Multivariate statistical analysis of a large odorants database aimed at revealing similarities and links between odorants and odors

Anne Tromelin, Claire Chabanet, Karine Audouze, Florian Koensgen, Elisabeth Guichard

Multivariate statistical analysis of a large odorants database aimed at revealing similarities and links between odorants and odors

Anne Tromelin[1]*, Claire Chabanet[1], Karine Audouze[2], Florian Koensgen[1], Elisabeth Guichard[1]

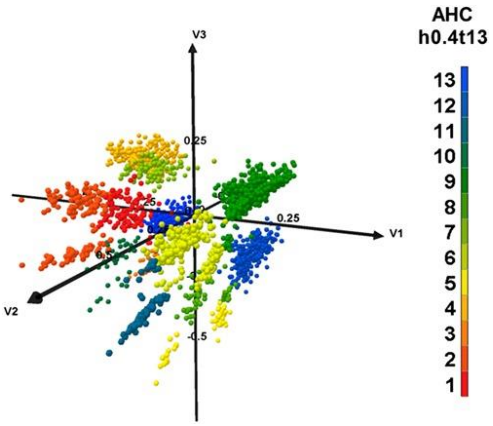
[1] UMR CSGA: CNRS, INRA, Université de Bourgogne Franche-Comté, 21000 Dijon, France

[2] Université Paris Diderot, MT*i*, Sorbonne Paris Cité, INSERM UMR-S 973, 75013 Paris, France

Keywords: odorants; odor notes; categorization; agglomerative hierarchical clustering; Kohonen self-organizing maps; multidimensional scaling

This is a pre-copy-edited, author-produced PDF of an article published in *Flavour and Fragrance Journal* 2018 33(1): 106-126 (doi/10.1002/ffj.3430).

Article first published online 07 Nov 2017



Categorizations of odorants and odor notes have been performed by Agglomerative Hierarchical Clustering and Kohonen's Self Organizing Maps. The analysis of results highlights several subsets of odorants that share similar groups of odor notes.

https://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

Correspondence to be sent to: Anne Tromelin, UMR CSGA: CNRS, INRA, Université de Bourgogne Franche-Comté, 21000 Dijon, France. email: anne.tromelin@inra.fr

Abstract

The perception of odor is an important component of smell; the first step of odor detection, and the discrimination of structurally diverse odorants depends on their interactions with olfactory receptors (ORs). Indeed, the perception of an odor's quality results from a combinatorial coding, in which the deciphering remains a major challenge.

Several studies have successfully established links between odors and odorants by categorizing and classifying data. Hence, the categorization of odors appears to be a promising way to manage odors.

In the proposed study, we performed a computational analysis using odor descriptions of the odorants present in Flavor-Base 9th Edition (2013). We converted the Flavor-Base data into a binary matrix (1 when the odor note appears in the odor description, 0 otherwise). We retained 251 odor notes and 3508 odorants, considering only the orthonasal perception. Two categorization methods were performed: agglomerative hierarchical clustering (AHC), and self-organizing map (SOM). AHC was based on a measure of the distance between the elements performed by multidimensional scaling (MDS) for the odorants, and correspondence analysis (CA) for the odor notes.

The results demonstrated that the SOM classes appeared to be less dependent on the frequency of the odor notes than those of the AHC clusters. SOMs are especially useful for identifying the associations between less than 4 or 5 odor notes within groups of odorants.

The obtained results highlight subsets of odorants sharing similar groups of odor notes, suggesting an interesting and promising way of using computational approaches to help decipher olfactory coding.

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

Introduction

Odor perception is an important component of the sense of smell, which results from an intricate phenomenon involving an ensemble of several perceptions [1-4]. There are two ways to detect the odorant stimuli: retronasal olfaction, in which the odorants are released in the nose from the mouth via the nasopharynx, and orthonasal olfaction, in which the route of odorants is the external environment. Retronasal olfaction refers to *flavor* perception, while *odor* perception relates to orthonasal olfaction.

In both cases, the first step of odor detection and the discrimination of structurally diverse odorants depends on their interactions with the olfactory receptors (ORs) in the nose [5], whereas the perception of an odor's quality results from a combinatorial coding [6], in which the deciphering remains poorly understood. The identification of the links between the chemical structure of odorants and the perceived odor constitutes an interesting challenge [7]. Nevertheless, making that link is difficult due to the diversity [8]and the flexibility of molecules [9] that can potentially be odorants [10]. Moreover, numerous steps are involved between OR activation and perception at higher levels by the brain, which make the olfactory signal too complex to be able to analytically link the structure of odorants to their perceived qualities [11]. Therefore, establishing such links can appear as an impossible goal [12].

Nevertheless, several structure-odor relationships have been successfully established, e.g., for rigid odorants [13-17], either to discriminate hedonic qualities [18] or to characterize the odor similarity of odorant mixtures [19]. A crucial point is that the use of sophisticated molecular descriptors is necessary to develop a metric for odorants [20].

In fact, most studies have applied categorization and classification [21,22]. Establishing a link between the chemical structure of odorants and their odors requires that the odors be clearly defined, which is also a challenge [23] due to the difficulty associated with identifying and describing odors [24-26].

The issue of organizing odors into categories underlies the controversial concept of "primary odors". This concept was first formulated on the basis of anosmia [27,28] but was then contested due to the numerous ORs [29]. Thereafter, some studies have disagreed with the concept of "primary odors" [30]; nevertheless, other studies have continued to support this idea by promoting classification into a limited number of odor classes, in which the number of "primary odors", or odor groups, is several tens [27,28,31,32].

The categorization of odors appears to be a promising way to manage odors [33,34]. Indeed, rather than a *chemotopic* organization, *tunotopic* (or odotopic) organization [35] has been observed at several levels of olfactory signal processing, such as in the olfactory bulb [36,37] and the brain [38-40].

Numerous studies have been performed on various sets of odors and odorants using diverse visualization and dimensionality reduction methods [22,30]. Principal component analysis (PCA) has been used in many studies [18,31,32,41-43]. Several distance-based methods have been applied, e.g., the multidimensional scaling (MDS) approach, which uses a distance representation by performing a nonlinear projection into a space of lower dimension [43-47]. Several clusterizations of odorants and odors have been performed by agglomerative hierarchical clustering (AHC) [43,48-50]. Pairwise calculation [32,43] or a co-occurrence square matrix [48,50] of the odor notes have been used to examine their associations and their categorization by AHC and/or by correspondence analysis (CA) [50]. CA offers the remarkable feature of jointly

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

representing individuals and categorical variables using Euclidean distances. As a result of such analyses, not only does one gain insight into the relationships among individuals and amongst variables but can also determine which variables are important in the description of individuals. Additionally, a non-linear classification method, i.e., self-organizing maps (SOMs), which belongs to the unsupervised Artificial Neural Network (ANN) approaches, is useful for visualizing data and reducing the analysis space; this technique has been used in several cases [47,51,52].

Previous studies have commonly used datasets of several dozen to several hundred odorants [22] and some tens to 150 odor notes, which are relatively few compared to the vast number of odorants. Larger sets of approximately 1000 to more than 2000 odorants have also been used [43,47,48], although the number of odor notes used in the statistical analysis has differed greatly, ranging from 47 [48] to 199 [43], while Mamlouk, Chee-Ruiter, Hofmann, Bower [47] used 851 odorants and 278 odor notes.

Furthermore, the use of large datasets is highly suitable to have a sample as representative as possible of the odorant space [33]. Thus, we aimed to perform an analysis of a large dataset of odorants described by several hundred odor notes. We selected the Flavor-Base [53], which is one of the largest collections of natural and synthetic aroma compounds, and an earlier version has been used in other works [43]. The odor and flavor descriptions reported in the Flavor-Base are based on bibliographic documents in which existing variations in organoleptic descriptions have been considered and used in previous studies [54,55].

We focused as much as possible on orthonasal perception of odors and did not account for the description of flavors. The retronasal perception of numerous odorants differs from their orthonasal perception [3,56], and some hypotheses have been proposed to explain this difference, which remains unclear [57-59].

After conversion of the Flavor-Base dataset into a binary matrix [43,48], the data were explored using statistical methods currently used in chemoinformatics [60,61] to explore odorants and odors spaces [22,30]. Due to the fuzzy nature of semantic odor descriptions [30,51] we aimed to categorize the odorants and odor notes by implementing various methods. We used three dimensionality reduction and visualization methods (MDS, HC, CA) and one classification method (SOM). This choice was based on the notion that the implementation of various categorization approaches could lead to various odorant and odor note patterns. Our assumption was that the comparison of these patterns could provide multiple means of exploration of odorants space, hoping that it would improve the enigmatic understanding of the relationships between the structure of odorants and odor perception.

Material and Methods

Data preparation

We used the 9th version of Flavor-Base [53], which is one of the largest collections of flavor and taste compounds, flavor and taste enhancers, and additives (4226 records) as a source of molecule, odor, flavor and taste descriptions. We selected the molecules with clear descriptions and identified all the odor, flavor and taste notes related to this set of molecules. Finally, we performed a final selection of odorant molecules (odorants) and odor notes limited to orthonasal perception and based on odor note frequency.

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

Odorants

Not all of the compounds described in Flavor-Base are used as flavor or taste odorants. Some are odorless and used as additives, solvents, emulsifiers, preservative agents, or used as reaction flavors, taste or flavor enhancers or modifiers (181 compounds). Moreover, no description is reported for 139 compounds, and their "Flavor description" refers to the "Comments" field. In most cases, the "Comments" field mentions "*The author is unfamiliar with this material…*"; we did not consider these compounds for analysis, even when a brief description was provided in the "Comments" field.

Consequently, the first step was to identify and remove the compounds without an intrinsic odor, flavor or taste quality as well as odorants described as "odorless" and "tasteless". We manually examined the cases of odorant mixtures to eliminate them.

We identified and excluded 451 compounds that meet at least one of these exclusion criteria, leading to a final subset of 3775 odorants with a clear odor, flavor, aroma, and/or taste description. The description of two of these 3775 odorants distinguishes two isomers (Dihydrocarvone: "*Minty rye like odor & taste for (-) isomers; spearmint for (+) isomers*" and Dihydroterpineol: "*Floral lilac (trans isomer); terpineol pine (cis isomer)*"). Therefore, we duplicated each odorant to separate the description of each of the isomers, resulting in a final dataset of 3777 odorants.

Odor, flavor, aroma and taste notes

The first identification of odors, flavors, aromas and taste notes was conducted on the final dataset. In a previous analysis of the Flavor-Base [43], the sensory lexicon of flavor descriptors and selected odorants registered in the FEMA GRAS™ Flavoring Substances List were used. We chose to offer another perspective and considered all odor notes used in the Description field (regardless of FEMA GRAS™ List).

Establishing the list of odor, flavor, aroma and taste descriptors is a difficult and challenging step. Because the descriptions are given as sentences, we first suppressed linking terms (e.g., "and" and "with"). Some other words do not refer to an odor description but provide information about the frequency of attribution of a note to an odorant (e.g., "somewhat", "reminiscent", connotations", and "like"). Such terms reflect the well-known subjective part of sensory perception [25,62]. Because we excluded the subjective role of memory in our analysis, all reported notes in the description were regarded at the same level. Similarly, the terms related to intensity were identified and excluded from the notes list [32,42,48].

We did not consider the intensities or the change in the odors according the concentrations. For example, butyl sulfide is described "*Green, onion-garlic-horseradish; violet green floral in dilution*" and we used the seven odors notes to create the binary matrix.

In the cases where the description related to the concentration of a molecule differs between the "Flavor Description" and the "Comments" fields, we chose to use only the text reported in the field "Description". For example, the "Flavor Description" of anisylidene acetone is "*In dilution sweet, floral and creamy*", while "Comment" indicates that "*Arctander describes this as "slightly pungent, but in dilution sweet, floral and creamy...*" and "*Boelens indicates: aromatic, somewhat fruity, slightly raspberry-like*". Accordingly to our choice, we use the odor notes "sweet", "floral" and "creamy" for the present analysis.

Another critical step was to consider the complexity of notes associated with several terms (e.g., "roasted meat", "unripe fruit", and "melon rind"). After final orthography homogenization, 782 notes were identified that referred to odor, flavor, aroma or taste.

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

Selection of odorants and odor notes intended for the analysis

Because the proposed study relates only to odorants, the molecules inducing only a taste and/or a trigeminal perception (177 molecules) were excluded. Moreover, we aimed to focus on orthonasal perception. Thus, we distinguished between the description related to odor (orthonasal stimulus) from the flavor or aroma (retronasal stimulus). Note that several "tastes" described in the Flavor-Base are probably related more to flavor and aroma than true tastes involving taste receptor activation in the mouth; consequently, we considered as flavors. We considered that the absence of precision (no indication of "odor", "flavor", "aroma", or "taste" in the description) meant that the perceived notes did not differ according to the mode of stimulus perception and, consequently, could be considered as orthonasal stimuli.

This procedure allowed us to retain 3571 odorants perceived as orthonasal stimuli and 725 odor notes. According to previous studies [43,48], we used these odor notes as fingerprints to create a matrix containing the 3571 odorants. Each element of the matrix was converted into binary values: 1 when the odor note appeared in the odor description, 0 otherwise.

Using this matrix, we counted the occurrences of the odor notes. According to Zarzo and Stanton, we did not consider the odor notes having fewer than five occurrences [42]. We identified 474 odor notes with fewer than 5 occurrences (282 odor notes with 1 occurrence, 110 odor notes with 2 occurrences, 45 odor notes with 3 occurrences, and 37 odor notes with 4 occurrences) (Figure S1). After removing the odor notes with fewer than 5 occurrences, we decided not to maintain the odorants described solely by "fruity" and/or "floral" notes because of the limited precision of these odor notes when used alone. This reduced the number of odorants by 63 odorants, with 3508 odorants remaining for the analysis. The database used in this work is called "FB9-3508" in the text.

We primarily focused on the four most frequent notes: "fruity", "floral", "sweet" and "green".

The terms "only fruity", "only floral", "only sweet" and "only green" designate odorants whose description included only one of these four most frequent notes. The terms "fruity-floral", "fruity-sweet", "floral-sweet", "fruity-floral-green", etc., indicate that several of the four most frequent notes were present together in the odor description of odorant(s). Conversely, the terms "fruity/floral", "floral/sweet", etc., refer to the presence of at least one of these notes in the odor description.

Additionally, we considered two other odor notes: "sulfuraceous", which has little association with "fruity" and "floral", and "rose", which is contrariwise frequently associated to these two frequent odor notes; additionally, we considered several other odor notes referring to various perceptual categories ("fatty", "herbaceous", "waxy", "woody", "balsamic", "onion", "phenolic", etc.).

Computational analysis and statistical methods and tools

Descriptive statistics were performed using XLStat (Addinsoft), and 3D graphical visualization was obtained using Miner3D (version 7).

Methods

Multidimensional scaling (MDS), agglomerative hierarchical clustering (AHC), multiple correspondence analysis (MCA), and Self-Organizing Map (SOM) [63-65] are methods based on the distance between objects; different metrics can be used, such as the Euclidean distance or Manhattan distance.

