OXFORD UNIVERSITY PRESS | Chemical Senses

Comparing manual counting to automated image analysis for the assessment of fungiform papillae density on human tongue

Journal:	*Chemical Senses*
Manuscript ID	CS-17-028.R1
Manuscript Type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Piochi, Maria; University of Florence, GESAAF Monteleone, Erminio; University of Florence, GESAAF Torri, Luisa; University of Gastronomic Sciences, University of Gastronomic Sciences Masi, Camilla; University of Florence, GESAAF Amli Lengarde , Valérie ; NOFIMA, AS, Postboks 210, NO-1431 Ås, Norway, NOFIMA, AS, Postboks 210, NO-1431 Ås, Norway Wold Petter, Jens ; Nofima AS, Postboks 210, NO-1431 Ås, Norway, Nofima AS, Postboks 210, NO-1431 Ås, Norway Dinnella, Caterina; University of Florence, GESAAF
Key Words:	density, individual differences, prediction, size, taste intensity

Comparing manual counting to automated image analysis for the assessment of fungiform papillae density on human tongue

Piochi, Maria[a]*, *Monteleone, Erminio*[a], *Torri, Luisa*[b], *Masi, Camilla*[a], *Almli, Valérie L.*[c], *Wold, Jens Petter*[c] *and Dinnella, Caterina*[a]

[a]*GESAAF, University of Florence, Via Donizetti, 6, 50144 Firenze, Italy*

[b]*University of Gastronomic Sciences, Piazza Vittorio Emanuele 9, 12060 Bra, CN, Italy*

[c]*NOFIMA, Postboks 210, NO-1431 Ås, Norway*

** Correspondence to be sent to: Maria Piochi, University of Florence, Via Donizetti, 6, 50144 Florence, Italy, maria.piochi@unifi.it*

Abstract

The density of fungiform papillae (FPD) on the human tongue is currently taken as index for responsiveness to oral chemosensory stimuli. Visual analysis of digital tongue picture and manual counting by trained operators represents the most popular technique for FPD assessment. Methodological issues mainly due to operator bias are considered among factors accounting for the uncertainty about the relationships between FPD and responsiveness to chemosensory stimuli.

The present study describes a novel automated method to count fungiform papillae from image analysis of tongue pictures. The method was applied to tongue pictures from 133 subjects. Taking the manual count as reference method, a PLRS model was developed to predict FPD from tongue automated analysis output. FPD from manual and automated count showed the same normal distribution and comparable descriptive statistic values. Consistent subject classifications as Low and High FPD were obtained according to the median values from manual and automated count. The same results on the effect of FPD variation on taste perception were obtained both using predicted and counted values.

The proposed method overcomes count uncertainties due to researcher bias in manual counting and is suited for large population studies. Additional information is provided such as FP size class distribution which would help for a better understanding of the relationships between FPD variation and taste functions.

Key words: density, individual differences, prediction, size, taste intensity

1. Introduction

The fungiform papillae (FP) are the anatomical structures involved in the detection and transduction of oral stimuli. Together with foliate and circumvallate papillae, FP are considered gustatory papillae since they carry taste receptors (Chen and Engelen, 2012).

FP are innervated by the Chorda Tympani (responsible for taste signals) and by the trigeminal nerve (associated to the somatosensory perception) (Whitehead et al., 1985; Prescott et al., 2004). Due to these double innervations, FP has been taken as a relevant oral responsiveness marker. Human subjects show large variations in FP density (FP/cm^2-FPD), from 0.0 (Webb et al., 2015) to 233.0 (Zhang et al., 2009). The fundamental assumption is that, the higher is the FPD, the more intense is the signal sent to the central system and the higher is the perceived intensity. Taste bud density varies among humans from 374 to 135 pores/cm^2 and not all FP bear taste buds (Miller and Reedy, 1990b; Segovia et al., 2002). Thus, even if significant associations have been reported between taste pores and FP densities (Miller and Reedy, 1990a, 1990b), the higher FPD values might not necessarily correspond to the more intense stimulation. Several studies confirmed the positive relationship between FPD and responses to taste (Miller and Reedy, 1990b; Bartoshuk, 2000; Delwiche et al., 2001; Yackinous and Guinard, 2002; Hayes et al., 2008) and somatosensations (Duffy et al., 2004a, 2004b; Hayes and Duffy, 2007; Nachtsheim and Schlich, 2013). On the other hand, more

recent studies failed to find a relationship between FPD and responsiveness to oral stimuli (Fischer et al., 2013).

Issues related to the methodology for FP identification and counting have been invoked among reasons responsible for controversial relationships found between FPD and oral responsiveness to chemosensory stimulation (Nuessle et al., 2015; Sanyal et al., 2016). Visual inspection of digital pictures of blue stained tongue, followed by manual counting by trained operators, represents the most popular technique for FPD assessment since when digital camera was validated as suitable substitute for videomicroscopy (Shahbake et al., 2005). The use of digital camera does not allow the taste bud detection, thus impairments in the identification of gustatory FP (carrying taste pores) and not gustatory FP (without taste pores) can occur and this might partially account for uncertainty of relationships between FPD assessed by visual digital picture inspection and taste responsiveness.

According to Miller and Ready (1990) description, FP are identified as round, elevated, and pink or stained lighter structures on the blue tongue background. However, FP identification suffers from researcher bias since often papillae can fail to meet every criterion and operators subjectively prioritize the importance of different characteristics leading to FP identification (Nuessle et al., 2015). Thus, highly variable counts can result from the same tongue image analyzed by different operators. A guideline called Denver Papillae Protocol has been developed to help in FP identification and to improve scoring consistency between operators (Nuessle et al., 2015). Bias related to the manual FP count can be even more severe in large population studies when thousands of pictures must be visually analysed and several operators, even working in different locations, are in charge for counting. The adoption of a shared standardized protocol to help in FP identification, together with a quite intensive operator training, can reduce but not fully remove the operator bias in FP count (Garneau et al., 2014).

Another limitation of manual counts relates to dimension and location of the considered tongue area. In fact, to simplify and speed the count, only restricted areas of the tongue picture are visually analysed and relevant counts used to infer the overall FPD value. FP are unevenly distributed all over the anterior two-third of the tongue (Jung et al., 2004). Wide differences between distribution of papillae of individuals have been reported, with some having high density on the tip whereas others exhibit more even distribution across the anterior area (Miller, 1986). Furthermore, the correlations amongst counts performed in small different area of the anterior part of the tongue are highly variable (Shahbake et al., 2005). All these aspects add variability in FPD visual estimation thus further impairing the investigation of relationships between FPD and taste function.

Automated image analysis could be a very useful tool to standardise FP count and to improve the consistency of data. Recently, two studies have been conducted to automatically count FP on human tongue (Sanyal et al., 2016; Valencia et al., 2016), demonstrating the increasing interest towards this issue. However, these methods have some limitations related to the need of manual intervention, to the restriction of tongue area suitable for the analysis (Valencia et al., 2016) and the relatively small number of pictures considered to test the correlation between automated and manual count (Sanyal et al., 2016).

This paper presents a novel automated procedure for FPD estimation based on the analysis of digital pictures taken with a digital microscope. The relationships between automated method response and manual counting were investigated. A multivariate model was proposed for FPD prediction from automated analysis outputs. The effect of the variation of FPD from manual and automated count on the perceived intensities of supra-threshold taste solutions was explored.

