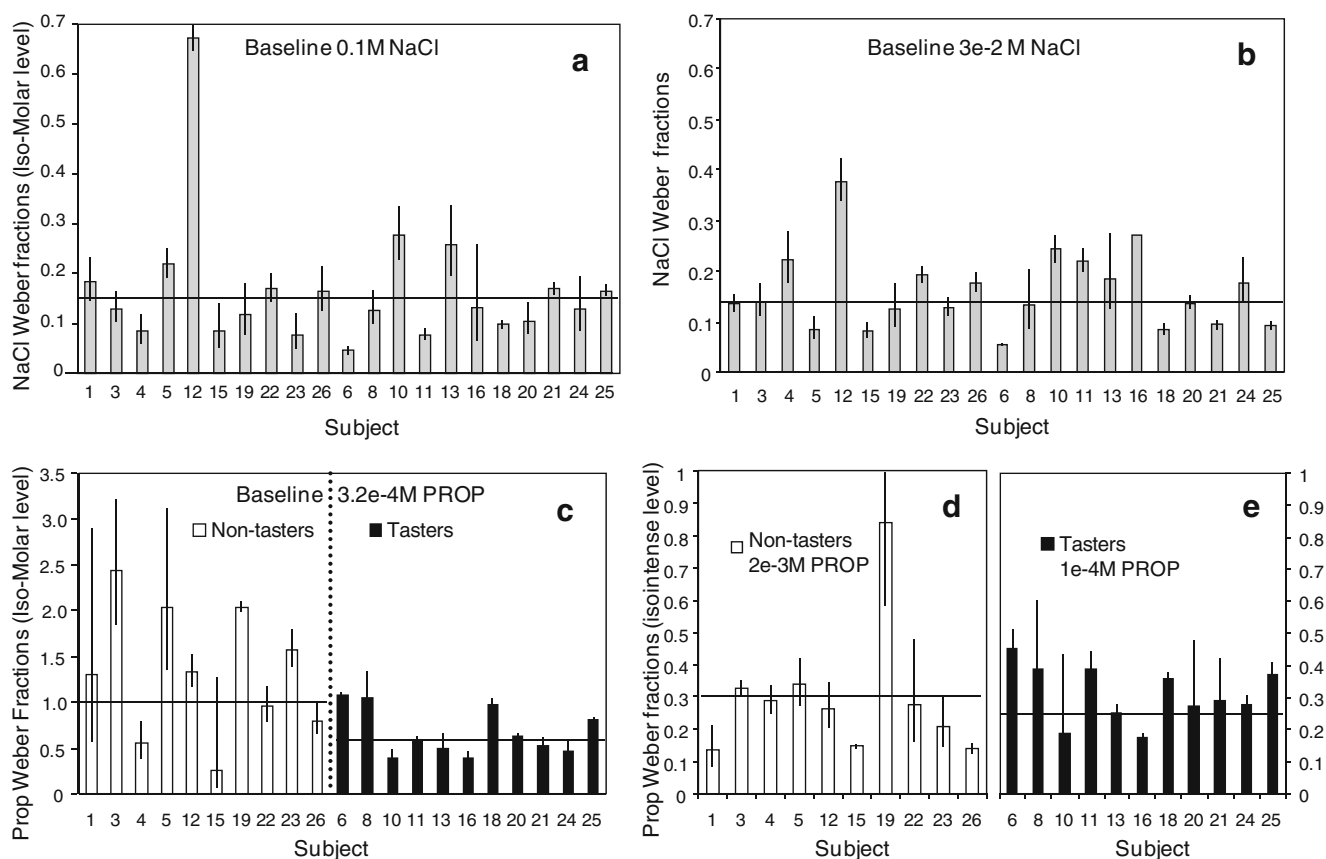

Fig. 3 Weber fractions for NaCl (**a** and **b**) and PROP (**c**, **d**, and **e**) for 21 subjects (ten non-tasters) at different intensity levels. **a** NaCl 100 mM (step 4 on gLMS). **b** NaCl 30 mM (step 3 on gLMS). **c** PROP 0.32 mM (step 4 on gLMS). **d** and **e** PROP Weber fractions obtained at isointense levels for tasters (0.1 mM PROP, step 3 gLMS) and non-tasters (2 mM PROP) based on individual psychophysical functions. Data are expressed as geometric means and geometric standard errors, *horizontal lines* indicate the average value for the group. Subjects are ordered according to PROP taster status

Table 3 Average ratings and confidence limits (lower LCL and upper UCL) per group/method and comparison of methods' accuracy for classifying people as tasters or non-tasters based on PROP sensitivity