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

The calculations were conducted using R version 3.0.1 [66,67]. The odor notes were considered to be variables in the context of odorants. Conversely, to perform the organization of the odor notes, we used the transposed matrix in which the odorants are the variables.

The MDS approach allows one to visualize the similarity between elements of a dataset by placing them in an N-dimensional space [68,69]. The coordinates of each element in this space are determined by the proximity matrix, which measures the similarity or the dissimilarity between two elements. In the present study, we used the Euclidian distance calculated using the Jaccard coefficient to obtain a similarity matrix between odorants (binary matrix 3508x251: 3508 odorants and 251 odor notes). Then, an MDS was performed on the matrix to derive Euclidian coordinates and distances.

MDS is one of the methods that allow dimensionality reduction in order to produce meaningful representations of high-dimensional data into a lower-dimensional space. [70]. Since the more relevant information is carried by the first dimensions, and the visualization and interpretation become difficult if the number of dimensions is higher than 2 or 3, the number of dimensions is usually 2 or 3. In most cases, only the first two dimensions are used: they allow visualizing the distance between the elements in the plane. Nevertheless, it is useful to consider more dimensions as long as their interpretation is possible. In some cases a significant part of information may be carried by the third dimension; in such cases, it is appropriate to visualize the data in three-dimensional space. In the present study, we used the coordinates of the first three dimensions to display the odorants in a three-dimensional scatterplot.

AHC is a "bottom up" approach that aims to group objects according to their similarity. Each element constitutes its own cluster; at each step, the two closest clusters are merged to create a new group. Lastly, all elements form a single cluster. The distance between two groups can be calculated using several linkage methods (simple linkage, average linkage, Ward's method, and complete linkage [68]. The successive clustering operations produce a dendrogram (binary clustering tree), which represents a hierarchy of partitions of the set. By truncating the tree at a given level (height in dendrogram) allows to obtain a partition into several clusters.

The clustering of odorants was performed using the coordinates in the first three dimensions of the MDS three-dimensional space (complete linkage). We determined the optimal number of clusters using the Kelley penalty score [71]; the calculation was performed by the ***kgs*** function of the R package ***maptree*** [66].

CA was used to examine the proximity between odor notes. The analysis was performed on the co-occurrence matrix of the odor notes, i.e., using the square 251x251 matrix in which the off-diagonal terms are the number of co-occurrences of two odors notes in an odorant description, while the diagonal terms are the number of all occurrences of each odor note. CA allocates the elements in a multidimensional space according to their similarity based on chi-square distances; we also selected the first three dimensions of this space to visualize the proximity of odors in a Euclidean space and to perform AHC.

SOM, also called Kohonen map, belongs to the unsupervised Artificial Neural Network (ANN) methods [63,64]. SOM is a useful approach to visualize the distribution of a dataset belonging to a large N-dimensional space in a 2D space without affecting the topology of the initial large space. Each element (e.g. here the odorants) of the dataset corresponds to a node, also called neuron, which is associated with a vector in the n-dimensional space. The SOM algorithm organizes the nodes to place similar nodes closer together. Then the nodes are connected to each other by a neighborhood relation via a two-dimensional regular spacing in

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

classes that are displayed on a bi-dimensional grid, which is usually rectangular or hexagonal. After classification, the nodes that are closer together in the n-dimensional space are placed to the same class or to neighboring classes of the grid (map). The map dimension defines the number of classes; the maps are most often (but not necessarily) square. For example, the nodes can be arranged into four classes on a map dimensions 2x2 (square), but also on a map dimensions 1x4 (“twine”). In the present study, we used the following as the datasets: (i) the binary matrix for the odorant classification; and (ii) the non-symmetrical square matrix of odor notes obtained from the co-occurrences symmetric square matrix by weighting the number of associations by the frequency of occurrences for the odor notes classification. Standard SOM form implemented in the ***som*** function of the Kohonen R package [67] was applied to the datasets using an hexagonal grid. The number of classes ranged from 2 to 100, varying the size of the map dimensions from 1x2 (2 classes) to 10x10 (100 classes).

Network visualization

We used Cytoscape [72] to build a network of the links between odor notes. This required the square matrix to be transformed into a two-way data table, which was performed via XLStat (Addinsoft). The links between odor notes were analyzed using Excel Boolean and statistics functions.

Results

Analysis of the distribution of odor notes carried by the odorants

Number of odor notes by odorant and frequency of odor notes occurrences

As observed by Martinez-Mayorga, Peppard, Yongye, Santos, Giulianotti, Medina-Franco [43], the frequency of odor notes by odorant follows slightly skewed distribution (Figure 1A).

The distribution is flatter than that of a normal distribution (kurtosis = -0.088) and concentrated to the right of the mean (skewness = 0.303). The mean and the median are 3.809 and 4.000, respectively. Only 1% of the odorants have eight or more descriptors, sightly more than 5% have one descriptor, and 50% of the odorants have 3 or 4 descriptors. There are 178 odorants described by one note without association with “fruity”, “floral”, “sweet” or “green” in the description. The number of these odor notes is 84, and the most frequent are “fatty”, “oily”, “sulfuraceous”, “meaty” and “balsamic”. Less than 10% of the odorants are described by one odor note, associated or not with "fruity", "floral", “sweet” or “green” in the description (about half of them are not associated with “fruity” and “floral” notes); there are 100 odor notes that meet this criterion, of which the most frequent are “fatty”, “balsamic”, “herbaceous”, “woody”, “oily”, “cheese”, “sulfuraceous” and “rose”. Conversely, 146 odor notes are never used alone in an odorant description, e.g., “nutty, “herbal”, “fermented”, “phenolic”, “cognac” and “melon”; these notes are not “isolated notes” according the term used by Chastrette, De Saint Laumer, Sauvegrain [73].

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

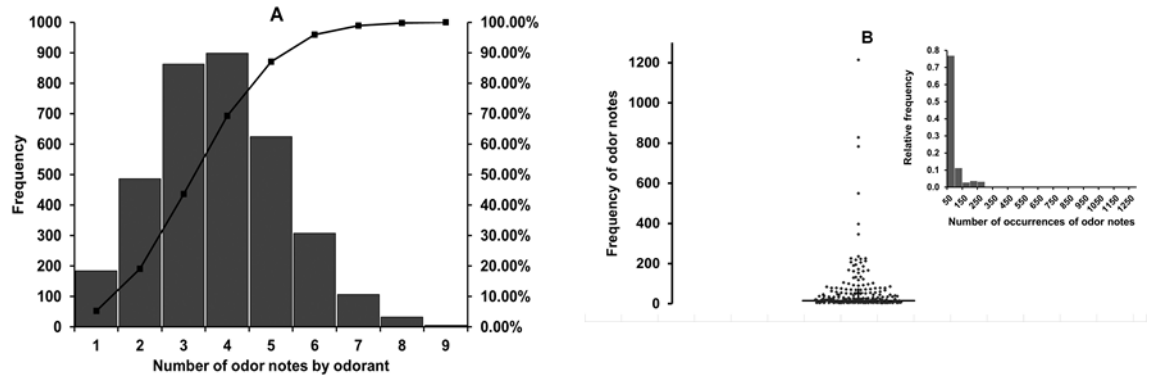


FIGURE 1. Distribution of odor notes associated with different odorants. (A) Histogram of the number of odor notes by odorant. (B) Scattergram of the distribution of the frequency of descriptor occurrences (mean is shown by the cross, and the median is denoted by the horizontal long bar); the histogram of relative frequency of occurrences is displayed in the inset.

Figure 1B shows the occurrence frequency of the 251 selected odor notes. The highest occurrences (> 300) are observed for only six odor notes: "fruity", "sweet", "green", "floral", "fatty" and "herbaceous" (1213, 827, 782, 549, 397 and 345 occurrences, respectively). The four most frequent notes are characterized by a general quality character [48]. Additionally, 55 notes have between 236 and 50 occurrences, which represent 22% of all odor notes; approximately 75% of the odor notes have fewer than 50 occurrences, 55% of them have 5-20 occurrences, and 30% have 10 or fewer occurrences.

Analysis by odor notes

There are 2212 odorants having at least one of the four most frequent notes in their odorant description; conversely, 1296 odorants lack any of these notes. Among the odorants with at least one of the four most frequent notes, 1249 have only one note: 487 are "only fruity", 118 are "only floral", 322 are "only sweet", and 322 are "only green". The association of two, three and four of the most frequent notes in their description occurs for 776, 178, and 9 odorants, respectively.

Clustering of odorants according to their odor notes

MDS of odorants

We made an MDS from the dissimilarity matrix obtained using the Jaccard coefficient to determine the level of similarity of odorants based on their odor notes. The maps obtained in two- and three-dimensional spaces are displayed in Figure S2. In the diagrams, the X, Y and Z axes correspond to the first three dimensions of the MDS space, i.e., V1, V2 and V3, and the locations of the odorants can be defined by their coordinates in this 3D space. The dimension V3 provided interesting information about the distances between odorants. Consequently, we used the coordinates of the first three components to visualize a 3D-scatterplot and explain how the odorants are distributed according to their odor notes.

The projection of the MDS 3D space of the odorants (Figure 2) separates the dataset into several zones, where "fruity", "sweet" and "green" odorants are segregated into several main zones [43]. The partition is less clear for "floral" odorants; nevertheless, the "floral" odorants form "islets" in the MDS projections.

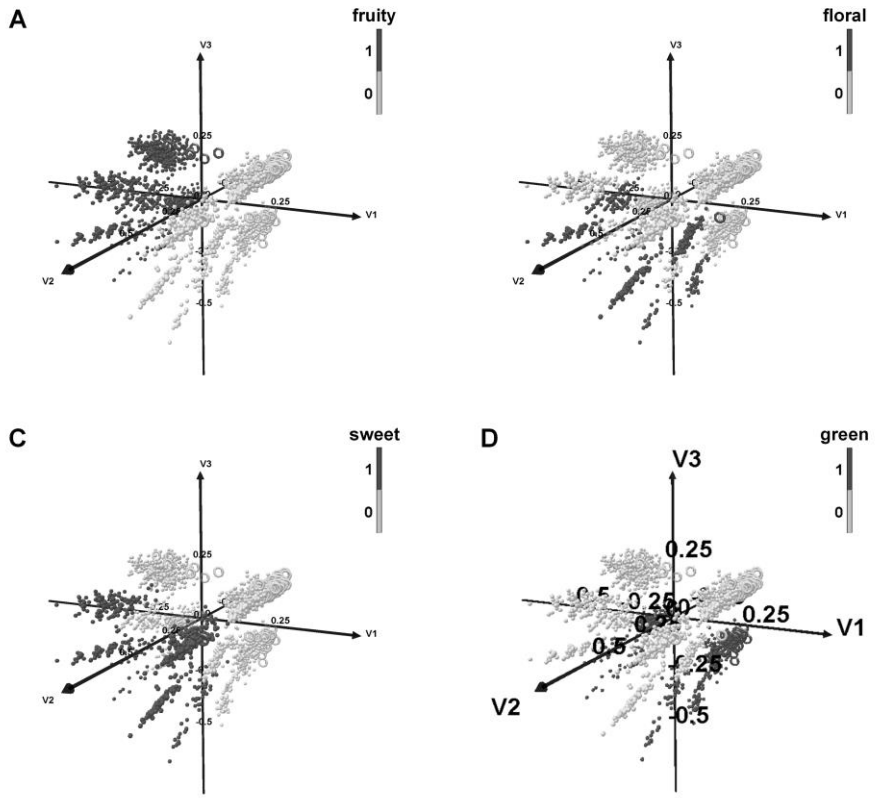


FIGURE 1. Allocation of the odorants associated with the most frequent notes. (A) "fruity", (B) "floral", (C) "sweet" and (D) "green" notes in the chemical space defined by MDS. Odorants associated with the related odor note are shown in dark grey; odorants not having the related note are depicted in light grey. Sulfuraceous molecules are represented as tori, while non-sulfuraceous molecules are represented as spheres.

All "fruity" odorants except one are grouped in the part of the space characterized by negative values for the V1 and V2 coordinates (Figure 2A). The exception is methyl furfurylthiopropionate ("*Sulfurous, meaty, earthy, fruity, coffee, roasted, nutty*", V1V2V3 0.03; -0.09; 0.18). Interestingly, another "fruity-sulfuraceous" odorant is very close to the V3 axis: ethanethioic acid ("*Sulfurous-meaty, savory, herb-fruity, or fresh-tropical-fruity in dilution*", V1V2V3: -0.03; -0.10; +0.15). The "only fruity" odorants are mainly grouped in the part of the space defined by negative coordinates along the V1 and V2 axes and positive coordinates along the V3 axis (Figure S3); five odorants are characterized by positive V2 coordinates and the "balsamic", "herbaceous", and "spicy" odor notes (Figure S4).

Conversely, "floral" odorants are located in several parts of the space characterized by negative V3 coordinates for most odorants (Figure 2B); nevertheless, 17 odorants that share both "fruity" and "floral" notes are in the space possessing a positive V3 coordinate. Most "only floral" odorants are clustered in a group characterized by negative coordinates along the V3 axis and positive coordinates along the V1 and V2 axes; four of the odorants have negative coordinates along the V2 axis and share both "fatty" and "waxy" notes. The coordinates of this peculiar fruity and floral odorants are reported in Table 1.

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

TABLE 1. Coordinates of peculiar fruity and floral odorants.

Common name	Description	V1	V2	V3
Phenethyl isovalerate	*Fruity, rose, **balsamic** odor; fruity pineapple, apple, tropical taste*	-0.26	-7.4810^{-05}	0.11
Tetrahydrocuminyl acetate	***Herbaceous**, woody, slight spicy fruity odor*	-0.19	7.4610^{-05}	0.13
Phenethyl propionate	*Spicy, **herbaceous**-rosy, deep-fruity; sweet green berry taste*	-0.16	0.01	0.11
Hexyl cinnamate	*Warm-**herbaceous**-balsamic odor with fruity balsamic notes*	-0.21	0.01	0.11
Menthyl isovalerate	***Balsamic**, **herbaceous**, fruity, minty odor; woody-bitter taste*	-0.21	0.01	0.11
Butyl benzoate	*Mild, fruity **balsamic** with a woody spicy note*	-0.18	0.02	0.13
Dodecyl alcohol	*Oily-fatty, slight waxy-floral-citrus odor; fatty waxy taste*	0.12	-0.04	-0.04
Benzyl decanoate	*Faint fatty-waxy, slightly floral*	0.12	-0.01	-0.08
Pentadecanol	*Weak, bland waxy, fatty-floral*	0.12	-0.01	-0.09
(4Z,7Z)-Tridecadienal	*Fatty, somewhat waxy, floral, herbal; Fatty, savory flavor*	0.11	-0.01	-0.07
Nonanal	*Fatty-floral-rose, waxy odor; citrus taste in dilution*	0.11	0.007	-0.09

The "sweet" odorants are divided into the four quadrants of the V1V3 plane; most of them have positive coordinates along the V2 axis. Moreover, 19 odorants that share the "green" note have negative coordinates along the V2 axis (Figure 2C). The "only sweet" odorants have positive coordinates on the V1 and V2 axes and are split along the V3 axis on both sides of the V1V2 plane.