Advantages are the complete automation of the procedure and the analysis of large portions of the tongue, thus overcoming the main factors responsible for bias in manual count; the device

for picture acquisition is portable and inexpensive and the time required to process the images and estimate FPD is strongly reduced, thus the method is suited to handle the large size sample from population studies aimed at investigating relationships between FPD and oral responsiveness. Finally, the proposed image analysis procedure adds information on FP size distribution that was not previously available with manual counting method.

2. Material and Methods

2.1 Subjects

One hundred thirty-three subjects (33% males; aged from 18 to 65 years, mean age=32) were recruited in two sensory analysis laboratories in Italy (University of Florence; University of Gastronomic Science in Pollenzo). Participants were part of the extended "Italian Taste" project, which envisaged the collection of a wide range of data, including pictures of their tongues (Monteleone et al., 2017). The whole procedure of the "Italian Taste" project was approved by the Ethical Committee of the IRCCS Burlo Garofolo Children Hospital of Trieste (Italy). The present study complies with the Declaration of Helsinki for Medical Research involving Human Subjects.Subjects had no history of disorders of oral perception. Written informed consent was obtained from each subject prior the experiment.

2.2 Acquisition of tongue images

Participants were asked to rinse their mouth before the beginning of the test. Subjects were seated with the tongue held by a holder. The anterior portion of the dorsal surface of the tongue was swabbed with household blue food coloring (F.lli Rebecchi), using a cotton-tipped applicator. Pictures of the tongue were recorded using a portable USB digital microscope (2.0 mega pixels' image sensor, MicroCapture version 2.0 bundle software, 20x to 400x magnification ratio)(Masi et al., 2015). Pictures captured both the anterior part of the tongue

and a ruler fixed behind the tongue which provided a spatial calibration. The picture acquisition had a duration of around 5-10 minutes per subject. From each picture a rectangle (400 x 200 pixels, area=1.125 cm^2), orthogonal to the median line and located 0.5 cm from the tongue tip, was selected. The selection was saved as image in JPG format (96 dpi) using the ImageJ software (ver. 1.50i, National Institutes of Health, USA). The selected area was chosen as representative of FPD on the whole tongue (Shahbake et al., 2005; Correa et al., 2013).

2.3 Manual count

Tongue images were modified with ImageJ (Color Inspector 3D plugin: saturation= x2.49, brightness=-23.0) to make the visual count easier. Two operators, blind to any data concerning subjects, trained according to the Denver Protocol (Nuessle et al., 2015) and with 1-year experience, independently counted FP. The counts from the two operators were submitted to one-way fixed ANOVA. Counts were considered valid if the operator effect was not significant ($p>0.05$). The mean FP number from valid counts was used for each image and expressed as density (FP/cm^2- FPD).

2.4 Automated count

A script was developed with the software Matlab (Mathsworks, U.S., ver. R2015a) (Appendix) based on the procedure used by Kraggerud and colleagues 2009 (Kraggerud et al., 2009). The script analyzed the image of each subject (Fig. 1a) in three automated steps: 1. correction of the background variation and graphical emphasis of the elevated structures providing an image with black background and white spots (Fig. 1b); 2. identification of circular-like elements amongst the white spots (Fig. 1c); 3. computing the frequency of circular-like elements in classes with varied Diameter Size (DS) (Fig. 1d). The script was set

up to return 11 classes in the range from 8 to 28 pixels (0.30-1.05 mm: DS 1=0.30-0.36, DS 2=0.37-0.43, DS 3=0.44-0.49, DS 4=0.50-0.56, DS 5= 0.57-0.63, DS 6= 0.64-0.70, DS 7= 0.71-0.77, DS 8= 0.78-0.84, DS 9= 0.85-0.91, DS 10= 0.92-0.98, DS 11= 0.99-1.05). The 11 DS classes covered a diameter's range slightly larger than the average variation of FP size (Segovia et al., 2002).

FIGURE 1

2.5 Sensory evaluations

Five water solutions, corresponding to five basic tastes, were rated for intensity. The concentration of the tastants was selected in order to obtain solutions equivalent to moderate/strong on a generalized Labelled Magnitude Scale-gLMS (sourness: 4.0 g/kg of citric acid, bitterness 3.0 g/kg caffeine, sweetness 200.0 g/kg sucrose, saltiness: 15.0 g/kg sodium chloride, umami 10.0 g/kg monosodium glutamate) (Monteleone et al., 2017). Subjects were trained to the use of gLMS (0: no sensation-100: the strongest imaginable sensation of any kind) following published standard procedure (Green et al., 1996; Bartoshuk, 2000). Subjects are instructed to treat the ‘‘strongest imaginable sensation" as the most intense sensation they can imagine that involves remembered/imagined sensations in any sensory modality. Water solutions (10 mL) were presented in 80cc plastic cups identified by a 3-digit code. Subjects were presented with a set consisting of the five water solutions. The presentation order of water solutions was randomized across subjects. Subjects were instructed to hold the whole water solution sample in their mouth for 10 s, then expectorate and evaluate the intensity of relevant target sensation on gLMS. After each sample, subjects rinsed their mouths with distilled water for 30 s had some plain crackers for 30 s and rinsed their mouths with water for a further 30 s. Evaluations were performed in individual booths under white lights. Data were collected with the software Fizz (ver.2.47.B, Biosystemes, Couternon, France).

2.6 Data analysis

The normality assumption of the FPD distributions from manual count (FPDm) and predicted from automated image analysis (FPDp) was tested by the Shapiro–Wilk W test (α=0.05) and by Pearson skewness test. The two distributions were compared with Kolmogorov-Smirnov test (α=0.05).

ANCOVA using Type III sum of square was performed to assess gender and age effects on FPDm and FPDp, independently (significant for $p \leq 0.05$).

Principal Component Analysis (PCA) was computed on frequencies of the 11 DS of each image. FPDm was included as supplementary variable. A visually oriented approach, based on the inspection of correlation loading plot, was used for grouping images and Y-axis was set as limit (Næs et al., 2010). The distribution along the PC2 of images on the left and on the right of the map was described by the box plots of their coordinate on the PC2.

A Partial Least Squares Regression (PLSR) model (full cross validation, Kernel Algorithm, 100 interactions) was applied to predict the FPD from the image analysis output, using the DS classes as explanatory variables (X) and the FPD from manual count as dependent variable(Y). In order to test the model, the image data set was split into a calibration (n=100) and a prediction (n=33) set. The observations for the prediction set were systematically selected to fully cover the FPDm variation across images. Three outliers were removed from the original calibration set, due their high residuals (2 observations) or high leverage value (1 sample). The model was full cross validated on 97 samples and then applied to the prediction set.

Images were split in low (L) and high (H) FPD according to the median of the FPDm and FPDp data sets. Two group of subjects were identified in each data set: L-FPDm (≤ FPDm

median value) and H-FPDm (>FPDm median value); L-FPDp (≤ FPDp median value) and H-FPDp (>FPDp median value).

Unpaired t-tests (significant for $p \leq 0.05$) were used to compare intensity ratings from Low-FPDm to Low-FPDp, and from High-FPDm to High-FPDp, for each stimulus.