Suprathreshold methods	Concentration PROP (mM)	Taster Intensity GeoMean	95% LCL-UCL	Non-taster Intensity GeoMean	95% LCL-UCL	Wilcoxon (p) Taster vs non-taster	Hubert-Arabie index
gLMS	0.032 mM	2.5	1.7–3.6	0.96	0.58–1.57	0.001*	0.14
gLMS	0.100 mM	12.3	10.2–14.7	1.5	0.9–2.6	<0.001*	0.64
gLMS	0.32 mM	26.2	23.2–29.6	2.0	1.2–3.4	<0.001*	1
gLMS	1.00 mM	40.2	36.0–44.8	6.7	4.0–11.1	<0.001*	0.64
gLMS	3.20 mM	54.1	49.4–59.3	24.5	17.1–35.2	0.02*	0.17
Weber fraction	PROP (mM) Baseline	Taster Weber Fraction	95% LCL-UCL	Non-taster Weber Fraction	95% LCL-UCL	Wilcoxon (p) Taster vs non-taster	Hubert-Arabie index
IsoMolar level	0.32 mM	0.59	0.51–0.69	1.02	0.69–1.52	<0.001*	0.14
Isointense level	0.10 mM	0.25	0.19–0.33	N/A	N/A	0.2	-0.02
	2.00 mM	N/A	N/A	0.30	0.25–0.36	(Two tailed)	

[a] Indicates significant differences at $\alpha = 0.05$. The Huber–Arabie index (−1 to 1) indicates the level of misclassification obtained when comparing clusters with and without a priori knowledge of taster status. An index of 0.64 indicates about 10% misclassification

Table 3, $p = 0.2$), despite having large variability in Weber fractions among subjects. We expected differences in Weber fractions between tasters and non-tasters; however, our results do not support this hypothesis. The mechanisms for this phenomenon are yet to be discovered.

gLMS Intensity Ratings

Individual psychophysical functions were obtained for PROP, PTC, and NaCl. Average individual data are represented in Fig. 4a, b, and c, respectively. Non-tasters can be distinguished from tasters by the shape of the function, which remains low in intensity throughout the concentration range for PTC and PROP. Although for PROP, the difference between tasters and non-tasters is optimized at the middle concentration point (0.32 mM; Hayes et al. 2008). There are clear differences in the functions for PTC as compared to PROP. For some non-tasters the intensity of PROP increases rapidly at high concentrations, showing an overlap with taster's functions, although responses for PTC for the same subjects remain low at high concentrations. No major differences were detected among NaCl functions. There were two major outliers in the NaCl distribution, subjects 12 and 26; the high ratings given for NaCl by subject 26 can be explained due to difference in scale usage.

Repeatability of Suprathreshold Methods

The same procedure used to compare repeatability of threshold was used for suprathreshold methods. We calculated and compared the individual coefficients of variation for both suprathreshold methodologies: Weber fractions and gLMS intensity ratings. Results show differences depending on the compound tested, the CVs for NaCl were significantly higher when using Weber fractions (average CV 0.227) than for gLMS ratings (average CV 0.115, $p = 0.01$). In the case of PROP, CVs for both methods were of similar magnitude (average CV 0.125, $p = 0.66$).

Accuracy of Methods to Classify Tasters and Non-tasters: Average PROP Responses per Group