The "green" odorants are distributed in several groups that are characterized by negative coordinates along the V3 axis (Figure 2D) and are spread in the four quadrants in the V1V2 plane. The "only green" odorants are in the quadrant of the V1V2 plane defined by positive V1 coordinates and negative V2 coordinates.

The "sulfuraceous" odorants are mainly in a no-fruity and no-floral odorants group, and several of the "sulfuraceous" odorants are also green odorants (Figure 2D). Interestingly, a "sulfuraceous" note is associated with a floral note for only one odorant in the FB9-3508 dataset, i.e., phenethyl mercaptan ("*Sulfurous, floral, tropical and meaty notes*"). Moreover, a "sulfuraceous" note is also associated with a "sweet" note for only one odorant, i.e. 4-(methylthio)-2-pentanone, ("*Sulfurous, sweet, vegetable, fruity ester like*").

Categorization of odorants

To define the clusters and the classes, we adopted the following notations:

- for AHC clusters: cl-XhYtZ, where X is the number of a cluster obtained by truncation at height Y, and Z is the number of clusters for this truncation. For example, if a truncation of the tree at height 100 gives a partition into 4 clusters, they are referred as cl-4h100t1, cl-4h100t2, cl-4h100t3 and cl-4h100t4 (arbitrary values);

- for SOM classes: cl-AmBxC, where A is the number of a class obtained for a BxC map. For example, the 4 classes on map dimensions 2x2 are noted cl-1m2x2, cl-2m2x2, cl-3m2x2 and cl-4m2x2.

AHC of odorants

We used the coordinates of the first three MDS components to perform AHC; the dendrogram is displayed in Figure 3.

The cut-off point for the clustering was determined by the Kelley penalty score calculation [71,74]. The minimum penalty score was obtained for 5 clusters (h = 0.65). Additionally, we selected several other levels of truncation ranging from 2 to 13 clusters to examine how clusters are agglomerated as one moves up the hierarchy.

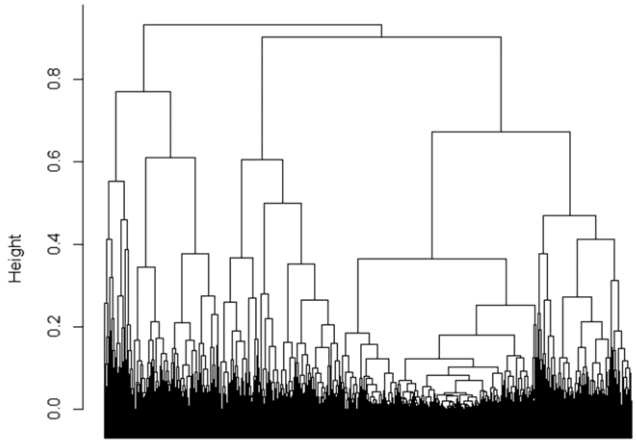


FIGURE 2. Dendrogram of odorants obtained by AHC on the first three dimensions of the MDS.

At each level of truncation, the odorants were distributed between a large cluster (1734 elements at the highest level, h = 0.925, and 1314 elements for h = 0.4) and several smaller clusters. The maximum number of elements in all other clusters was 797 elements (cl-2h0.65t5). The graphical presentations of the distribution of the clusters obtained by truncation at four different levels are reported in Figure S5.

At high levels of agglomerative clustering, the highest truncation (h = 0.925) separates the odorants into two classes with 2734 and 774 elements. This split leads to an allocation of odorants related primarily to the "green" note.

At height 0.65, the odorants are allocated in five clusters: cl-1h0.65t5 (574 elements) and cl-3h0.65t5 (200 elements) merge to form the cluster cl-1h0.925t2, while cl-2h0.65t5 (797 elements), cl-4h0.65t5 (1314 elements) and cl-5h0.65t5 merge to form the cluster cl-2h0.925t2.

At lower heights, the size of the clusters ranges between 81 to 1314 at height h0.5t8 and 35 to 1314 at height h0.4t13. At these two levels of truncation, respectively cl-1h0.5t8 and cl-1h0.4t13 (246 elements), cl-8h0.5t8 and cl-12h0.4t13 (263 elements) cl-5h0.5t8 and cl-5h0.4t13 (1314 elements) are identical.

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

Odorants associated with the "fruity", "sweet" and "green" notes are grouped in a few clusters, of which several are entirely composed of odorants sharing at least one of these notes, which are absent from several other clusters. At height h0.5t8, several clusters are completely or partially composed of odorants carrying only one of the most frequent notes.

Conversely, "floral" odorants are dispersed over several clusters from truncation at height h0.925t2 to h0.5t8, although they form entirely two clusters obtained at height h0.4 (cl-9h0.4t13: "floral-green" odorants; cl-11h0.4t13: "fruity-floral-green" odorants).

A detailed description of the clusters is reported in Supplemental results 1.1.

SOMs of odorants

SOM classification was directly performed on the binary matrix using several sizes of maps: map 1x2, map 2x2, map 4x4, map 7x7 and map 10x10, leading to maxima of 2, 4, 16, 49 and 100 classes, respectively; in all cases, the odorants were spread over the maximum number of classes (no empty class). The choice to examine the partition into two classes (map 1x2) was guided by the desire to compare the separation of odorants obtained by AHC into 2 clusters.

The graphs presented in Figure S6 show that one class of each map is characterized by numerous elements: 2295 for map 1x2, 1576 for map 2x2, 707 for map 4x4, 345 for map 7x7, and 221 for map 10x10. In addition to the largest and the smallest classes, two classes of map 2x2 have a balanced number of elements (774 and 781 elements); more than half of the odorants are in classes of 125-250 elements of map 4x4, and approximately 70% of the odorants are located in classes of 100-300 elements. The smallest class is composed of 70 elements; for map 7x7, more than half of the odorants belong to classes of 50-100 elements. For map 10x10, more than half of odorants belong to classes of 25-50 elements. Finally, three classes of map 10x10 contain only 2 elements (cl-26m10x10, cl-48m10x10, and cl-61m10x10).

There are no hierarchical relationships between the classes of map 1x2, map 2x2, map 4x4, map 7x7 and map 10x10. Each class obtained by numerous classes can be formed with elements that belong to several classes of a map having fewer classes. The 1576 odorants of cl-1m2x2 are distributed into 9 classes of map 4x4, mainly in the largest class cl-1-m4x4x, whose 707 elements come only from cl-1-m2x2. The elements of the largest class cl-1m4x4 are distributed into 19 classes of map 7x7, of which the largest class is cl-2m7x7 (345 elements). However, 327 elements of cl-2m7x7 come from cl-m4x4, and 18 elements come from the 2 other classes of map 4x4 (cl-3m4x4 and cl-5m4x4). Similarly, the elements of the largest class, cl-2m7x7, are distributed over 15 classes of map 10x10, primarily in cl-56m10x10. Moreover, 206 of the 221 odorants of cl-56m10x10 come from cl-2m7x7, 15 odorants come from three other classes of map 7x7 (2 odorants from cl-4m7x7, 8 odorants from cl-26m7x7, and 5 odorant from cl-33m7x7).

Odorants sharing at least one of the four most frequent notes ("fruity", "floral", "sweet" and "green") are distributed over all classes of map 1x2 and map 2x2, 11 of the 16 classes of map 4x4, approximately 80% of the classes of map 7x7x, and 75% of the classes of map 10x10. Moreover, 2 classes of map 2x2, 4 classes of map 4x4, 14 classes of map 7x7x and 35 classes of map 10x10 are entirely composed of odorants described with at least one of the four most frequent notes.

A detailed description of the analysis of the distribution of the odorants among the SOM classes is reported in Supplemental results 1.2.

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

Comparing the spread of odorants and their odor notes using two categorization approaches

Because of the fundamental differences between the two approaches, i.e., AHC and SOM, the lack of a strict equality between the number of clusters and the number of classes is not of primary importance, allowing a comparative study of the grouping of odorants. It appears that the clusters obtained by AHC fit the "islets" of the MDS space, considering that the coordinates on MDS were used to perform the ACH clustering. However, the spread in the SOM classes differs at some points. The distribution of the elements of AHC clusters and SOM classes for splitting into 4 and 13 AHC clusters (height h0.7t4 and h0.4t13) and 4 and 16 SOM classes (map 2x2 and map 4x4) is displayed in Figure 4.

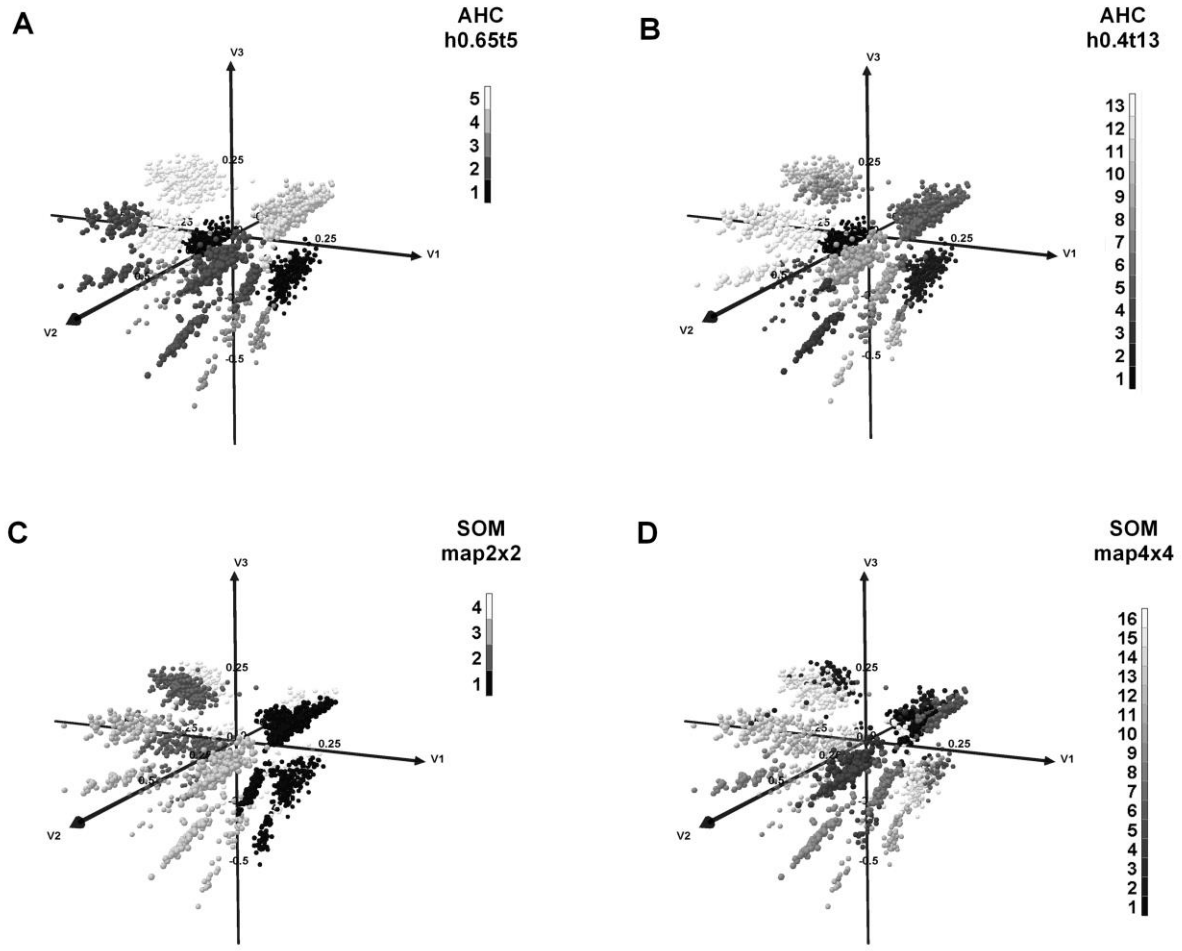


FIGURE 3. Projection of the MDS 3D space of odorant groups obtained via the AHC and SOM categorization techniques. (A) AHC height h0.65t5; (B) AHC height h0.4t13; (C) SOM map 2x2; and (D) SOM map 4x4. The elements are depicted in grey levels according the various groups.

Interestingly, several overlaps between the AHC clusters and the SOM classes exist, which are reported in Table 2, and allows us to appreciate the correspondence between the two clustering approaches.

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

TABLE 2. List of the overlaps of odorant groups between AHC clusters and SOM classes.

AHC cluster*	SOM class *	Number of elements in common
cl-1t4 (874)	cl-1m2x2 (1576)	267
cl-1t4 (574)	cl-2m2x2 (774)	202
cl-4t4 (1937)	cl-1m2x2 (1576)	1145
cl-4t4 (1937)	cl-2m2x2 (774)	518
cl-4t4 (1937)	cl-4m2x2 (377)	252
cl-1t13 (246)	cl-13m4x4 (245)	236
cl-2t13 (328)	cl-14m4x4 (211)	203
cl-3t13 (134)	cl-8m4x4 (192)	99
cl-5t13 (1314)	cl-1m4x4 (707)	707 (all cl-1m4x4 in cl-5t13)
cl-5t13 (1314)	cl-2m4x4 (226)	140
cl-5t13 (1314)	cl-5m4x4 (212)	191
cl-5t13 (1314)	cl-6m4x4 (142	102
cl-8t13 (400)	cl-4m4x4 (314)	253
cl-8t13 (400)	cl-7m4x4 (112)	103
cl-9t13 (66)	cl-10m4x4 (127)	64
cl-10t13 (352)	cl-15m4x4 (344)	270
cl-11t13(35)	cl-10m4x4 (127)	33
cl-11t13 (35)	cl-10m4x4 (127)	33
cl-13t13 (135)	cl-11m4x4 (105)	105 (all cl-11m4x4)

* The number of elements in the AHC cluster or SOM class is in brackets

Conversely, for the SOM class cl-8m4x4, whose elements belong to two regions of the MDS space and are mainly allocated between cl-3h0.4t13 and cl-12h0.4t13 (Figure S7), this overlap is not present. Moreover, the “sweet”, “floral”, “balsamic” and “herbaceous” notes are part of the most frequent odor notes, although cl-3t13 does not include “fruity” odorants, unlike cl-12t13 and cl-8m4x4 (Table 3).

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

TABLE 3. List of the ten most frequent odor notes associated with the odorants belonging to cl-3h0.4t13, cl-12h0.4t13 and cl-8m4x4.

AHC cluster		
cl-3t13 (16*/134**)	cl-12t13 (19*/263**)	cl-8m4x4 (27*192**)
sweet	sweet	floral
floral	fruity	sweet
balsamic	floral	fruity
rose	herbaceous	balsamic
spicy	balsamic	rose
woody	pineapple	herbaceous
herbaceous	caramellic	woody
honey	apple	spicy
anisic	berry	honey
vanilla	rose	vanilla

The number of elements in each cluster or class is shown in brackets; * number of "rose" odorants; ** total number of odorants in this group

The elements of cl-8m4x4 are distributed in two different parts of the MDS space, especially in cl-3t13 (all odorants are "sweet", and 74% are "floral") and cl-12t13 (all odorants are "fruity-sweet").

We focused on the distribution of the odorants with the four most frequent odor notes and "sulfuraceous" for the AHC clusters and the SOM classes. The principal results are reported in Table 4.