ANCOVA models using Type III sum of square with FPD variation as main factor (2 levels: H and L) and age as covariate were applied on intensity ratings, for each stimulus independently (significant for $p \leq 0.05$).

H-FPDp subjects were categorized as mainly associated to DS with smaller diameter (DS 1-4) and mainly associated to DS with larger diameter (DS 7-11) based on the characteristic values of the percentile distribution of their coordinate values on PC2 (Small Size ≤first tertile; Large Size ≥ second tertile). Unpaired t-tests (significant for $p \leq 0.05$) were used to compare intensity ratings from Small Size to Large Size subjects.

All data analysis were performed with XLStat 2016.05 (Addinsoft). PLSR model was computed using The Unscrambler ® (ver. 10.4 – © 2016 CAMO Software AS, Oslo Norway).

3.Results

3.1 Manual count

The manual count had an error of 2.3 FPD, measured as mean of standard deviations given by the two operators for each image. The distribution of FPD from manual count (FPDm) across the 133 subjects tended to a normal distribution ($W=0.968$; $p=0.004$) with data skewed to the right (Fig. 2a).

FIGURE 2

Descriptive statistic of FPDm is reported in Tab.1, with a mean value of 37.2 and limits of the percentile distribution for 1st and 3rd quartile of 23.1 and 46.2, respectively. No significant effect of gender on FPDm was found ($F=1.13$; $p=0.29$); FPDm significantly decreased with

aging (F=16.53, p<0.0001). No significant interaction gender*age were found (F=1.49; p=0.22).

TABLE 1

3.2 Image analysis output

Similarities and differences among images in frequencies of DS classes are visualized in the correlation loading plot from PCA (Fig. 3). The first two principal components accounted for 66.9% of the total variability (PC1 contributing with 46.5%). Tongue images were evenly spread across the bi-dimensional space. Image positioning along the first component was positively associated to the increase of frequencies of all DS classes. PC2 contributed to separate images according to the size of the classes. Images positioned on the bottom of the bi-dimensional space were mainly associated to the smaller size DS classes (DS 1-5, 0.30 to 0.63 mm) while images positioned on the top of the map were associated to the larger size DS classes (DS 7-11, 0.71 to 1.05 mm).

FIGURE 3

The projection of FPDm on the map indicated a positive association to PC1, thus tongue images positioned on the left were characterized by a lower FPDm than images positioned on the right. The map visual inspection indicated that images positioned on the right were more spread along the PC2 than images on the left, thus indicating a wider diameter variation (Fig. 4).

FIGURE 4

Four image groups were tentatively identified according to their position on the map (Fig. 5): group 1 (left-top) negatively related to both FPDm and frequencies of DS classes and mainly associated to DS classes with the large diameter, group 2 (right-top) positively associated to both FPDm and frequencies of DS classes and mainly associated to DS classes with large diameter; group 3 (right-bottom) positively associated to both FPDm and frequencies of DS

classes and mainly associated to DS classes with small diameter; group 4 (left-bottom) negatively associated to FPDm and frequencies of DS classes and mainly associated to DS classes with small diameter.

FIGURE 5

3.3 Prediction of FPD from automated analysis output

The PLSR was full-cross validated. The calibration (RMSEC) and cross-validation (RMSECV) errors were respectively 12.4 and 13.9 FPD. Calibration and validation R values were 0.7 and 0.6, respectively. The first PLSR component explained 46% of the X variables (DS frequencies) and 31% of the Y variable (FPDm). The second PLSR component explained 8% of the X variables and 14% of the Y variable. The first PLSR dimension separated observations based on the frequencies of DS classes. The opposition of DS 5-7 versus DS 1-4 was responsible for sample separation along the second dimension. The regression of predicted versus manually counted FPD for the validation of the training model is shown in Figure 6.

FIGURE 6

To test the model's predictive ability, the model was run on the prediction set, showing an error of prediction (RMSEP) of 13.9 FPD, in line with that found in cross-validation.

The distribution of predicted FPD (FPDp) across the 130 subjects followed a normal distribution (W=0.99; p=0.46) (Fig. 2b). Descriptive statistic of FPDp is reported in Tab.1, with a mean value of 37.1 and limits of the percentile distribution for 1st and 3rd quartile of 29.6 and 44.9, respectively. No significant differences were found between distributions from manual and automated count (D=0.15; p=0.12). No significant effect of gender on FPDp was found (F=1.99; p=0.16); FPDp significantly decreased with aging (F=5.52, p<0.02). No significant interaction gender*age was found (F=2.28; p=0.13).

3.4 Comparison between counted and predicted FPD as indicators for taste functions

Taste solutions were all rated almost at strong intensity on the gLMS (mean value and standard error: sourness 31.2±1.7; bitterness 31.1±1.8; sweetness 40.1 ±1.5; saltiness 35.6±1.8; umami 30.0±1.8).

Ratings by subjects grouped as L and H according to the median of manually counted (L-FPDm from 3.6 to 37.3, n=68; H-FPDm from 38.0-101.3, n=65) and predicted FPD (L-FPDp from 11.8 to 38.1, n= 66; H-FPDp from 39.0 to 68.4, n=64) were independently compared. No significant intensity differences were found comparing L-FPDm to L-FPDp ($p \geq 0.63$) and H-FPDm to H- FPDp ($p \geq 0.54$).

The effect of FPD variation on perceived taste intensity was assessed comparing ratings from L and H groups. A significant effect of FPD variation was found for saltiness ratings. L-FPD rated saltiness higher than H-FPD (L vs H FPDm: $F=4.50$; $p=0.03$; L vs H FPDp: $F=6.46$; $p=0.01$). No significant effect of FPD variation was found on perceived intensity of sourness, bitterness, sweetness, and umami ($p \geq 0.218$). Age did not significantly influence taste ratings ($p \geq 0.140$).

The effect of variation in FP size on the perceived taste intensity was assessed within H-FPDp group. H-FPDp subjects with small size FP (coordinate value on PC2$\leq$-0.884; n=16) tended to rated intensity of taste solutions significantly higher than subjects with large size FP (coordinate value on PC2 $\geq$0.418; n=17) ($t_{163;197}=1.85$; $p=0.06$).

4. Discussion

In the present study, a novel automated procedure for FPD estimation based on the analysis of tongue pictures taken with a digital microscope is described. Results from automated image analysis were compared to those from manual count taken as reference.

The FPDm distribution across observations tended to a normal distribution (Segovia et al., 2002; Zhang et al., 2009; Webb et al., 2015). The mean was similar to values reported in studies using analogous counting procedures on the same portion of the tongue (Segovia et al., 2002; Shahbake et al., 2005; Correa et al., 2013; Feeney and Hayes, 2014a; Webb et al., 2015). Aging confirms as negative predictor of papillae density (Correa et al., 2013; Fischer et al., 2013; Pavlidis et al., 2013). No effect of sex on FPD was found, in agreement with studies performed on similar sample size and females/males ratio (Bajec and Pickering, 2008; Feeney and Hayes, 2014a). In general, results from manual count were in line with existing findings, thus supporting the reliability of the data set taken as reference.