Tables 2 and 3 show average psychophysical responses to PROP stimuli by tasters and non-tasters, classified based on their genetic haplotype and PTC strip classification, for the four methods in question. The seventh column shows statistical results from the Wilcoxon signed rank test for testing differences between the pre-defined groups of tasters vs. non-tasters. Significant differences between taster and non-tasters were obtained for all methods, except when testing Weber fractions at isointense levels. The last column represents the Hubert–Arabie index used when comparing two clustering methods, with and without a priori knowledge of subject taster status. Values closer to 1 reflect perfect match between clusters, while values closer to −1 indicate no congruency between clusters. Figure 5 shows the box-plots results for PROP for each individual methodology based on pre-determined taster status. Both threshold methodologies show good separation between tasters and non-tasters (Fig. 5a and b) and have on average higher Hubert–Arabie
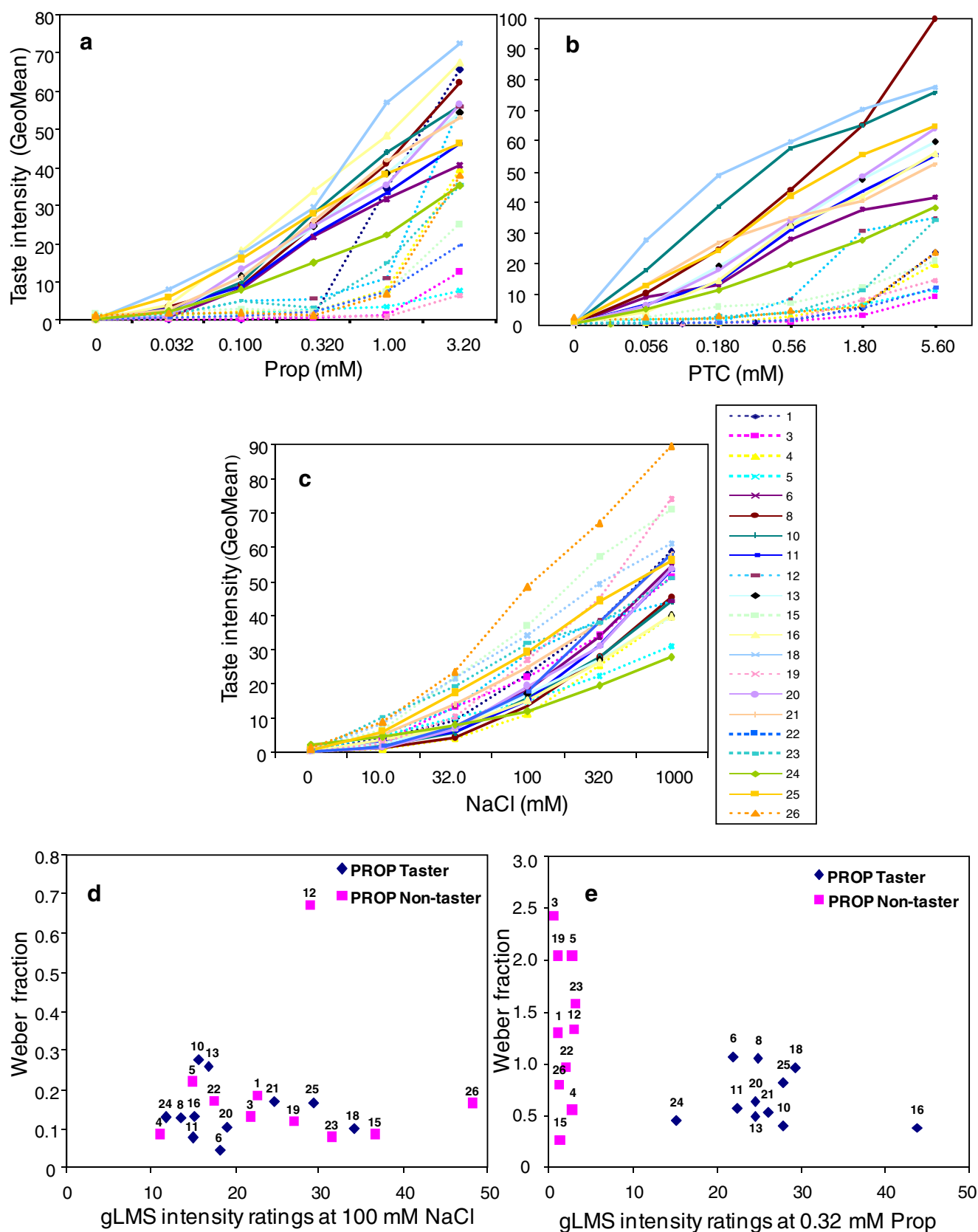
Fig. 4 Comparison of methods for measuring suprathreshold sensitivity and individual psychophysical functions for **a** PROP, **b** PTC, and **c** NaCl obtained with gLMS; GeoMeans, *n*=3 reps, *dotted lines* indicate PROP non-tasters. Distribution of positive Weber fractions versus gLMS intensity ratings for **d** NaCl 100 mM baseline concentration or **e** PROP 0.32 mM baseline concentration

indexes as compared to suprathreshold methodologies. gLMS intensity ratings show high indexes in the mid-range of the psychophysical function where good separation of groups is visually apparent (Fig. 5c), while JND/Weber fractions show the lowest index reflecting the difficulty in classifying subjects without preliminary information on the subject taster status. The distributions of Weber fractions at iso-Molar and isointense levels show a clear overlap between groups and hence are the least accurate method for phenotyping people (Fig. 5d and e). Figure 4d and e show a direct comparison of the Weber fractions vs. gLMS ratings when using Iso-Molar concentrations. It is clear from the results that clustering of subjects as taster/non-tasters is more

reliable for the gLMS method as Weber fractions from tasters and non-tasters show an overlap.

Time Evaluation

As testing time becomes crucial when running large cohort studies, we calculated the time needed to complete each test as an extra parameter. The average time to complete the modified Harris–Kalmus test was of 15 min while the staircase evaluation required an average of 30 min per compound.