The clusters obtained by AHC spread the "fruity" odorants over the two clusters obtained at height h0.925t2, while cl-1m1x2 is entirely composed of "fruity" odorants; consequently all "floral", sweet" and "green" odorants whose description does not include the "fruity" note are contained within cl-2m1x2. The "sweet" and "floral" odorants are spread over the two AHC clusters and the two SOM classes. However, the distribution is better balanced between the SOM classes than between the AHC clusters. This is particularly true in the case of "green" odorants. Here, 99% of "green" odorants are in cl-1h0.925t2, and only 8 odorants are in cl-2t2, although they are balanced between cl-1m1x2 (43%) and cl-2m1x2 (57%).

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

TABLE 4. Summary of the distribution of odorants with most frequent notes across the AHC clusters and SOM classes.

Odor note	AHC cluster	SOM class
"fruity"	cl-2t2 (886); cl-1t2 (327)	cl-1m1x2 (1213)
	cl-4t4 (623); cl-2t4 (263); cl-1t4 (246); cl-3t4 (81)	cl-2m2x2 (774); cl-3m2x2 (304); cl-4m2x2 (135)
	cl-7t8 (623); cl-8t8 (263); cl-1t8 (246); cl-4t8 (81)	
	cl-10t13 (352); cl-12t13 (263); cl-1t13 (246); cl-7t13 (136); cl-13t13 (135); cl-4t13 (46); cl-11t13 (35)	cl-15m4x4 (344); cl-13m4x4 (244); cl-12m4x4 (239); cl-11m4x4 (105); cl-8m4x4 (81); cl-2m4x4 (71); cl-16m4x4 (61); cl-10m4x4 (40); cl-3m4x4 (19); cl-9m4x4 (5); cl-6m4x4 (3); cl-5m4x4 (1)
"floral"	cl-2t2 (414); cl-1t2 (135)	cl-1m1x2 (231); cl-2m1x2 (318)
	cl-2t4 (293); cl-3t4 (128); cl-4t4 (121); cl-1t4 (7)	cl-3m2x2 (216); cl-1m2x2 (171); cl-2m2x2 (131); cl-4m2x2 (31)
	cl-3t8 (213); cl-7t8 (106); cl-6t8 (84); cl-8t8 (80); cl-4t8 (44); cl-5t8 (15); cl-2t8 (6); cl-1t8 (1)	
	cl-8t13 (114); cl-13t13 (105); cl-3t13 (99); cl-12t13 (80); cl-9t13 (66); cl-11t13 (35); cl-6t13 (18); cl-5t13 (15); cl-4t13 (9); cl-2t13 (6); cl-1t13 (1); cl-7t13 (1)	cl-8m4x4 (192); cl-10m4x4 (127); cl-7m4x4 (112); cl-11m4x4 (105); cl-2m4x4 (8); cl-13m4x4 (2); cl-9m4x4 (1); cl-12m4x4 (1); cl-14m4x4 (1)
"sweet"	cl-2t2 (726); cl-1t2 (101)	cl-1m1x2 (338); cl-2m1x1(489)
	cl-2t4 (694); cl-3t4 (101); cl-4t4 (32)	cl-3m2x2 (781); cl-2m2x2 (25); cl-4m2x2 (21)
	cl-3t8 (431); cl-8t8 (263); cl-6t8 (55); cl-4t8 (46); cl-7t8 (29); cl-5t8 (3)	
	cl-8t13 (297); cl-12t13 (263); cl-3t13 (134); cl-4t13 (46); cl-6t13 (38); cl-13t13 (29); cl-9t13 (17); cl-5t13 (3)	cl-4m4x4 (314); cl-12m4x4 (239); cl-8m4x4 (192); cl-10m4x4 (25); cl-3m4x4 (24); cl-6m4x4 (9); cl-2m4x4 (8); cl-14m4x4 (7); cl-13m4x4 (6); cl-9m4x4 (3)

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

Odor note	AHC cluster	SOM class
"green"	cl-1t2 (774); cl-2t2 (8)	cl-1m1x2 (333); cl-2m1x2 (449)
	cl-1t4 (574); cl-3t4 (200); cl-4t4 (6); cl-2t4 (2)	cl-1m2x2 (329); cl-2m2x2 (254); cl-4m2x2 (117); cl-3m2x2 (82)
	cl-2t8 (328); cl-1t8 (246); cl-6t8 (119); cl-4t8 (81); cl-7t8 (6); cl-3t8 (2)	
	cl-2t13 (328); cl-1t13 (246); cl-9t13 (66); cl-6t13 (53); cl-4t13 (46); cl-11t13 (35); cl-13t13 (6); cl-8t13 (2)	cl-13m4x4 (245); cl-14m4x4 (211); cl-10m4x4 (127); cl-9m4x4 (42); cl-12m4x4 (37); cl-6m4x4 (28); cl-16m4x4 (28); cl-4m4x4 (27); cl-5m4x4 (20); cl-2m4x4 (14); cl-8m4x4 (3)
"sulfuraceous"	cl-2t2 (182); cl-1t2 (43)	cl-2m1x2 (212); cl-1m1x2 (13)
	cl-4t4 (181); cl-1t4 (43); cl-2t4 (1)	cl-1m2x2 (199); cl-2m2x2 (13); cl-4m2x2 (12); cl-3m2x2 (1)
	cl-5t8 (172); cl-2t8 (39); cl-7t8 (9); cl-1t8 (4); cl-3t8 (1)	
	cl-5t13 (172); cl-2t13 (39); cl-7t13 (5); cl-1t13 (4); cl-10t13 (4); cl-8t13 (1)	cl-5m4x4 (161); cl-9m4x4 (48); cl-2m4x4 (5); cl-15m4x4 (5); cl-6m4x4 (2); cl-14m4x4 (2); cl-4m4x4 (1); cl-13m4x4 (1)

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

To illustrate the distribution of odorants, we selected three distributions of odor notes as examples. We focused on odorants with "green", "sulfuraceous", which is frequently associated with "green", and "rose", which is never associated with "sulfuraceous" and has been examined in other studies [32,73].

"Green" odorants

An exploration of the distribution of "green" odorants shows that 6 of the 13 AHC clusters obtained by truncation at height 0.4 (cl-1h0.4t13, cl-2h0.4t13, cl-4h0.4t13, cl-6h0.4t13, cl-9h0.4t13 and cl-11h0.4t13) are entirely composed of "green" odorants, primarily associated with "fruity", "floral" and/or "sweet" notes. Similar results were obtained for "sulfuraceous" odorants of cl-2h0.4t13 (Table 5). Conversely, only 3 classes of map 4x4 (cl-10m4x4, cl-13m4x4 and cl-14m4x4, in which the latter comprises more than half of the odorants in cl-2h0.4t13) are entirely composed of "green" odorants, and the three other most frequent odor notes are seldom associated using AHC clustering. Moreover, fewer than 2 "sulfuraceous" odorants are in these classes.

TABLE 5. List of the ten most frequent notes associated with odorants belonging to the "100% green" AHC height h0.4t13 and SOM map 4x4 (the "green" note is not reported).

AHC cluster						**SOM class**		
cl-1t13 (246)	cl-2t13 (328)	cl-4t13 (46)	cl-6t13 (53)	cl-9t13 (69)	cl-11t13 (35)	cl-10m4x4 (127)	cl-13m4x4 (245)	cl-14m4x4 (211)
fruity	fatty	**sweet**	**sweet**	**floral**	**floral**	**floral**	**fruity**	alcoholic
apple	alcoholic	**fruity**	**floral**	**sweet**	**fruity**	**fruity**	apple	fatty
waxy	nutty	apple	earthy	hyacinth	herbaceous	**sweet**	waxy	cognac
fatty	tropical fruit	floral	citrus	herbaceous	fresh	herbaceous	pear	ethereal
pear	***sulfuraceous***	tropical fruit	fresh	rose	jasmine	rose	fatty	fermented
tropical fruit	cognac	pineapple	herbaceous	leafy	waxy	citrus	fresh	vegetable
fresh	ethereal	waxy	woody	balsamic	rose	waxy	tropical fruit	citrus
pineapple	vegetative	fresh	fatty	woody	earthy	fresh	pineapple	melon
winey	earthy	winey	rose	citrus	citrus	hyacinth	ethereal	earthy
ethereal	vegetative	herbaceous	tropical fruit	melon	winey	earthy	winey	oily

"Sulfuraceous" odorants

"Sulfuraceous" odorants are mainly contained in a single AHC cluster obtained at the different levels of truncation (cl-2h0.925t2, cl-40.7t4, cl5h0.5t8, and cl-5h0.4t13) and one of the SOM classes from m1x2 up to and including m4x4 (cl-2m1x2, cl-1m2x2, an d cl-5m4x4). Significant overlaps exist between several clusters and classes, especially between cl-5h0.4t13 and cl-5m4x4.

Conversely, the distribution of approximately 95% of the "sulfuraceous" odorants is balanced over 5 classes of map 7x7 and 7 classes of map 10x10. Moreover, 2 classes of map 7x7 and 6 classes of map 10x10 are entirely composed of "sulfuraceous" odorants, while no AHC cluster is entirely composed of "sulfuraceous" odorants.

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

"Rose" odorants

The 162 "rose" odorants are distributed over all of the AHC classes up to and including AHC h0.4t13 and SOM map 2x2. Conversely, 4 classes of map 4x4 and more than half of the classes of map 7x7 and map 10x10 (19 and 38 classes, respectively) do not contain "rose" odorants. Moreover, "rose" odorants completely comprise cl-14m7x7, cl-42m7x7 and cl-46m7x7. Interestingly, all odorants of cl-42m7x7 are "fruity", whereas this odor note is absent from the odor descriptions of the odorants of cl-14m7x7 and cl-46m10x10. Additionally, approximately 25% of the odorants of cl-14m7x7 and cl-46m10x10 have the "honey" note, which characterizes only 5% of the odorants of cl42m7x7.

We additionally observed that several "rose" odorants belonging to cl-8m4x4 are allocated over two separate regions of the MDS space, corresponding mainly to cl-3h0.4t13 and cl-12h0.4t13. Both cl-3h0.4t13 (in which all odorants are "sweet" and 74% are "floral") and cl-12h0.4t13 (in which all odorants are "fruity-sweet") encompass odorants with the "sweet", "balsamic", "herbaceous", "woody", "vanilla", and/or "spicy" notes, whereas the absence of "fruity" characterizes the odorants of cl-3h0.4t13 (Table 3).

Analysis of odor notes: pairs of odor notes, links between notes and clustering

Associations and links between odor notes

An exploration of pairs of odor notes was already performed by [43]. Here, we aimed to consider how the odor notes are linked to each other using a network of odor notes.

The networked structures are characterized in terms of nodes (people, or things within the network), and edges, or links, or ties (relationships or interactions) that connect them. In the case of our work, the nodes are the odor notes and the links are the odorants. The network analysis involves the graph theory. The graph theory is now largely used to study social networks, but also in the area of biology to analysis the links between biological targets and their ligands (drugs, or endogenous ligands), in neurosciences, material science (porous material), and numerous other domains. Our approach of the links between odor notes can be can be compared to the analysis of a social network between peoples.

The term "social network" is used in the social sciences to study relationships between individuals or groups, and to describe the structure defined by these relationships.

Several notions are used in the study of social networks, for example:

- in a social network, two individuals are linked when communicate directly with each other. In other words, they are linked by one tie (or pathway);
- there is a "structural hole" if two individuals are not directly connected. When a third individual provides the only link between these two no-connected individuals, such third individual is a "bridge" that fills the structural hole, and there are two ties;
- the distance between two individuals is the minimum number of links required to connect them. According to the Stanley Milgram's "small-world" experiment 75, there would be only 'six degrees of separation' between all individuals in the world.

We used these notions to describe the network of odor notes. In this way, if two notes co-exist in the odorant description of one or several odorants they are "associated", or linked by a tie. We adopt the terminology "associated at Level L1" for such relationships.

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

If no link at Level L1 exists, two notes can be associated with the same note: they are linked through one "bridge" (two ties, or two "intermediates notes"); in other words they are "linked at Level L2".

Similarly, two odor notes can be linked at Level L3 (linked through two "bridge" notes: three ties, or three "intermediates notes"), L4 (three "bridge" notes: four ties), etc. We attempted to identify how the 251 odor notes of the Flavor-Base are linked with each other and determine the longest path between two notes (i.e., the maximum ties of connection).

Associations of odor notes at Level L1 (co-occurrence matrix)

To study the links between odor notes, we considered the notes in the space of the odorants, i.e., the transpose of the binary matrix, in which the odor notes are the observations (rows) and the odorants are the variables (columns).

We calculated the symmetric square matrix of two-way cross-tabulations between the variables. This requires providing the number of odorants for each possible pair of odor notes in which the two odor notes appear together in the odorant description, which can be called the "frequency of association" between two odor notes.

We obtained a list of 31,375 pairs of notes by stacking the co-occurrence matrix after excluding the diagonal elements and the entries below the main diagonal because for any X and Y odor notes, the pairs XY and YX are equivalent. The number of associated odor notes at Level L1 is provided by the pairs characterized by non-zero values in the co-occurrence matrix and is 4675.

The results provided by the co-occurrence matrix can be exploited in several ways. In all cases, the values of the diagonal are excluded (the upper-left to lower-right diagonal shows the frequency of each odor note).

It could be advantageous to consider the ratio between the number of associations and the occurrence frequency of each odor note. The obtained asymmetrical matrix gives the relative frequency of associations for all pairs of descriptors. For example, the "fruity" and "floral" notes appear together in the odorant description of 231 odorants, i.e., "fruity-floral" comprise 19% of the occurrences of "fruity" and 42% of the occurrences of "floral".

The number of associated odor notes ranges between 2 and 203. Most of the odor notes are associated with approximately 15 to 50 other odor notes (Figure 5).

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

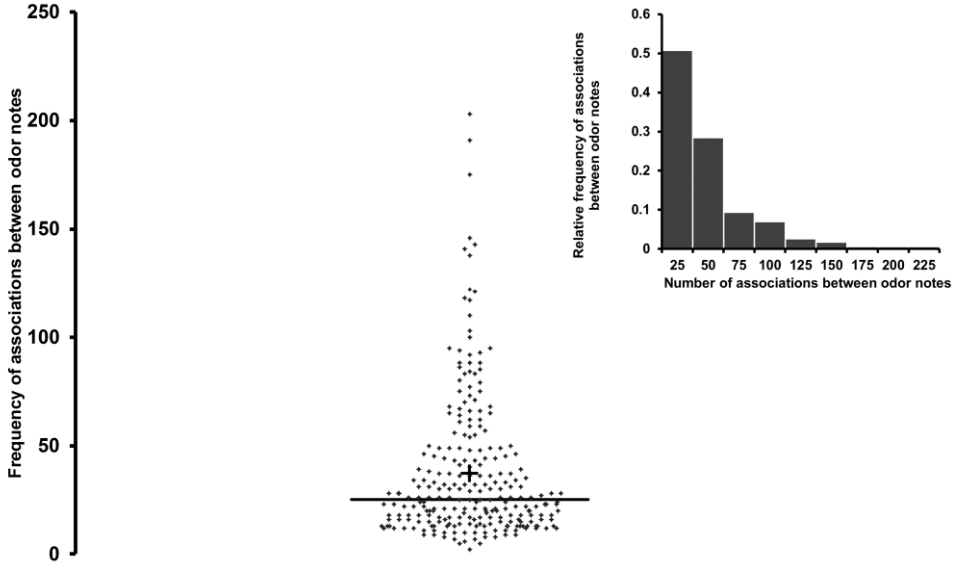


FIGURE 4. Scattergram of the frequency distribution of the number of associations between odor notes. The inset displays the relative frequency of the distribution with intervals of 25.