The script used to analyse images identifies circular elements in a diameter ranging from 0.30 to 1.05 mm and covers the expected variation of fungiform papillae diameter on tongue of adults (Essick et al., 2003). PCA confirmed the positive association between the number of circular elements and the papillae density assessed by manual count. The association to classes of circular elements with varied diameters contributed to discriminate amongst tongue images. The variation of diameter size was more evident in images associated to high than low papillae density. Automated analysis outputs allowed a tentative visual image classification based on the variation of both density and size of fungiform papillae.

Automated image analysis output was significantly related to papillae density variation. The predictive model explained 60% of variance among images.

The images used to build the predictive model can be considered as representative of field experimental data set since no inclusion criteria were adopted for the picture clarity and uniformity of tongue blue coloring. The only condition was that the two operators independently agreed on the papillae count. Thus, despite a prediction error of 13.9 FPD, the reliability of the model is considered encouraging.

In general, results from predicted papillae density matched those from manual count. The influence of the population demographics (age and sex) on the variation of papillae density predicted by the model was coherent with findings observed on data from manual count. Predicted values showed a normal distribution as expected for the variation of papillae density across adult individuals and superimposed the distribution of data from manual count. Median, mean values and limits of percentile distribution are widely used to categorize subjects as Low and High papillae density in studies aimed to investigate the relationships between papillae density and taste functions (Hayes and Duffy, 2008; Bakke and Vickers, 2011; Masi et al., 2015). Descriptive statistics values of FPDm and FPDp were in good agreement thus providing very similar subject segmentation according to FPD variation. The consistency in subject classification was further highlighted by the same mean ratings for taste solutions observed in subject groups classified as Low or High papillae density according to the median value of counted and predicted FPD. The same results on the effect of FPD variation on taste perception were obtained both using predicted and counted values. FPD variation failed to explain perceived intensity of bitterness, sourness, sweetness and umami in line with recent studies (Fischer et al., 2013). Only the perception of saltiness intensity was significantly affected by the variation of papillae density. Subjects categorized as High FPD rated saltiness lower than subjects categorized as Low FPD both using the median of counted and predicted density. The influence of papillae density on the perceived intensity of saltiness from sodium chloride is still controversial. Fungiform papillae associated to heightened saltiness perception on the tongue tip (Miller and Reedy, 1990b; Doty et al., 2001) but may not explain whole mouth saltiness (Hayes et al., 2008). Hayes and co-workers (2010) already reported an inverse relationship between saltiness perception and papillae density in complex stimuli (Hayes et al., 2010). Intensity ratings from whole-mouth and regional stimulation are significantly correlated even if at varying extent for different

prototypical tastes (Feeney and Hayes, 2014b). The lack of uniformity in the procedures adopted for stimulation can be seen as a further reason for uncertainty of association between FPD and taste responsiveness in the existing literature. However, the variation of responsiveness to different tastes across different regions of the tongue is still controversial and other indices of oral responsiveness (e.g. thermal taste) appear to be involved in regional responsiveness (Cruz and Green, 2000). Intensity responses from whole-mouth stimulation are considered reliable proxy of the average individual oral responsiveness and still appear the most appropriate and ecological stimulation procedure in studies aimed at investigating association between food perception and preference (Törnwall et al., 2012; Monteleone et al., 2017). Investigating the relationships between FPD variation and taste functioning is behind the aim of the present study. The study rather focuses on the comparison between methods. The proposed automated image analysis of tongue pictures appears a reliable substitute for manual counting when the purpose is subject classification according the papillae density.

It is worthy to note that the proposed automated analysis allowed an explorative analysis on the role of papillae size in taste function. High papillae density seemed to be associated to a wider size variation. Subjects with small size papillae perceived higher taste intensity than large size subjects. This result need to be further confirmed in a larger size population. The variation of papillae functionality according to diameter supports the hypothesis that size other than density is a relevant feature for oral chemosensory acuity. Small papillae diameter has been positively related to tongue tactile acuity (Essick et al., 2003), PROP responsiveness and gustin expression (Melis et al., 2013). Thus, the variation in papillae functionality according to their size might be a further bias impairing investigations on the association between papillae density and perceived taste intensity. The use of automated analysis with the possibility to estimate the size distribution may help to clarify these associations.

Some considerations can be done considering strengths and weaknesses of the presented method. The distortion degree has previously been suggested as potentially having an effect on taste function (Melis et al., 2013) and could further contribute to explain the association between FP density and taste perception. Other proposed methods for automated papillae detection make this measure available (Sanyal et al., 2016) while the script adopted in the present study did not. The possibility to include the detection of distortion degree in circular-like elements detection deserves further investigations. Moreover, the script may be further developed to handle unstained tongues, in order to eliminate this step which is somewhat annoying for subjects and to avoid technical issues due to the lack of background uniformity (Valencia et al., 2016). The number of observations higher than in the previous studies on methods alternative to manual counting (Sanyal et al., 2016; Valencia et al., 2016) represents one of strength points of the present study. Another positive aspect is that the area to be analysed can be easily changed (extended/reduced or moved) allowing to investigate different areas and improving reliability of the count as representative of the whole tongue. The developed approach is well suited for large field experiments, even involving different teams in different locations, for the following reasons: 1. the device for pictures acquisition is really inexpensive and can be afforded even by relatively small laboratories, 2. the script is not limited in the number of pictures that can be handled, 3. apart from the selection of the area to be analysed, the whole procedure is completely automated and takes a few seconds per picture, 4. image analysis can be easily centralized with a core team appointed for the image analysis, without overworking as in the case of manual count where several operators are needed. Further future applications could combine outputs from the proposed technique to in-vivo methods (e.g. video microscopy and confocal endomicroscopy) that allow the identification of taste pores or gustatory organs, to gain knowledge on associations between

papillae morphological characteristics (e.g. size and relevant distributions) and taste functionality.

5.Conclusions

The present paper describes a novel procedure to count fungiform papillae based on the automated analysis of tongue pictures. FPD predicted from automated analysis output are in good agreement with data from manual count. The proposed method appears a reliable and easy to handle substitute for manual counting when the purpose is subject classification according to FPD variation. The method fits the requirements of field researches aimed to investigate the relationships between FPD and taste functions in large size population studies. Furthermore, the new method makes available the estimation of the number of papillae for different diameter classes. Future research on larger sample would address the relevance of papillae size on taste functions.

6. Appendix

The Matlab script (1) and the additional *FindCircleFast* function (2)adopted in the present study are provided below. Both scripts are necessary to properly run the analysis. Scripts must be put in the same folder of images. To run the script, open it in Matlab and press run. The script will automatically stop at the end of operation and provide a table with the frequenciesof all RS for all subjects under the section "SizeHist". Frequencies values can be directly exported and used for the analysis.