For Weber fractions, subjects took an average of 30 min to complete the evaluation, after 30 min we were compelled to stop the evaluation, as people were fatigued. In the case

Fig. 5 *Box-plots* comparing data distribution of tasters (*gray box*) vs. non-tasters (*white box*) as they respond to PROP stimuli using different threshold and suprathreshold methodologies, *n*=21 (11 tasters, three reps). The *length of the box* represents the interquartile range (IQR, middle 50% of the data); the *middle line* represents the group median. *Y*- and *X*-axes scales vary according to the type of test. The methods consist of: **a** bitter recognition threshold for PROP by modified Harris–Kalmus method. **b** Detection threshold by staircase method. **c** Average gLMS intensity ratings for five different concentrations of PROP. **d** Weber fractions measured at a single PROP concentration (0.32 mM). **e** Weber fractions measured at isointense (weak) level based on individual psychophysical functions (0.1 mM for tasters and 2.0 mM PROP for non-tasters)
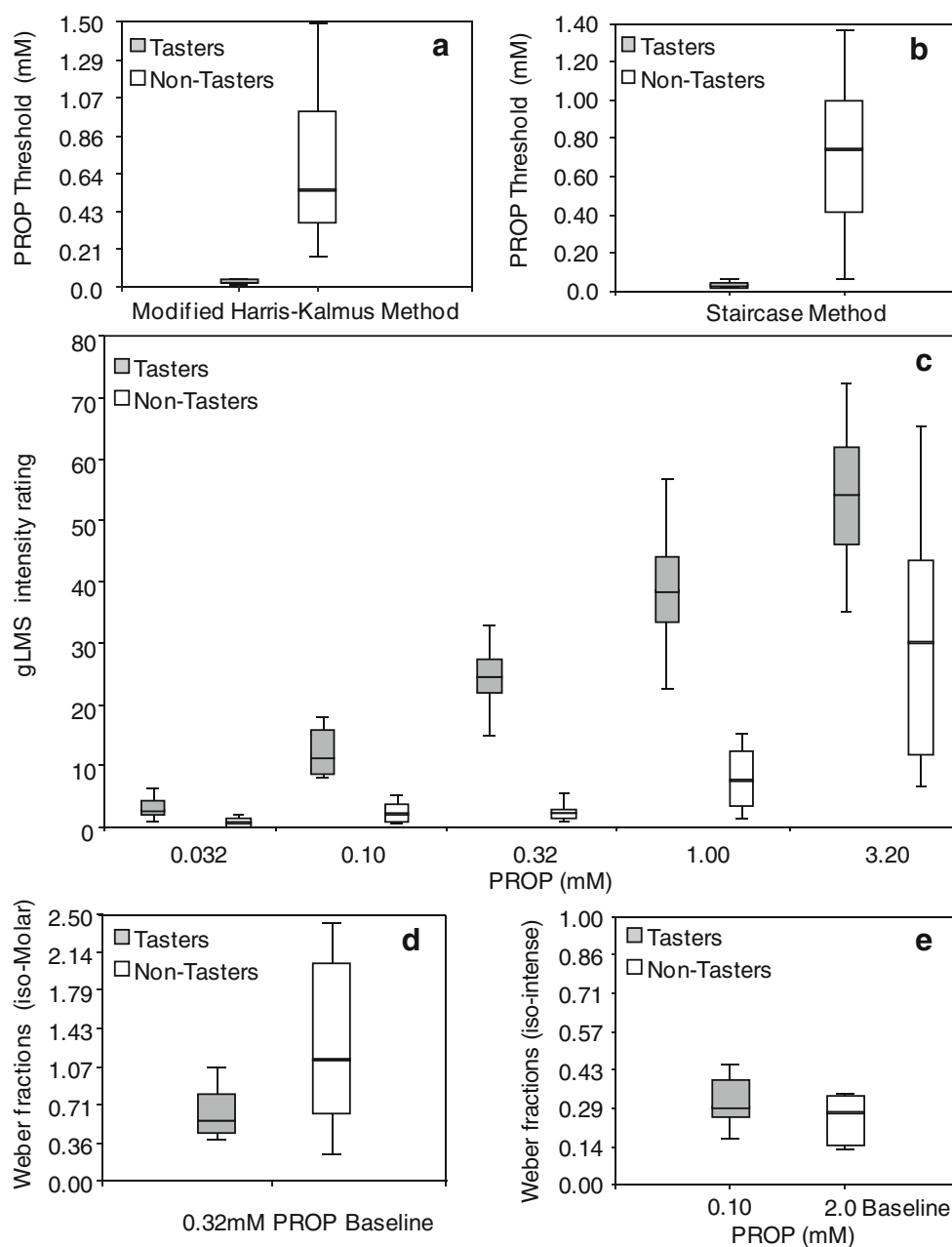
of intensity ratings, the average time needed to complete six ratings using gLMS was 10 min, which makes the gLMS and the modified Harris–Kalmus very fast tests.