The number of associated odor notes compared to the occurrence frequency increases until approximately 300 occurrences, and then varies between 141 and 203. The number of associated odor notes reaches 203 for "green" (782 occurrences), which is followed by "sweet" (191 associated notes for 827 occurrences), "fruity" (175 associated notes for 1213 occurrences), and "floral" (146 odor notes for 549 occurrences).

The frequent odor notes (occurrence frequency of at least 200) are strongly associated with themselves. Nevertheless, there is no Level L1 association between "sulfuraceous" and "apple", "fresh" and "nutty", and "apple" and "nutty".

The least associated odor note at Level L1 is "gasoline", which is associated with "gassy" in the odorant description of 3 odorants and with "hydrocarbon" in the odorant description of 8 odorants ("gasoline" is present in the odorant description of 8 odorants and is always associated with "hydrocarbon", which has 32 occurrences). The most largely associated odor note compared to its frequency is "clam", which contributes to the odorant description of 6 odorants in concert with another 22 odor notes.

Links of odor notes at Level L2

Two notes that are not associated at Level L1 but are linked to a common note are regarded as linked at Level L2. These notes are linked by at least one intermediate note, although they can also be linked by several intermediates, which we call "pathways". The procedure used to identify the links at Level L2 is explained in Supplemental results 3.1.

We obtained 25,341 pairs of odor notes connected at Level L2. For example, the "fruity" and "garlic" notes never appear together in an odorant description; nevertheless, "fruity" and "garlic" belong to the list of notes connected at Level L1 to "floral" "sweet", "green" and "fatty" notes. Consequently, "fruity" and "garlic" are connected at Level L2 according to the paths "fruity-floral-garlic", "fruity-sweet-garlic", "fruity-green-garlic", "fruity-fatty-garlic", and "fruity-sulfuraceous-garlic". The same applies for "fruity" and "hydrocarbon" ("fruity-sulfuraceous-hydrocarbon"), "floral" and "alliaceous" ("floral-alliaceous-hydrocarbon"), and "floral" and "hydrocarbon" ("floral-sulfuraceous-hydrocarbon").

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

The number of intermediate odor notes between two notes connected at Level L2 ranges between 1 and 69. More precisely, 1971 pairs of odor notes are connected at Level L2 by only one intermediate. For example, “fruity” and “floral” are linked to “gasoline” only via “gassy”. Conversely, “fresh” and “nutty” constitute a unique pair linked by 69 paths (intermediate notes). Approximately half of the L2 pairs are linked by 1 to 7 intermediates. “Green” is the most frequent “intermediate” (in 16,697 pairs); conversely, “gasoline” and “heliotrope” are never involved as intermediates at Level L2. Most of the notes play the intermediate role less than 800 times.

Finally, except for the six most frequent notes and “nutty”, all of the notes are more frequently connected at Level L2 than at Level L1; 28 odor notes, included in which are the four most frequent notes and “sulfuraceous”, are connected to all other notes at levels L1 and/or L2, meaning that they are connected at most by one intermediate and do not appear in the list of pairs at the highest level.

Links of odor notes at Level L3

At this level, two intermediate odor notes are required to link the two odor notes of each pair. We used a similar approach to the one used for Level L2 to identify the links at Level L3, which is detailed in Supplemental results 3.2.

As observed for Level L2, “green” is the most frequent intermediate in Level L3 pairs (appears in 1073 pairs). Half of the notes serve as intermediates at most 50 times; approximately 90% serve this role less than 400 times. Moreover, 10 notes never play an intermediate role.

Conversely, “gasoline” is the least associated odor note at Level L1 (associated 8 and 3 times with “hydrocarbon” and “gassy”, respectively); “gasoline” is associated with 43 odor notes at Level L2, and 205 at Level L3. The maximum number of links for “gasoline” is 35 at Level L3 with “herbaceous”. By analogy with a social network, “herbaceous” and “gasoline” do not “know” one another, although they have many “common relationships”.

The frequency of links at each level is illustrated by the histogram displayed in Figure 6A. It appears that 15% of the odor note pairs co-occur in the same odorant description (“direct link”, i.e., Level L1); 81% of the odor notes are not in the same odorant description but co-occur in at least one description with a common odor note that is directly linked (“one intermediate”, i.e., Level L2); and only 4% of the odor notes have no mutual “intermediate” but are linked by “two intermediates” (Level L3).

To facilitate the visualization of the links between the odor notes, we constructed an odor note network using Cytoscape [72]. This software was initially conceived for biological research and to analyze the links between biological stimuli and their targets, although it is now largely used for visualization of complex networks. In the case of odor notes, it allows us to visualize various information, such as the frequency of occurrences, the number of odorants in which two descriptors co-occur, and the relative frequency of these co-occurrences. In this network, the odor notes are the nodes, and the odorant descriptions are the edges. We used the unstacked co-occurrence matrix as the input data. The representation of the entire network displayed in Figure 6B reflects its high compactness and reveals the absence of isolated groups of odor notes.

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

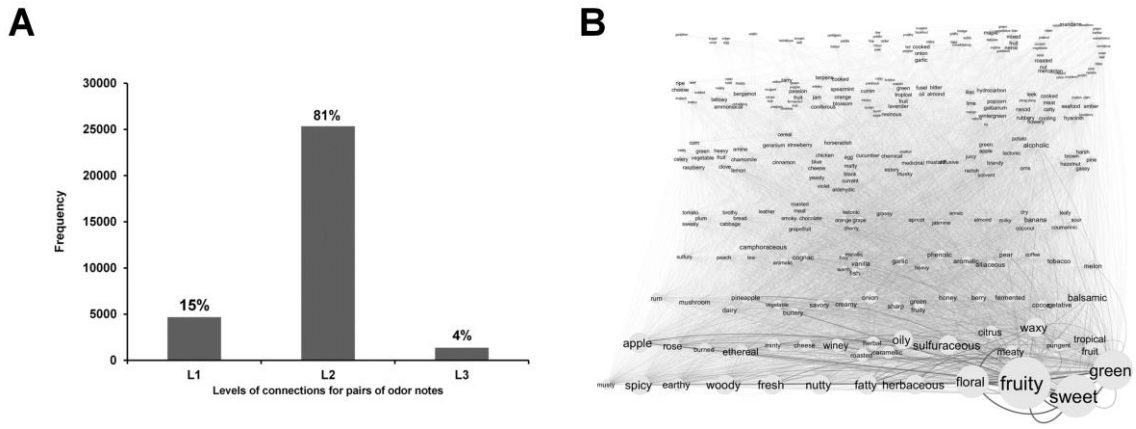


FIGURE 5. Associations and network of odor notes. (A) Histogram of the levels of connections between odor notes. (B) Visualization of the odor note network: the odor notes (nodes of the network) are represented by circles whose size depends on their frequency; the number of association at Level L1 between the odor notes (edges of the network) are depicted as lines whose width and grey levels depends of the number of associations (the more frequent the associations, the more dark and wide the lines).

Correspondence analysis

To examine the associations between odor notes, we performed a correspondence analysis to obtain a map showing the distances between groups of odor notes. The correspondence analysis was performed with the co-occurrence matrix as the input information. Visualization of the graph obtained by projection on the 3 first dimensions is displayed in Figure 7. Three notes are characterized by high V2 coordinate values: “gasoline”, “hydrocarbon” and “gassy”.

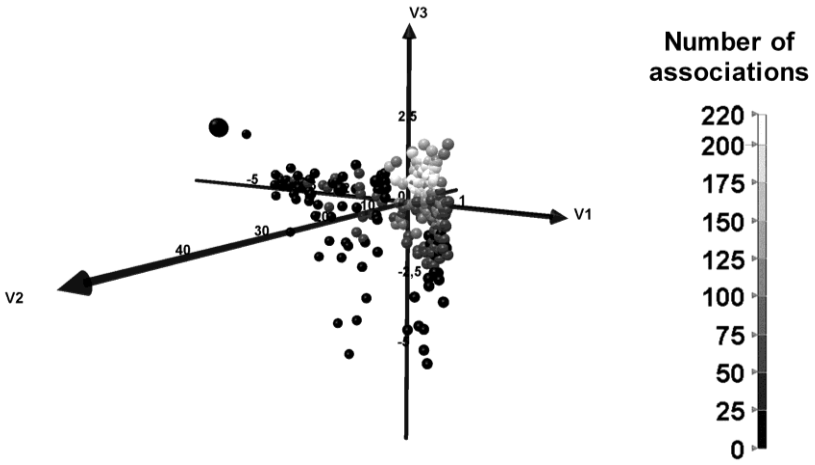


FIGURE 6. Map obtained by correspondence analysis of odor notes. The projection is on the first three dimensions (V1V2V3). The elements (odor notes) are depicted in grey levels according their number of associations.

Spreading of odor notes between the AHC clusters and SOM classes

AHC was performed using the first three dimensions of the correspondence analysis (complete linkage method). The dendrogram of the clusters is displayed in Figure 8.

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

In the same way as for the odorants, we determined the cut-off point for the clustering by the Kelley penalty score calculation [71,74]. The minimum penalty score was obtained for 5 clusters (h = 7). Additionally, we examined the clusters obtained at four other truncation levels: h5 (6 clusters), h4 (7 clusters), h3 (11 clusters), and h2 (17 clusters).

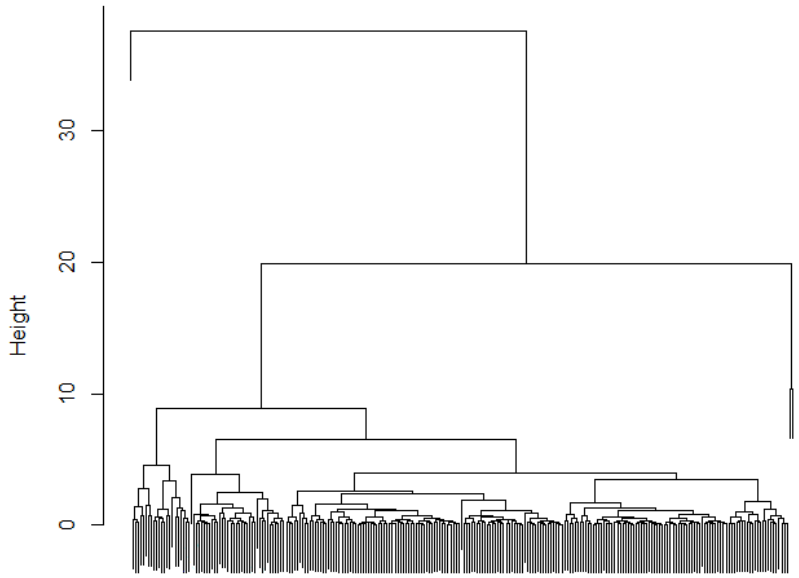


FIGURE 7. Dendrogram of odor notes obtained by AHC on the first three dimensions based on CA.

We used the dissymmetric matrix obtained from the co-occurrence matrix via frequency weighting to classify the odor notes using SOMs for all map sizes, i.e., map 2x2 (4 classes) up to map 10x10 (100 classes). In all cases, the odor notes were spread over the maximum number of classes except for map 10x10, for which only 92 classes were occupied.

Detailed descriptions of the odor note groups are reported in Supplemental results 4.

Contrary to what we observed for the clustering of odorants, a global comparison between the odor note groups obtained by AHC and SOMs is problematic, especially if poor overlaps between groups exist. The graphical view of the co-occurrence matrix sorted according to AHC h2t17 (Figure 9A) and SOM map 4x4 (Figure 9B) shows an organization of AHC clusters in clearly discernable regions. Conversely, the largest unorganized class (178 elements) and the smallest SOM classes are different due to the strongly associated odor notes. Nevertheless, similar associations of some odor notes can be observed in AHC clusters and SOM classes.

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

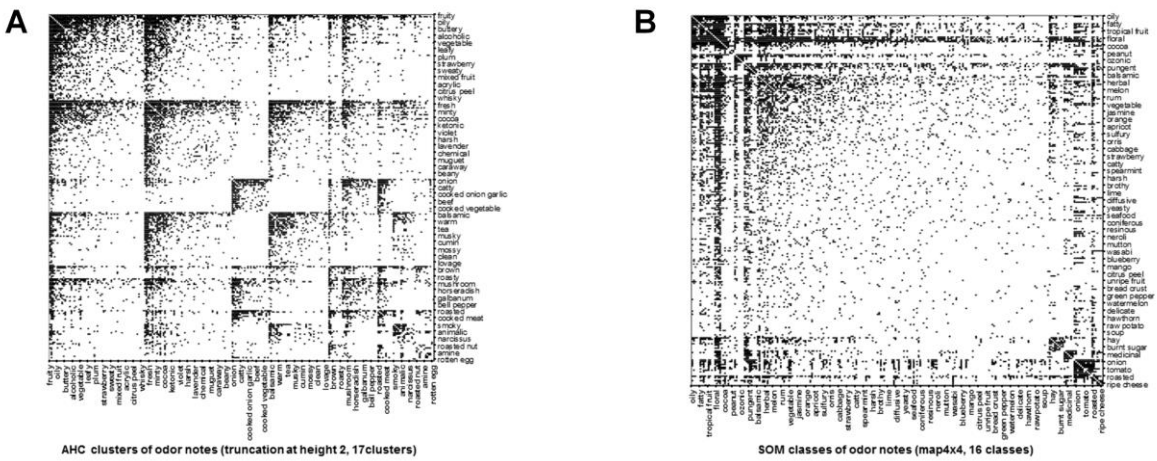


FIGURE 8. Square matrix showing the associations between odor notes organized by (A) AHC height h2t17 and (B) SOM map 4x4.

The "gasoline" and "rotten egg" notes appear to be the least connected to the other odor notes of FB9-3508. The graphical representation of the odor note space based on correspondence analysis shows that the "gasoline" and "rotten egg" notes are the most distant elements in the space defined by the correspondence analysis based on the co-occurrence matrix (Figure 7). Two other notes characterized by a high value along the V2 axis, "hydrocarbon" and "gassy", are the only two odor notes connected to "gasoline" at Level L1. The AHC approach puts the "gasoline", "hydrocarbon" and "gassy" notes into three one-element clusters at height h9t4. Conversely, the three notes constitute a class for SOMs from map 5x5 to map 9x9.

The "phenolic", "smoky" "medicinal", "leather", "tarry", "fish", "animalic", "amine", "ammoniacal" and "ripe cheese" odor notes form interesting connections; several of which already observed by other authors in previous works [32,43,48,73]. The cluster cl-2h7t5 (22 elements) brings together the "phenolic", "fish", "animalic", "smoky", "medicinal", "leather" and "tarry" odor notes, and these notes (except for "fish") form the classes cl-1m3x3 and cl-12m4x4. Furthermore, a more discriminating distribution puts together "amine", "ammoniacal" and "ripe cheese" in the same cluster at height h3t11 and height h2t17 while "fish", "amine", "ammoniacal" and "ripe cheese" form 7 four-elements classes from map 4x4 to map 10x10. Thus, "phenolic", "smoky", "medicinal", "leather" and "tarry" on one hand and "amine", "ammoniacal" and "ripe cheese" on the other hand form stable sets by both AHC and SOM clustering; conversely, the "fish" and "animalic" allocations vary.