1. Matlab script

```
Dr=dir('C:\..... \*.jpg');
[ant,dummy]=size(Dr);
texture=zeros(ant,200);
SizesHist=zeros(ant,11);
FileNames=struct2cell(Dr);
FileNames=FileNames(1,1:end);
Sizes=zeros(ant,2);
%%
i_fig = 1;
for K= 1:ant
   filename=[Dr(K).name];
   a=imread(filename,'jpg');
%a=imread('43 (2) contrast.jpg','jpg');
figure(i_fig), i_fig = i_fig + 1;
imagesc(a)
   title(filename)
   figure(i_fig), i_fig = i_fig + 1;  imagesc(a(:,:,1));
   a=a(:,:,1);
   D= imresize(a, [260 560]);
   figure(i_fig), i_fig = i_fig + 1;  imagesc(D);colormap('gray')

   D=double(D(:,:,1));

   background = imopen(D,strel('disk',15));
```

```
D2 = imsubtract(D,background);
   title(filename)
figure(i_fig), i_fig = i_fig + 1;
imagesc(D2)
   title(filename)
eval(['Im', num2str(K),'=D2;']);
D3=D2/max(max(D2));
   D3BW = im2bw(D3,0.3);
   title(filename)
   figure(i_fig), i_fig = i_fig + 1;
imagesc(D3BW)
eval(['ImBW', num2str(K),'=D3BW;']);
S=svd(D2);
   [L,d]=size(S);
   figure(i_fig), i_fig = i_fig + 1;hold on
title(filename)
   plot(log(S))
   texture(K,1:L)=log(S);
   [totVol, radHist] = findCirclesFast(D3BW, K );
title(filename)
   figure(i_fig), i_fig = i_fig + 1;bar(radHist)
   title(filename)
SizesHist(K,:)=radHist;
   pause(1)
end
figure(6);hold off
```

2. FindCircleFast function:

```
function [totVol, radHist] = findCircles(img, imgName)

% Correlation threshold for identification of holes
corrThres = 0.51;
rMin=4;rMax=14;
[M,N] = size(img);
corrMat = zeros(rMax,M,N);

% Calculate correlation images for each radius
for r = rMin:rMax
   circle = getnhood( strel('disk', r, 0) );
```

```
c = normxcorr2(circle, img);
corrMat(r,:,:) = c(r+1:end-r,r+1:end-r);
end

% Find pixels and corresponding radii with highest correlation
[maxCorr, maxRadius] = max(corrMat,[],1);
maxCorr = squeeze(maxCorr);
maxRadius = squeeze(maxRadius);

% Threshold max-correlation image and identify centroids
maxCorr(maxCorr<corrThres) = 0;
L = bwlabel(maxCorr);
s = regionprops(L, 'Centroid', 'Area');
if (numel(s) == 0)
errordlg('Beklager, ingen hull funnet')
totVol = 0;
radHist = zeros(1,rMax-rMin+1);
return
end
centroids = round(cat(1, s.Centroid));

% Calculate total hole-volume and distribution of hole-sizes
radii   = maxRadius(sub2ind(size(maxRadius), centroids(:,2), centroids(:,1)));
totVol  = sum( 4/3*pi*radii.^3 ) / 1000;
radHist = hist(radii, rMin:rMax);

% Optional plotting for debugging purposes
%if (opts.debugplot)
   figure(11)
imagesc(img), colormap(gray)
   hold on
%     plot(centroids(:,1), centroids(:,2), 'b*');
fori = 1:size(centroids,1)
drawCircle(centroids(i,1), centroids(i,2), radii(i), 20, 'r');
end
   hold off
   title(imgName, 'Interpreter', 'None')
%end
end
```

```
function h = drawCircle(x, y, r, nseg, S)

theta = 0 : (2 * pi / nseg) : (2 * pi);
pline_x = r * cos(theta) + x;
pline_y = r * sin(theta) + y;

h = plot(pline_x, pline_y, S, 'LineWidth', 2);
end
```

References

Bajec, M.R., and Pickering, G.J. 2008. Thermal taste, PROP responsiveness, and perception of oral sensations. Physiol Behav. 95:581–590.

Bakke, A., and Vickers, Z. 2011. Effects of bitterness, roughness, PROP taster status, and fungiform papillae density on bread acceptance. Food Qual Prefer. 22:317–325.

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

Correa, M., Hutchinson, I., Laing, D.G., and Jinks, A.L. 2013. Changes in Fungiform Papillae Density During Development in Humans. Chem Senses. 38:519–527.

Cruz, a, and Green, B.G. 2000. Thermal stimulation of taste. Nature. 403:889–892.

Delwiche, J.F., Buletic, Z., and Breslin, P. a S. 2001. Relationship of papillae number to bitter intensity of quinine and PROP within and between individuals. Physiol Behav. 74:329–337.

Doty, R.L., Bagla, R., Morgenson, M., and Mirza, N. 2001. NaCl thresholds : relationship to anterior tongue locus , area of stimulation , and number of fungiform papillae. 72:373–378.

Duffy, V.B., Davidson, A.C., Kidd, J.R., Kidd, K.K., Speed, W.C., Pakstis, A.J., Reed, D.R., Snyder, D.J., and Bartoshuk, L.M. 2004a. Bitter receptor gene (TAS2R38), 6-n-propylthiouracil (PROP) bitterness and alcohol intake. Alcohol Clin Exp Res. 28:1629–1637.

Duffy, V.B., Peterson, J.M., and Bartoshuk, L.M. 2004b. Associations between taste genetics, oral sensation and alcohol intake. Physiol Behav. 82:435–445.

Essick, G.K., Chopra, A., Guest, S., and McGlone, F. 2003. Lingual tactile acuity, taste perception, and the density and diameter of fungiform papillae in female subjects. Physiol Behav. 80:289–302.

Feeney, E.L., and Hayes, J.E. 2014a. Exploring associations between taste perception, oral anatomy and polymorphisms in the carbonic anhydrase (gustin) gene CA6. Physiol Behav. 128:148–154.

Feeney, E.L., and Hayes, J.E. 2014b. Regional Differences in Suprathreshold Intensity for Bitter and Umami Stimuli. Chemosens Percept. 147–157.

Fischer, M.E., Cruickshanks, K.J., Schubert, C.R., Pinto, A., Klein, R., Pankratz, N., Pankow, J.S., and Huang, G.H. 2013. Factors related to fungiform papillae density: The beaver dam offspring study. Chem Senses. 38:669–677.

Garneau, N.L., Nuessle, T.M., Sloan, M.M., Santorico, S. a, Coughlin, B.C., and Hayes, J.E. 2014. Crowdsourcing taste research: genetic and phenotypic predictors of bitter taste perception as a model. Front Integr Neurosci. 8:33.

Green, B.G., Dalton, P., Cowart, B., Shaffer, G., Rankin, K., and Higgins, J. 1996. Evaluating the "labeled magnitude scale" for measuring sensations of taste and smell. Chem Senses. 21:323–334.

Hayes, J.E., Bartoshuk, L.M., Kidd, J.R., and Duffy, V.B. 2008. Supertasting and PROP bitterness depends on more than the TAS2R38 gene. Chem Senses. 33:255–265.

Hayes, J.E., and Duffy, V.B. 2007. Revisiting sugar-fat mixtures: Sweetness and creaminess vary

with phenotypic markers of oral sensation. Chem Senses. 32:225–236.

Hayes, J.E., and Duffy, V.B. 2008. Oral sensory phenotype identifies level of sugar and fat required for maximal liking. Physiol Behav. 95:77–87.

Hayes, J.E., Sullivan, B.S., and Duffy, V.B. 2010. Explaining variability in sodium intake through oral sensory phenotype, salt sensation and liking. Physiol Behav. 100:369–380.

Jung, H.S., Akita, K., and Kim, J.Y. 2004. Spacing patterns on tongue surface-gustatory papilla. Int J Dev Biol. 48:157–161.