Discussion

Taster Status

The three most common polymorphisms of *TAS2R38* bitter receptor responsible for the perception of a bitter taste from PTC and PROP occur at amino acid positions 49 (encoding for proline or valine), 262 (alanine or valine), and 296 (valine or isoleucine). Among the different combinations, the most frequent are PAV and AVI corresponding to taster and non-taster haplotypes; however, other haplotypes AAI, PVI, and AAV have also been encountered (Kim et al. 2003; Bufe et al. 2005). Table 1 shows subjects' individual haplotypes for this gene, which closely resemble what has been previously found in European communities and all haplotypes corroborate results from the initial PTC rating (Bufe et al. 2005). In our study, we did not determine the third amino acid; however, based on previous correlations done with only the first two amino acids and the incidence in the European population of the different haplotypes, we are confident of the classification as taster or non-tasters (Kim et al. 2003; Timpson et al. 2007). Psychophysical responses obtained for PROP (Fig. 4) at threshold and suprathreshold levels agree with previous data reported for the different haplotypes (Duffy et al. 2004; Bufe et al. 2005). Subjects that had at least one copy of PA were sensitive to the taste of PROP and PTC as shown by their threshold and suprathreshold ratings (Table 1, Figs. 2 and 4).

Controversy has risen from the accuracy of the paper strips to classify people as tasters and non-tasters, especially when using PROP as stimulus (Lawless 1980), although in his same manuscript Lawless reports that PTC paper test generally agreed with the classification done with a forced choice method. We found the use of commercial PTC strips a good and fast method for rough classification of subjects; however, given that it is difficult to ensure the exact concentration of PTC on the paper, it should not be a definite test for evaluating the phenotype. In this study the individual ratings for 0.56 mM PTC solution were in agreement with the PTC strip test and the individual genetic status for *TAS2R38* (Table 1). Individual phenotype was confirmed by analyzing the individual psychophysical functions for PTC (Fig. 4b) and comparing them to previous reports in the literature (Bufe et al. 2005).

We chose PROP and NaCl as taste exemplars, first due to the lower toxicity given by PROP as compared to PTC (Fischer 1971) and also given the amount of information

available on these tastants from previous psychophysical studies. PROP is the best characterized compound from a psychophysical, molecular, and genetic point of view. Having all background information makes it easier to compare the accuracy and reproducibility of sensory methods. We divided subjects in two groups of tasters and non-tasters, based on their sensitivity to PTC and their individual haplotype (Table 1), instead of three groups that would include supertasters. The reason is directly related to our objectives, which consist in the identification of reliable methods to measure taste sensitivity that can be then applied to compounds other that PROP. Furthermore, the classification of supertasters based on sensitivity to PROP has been recently challenged (Reed 2008) based on the discovery of Lim et al. (2008) that PROP is not the most predictive compound to detect higher responses to other tastes.

Repeatability of Methods

In terms of repeatability of threshold methods, based on the individual standard errors, the staircase (for detection threshold) show less variability among subjects as compared to the modified Harris–Kalmus method (recognition threshold; Fig. 2a and b). As expected, data dispersion is larger for recognition thresholds (Fig. 2a and b, gray bars), and this difference is likely due to the difficulty that subjects have to correctly identify the taste quality in the modified Harris–Kalmus method (e.g. confusing sweet for salty in the case of NaCl, 11 times out of 63). For some subjects, this quality confusion or mislabeling results in higher recognition thresholds, which would suggest less sensitivity to salty taste (Fig. 2a, subjects 3 and 4); however, results from the other threshold and suprathreshold methodologies do not support this lack of sensitivity by the same subjects (Figs. 2a and 4c). When evaluating individual variability between methods taking into account the differences in average through CV calculation, we found no significant differences between the two threshold methodologies with 95% confidence level (Fig. 2a and b), only a tendency to have higher variability of thresholds using the m-HK method, especially when testing PROP (average CV=0.046 SC vs. CV=0.071 m-HK). Despite not having differences on individual CV, we believe SC is a more reliable method than m-HK as the former is an objective measurement, while m-HK depends partly on the subjects identifying the taste quality, which requires a more integrated cognitive process.

When comparing the repeatability of suprathreshold methods, results show differences depending on the compound tested. We found that individual CVs for NaCl gLMS ratings were significantly lower than for Weber fractions ($p=0.01$); in contrast, there was no statistical differences between the two suprathreshold methods for

PROP ($p=0.66$). The difference found between compounds is mainly due to taste adaptation encountered when testing NaCl Weber fractions, as verbalized by some subjects. This problem was not as evident with PROP, as for non-tasters the bitter intensity perceived by the PROP baseline solution was very low compared to the saltiness perceived for NaCl by the same subjects. gLMS ratings on the contrary, were very repeatable within subjects, and therefore are the preferred method in terms of repeatability of data.

Accuracy of Methods for Phenotyping People

When comparing the accuracy of methods for phenotyping people based on their psychophysical response, we can see that both threshold methods (the staircase for detection threshold and modified Harris–Kalmus for recognition threshold) are adequate to separate tasters from non-tasters with a low rate of misclassification (Table 2 Hubert–Arabie index, Fig. 5a and b). Nevertheless, when comparing results among subjects we can see that, similar to NaCl, there is higher variability of threshold values when using the m-HK as compared to staircase method (Figure 2a and b). This large variability could become an issue when phenotyping people using other taste compounds for which differences in sensitivity are not as pronounced as PROP. Based on all combined results, we recommend the staircase method as the preferred method for phenotyping people based on threshold sensitivity.