Similar observations can be made for "sulfuraceous", "meaty", "onion", "savory", "garlic" and "alliaceous" notes. These six odor notes are gathered in all AHC clusters except "meaty", which is associated with roasted notes at height h2t17. Conversely, SOM classification separates "sulfuraceous-meaty" from "savory-garlic-alliaceous", whereas "onion" shifts from one group to the other according the size of the map.

Among the 80 odor notes in which "rose" is associated at Level L1, "waxy" and "honey" belong to the ten most frequently associated. Moreover, "rose" and "honey" belong to a 5-element class (cl-12m5x5) and a 2-element class (cl-15m7x7). Additionally, the "rose", "geranium" and "metallic" notes form the class cl-66m10x10. The "waxy" and "honey" notes on one hand and "geranium" and "metallic" on the other hand are associated at Level L1. The associations between these four odor notes and "rose" are summarized in Table 6, which indicates that "geranium" and "metallic" are not associated to "honey", and each is associated

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

with “waxy” only in the description of one odorant. However, “honey” and “metallic” and “honey” and “geranium” are linked at Level L2 by 21 and 10 intermediates, respectively.

TABLE 6. Relative frequency of associations at Level L1 between “rose”, “waxy”, “honey”, “metallic” and “geranium”.

Number of associated notes at Level L1	Frequency	Odor note	rose	waxy	honey	metallic	geranium
80	162	rose		14.2% (23)	11.73% (19)	2.47% (4)	5.56% (9)
95	236	waxy	9.75% (23)		2.97% (7)	0.42% (1)	0.42% (1)
70	84	honey	22.62% (19)	8.33% (7)		0% (0)	0% (0)
49	24	metallic	16.67% (4)	4.17% (1)	0% (0)		12.5% (3)
25	15	geranium	60% (9)	6.67% (1)	0% (0)	20% (3)	

The number of odorants for which two notes are associated in the odor description are reported in brackets.

Global visualization: graphical representation(s) of the sorted matrix of odorants and odor notes

We incorporated both the classifications of odorants and the classifications of odor notes to examine the possibility of the emergence of particular groups of odorants sharing common odor notes.

Sorting can be performed based on the clusters obtained from AHC and SOM. An example of the binary matrix obtained after classification of both odorants and odor notes according to AHC clusters (AHC h0.4t13 for the odorants and AHC h2t17 for the odor notes) is displayed in Figure 10. It is also possible to associate the classification obtained by SOMs for the odorants to the classification obtained by AHC for the descriptors. Each allows the visualization of several groups of odorants and descriptors. However, the classification by SOMs appears to be visually more complicated than the classification by AHC because the odorants with the most frequent odor notes, i.e., “fruity”, “floral”, “sweet” and “green”, are distributed across several groups.

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

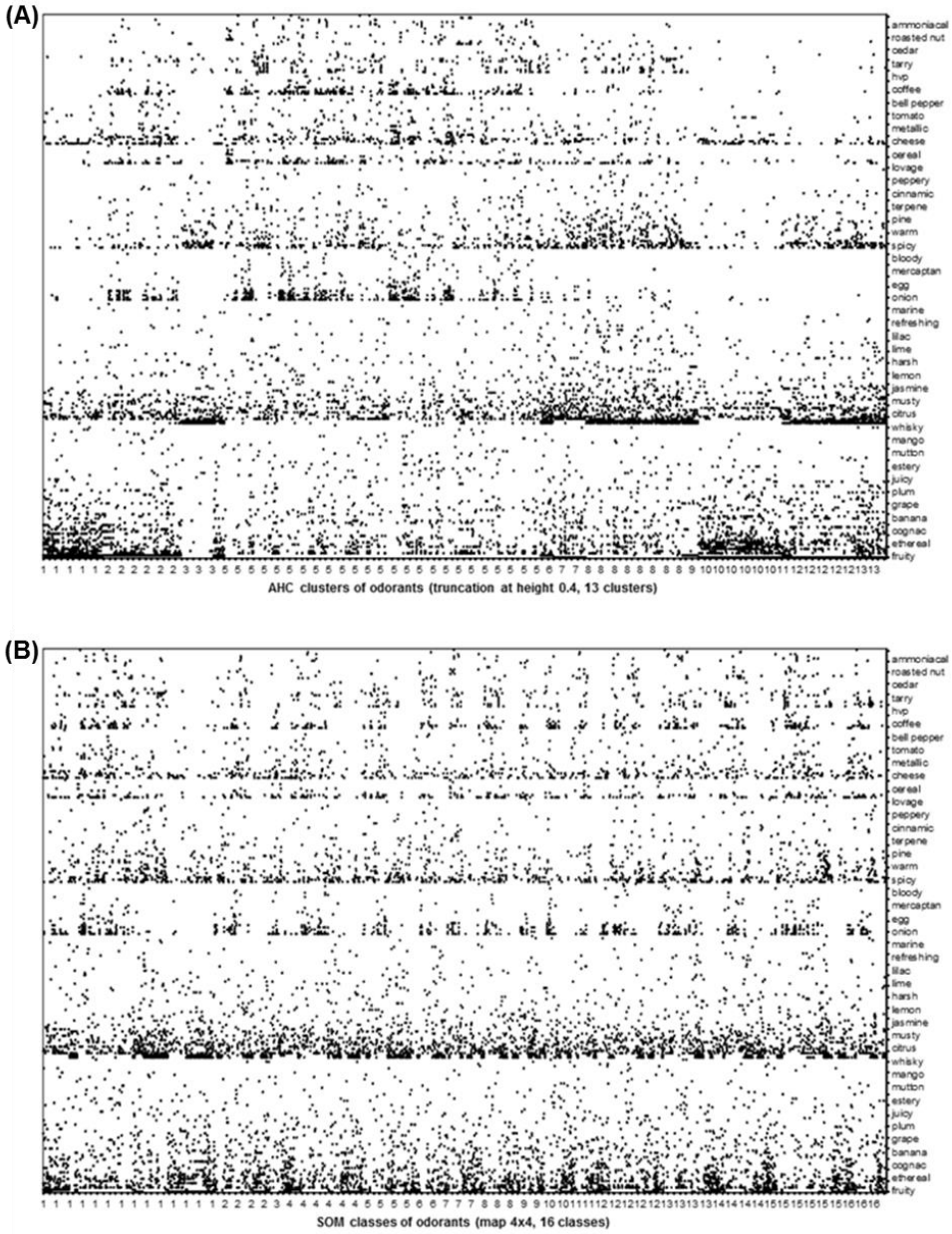


FIGURE 9. Visualization of two examples of the binary matrix organized according to the classifications of odorants and odor notes. (A) Odorants are grouped according to AHC height h04t13 and odor note AHC height h2t17. (B) Odorants are grouped according to SOM map 4x4 and odor note AHC height h2t17.

Discussion

In this study, we investigated links between odor notes resulting in the emergence of clusters of odorants. Although some odor notes can be considered individually, we focused on the association between the odor notes.

Although the dataset used in the present work differs in some aspects from the dataset used by Martinez-Mayorga, Peppard, Yongye, Santos, Giulianotti, Medina-Franco [43], especially concerning the number of odorants and the selection of odor notes, the descriptive statistics

performed on odorants and odor notes displays similar results, indicating that the choice of a larger set of odorants and odor notes does not have unwanted consequences.

The categorization based on a large database facets some limits inherent to such approaches. The verbal description implies obviously some linguistic problems resulting in a quite imprecise description of the odors, which constitute a main drawback. [22,30,76]. Nevertheless, the compilation of numerous results from different studies in a database can be viewed as a way to reduce these effects.

As presented in the Material and Methods part, we have taken into account all the odor notes reported in the odor description at the same level. Neither did we consider the role of the concentration [22]. The changes in the descriptions according to the concentrations are probably linked to the ORs activated by the odorant.

At higher concentrations, a largest number of ORs can be implicated. Nevertheless, ORs which are targeted at weak concentration must be also targeted at high concentration. In other words, a common pattern exist for an odorant whatever its concentration, but the global pattern of activated ORs would be larger at higher concentration. Nevertheless, such issue is beyond the scope of the present study.

Distribution and organization of odorants and their odor notes

It appears that the odor description of an odorant typically requires three to four odor notes, which are called "objects" [30,77,78], to specify the global odor of any odorant. This notion refers to the difficulty in describing an odor and to the crucial role of semantics in odor representation [26,79,80]. It has been already suggested that Western languages are poorly adapted to odor description [81,82]. This suggests the need to consider at least three or four odor notes rather than a shorter description to reliably define an odor. Nevertheless, we observed approximately 100 odor notes that can fully describe an odorant and be considered "pure" odors.

The analysis of the 251 odor notes used in the dataset shows an over-representation of the least frequent odor notes (less than 50 occurrences) compared to the frequency of the other notes (Figure 1). The infrequent notes have very specific odor qualities and are the most abundant in diversity. Conversely, the less specific notes ("fruity", "floral", "sweet", and "green") are the most frequently used in the odor descriptions [48].

After MDS clustering, the distribution of odorants is very similar to the results obtained by PCA on a smaller set of odorants and odor notes [43], separating the space into main zones in which the odorants are primarily allocated according to the "fruity", "sweet" and "green" notes, which refer more to a semantic association than to a precise quality [48]. Interestingly, this distribution into several zones of the MDS space can evoke the sparse response of glomeruli of the olfactory bulb, which has been observed using small concentrations of odorants [83-87]. This result is consistent with hypotheses concerning the spatial organization of glomeruli in the olfactory bulb. Indeed, several studies have shown the absence of *chemotopic* organization of glomeruli, i.e., no spatial organization occurs related to the chemical features of odorants [87,88]. Nevertheless, glomeruli can present an *odotopic* (or *tunotopic)* organization, i.e., an assembly into clusters related to their odor-tuning similarities [35].

Given that most frequent notes govern the split according to the odor distances, an issue was whether the use of "fruity" and "floral" notes would muddle the clusters because of the very general nature of these notes and the high frequency, especially for "fruity". Having examined and tested this aspect, we concluded that despite an apparent improvement in the separation

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

into several large clusters, the removal of “fruity” and “floral” did not improve the clustering results, resulting in more clouded conclusions (results not reported). Consequently, “fruity” and “floral” can be regarded as fully fledged odor notes and are necessary for the analysis. Nevertheless, it remains a point of ambiguity when an odor description uses several names of fruits or flowers without using the “fruity” or “floral” notes. We have been able to fortuitously test such cases. Cinnamyl isobutyrate is described as “*fruity, balsamic-spicy, sweet taste; plum, pineapple, apple, banana notes*”. When the notes in the taste description are used, cinnamyl isobutyrate is located in clusters of odorants lacking the “fruity” and “sweet” notes and instead characterized by balsamic and spicy notes (results not reported). Thus, cinnamyl isobutyrate is an “intruder” in these clusters due to the “balsamic” and “spicy” notes, which means that the “fruity” note is not systematically more influential. Conversely, using the odor description (we considered that “*plum, pineapple, apple, banana notes*” refers to the odor description), cinnamyl isobutyrate belongs to classes with odorants sharing the “apple”, “banana”, “pineapple”, and, for some, “fruity” and/or “sweet” notes (cl-45m7x7 and cl-93m10x10).

The two organization approaches show different and complementary results. These differences are illustrated by the splits into two groups of odorants obtained by MDS and AHC in the high-level dendrogram. For example, the “fruity” odorants are distributed over two AHC clusters but constitute only one SOM class. Interestingly, the divergence between the AHC and SOM clustering process is well perceived for a split into 2 and 4 groups, but less obviously for the truncation at height h0.7t13 clusters and map 4x4.

The large overlaps between the AHC clusters and the SOM classes reveal a good coherence in the results for numerous odorants (Table 2). Nevertheless, several SOM classes are spread over distinct regions of the MDS space and, consequently, over several AHC clusters. For example, the class cl-8m4x4 is separated into two AHC clusters (cl-3h0.4t13 and cl-12h0.4t13) that differ with regard to the presence of the "fruity" note. Moreover, examples of allocations in the AHC and SOM groups of “green” odorants, “sulfuraceous” odorants and “rose” odorants suggest than AHC clustering of odorants is more dependent on the four most frequent notes than SOM classification, which appears to be based on less frequent notes, e.g., "balsamic", spicy", and "herbaceous" in the case of “rose” (Table 3).

An analysis of the composition of the groups obtained by AHC and SOMs suggests that the SOM classes are less dependent on the frequency of odor notes than the AHC clusters, especially regarding the classification of odor notes based on the co-occurrence matrix and correspondence analysis of odor notes. However, in good agreement with the specific nature of these two methods, AHC is more efficient at identifying large groups, whereas SOMs are especially useful to identify the associations inside odorants groups with fewer than 4 or 5 odor notes. This last point permits extending and supplementing the identification of small sets of odor notes beyond the pairs that have already been identified [32,43].

Links between odor notes

The categorization of odorants was completed by examining the links between odor notes and the categorization of odor notes.

The links between odor notes reveal a dense and intricate network in which more than 95% of all possible odor note pairs are connected by at least one intermediate, and a maximum of two intermediates is necessary to connect all odor notes together (Figure 6). This density of connections depends partly on the fact that the four most frequent odor notes were considered. Consequently, there is no insular zone of notes existing in the developed network. Thus, the

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

odor space resembles a quasi-continuum space, which is coherent with a non-hierarchical organization [22,30].

Several pairs of odor notes appear in the description of a single odorant; these rare associations suggest that the two odor notes of such pairs refer to very dissimilar odorant types, e.g., "floral" and "sulfuraceous", "floral" and "meaty" (Phenethyl mercaptan, "*Sulfurous, floral, tropical and meaty notes*") or "citrus" and "roasted" (Methylthioheptanal, "*Sulfury, green, citrus & tropical; roasted-savory at high levels*"), "sweet" and "cumin" (Dihydrocarveol, "*Sweet, woody-cuminic, floral odor; minty, sweet weedy spicy taste*"), and "herbaceous" and "banana" (Hexyl methyl ether, "*Sweet, herbaceous floral notes of lavender, pear, banana, green tonalities*"). There are 2228 pairs at Level L1 that meet this criterion, which correspond to nearly half of all L1 pairs and 7% of all possible pairs.

"Green" is associated with numerous odor notes (203 odor notes); in fact, it is the more associated odor, and more than 80% of odor notes are associated to "green". Although the associations with "fruity" and "floral" are the most frequent, "green" is also associated with "sweet", "fresh", and also "fatty", "waxy", "sulfuraceous", "fermented", "vegetative" and "winey", which refers to another olfactory universes (**Erreur ! Source du renvoi introuvable.**). As consequence, "green" is the most frequent intermediate note at both Levels L2 (involved in L2 16,697 pairs) and L3 (involved in 1073 L3 pairs). For example, "green" in a "bridge" between "violet" and "camphoraceous", "garlic" and "cucumber", "malty" and "hyacinth".