Kraggerud, H., Wold, J.P., HØy, M., and Abrahamsen, R.K. 2009. X-ray images for the control of eye formation in cheese. Int J Dairy Technol. 62:147–153.

Masi, C., Dinnella, C., Monteleone, E., and Prescott, J. 2015. The impact of individual variations in taste sensitivity on coffee perceptions and preferences. Physiol Behav. 138:219–226.

Melis, M., Atzori, E., Cabras, S., Zonza, A., Calò, C., Muroni, P., Nieddu, M., Padiglia, A., Sogos, V., Tepper, B.J., et al. 2013. The Gustin (CA6) Gene Polymorphism, rs2274333 (A/G), as a Mechanistic Link between PROP Tasting and Fungiform Taste Papilla Density and Maintenance. PLoS One. 8:1–15.

Miller, I., and Reedy, F. 1990a. Quantification of fungiform papillae and taste pores in living human subjects. Chem Senses. 15:281–294.

Miller, I.J., and Reedy, F.E. 1990b. Variations in human taste bud density and taste intensity perception. Physiol Behav. 47:1213–1219.

Monteleone, E., Spinelli, S., Dinnella, C., Endrizzi, I., Laureati, M., Pagliarini, E., Sinesio, F., Gasperi, F., Torri, L., Aprea, E., et al. 2017. Exploring influences on food choice in a large population sample: the Italian Taste Project. Food Qual Prefer.

Nachtsheim, R., and Schlich, E. 2013. The influence of 6-n-propylthiouracil bitterness, fungiform papilla count and saliva flow on the perception of pressure and fat. Food Qual Prefer. 29:137–145.

Næs, T., Brockhoff, P. B., Tomic, O. 2010. Statistics for sensory and consumer science. Chichester, UK: John Wiley and Sons.

Nuessle, T.M., Garneau, N.L., Sloan, M.M., and Santorico, S. a. 2015. Denver Papillae Protocol for Objective Analysis of Fungiform Papillae. J Vis Exp.

Pavlidis, P., Gouveris, H., Anogeianaki, A., Koutsonikolas, D., and Koblenz, K.K. 2013. Age-related Changes in Electrogustometry Thresholds, Tongue Tip Vascularization, Density, and Form of the Fungiform Papillae in Humans. Chem Senses. 38:35–43.

Prescott, J., Bartoshuk, L.M., Prutkin, J. 2004. PROP tasting and the perception of non-taste oral sensations. In: Prescott, J. & Tepper, B. (eds) Genetic Variation in Taste Sensitivity. NY: Marcel Dekker, pp. 89 – 104.

Sanyal, S., O'Brien, S.M., Hayes, J.E., and Feeney, E.L. 2016. TongueSim: Development of an Automated Method for Rapid Assessment of Fungiform Papillae Density for Taste Research. Chem Senses. 0:1–9.

Segovia, C., Hutchinson, I., Laing, D.G., and Jinks, A.L. 2002. A quantitative study of fungiform papillae and taste pore density in adults and children. Dev Brain Res. 138:135–146.

Shahbake, M., Hutchinson, I., Laing, D.G., and Jinks, A.L. 2005. Rapid quantitative assessment of fungiform papillae density in the human tongue. Brain Res. 1052:196–201.

Törnwall, O., Silventoinen, K., Kaprio, J., and Tuorila, H. 2012. Why do some like it hot? Genetic and environmental contributions to the pleasantness of oral pungency. Physiol Behav. 107:381–389.

Valencia, E., Rios, H. V., Verdalet, I., Hernández, J., Juárez, S., Herrera, R., and Silva, E.R. 2016. Automatic counting of fungiform papillae by shape using cross-correlation. Comput Biol Med. 76:168–172.

Webb, J., Bolhuis, D.P., Cicerale, S., Hayes, J.E., and Keast, R. 2015. The Relationships Between Common Measurements of Taste Function. 11–18.

Whitehead, M. C., Beeman, C. S., Kinsella, B. A. 1985. Distribution of taste and general sensory nerve endings in fungiform papillae of the hamster. Dev Dynam. 173(3):185-201.

Yackinous, C. a, and Guinard, J.-X. 2002. Relation between PROP (6-n-propylthiouracil) taster status, taste anatomy and dietary intake measures for young men and women. Appetite. 38:201–209.

Zhang, G.H., Zhang, H.Y., Wang, X.F., Zhan, Y.H., Deng, S.P., and Qin, Y.M. 2009. The relationship between fungiform papillae density and detection threshold for sucrose in the young males. Chem Senses. 34:93–99.

Captions

Fig. 1: Scheme of automated analysis steps operated by Matlab script.
DS= Diameter Size.

Fig. 2: Distribution and q-q-plots of papillae density from manual count (FPDm) and predicted from automated analysis outputs (FPDp).

Fig. 3: Bi-plot from Principal Component Analysis on frequency values of Diameter Size classes (DS 1-11) from 133 observations.
Papillae density from manual count (FPDm) is plotted as supplementary variable (dotted line).

Fig. 4: Box plots of coordinate on PC2 of images positioned on the left (L) and on the right (R) of the PCA. Median (line) and mean (cross) values.

Fig. 5: Images representative of 4 groups with varied FP density and diameter, according to the positioning on PCA: group 1 low density and large diameter; group 2 high density and large diameter; group 3 high density and small diameter; group 4 low density and small diameter. Arrows indicate the increase of the observed characteristics.

Fig. 6: Relationships between FPD from manual count (FPDm) and predicted by PLSR model from automated analysis output (FPDp). Model was build using 11 Diameter Size (DS) classes as explanatory variables (X) and the FPDm as dependent variable (Y).
RMSE= Root Mean Square Error