When comparing accuracy of suprathreshold phenotyping methods, results differ depending on the method. When using intensity scores through gLMS, our results show good clustering of data in the mid-range of the psychophysical function, although at low and high concentrations the degree of misclassification increases considerably (Table 3, Hubert–Arabie indexes and Fig. 5c). Misclassification at high concentrations could be related to the individual use of the gLMS scale. This difference in scale usage is what encourages scientists to standardize the gLMS with other non-oral sensations such as weights and tones; however, for the present study we decided not to collect standardization data as we wanted to assess the reliability of the method for large cohort studies, where standardization might not be feasible. The second source of misclassification is the potential activation of other bitter receptors such as TAS2R4 by high concentrations of PROP (Chandrashekar et al. 2000; Reed et al. 2006). The second receptor hypothesis becomes evident when comparing individual functions for PROP vs. PTC (Fig. 4b and c), PTC functions shows no crossover throughout the concentration range while PROP functions show overlap between tasters and non-tasters at high concentrations.

A common method to measure sensitivity relies on JND/ Weber fractions. According to Weber–Fechner law (Harris

and Kalmus 1949c; Fechner 1987), Weber fractions remain constant throughout the dynamic range of the psychophysical function; however, exceptions appear when testing closer to threshold and saturation levels. We hypothesized that PROP tasters would have higher sensitivity (lower Weber fractions) as compared to non-tasters, and that this difference would be large enough to be used as a parameter to classify people as tasters and non-tasters. This hypothesis is based on the idea that sensitivity is directly related to receptor functionality and given that non-tasters lack proper functionality of TAS2R38 receptor, we expect lower sensitivity to PROP in these subjects. Our results show significant differences on Weber fractions between tasters and non-tasters at iso-Molar concentrations (0.32 mM baseline), non-tasters showed on average larger Weber fractions as compared to tasters. However, when measuring accuracy of the method to separate tasters from non-tasters, we can see a clear overlap between groups (Figs. 4e and 5d and e), making it difficult to classify people based solely on this parameter. There are several reasons that could explain this overlap. First, we cannot rule out the activation of alternative bitter receptors in the case of non-tasters. Large individual variations have been observed in bitter taste perception presumably due to environmental, cellular mechanisms, and genetic factors affecting bitter perception (Cowart et al. 1994; Bartoshuk et al. 1998; Keast and Breslin 2002). These factors could render non-tasters a similar level of discrimination as tasters despite low bitter perception. Second, the delayed onset of PROP taste makes it difficult for subjects to quickly discriminate between solutions, hence extending the time to test each solution and increasing the fatigue experienced by the subject. Third, the baseline concentration (Iso-Molar) could either be too close to threshold for non-tasters or to saturation levels for some tasters. This would cause deviations from Weber's law (Schutz and Pilgrim 1957), and could potentially explain the large range of Weber fractions found for non-tasters and tasters (Table 3).

It has been suggested that in order to better compare differences in sensitivity among people, JNDs should be obtained at equivalent perceived intensity levels, as the JND will change depending where in the psychophysical function they are measured. Based on this idea, we decided to also determine JND/Weber fractions at isointense bitter levels (based on the average gLMS ratings per group). It is interesting to mention that when tested at isointense levels (Table 3); there are no differences in Weber fractions between tasters and non-tasters. Average Weber fractions were 0.34 for tasters and 0.31 for non-tasters. These results are more in line with what has previously been found for bitter taste (Schutz and Pilgrim 1957) and once more proved Weber's law which states that fractions may be constant over several orders of magnitude. The question

remains if the response we are measuring for non-tasters at such high levels of PROP is still due to the response of TAS2R38 receptor, or is there an alternative receptor responding to the same stimuli, such as hTAS2R4 receptor (Chandrashekar et al. 2000; Reed et al. 2006; Hayes et al. 2008). Furthermore, this finding gives rise to new ideas related to the process of differential sensitivity; our results suggest that at isointense levels PROP intensity discrimination is independent of phenotype and even genotype, and that other transduction mechanisms might be responsible for the differentiation. Further research is necessary to address these mechanisms but the idea opens a new area for research. Overall, our results show that although JNDs/Weber fractions are good measurements of sensitivity, they are not good criteria for classifying people as tasters and non-tasters. Hence, for the purpose of classifying people as tasters and non-tasters, our results show that gLMS intensity ratings are more accurate and reproducible than Weber fractions at the tested levels.