The classification based on the associations between odor notes is not as clear as the classification of odorants. The AHC results obtained from the associations of odor notes shows for several of them a very similar profile to that observed in other studies, indicating that the odor note space is of less significance than the odorant space [32,43,73]. Nevertheless, the examination of the network of odors and comparison of AHC results against the SOM results provides interesting and complementary findings.

The AHC clustering results in several very large groups. At height h5t6, the odor space is divided between three major groups characterized by fruity/floral/sweet/green, "sulfuraceous" and "phenolic" notes. These three domains are considered to be the three main odorant spaces, in which the first refers primarily to pleasant odors, whereas the two others are frequently recognized as unpleasant. Nevertheless, the odor network shows that numerous links exist between these odor categories.

The SOM classification based on the dissymmetric matrix tends to favor the small classes, while the largest SOM classes contain little information. The notes belonging to a large class exhibit little mutual association, while the notes belonging to small class are significantly associated. The grouping of the four most frequent notes in the 4-element class cl-1m2x2 indicates that these notes do not interfere with the classification of the other odor notes. This shows an important difference between the SOM and AHC techniques, in which the latter defines classes based on "fruity", "floral", "sweet" and "green" notes being dominated (cl-1h4t7, 190 elements). The distribution of the four most frequent notes is consistent in the AHC and SOM results such that "fruity" and "green" and "floral" and "sweet" belong to two distinct clusters using truncation at height h3t11 or division according to map 7x7.

We focused on the associations of odor notes that have been highlighted in several previous studies [32,43,48,73]. Several groups defined by AHC and SOM categorization, especially for the sets "gasoline-gassy-hydrocarbon", "phenolic-smoky-medicinal-animalic", "fish-amine-ammoniacal-ripe cheese", and "sulfuraceous-meaty-onion-savory-garlic-alliaceous", with

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

good accordance with previous observations are found considering the links at several levels. We found that both the AHC and SOM results have close associations between “onion” and “garlic”. Additionally, these two odor notes are also associated to “savory” and “alliaceous” in several clusters. Similarly, the association of “phenolic” and “smoky” (i.e., “smoke-like”) is completed by “medicinal”, “tarry”, and “leather”.

The examination of “rose” from our analysis shows that it is a versatile note, which agrees well with the earlier observation that the "rose" odor notes share common features with several other odor groups [22]. The “rose” odorants in the SOM class cl-8m4x4 corresponds to two distinct regions of the MDS space related to the presence or absence of “fruity” odor notes. Moreover, “rose” is mainly located in large AHC clusters and SOM classes obtained based on the correspondence analysis and co-occurrence matrix, respectively.

Nevertheless, an examination of the most frequent notes associated with odorants that form the SOM classes entirely composed of “rose” odorants (“100% rose” classes) reveals recurring associations with “herbaceous”, “honey”, “balsamic”, “fresh”, “geranium” and “waxy” odor notes. After the four most frequent notes, “waxy” [55], “herbaceous” and “honey” are the most common associations with "rose" [32,73].

According to the SOM classification based on the co-occurrence matrix of odor notes, interesting associations of “rose” to “honey”, “metallic”, and “geranium” can be identified [48]. However, “honey” and “geranium” never coexist in an odor description. Thus, “rose-honey” and “rose-geranium-metallic” are two different odors related to the perception of the odor of flowers called “roses”. Furthermore, “rose” when used alone, could be a third “rose” nuance.

Groups of odorants versus “primary odors”

The 2D representation of the binary matrix and the co-occurrence matrix sorted according to the AHC classifications of odorants and odor notes display the existence of several odorant groups sharing similar descriptor sets and large domains of associated odor notes. Because AHC clustering is more dependent on the most frequent notes, the graphical representations are clearer than those obtained using the SOM classes. Thus, the smaller groups of odorants and odor notes defined by the SOM classification provide useful and complementary information. The visualization of the AHC results also suggests the existence of smaller subsets within relatively large groups defined by the AHC classes. This finding hints at the hypothetical existence of “primary odors”. This hypothesis was first proposed by Amoore [27] and was later discredited [29,30]. Other studies have contributed to the discussion on “primary odors” [32,47], which remains unresolved. An underlying issue is related to both the maximum detectable nuances and the main odor qualities related to these nuances [10,89-92]. Based on the number of olfactory receptors (ORs), assuming that an odorant can activate several ORs [6] and that each activation pattern corresponds to one odor [93], it is possible to estimate the number of potential distinctive odors by calculating the number of *k*-combinations among *n*, if *k* ORs are activated in the space of 380-400 human ORs, which is equal to the binomial coefficient $n!/(k!(n-k)!)$, where *n* is the total number of ORs, *k* is the number of ORs activated by one odorant, and “*n!*” and “*k!*” are the factorials (the product of all positive integers less than or equal to n or to k) of *n* and *k*, respectively. For example, considering a total of 390 ORs, there are approximately 8.10^4 combinations of two ORs, and there are 5.10^{12} combinations of 6 ORs. However, the human perception of olfactory signals processed at higher levels by the brain could correspond to more limited “types of odors” and appear as “primary odors” [33,79,94,95].

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

From studying the links between odor notes and groups of odor notes, two points related to the controversial issues of the existence of "primary odors" and hierarchical organization of odors can be mentioned:

- the number of "simple odors", i.e., the number of odorants in the FB9-3508 dataset that can be described via a single odor note, is approximately 100, although these odor notes are not always isolated;
- the odor space is a quasi-continuum in which no set of odor notes is totally isolated.

The first point would be in accordance with the existence of peculiar odors. Nevertheless, the existence of odorants described by a single odor -and not by an association of odor notes- does not necessarily mean that such "simple odors" are "primary odors". At least in some cases, it can rather reveal a semantic factor due to the association of a familiar odor with a well-defined object. Moreover, the presence of "simple odors" could be amplified by the size or construction of the dataset, which can promote certain odorant types, and well-known odors described with a precise terminology, and this brings back to the semantic problem.

The second point indicates that no clear discrepancy and grading exist between the odor spaces, agreeing with the absence of a hierarchical organization of odors [30,48]. Exploring the network of odor notes shows clearly the complex links existing between all odors, even between odors evoking very different qualities (for example at Level L2 "fruity" and "mercaptan", "rose" and "beef", "meaty" and "honey"; at Level L3 "rose" and "gasoline", "banana" and "liver, "anisic and "bloody"). In this way, there is no juxtaposition of "isolated worlds" of odors.

Nevertheless, therein lies a paradox between these two statements, which leaves the issue unresolved.

Odor-structure relationships

In the proposed approach, the structural properties of the studied odorants were not used. Our objective was to develop an innovative strategy to decipher new connections. Regarding the structural information, we expect to obtain similar chemical spaces for the same large set of odorants using odor notes and molecular features. In practice, the results are likely to depend on the choice of the molecular features used as variables [96-99]. Nevertheless, there are several biases or difficulties concerning the categorization approaches because dimensionality reduction is a source of bias, and AHC requires determining the most appropriate number of groups [30]. Moreover, achieving relationships between an odor and a molecular structure raises the issue of establishing molecular similarity in a large chemical space, which is not easy [100-104].

It would be interesting to focus on "unusual" (rare) associations, e.g., "floral-sulfuraceous", "floral-meaty", "woody-ethereal", etc. [73]. Through the 2228 pairs, it would be preferable to focus initially on the most frequent notes.

Some of these rare associations are easy to explain, e.g., phenethyl mercaptan ("*Sulfurous, floral, tropical and meaty notes*"). Indeed, the phenethyl group is frequently associated with "floral" notes because it is a phenethyl alcohol ("*Floral, rose-like odor; floral taste*"), whereas the thiol group is present in the odorants described with sulfuraceous/sulfury notes.

Nevertheless, the presence of a sulfur atom is not always associated with "sulfuraceous", "alliaceous", and "sulfury" notes or to any frequently associated notes belonging to the same odorant group (e.g., "onion", "cooked onion", and "garlic"). This is particularly the case for

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

some terpenic odorants, e.g., menthenethiol ("*Grapefruit, fresh, tropical, juicy and mango*"), methylthiomenthone ("*Catty-ribes like, buchu, blackcurrant odor in dilution*"), thiolinalool ("*Fresh, fruity, bitter grapefruit note in dilution*"), and thiomenthone ("*Catty aroma; dull fruity black currant flavor")*. Several key structures linked to the odor quality of mercapto terpenoids have recently been examined [105], highlighting the role of an intracyclic double bond and a tertiary thiol group in the side chain.

The various "rose" notes suggest that several molecular structures should be related to "rose-metallic/rose-geranium" "rose-honey", "rose-waxy" and "pure rose". Four odorants share the "metallic" note: biphenyl ("***Metallic** geranium-rose odor; neroli, bergamot & cinnamon at 2 ppm taste*"), diphenylmethane ("*Geranium-leaf, **metallic** orange-blossom, rose note*"), methyl diphenyl ether ("*Mild rosy with green aspects, slightly **metallic***") and benzyl methyl ether ("*Strange **metallic**, fruity-rose odor*"). The presence of two aromatic cycles clearly characterizes the first three, although the "strange" benzyl methyl ether possesses only one aromatic ring. Moreover, the eight odorants described by "rose-geranium" without the "metallic" note have various structures, e.g., aromatic rings, unsaturated carbon chains, and other saturated cyclic forms. Although the moieties phenethyl, geranyl and citronellyl are largely represented, a similar diverse structure characterizes the "rose-honey" and "pure-rose" odorants. Finally, "waxy-rose" odorants are generally unsaturated alcohols or aldehydes with C9 to C11 carbon chains, although several are also phenethyl esters.

It appears that it is very difficult to reduce the possible common feature(s) to a chemical group based on the number of carbon atoms, heteroatoms, or unsaturated bonds. Accounting for the distribution of the chemical structure in the tridimensional space and/or the use of sophisticated combinations of molecular descriptors is needed [98,99,106-109]. The integrated view of categorization of odorants and odor notes provides an approach that can be completed by *in silico* assays e.g. the pharmacophore approach and multivariable statistical analysis of molecular properties, leading to odorant categorizations based on molecular descriptors. The clustering of odorants based on molecular properties raises the problem of choosing molecular descriptors. Focusing on small groups of odorants that account for pairs, triplets or quadruplets of odor notes rather than a single odor note could provide a promising means to identify the pertinent molecular descriptors and the critical range of these descriptors.

Among the possible approaches, it would be beneficial to focus on the characteristics of the odorants belonging to an overlap of clusters and compare with the odorants excluded from these overlapping groups. The aim would be to identify the similarities in terms of the odor notes and determine the possibilities of linking other molecular characteristics. Special attention should be devoted to the isolated position of several odorants in groups, i.e., "unique odorants" having a peculiar odor note in a group lacking this note or odorants with rare association of odor notes [73].

Conclusion

The groups of odorants and odors presented in this work should be considered as models, meaning stimulating the thinking and providing original directions for further studies with the aim to improve the understanding of odor coding. The major outcomes take advantage of using a large dataset compared to existing studies, considering both the number of odorants and odor notes [32,43]. The categorization methods AHC and SOMs appear to be complementary in that the classification with SOMs is less driven by the most frequent odor notes than AHC. As results, the performed categorizations highlight various groups of odorants related to

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

specific associations of odor notes while the characterization of the network of odor notes affords a new exploration of their associations. The identification of numerous pairs, triplets or quadruplets of odor notes obtained by SOM constitutes especially a promising avenue to explore the space of odorants [73]. Next studies will be implemented on the basis of these results, in particular via *in silico* experiments involving chemoinformatics and bioinformatics methods.

Additional Note

The detailed tables of the results are available under request to the corresponding author.

Acknowledgements

This work has been done as part of the "CAMelIA" project with financial support of INRA-CEPIA department.

Abbreviations

AHC: agglomerative hierarchical clustering
CA: correspondence analysis
OR: olfactory receptor
MDS: multidimensional scaling
PCA: principal component analysis
SOM: self-organizing map (Kohonen map)

References

1. Stevenson RJ. Flavor binding: Its nature and cause. *Psychol Bull* 2014;140:487-510.
2. Stevenson RJ, Mahmut MK. Experience dependent changes in odour-viscosity perception. *Acta Psychol* 2011;136:60-66.
3. Bult JHF, de Wijk RA, Hummel T. Investigations on multimodal sensory integration: Texture, taste, and ortho- and retronasal olfactory stimuli in concert. *Neurosci Lett* 2007;411:6-10.
4. Thomas-Danguin T, Sinding C, Tournier C, Saint-Eve A. Multimodal interactions. In: Etiévant P, Guichard E, Salles C, Voilley A, eds. *Flavor: From food to behaviors, wellbeing and health.* Oxford, UK: Elsevier; 2016:121-141.
5. Buck L, Axel R. A novel multigene family may encode odorant receptors: A molecular basis for odor recognition. *Cell* 1991;65:175-187.
6. Malnic B, Hirono J, Sato T, Buck LB. Combinatorial receptor codes for odors. *Cell* 1999;96:713-723.
7. Poivet E, Peterlin Z, Tahirova N, et al. Applying medicinal chemistry strategies to understand odorant discrimination. *Nat Commun* 2016;7:11157.

8. Virshup AM, Contreras-Garcia J, Wipf P, Yang WT, Beratan DN. Stochastic voyages into uncharted chemical space produce a representative library of all possible drug-like compounds. *J Am Chem Soc* 2013;135:7296-7303.

9. Peterlin Z, Li YD, Sun GX, Shah R, Firestein S, Ryan K. The importance of odorant conformation to the binding and activation of a representative olfactory receptor. *Chem Biol* 2008;15:1317-1327.

10. Gerkin RC, Castro JB. The number of olfactory stimuli that humans can discriminate is still unknown. *eLife* 2015;4:e08127.

11. Triller A, Boulden EA, Churchill A, et al. Odorant-receptor interactions and odor percept: A chemical perspective. *Chem Biodivers* 2008;5:862-886.

12. Sell CS. On the unpredictability of odor. *Angew Chem-Int Edit* 2006;45:6254-6261.

13. Takane SY, Mitchell JBO. A structure-odour relationship study using EVA descriptors and hierarchical clustering. *Org Biomol Chem* 2004;2:3250-3255.

14. Klocker J, Wailzer B, Buchbauer G, Wolschann P. Aroma quality differentiation of pyrazine derivatives using self-organizing molecular field analysis and artificial neural network. *J Agric Food Chem* 2002;50:4069-4075.

15. Kovatcheva A, Golbraikh A, Oloff S, et al. Combinatorial QSAR of ambergris fragrance compounds. *J Chem Inf Comput Sci* 2004;44:582-595.

16. Zakarya D, Chastrette M, Tollabi M, Fkih-Tetouani S. Structure-camphor odour relationships using the generation and selection of pertinent descriptors approach. *Chemometr Intell Lab Syst* 1999;48:35-46.

17. Chastrette M, de Saint Laumer J. Structure-odor relationships using neural networks. *Eur J Med Chem* 1991;26:829-833.

18. Khan RM, Luk CH, Flinker A, et al. Predicting odor pleasantness from odorant structure: Pleasantness as a reflection of the physical world. *J Neurosci* 2007;27:10015-10023.