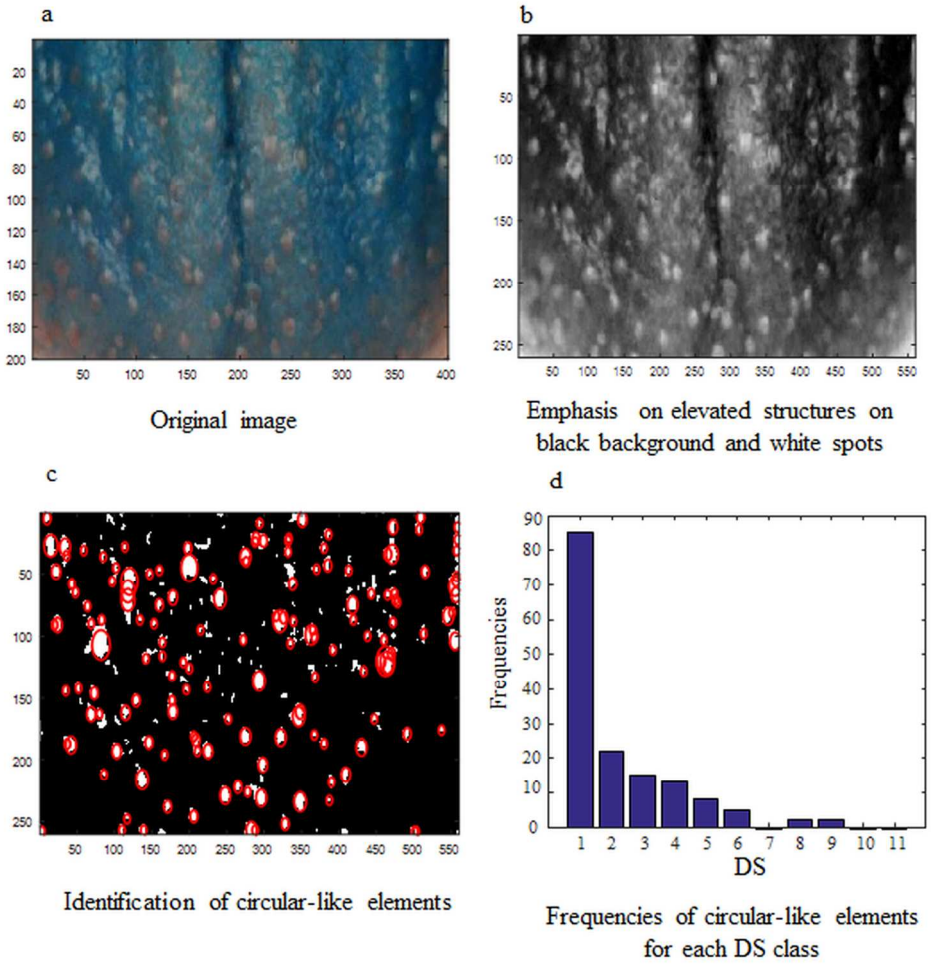

Fig. 1: Scheme of automated analysis steps operated by Matlab script.
DS= Diameter Size.

184x183mm (300 x 300 DPI)

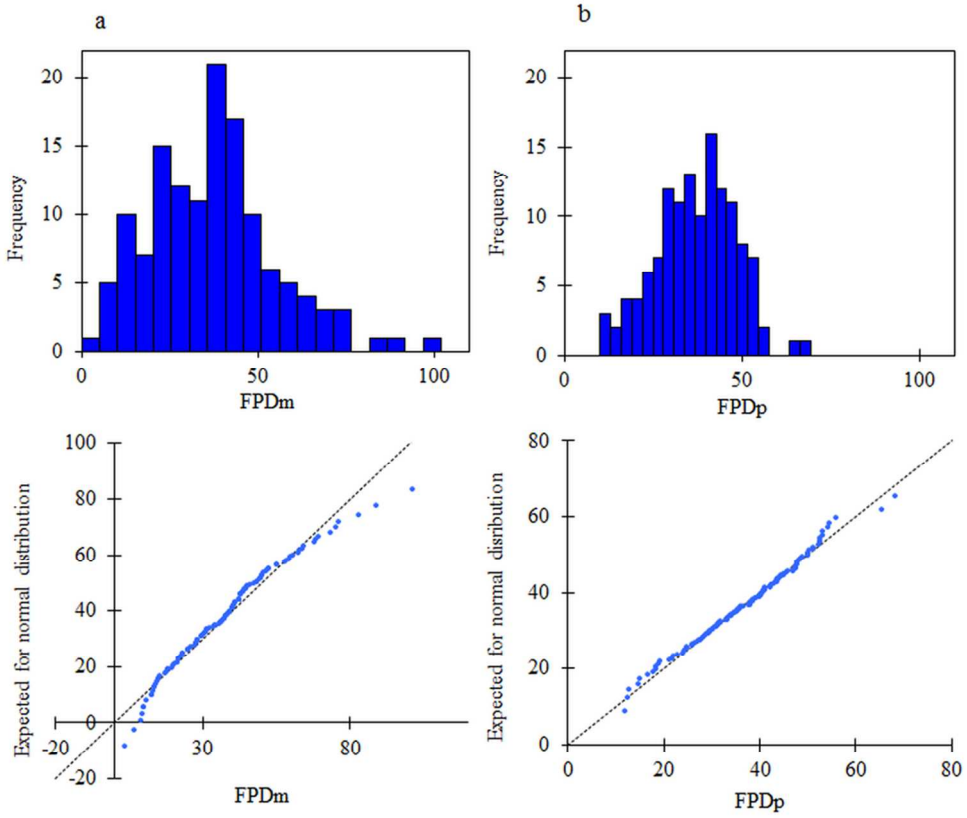

Fig. 2: Distribution and q-q-plots of papillae density from manual count (FPDm) and predicted from automated analysis outputs (FPDp).

88x76mm (300 x 300 DPI)

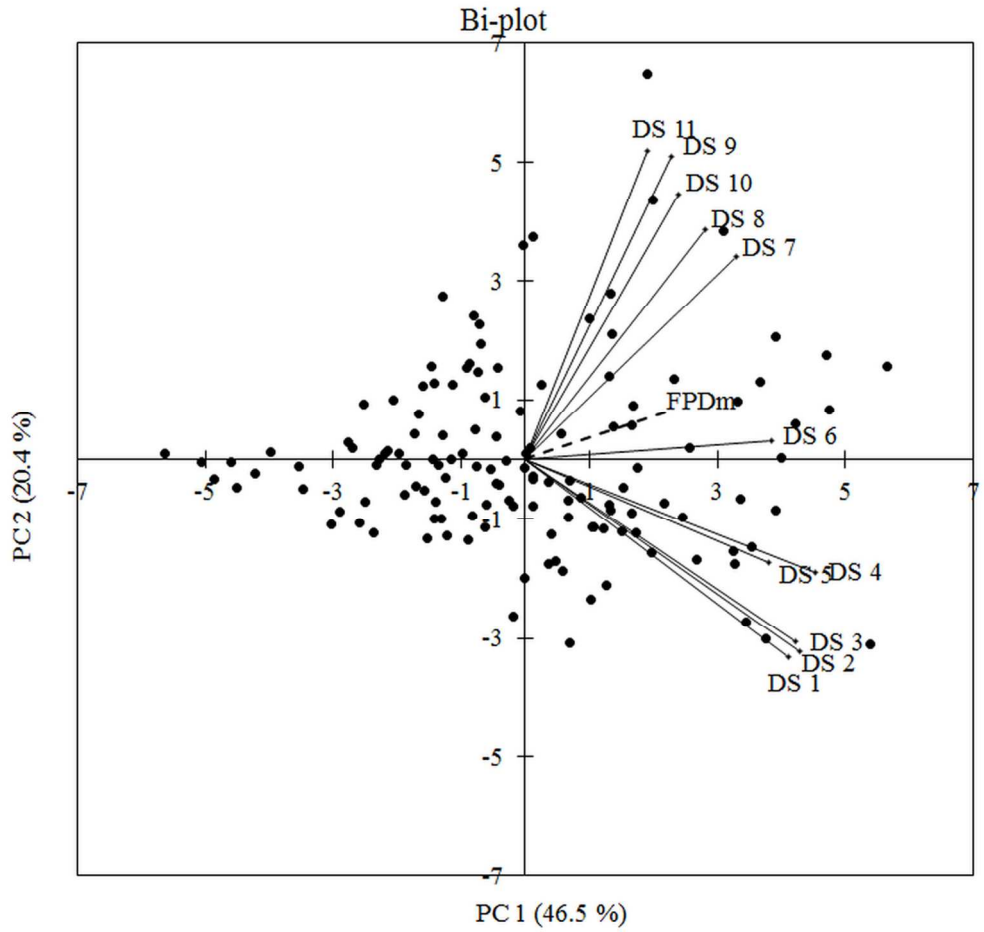

Fig. 3: Bi-plot from Principal Component Analysis on frequency values of Diameter Size classes (DS 1-11) from 133 observations. Papillae density from manual count (FPDm) is plotted as supplementary variable (dotted line).