Time Considerations

Time expenditure as well as reliability of methods are key factors for the choice of method when working with large-scale studies, and constitute a balance which needs to be weighted depending on the overall objectives of the study. In terms of time expenditure, the m-HK, although less reliable, is a quick threshold method as it requires about 15 min per test as opposed to ~30 to 45 min needed for SC determinations. Furthermore, m-HK method can be applied simultaneously to a group of people as compared to SC, which requires a one-to-one session. gLMS is also a quick suprathreshold method based on time expenditure (10 min) and is the preferred method when compared to JND/Weber fractions, which also requires a one-to-one session.

Conclusions

Based on the accuracy of clustering subjects and the repeatability of the methods presented here, we found no significant differences between the two threshold methodologies for classifying subjects as tasters or non-tasters, as both methods showed high accuracy >95% for clustering subjects. No significant differences were found on repeatability between threshold methods. However, the staircase methodology for detection threshold shows a tendency for lower variability as compared to modified Harris–Kalmus method, and is recommended as an objective method that does not require subjects to identify the taste quality. On the suprathreshold front the gLMS intensity ratings is the most suitable method for phenotyping people based on genetic status, especially when using the mid-range concentrations of the psychophysical function. JNDs/Weber fractions are not good predictors of taster status. Based on the results exposed in this study we recommend the staircase method for detection threshold and gLMS as reliable methods to be used in large cohort studies to phenotype people based in individual sensitivity to different tastants.

References

Barnicot NA et al (1951) Taste thresholds of further eighteen compounds and their correlation with P.T.C thresholds. Annals of Eugenics 16:119–128

Bartoshuk LM (1978) The psychophysics of taste. Am J Clin Nutr 31:1068–1077

Bartoshuk LM (1991) Sweetness: history, preference, and genetic variability. Food Technol 45:108–113

Bartoshuk LM (2000) Comparing sensory experiences across individuals: recent psychophysical advances illuminate genetic variation in taste perception. Chem Senses 25:447–460

Bartoshuk LM et al (1998) PROP (6-n-propylthiouracil) supertasters and the saltiness of NaCl. Ann N Y Acad Sci 855:793–796

Bartoshuk LM et al (1994) PTC/PROP tasting: anatomy, psychophysics, and sex effects. Physiol Behav 56:1165–1171

Blakeslee AF (1932) Genetics of sensory thresholds: taste for phenyl thio-carbamide. Proc Natl Acad Sci U S A 18:120–130

Blakeslee AF, Salmon TN (1935) Genetics of sensory thresholds: individual taste reactions for different substances. Proc Natl Acad Sci U S A 21:84–90

Bufe B et al (2005) The molecular basis of individual differences in phenylthiocarbamide and propylthiouracil bitterness perception. Curr Biol 15:322–327

Chandrashekar J et al (2000) T2Rs function as bitter taste receptors. Cell 100:703–711

Cowart BJ et al (1994) Bitter taste in aging: compound-specific decline in sensitivity. Physiol Behav 56:1237–1241

Delwiche JF et al (2001) Covariation in individuals' sensitivities to bitter compounds: evidence supporting multiple receptor/transduction mechanisms. Percept Psychophys 63:761–776

Dinehart ME et al (2006) Bitter taste markers explain variability in vegetable sweetness, bitterness, and intake. Physiol Behav 87:304–313

Drayna D et al (2003) Genetic analysis of a complex trait in the Utah Genetic Reference Project: a major locus for PTC taste ability on chromosome 7q and a secondary locus on chromosome 16p. Hum Genet 112:567–572

Drewnowski A (1997) Genetic sensitivity to 6-n-propylthiouracil (PROP) and hedonic responses to bitter and sweet tastes. Chem Senses 22:27–37

Drewnowski A (1998) Genetic sensitivity to 6-n-propylthiouracil and sensory responses to sugar and fat mixtures. Physiol Behav 63:771–777

Drewnowski A et al (1997) Taste responses to Naringin, A flavonoid, and the acceptance of grapefruit juice are related to genetic sensitivity to 6-n- propylthiouracil. Am J Clin Nutr 66:391–397

Drewnowski A et al (2001) Genetic taste responses to 6-n-propylthiouracil among adults: a screening tool for epidemiological studies. Chem Senses 26:483–489

Duffy VB et al (2004) Bitter receptor gene (TAS2R38), 6-n-propylthiouracil (PROP) bitterness and alcohol intake. Alcohol Clin Exp Res 28:1629–1637

Fechner GT (1987) My own viewpoint on mental measurement (1887). Psychol Res 49:213–219

Fischer R (1971) Gustatory, behavioral and pharmacological manifestations of chemoreception in man. In: Ohloff G, Thomas AF (eds) Gustation and olfaction. Academic, New York, pp 187–237

Fischer R, Griffin F (1963) Quinine dimorphism: a cardinal determinant of taste sensitivity. Nature 200:343–347