19. Snitz K, Yablonka A, Weiss T, Frumin I, Khan RM, Sobel N. Predicting odor perceptual similarity from odor structure. *PLoS Comput Biol* 2013;9:e1003184.

20. Haddad R, Khan R, Takahashi YK, Mori K, Harel D, Sobel N. A metric for odorant comparison. *Nat Methods* 2008;5:425-429.

21. Harper R. On odour classification. *Int J Food Sci Technol* 1966;1:167-176.

22. Chastrette M. Data management in olfaction studies. *SAR QSAR Environ Res* 1998;8:157-181.

23. Auffarth B. Understanding smell-the olfactory stimulus problem. *Neurosci Biobehav Rev* 2013;37:1667-1679.

24. Cleary AM, Konikel KE, Nomi JS, McCabe DP. Odor recognition without identification. *Mem Cogn* 2010;38:452-460.

25. Stevenson RJ. The acquisition of odour qualities. *Q J Exp Psychol Sect A-Hum Exp Psychol* 2001;54:561-577.

26. Stevenson RJ, Mahmut MK. The accessibility of semantic knowledge for odours that can and cannot be named. *Q J Exp Psychol* 2013;66:1414-1431.

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

27. Amoore JE. Primary odor correlated with molecular shape by scanning computer. *J Soc Cosmet Chem* 1970;21:99-106.

28. Amoore JE. Specific anosmia and concept of primary odors. *Chem Senses* 1977;2:267-281.

29. Weyerstahl P. Odor and structure. *J Prakt Chem-Chem Ztg* 1994;336:95-109.

30. Kaeppler K, Mueller F. Odor classification: A review of factors influencing perception-based odor arrangements. *Chem Senses* 2013;38:189-209.

31. Mamlouk AM, Martinetz T. On the dimensions of the olfactory perception space. In: DeSchutter E, ed. *Computational neuroscience: Trends in research 2004.* Amsterdam: Elsevier Science Bv; 2004:1019-1025.

32. Zarzo M, Stanton DT. Identification of latent variables in a semantic odor profile database using principal component analysis. *Chem Senses* 2006;31:713-724.

33. Castro JB, Ramanathan A, Chennubhotla CS. Categorical dimensions of human odor descriptor space revealed by non-negative matrix factorization. *PLoS One* 2013;8:e73289.

34. Zarzo M. A sensory 3D map of the odor description space derived from a comparison of numeric odor profile databases. *Chem Senses* 2015;40:305-313.

35. Ma LM, Qiu Q, Gradwohl S, et al. Distributed representation of chemical features and tunotopic organization of glomeruli in the mouse olfactory bulb. *Proc Natl Acad Sci U S A* 2012;109:5481-5486.

36. Auffarth B, Gutierrez-Galvez A, Marco S. Statistical analysis of coding for molecular properties in the olfactory bulb. *Front Syst Neurosci* 2011;5:62.

37. Auffarth B, Gutierrez-Galvez A, Marco S. Continuous spatial representations in the olfactory bulb may reflect perceptual categories. *Front Syst Neurosci* 2011;5:82.

38. Rolls ET, Kringelbach ML, de Araujo IET. Different representations of pleasant and unpleasant odours in the human brain. *Eur J Neurosci* 2003;18:695-703.

39. Howard JD, Plailly J, Grueschow M, Haynes JD, Gottfried JA. Odor quality coding and categorization in human posterior piriform cortex. *Nat Neurosci* 2009;12:932-U158.

40. Qu LP, Kahnt T, Cole SM, Gottfried JA. De novo emergence of odor category representations in the human brain. *J Neurosci* 2016;36:468-478.

41. Zarzo M. Relevant psychological dimensions in the perceptual space of perfumery odors. *Food Qual Prefer* 2008;19:315-322.

42. Zarzo M, Stanton DT. Understanding the underlying dimensions in perfumers' odor perception space as a basis for developing meaningful odor maps. *Atten Percept Psychophys* 2009;71:225-247.

43. Martinez-Mayorga K, Peppard TL, Yongye AB, Santos R, Giulianotti M, Medina-Franco JL. Characterization of a comprehensive flavor database. *J Chemometr* 2011;25:550-560.

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

44. Schiffman SS. Characterization of odor quality utilizing multidimensional scaling techniques. In: *Odor quality and chemical structure.* Vol 148. American Chemical Society; 1981:1-21.

45. Youngentob SL, Johnson BA, Leon M, Sheehe PR, Kent PF. Predicting odorant quality perceptions from multidimensional scaling of olfactory bulb glomerular activity patterns. *Behav Neurosci* 2006;120:1337-1345.

46. Higuchi T, Shoji K. Multidimensional scaling of fragrances: A comparison between the verbal and non-verbal methods of classifying fragrances. *Jpn Psychol Res* 2004;46:10-19.

47. Mamlouk AM, Chee-Ruiter C, Hofmann UG, Bower JM. Quantifying olfactory perception: Mapping olfactory perception space by using multidimensional scaling and self-organizing maps. *Neurocomputing* 2003;52-4:591-597.

48. Chastrette M, Elmouaffek A, Sauvegrain P. A multidimensional statistical study of similarities between 74 notes used in perfumery. *Chem Senses* 1988;13:295-305.

49. Lawless HT. Exploration of fragrance categories and ambiguous odors using multidimensional-scaling and cluster-analysis. *Chem Senses* 1989;14:349-360.

50. Prost C, Le Guen S, Courcoux P, Demaimay M. Similarities among 40 pure odorant compounds evaluated by consumers. *J Sens Stud* 2001;16:551-565.

51. Audouze K, Ros F, Pintore M, Chretien JR. Prediction of odours of aliphatic alcohols and carbonylated compounds using fuzzy partition and self organising maps (SOM). *Analusis* 2000;28:625-632.

52. Harada Y, Tomoki K, Kanzaki R, Nakamoto T. Response prediction of an insect's olfactory receptor neuron by using structural parameters of odorant and self-organizing map. *IEEE Sens J* 2016;16:580-585.

53. Flavor-Base 9th Ed. (2013). Leffingwell & Associates, http://www.leffingwell.com/flavbase.htm.

54. Audouze K, Tromelin A, Le Bon AM, et al. Identification of odorant-receptor interactions by global mapping of the human odorome. *PLoS One* 2014;9:e93037.

55. Sanz G, Thomas-Danguin T, Hamdani EH, et al. Relationships between molecular structure and perceived odor quality of ligands for a human olfactory receptor. *Chem Senses* 2008;33:639-653.

56. Hummel T. Retronasal perception of odors. *Chem Biodivers* 2008;5:853-861.

57. Scott JW, Acevedo HP, Sherrill L, Phan M. Responses of the rat olfactory epithelium to retronasal air flow. *J Neurophysiol* 2007;97:1941-1950.

58. Gautam SH, Verhagen JV. Retronasal odor representations in the dorsal olfactory bulb of rats. *J Neurosci* 2012;32:7949-7959.

59. Furudono Y, Cruz G, Lowe G. Glomerular input patterns in the mouse olfactory bulb evoked by retronasal odor stimuli. *BMC Neurosci* 2013;14:45.

60. Abdolmaleki A, Ghasemi JB, Shiri F, Pirhadi S. Application of multivariate linear and nonlinear calibration and classification methods in drug design. *Comb Chem High Throughput Screen* 2015;18:795-808.

61. Pirhadi S, Shiri F, Ghasemi JB. Multivariate statistical analysis methods in QSAR. *RSC Adv* 2015;5:104635-104665.

62. Stevenson RJ. Phenomenal and access consciousness in olfaction. *Conscious Cogn* 2009;18:1004-1017.

63. Kohonen T. Self-organized formation of topologically correct feature maps. *Biol Cybern* 1982;43:59-69.

64. Kohonen T. Essentials of the self-organizing map. *Neural Netw* 2013;37:52-65.

65. Melssen W, Wehrens R, Buydens L. Supervised Kohonen networks for classification problems. *Chemometr Intell Lab Syst* 2006;83:99-113.

66. R Core Team (2013). R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

67. Wehrens R, Buydens LMC. Self- and super-organizing maps in R: The kohonen package. *J Stat Softw* 2007;21:1-19.

68. Timm NH. Cluster analysis and multidimensional scaling. In: Timm NH, ed. *Applied multivariate analysis.* New York, NY: Springer New York; 2002:515-555.

69. France SL, Carroll JD. Two-way multidimensional scaling: A review. *IEEE Trans Syst Man Cybern Part C-Appl Rev* 2011;41:644-661.

70. Lee JA, Renard E, Bernard G, Dupont P, Verleysen M. Type 1 and 2 mixtures of Kullback-Leibler divergences as cost functions in dimensionality reduction based on similarity preservation. *Neurocomputing* 2013;112:92-108.

71. Kelley LA, Gardner SP, Sutcliffe MJ. An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. *Protein Eng* 1996;9:1063-1065.

72. Shannon P, Markiel A, Ozier O, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498-2504.

73. Chastrette M, De Saint Laumer JY, Sauvegrain P. Analysis of a system of description of odors by means of 4 different multivariate statistical-methods. *Chem Senses* 1991;16:81-93.

74. Yongye AB, Bender A, Martinez-Mayorga K. Dynamic clustering threshold reduces conformer ensemble size while maintaining a biologically relevant ensemble. *J Comput-Aided Mol Des* 2010;24:675-686.

75. Milgram S. The small-wordl problem. *Psychol Today* 1967;1:61-67.

76. Chastrette M. Trends in structure-odor relationships. *SAR QSAR Environ Res* 1997;6:215-254.

77. Rouby C, Thomas-Danguin T, Sicard G, et al. Influence du contexte sémantique sur la performance d'identification d'odeurs. *Psychol Fr* 2005;50:225-239.

78. Castro JB, Seeley WP. Olfaction, valuation, and action: Reorienting perception. *Front Psychol* 2014;5:299.

79. Urdapilleta I, Giboreau A, Manetta C, Houix O, Richard JF. The mental context for the description of odors: A semantic space. *Eur Rev Appl Psychol-Rev Eur Psychol Appl* 2006;56:261-271.

80. Chrea C, Valentin D, Sulmont-Rosse C, Nguyen DH, Abdi H. Semantic, typicality and odor representation: A cross-cultural study. *Chem Senses* 2005;30:37-49.

81. Wnuk E, Majid A. Revisiting the limits of language: The odor lexicon of maniq. *Cognition* 2014;131:125-138.

82. Agapakis CM, Tolaas S. Smelling in multiple dimensions. *Curr Opin Chem Biol* 2012;16:569-575.

83. Egana JI, Aylwin ML, Maldonado PE. Odor response properties of neighboring mitral/tufted cells in the rat olfactory bulb. *Neuroscience* 2005;134:1069-1080.

84. Davison IG, Katz LC. Sparse and selective odor coding by mitral/tufted neurons in the main olfactory bulb. *J Neurosci* 2007;27:2091-2101.

85. Fantana AL, Soucy ER, Meister M. Rat olfactory bulb mitral cells receive sparse glomerular inputs. *Neuron* 2008;59:802-814.

86. Soucy ER, Albeanu DF, Fantana AL, Murthy VN, Meister M. Precision and diversity in an odor map on the olfactory bulb. *Nat Neurosci* 2009;12:210-220.

87. Cleland TA. Early transformations in odor representation. *Trends Neurosci* 2010;33:130-139.

88. Shepherd GM, Chen WR, Willhite D, Migliore M, Greer CA. The olfactory granule cell: From classical enigma to central role in olfactory processing. *Brain Res Rev* 2007;55:373-382.

89. Mombaerts P. The human repertoire of odorant receptor genes and pseudogenes. *Annu Rev Genomics Hum Genet* 2001;2:493-510.

90. Bushdid C, Magnasco MO, Vosshall LB, Keller A. Humans can discriminate more than 1 trillion olfactory stimuli. *Science* 2014;343:1370-1372.

91. Meister M. On the dimensionality of odor space. *eLife* 2015;4:e07865.

92. Secundo L, Snitz K, Weissler K, et al. Individual olfactory perception reveals meaningful nonolfactory genetic information. *Proc Natl Acad Sci U S A* 2015;112:8750-8755.

93. Malnic B. Searching for the ligands of odorant receptors. *Mol Neurobiol* 2007;35:175-181.

94. Thomas-Danguin T, Sinding C, Romagny S, et al. The perception of odor objects in everyday life: A review on the processing of odor mixtures. *Front Psychol* 2014;5:504.

95. Zarzo M. Underlying dimensions in the descriptive space of perfumery odors: Part ii. *Food Qual Prefer* 2015;43:79-87.

96. Shahlaei M. Descriptor selection methods in quantitative structure-activity relationship studies: A review study. *Chem Rev* 2013;113:8093-8103.

97. Bajorath J. Selected concepts and investigations in compound classification, molecular descriptor analysis, and virtual screening. *J Chem Inf Comput Sci* 2001;41:233-245.

Flavour and Fragrance Journal, John Wiley & Sons, Ltd, http://onlinelibrary.wiley.com/doi/10.1002/ffj.3430

98. Tseng YJ, Hopfinger AJ, Esposito EX. The great descriptor melting pot: Mixing descriptors for the common good of QSAR models. *J Comput-Aided Mol Des* 2012;26:39-43.

99. Shao CY, Chen SZ, Su BH, Tseng YFJ, Esposito EX, Hopfinger AJ. Dependence of QSAR models on the selection of trial descriptor sets: A demonstration using nanotoxicity endpoints of decorated nanotubes. *J Chem Inf Model* 2013;53:142-158.

100. Maggiora G, Vogt M, Stumpfe D, Bajorath J. Molecular similarity in medicinal chemistry. *J Med Chem* 2014;57:3186-3204.

101. Maldonado AG, Doucet JP, Petitjean M, Fan BT. Molecular similarity and diversity in chemoinformatics: From theory to applications. *Mol Divers* 2006;10:39-79.

102. Willett P. The calculation of molecular structural similarity: Principles and practice. *Mol Inf* 2014;33:403-413.

103. Khanna V, Ranganathan S. Molecular similarity and diversity approaches in chemoinformatics. *Drug Dev Res* 2011;72:74-84.

104. Bajusz D, Racz A, Heberger K. Why is tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J Cheminformatics* 2015;7:20.

105. Schoenauer S, Schieberle P. Structure-odor activity studies on monoterpenoid mercaptans synthesized by changing the structural motifs of the key food odorant 1-p-menthene-8-thiol. *J Agric Food Chem* 2016;64:3849-3861.

106. Koutsoukas A, Paricharak S, Galloway W, et al. How diverse are diversity assessment methods? A comparative analysis and benchmarking of molecular descriptor space. *J Chem Inf Model* 2014;54:230-242.

107. Keefer CE, Kauffman GW, Gupta RR. Interpretable, probability-based confidence metric for continuous quantitative structure-activity relationship models. *J Chem Inf Model* 2013;53:368-383.

108. Tomal JH, Welch WJ, Zamar RH. Exploiting multiple descriptor sets in QSAR studies. *J Chem Inf Model* 2016;56:501-509.

109. Maggiora GM, Bajorath J. Chemical space networks: A powerful new paradigm for the description of chemical space. *J Comput-Aided Mol Des* 2014;28:795-802.