184x177mm (300 x 300 DPI)

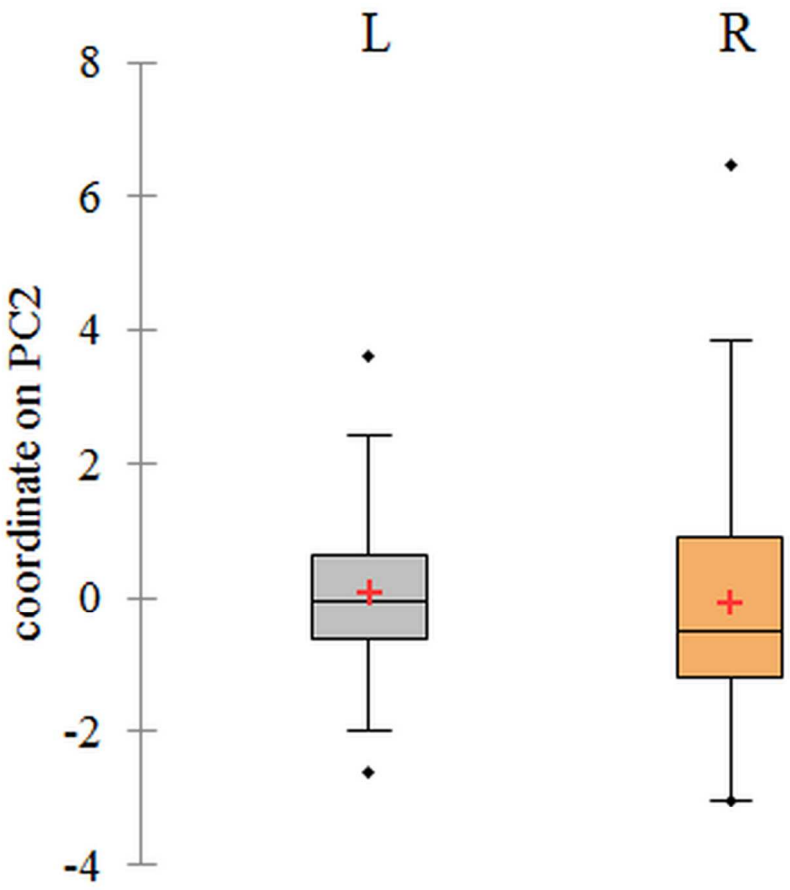

Fig. 4: Box plots of coordinate on PC2 of images positioned on the left (L) and on the right (R) of the PCA. Median (line) and mean (cross) values.

88x86mm (300 x 300 DPI)

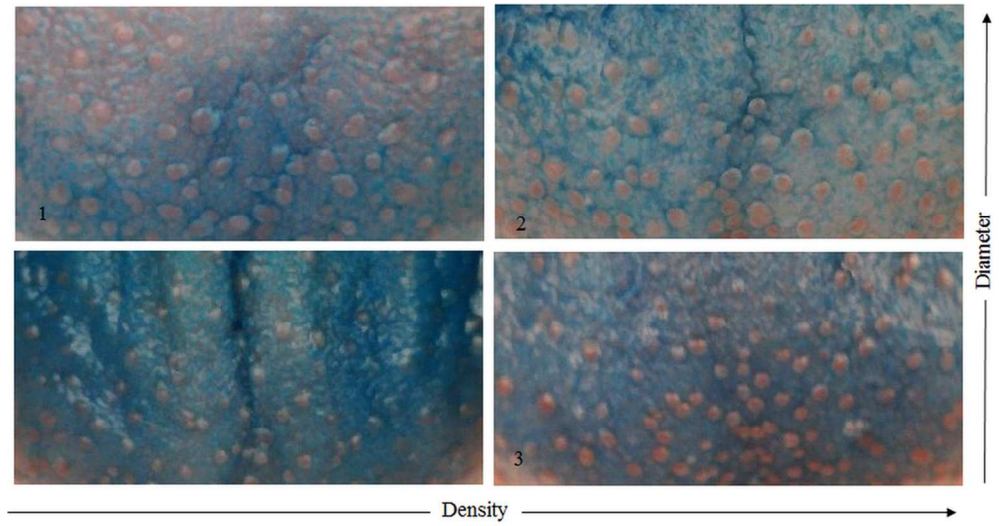

Fig. 5: Images representative of 4 groups with varied FP density and diameter, according to the positioning on PCA: group 1 low density and large diameter; group 2 high density and large diameter; group 3 high density and small diameter; group 4 low density and small diameter. Arrows indicate the increase of the observed characteristics.

184x98mm (300 x 300 DPI)

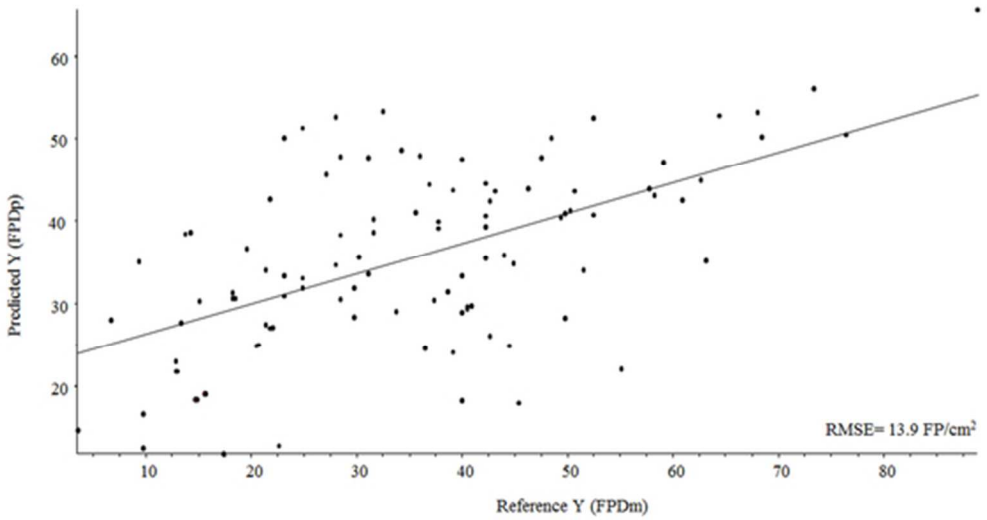

Fig. 6: Relationships between FPD from manual count (FPDm) and predicted by PLSR model from automated analysis output (FPDp). Model was build using 11 Diameter Size (DS) classes as explanatory variables (X) and the FPDm as dependent variable (Y).
RMSE= Root Mean Square Error

48x26mm (300 x 300 DPI)

Descriptive statistics	FPDm	FPDp
Observations (n)	133	130
Min	3.56	11.8
Max	101.33	68.4
1° Quartile	23.11	29.6
Median	37.33	38.1
3° Quartile	46.22	44.9
Mean	37.25	37.1
Standard deviation (n-1)	17.96	11.1