Fischer R et al (1961) Taste thresholds and food dislikes. Nature 191:1328

Fox AL (1931) Six in ten "Tasteblind" to bitter chemical. Sci Newsl 9:249

Glanville EV, Kaplan AR (1965) Food preference and sensitivity of taste for bitter compounds. Nature 205:851–853

Green BG et al (1993) Derivation and evaluation of a semantic scale of oral sensation magnitude with apparent ratio properties. Chem Senses 18:683–702

Green BG et al (1996) Evaluating the 'labeled magnitude scale' for measuring sensations of taste and smell. Chem Senses 21:323–334

Hall MJ et al (1975) PTC taste blindness and the taste of caffeine. Nature 253:442–443

Hansen JL et al (2006) Heritability and genetic covariation of sensitivity to PROP, SOA, quinine HCl, and caffeine. Chem Senses 31:403–413

Harris H, Kalmus H (1949a) Chemical specificity in genetical differences of taste sensitivity. Annals of Eugenics 15:32–45

Harris H, Kalmus H (1949b) Genetical differences in taste sensitivity to phenylthiourea and to anti-thyroid substances. Nature 163:878–879

Harris H, Kalmus H (1949c) The measurement of taste sensitivity to phenylthiourea. Annals of Eugenics 15:24–31

Hayes JE et al (2008) Supertasting and PROP bitterness depends on more than the TAS2R38 gene. Chem Senses 33:255

Hubert L, Arabie P (1985) Comparing partitions. J Classif 2:193–218

Jian BI, Ennis DM (1998) Sensory thresholds: concepts and methods. J Sens Stud 13:133–148

Keast RSJ, Breslin PAS (2002) Modifying the bitterness of selected oral pharmaceuticals with cation and anion series of salts. Pharm Res 19:1019–1026

Keast RSJ, Roper J (2007) A complex relationship among chemical concentration, detection threshold, and suprathreshold intensity of bitter compounds. Chem Senses 32:245–253

Kim UK et al (2003) Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. Science 299:1221–1225

Lawless H (1980) A comparison of different methods used to assess sensitivity to the taste of phenylthiocarbamide (PTC). Chem Senses 5:247–256

Leach EJ, Noble AC (1986) Comparison of bitterness of caffeine and quinine by a time—intensity procedure. Chem Senses 11:339–345

Lim J et al (2008) Measures of individual differences in taste and creaminess perception. Chem Senses 33:493–501

Lucchina LA et al (1998) Psychophysical measurement of 6-n-propylthiouracil (PROP) taste perception. Ann N Y Acad Sci 855:816–819

Ly A, Drewnowski A (2001) PROP (6-n-propylthiouracil) tasting and sensory responses to caffeine, sucrose, neohesperidin dihydrochalcone and chocolate. Chem Senses 26:41–47

Mojet J et al (2005) Taste perception with age: pleasantness and its relationships with threshold sensitivity and supra-threshold intensity of five taste qualities. Food Qual Prefer 16:413–423

Pangborn RM, Pecore SD (1982) Taste perception of sodium chloride in relation to dietary intake of salt. Am J Clin Nutr 35:510–520

Prescott J et al (2004) Responses of PROP taster groups to variations in sensory qualities within foods and beverages. Physiol Behav 82:459–469

Reed DR (2008) Birth of a new breed of supertaster. Chem Senses 33:489–491

Reed DR et al (1995) Propylthiouracil tasting: determination of underlying threshold distributions using maximum likelihood. Chem Senses 20:529–533

Reed DR et al (1999) Localization of A gene for bitter-taste perception to human chromosome 5p15. Am J Hum Genet 64:1478–1480

Reed DR et al (2006) Diverse tastes: genetics of sweet and bitter perception. Physiol Behav 88:215–226

Salmon TN, Blakeslee AF (1935) Genetics of sensory thresholds: variations within single individuals in taste sensitivity for PTC. Proc Natl Acad Sci U S A 21:78–83

Sandell MA, Breslin PAS (2006) Variability in A taste-receptor gene determines whether we taste toxins in food. Curr Biol 16:R792–R794

Schutz HG, Pilgrim FJ (1957) Differential sensitivity in gustation. J Exp Psychol 54:41–48

Tepper BJ, Nurse RJ (1997) Fat perception is related to PROP taster status. Physiol Behav 61:949–954

Tepper BJ et al (2001) Development of brief methods to classify individuals by PROP taster status. Physiol Behav 73:571–577

Timpson N et al (2007) Refining associations between TAS2R38 diplotypes and the 6-n-propylthiouracil (PROP) taste test: findings from the Avon Longitudinal Study of Parents and Children. BMC Genetics 8:51

Weber EH (1834) De pulso, resorptione, auditu et tactu Annotationes anatomicae et physiologicae. Koehler, Leipzig, pp 252–274

Wise PM et al (2007) Twin study of the heritability of recognition thresholds for sour and salty taste. Chem Senses 32:749–754