 doi:10.20944/preprints202106.0278.v1

Type of the Paper (Article, Review, Communication, etc.)

Sniff species - SURMOF based sensor array discriminate aromatic plants beyond the genus level

Salih Okur[1,*], Chun Li [1], Zejun Zhang[1], Sahi Vaidurya Pratap[2], Mohammed Sarheed[2], Adnan Kanbar[2], Leonardo Franke[3], Felix Geislhöringer[3], Lars Heinke[1], Uli Lemmer[3,4], Peter Nick[2], Christof Wöll[1]

[1] Institute of Functional Interfaces (IFG), Karlsruhe Institute of Technology (KIT), Hermann-von-Helmholtz-Platz 1, 76344, Eggenstein-Leopoldshafen, Germany; salih.okur@kit.edu, zejun.zhang@partner.kit.edu, lars.heinke@kit.edu, christof.woell@kit.edu

[2] Karlsruhe Institute of Technology, Botanical Institute, Molecular Cell Biology, Fritz-Haber-Weg, 76131 Karlsruhe, Germany; sahi.vaidurya@kit.edu, mu.sarheed@gmail.com, adnan.kanbar@kit.edu, peter.nick@kit.edu

[3] Light Technology Institute, Karlsruhe Institute of Technology, Engesserstraße 13, 76131 Karlsruhe, Germany; leonard.franke@kit.edu, , felix.geislhoeringer@kit edu uli.lemmer@kit.edu

[4] Institute of Microstructure Technology, 223, Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, Eggenstein-Leopoldshafen, Karlsruhe, 76344, Germany; uli.lemmer@kit.edu

* Correspondence: salih.okur@kit.edu; Tel.: (+49-721-608-26078)

Abstract: The Lamiaceae belong to the species-richest families of flowering plants and harbor many species used as herbs or for medicinal applications, such as Basils or Mints. Evolution of this group has been driven by chemical speciation, mainly of Volatile Organic Compounds (VOCs). The commercial use of these plants is characterized by a large extent of adulteration and surrogation. To authenticate and discern the species, is, thus, relevant for consumer safety, but usually requires cumbersome analytics, such as Gas Chromatography, often to be coupled with Mass Spectroscopy. We demonstrate here that quartz-crystal microbalance (QCM)-based electronic noses provide a very cost-efficient alternative, allowing for a fast, automated discrimination of scents emitted from leaves of different plants. To explore the range of this strategy, we used leaf material from four genera of Lamiaceae along with Lemongrass as similarly scented, but non-related outgroup. In order to unambiguously differentiate the scents from the different plants, the output of the 6 different SURMOF/QCM sensors was analyzed using machine learning (ML) methods, together with a thorough statistical analysis. The exposure and purging datasets (4 cycles) obtained from a QCM-based, low-cost homemade portable e-Nose were analyzed with Linear Discriminant Analysis (LDA) classification model. Prediction accuracies with repeating test measurements reached values of up to 90%. We show that it is not only possible to discern and identify plants on the genus level, but even to discriminate closely related sister clades within a genus (Basil), demonstrating that e-Noses are a powerful technology to safeguard consumer safety against the challenges of globalized trade.

Keywords: Basil, Mint, Plant volatiles, Electronic Nose, Principal Component Analysis, Linear Discriminant Analysis, k-Nearest-Neighbors Analysis.

1. Introduction

Plants have developed subtle mechanisms to defend and adapt themselves against biotic and abiotic stress factors. One of the ways plants have evolved to protect themselves is by producing volatile organic compounds (VOCs) as part of their essential oils [1]. These plant essential oils, often with monoterpenes as primary components, accumulate in different organs such as leaves, bark, wood, roots, in flowers or fruits, sometimes in specialized glands, but also in lysogenic or schizogenic oil ducts [2]. These VOCs are the base for

 doi:10.20944/preprints202106.0278.v1

human use of aromatic plants, both as spices and for medicinal applications. The multitude of VOC profiles has been shaping entire cultures, cuisines, and medical traditions.

One of the most prodigious plant families in this context are the Lamiaceae. With more than 7,000 species belonging to more than 200 genera, they belong to the taxonomically most challenging and diverse groups in the flowering plants [3]. They secrete complex bouquets of VOCs from their glandular hairs and scales. These bouquets are often specific for a given species, and due to their interaction with pollinator insects, might have been even one of the drivers for the immense complexity of this family. Sometimes even within a species, different chemotypes exist. Commercially relevant plants, such as the Mints or the Basils, belong to this group, and are often part of novel food trends that are fueled by their reputation to be so called "superfoods" [4].

With growing concerns for holistic approaches towards health, there has been a trend in usage of supplements and plant-based products which have their roots in traditional medical systems like Ayurveda or Traditional Chinese medicine. With the growing popularity of Ayurveda in Europe, products containing *Ocimum tenuiflorum* L. are readily available in supermarkets. *O tenuiflorum* (Holy Basil or Tulsi) has been used for treating ailments like pains in the joints, headache, cold, fever, and also insect bites [5-9]. In addition, Holy Basil has also been recommended to relieve stress [10], and to reduce the effects of diabetes mellitus [11]. Due to the benefits attributed to Holy Basil, the volume of its market in the West is progressively increasing [12]. This accentuates the problem of authentication and identification of commercial products declared to contain Tulsi [13]. The genus *Ocimum* is composed of many species, several of which are commonly traded. However, each species is endowed with a unique chemical profile that is mostly genetically determined [12]. To authenticate *O. tenuiflorum* by microscopic diagnostic is possible [5], although limited, especially in commercial products that are often processed [14]. In addition, it is possible to discriminate true *O. tenuiflorum* from other Basils on the basis of DNA barcoding [12], a rather expensive and time-consuming process. An alternative would be the detection of the different odorous contents, since the spectrum of volatile organic compounds (VOCs) emitted by *Ocimum* species is unique [15]. In fact, a trained human nose is able to distinguish *O. tenuiflorum* from other Basils, due to the emission of specific patterns of volatile phenylpropanoids [16].

The profile of VOCs emitted by a plant is unique to a particular species and provides a unique way to identify plant species contained in commercial products. However, the chemical analysis necessary for such an identification of gaseous compounds, is a time consuming and costly undertaking, requiring gas chromatography (GC). Since a human nose can discriminate those species, sensor arrays, also referred to as electronic noses, might offer a cost-efficient, convenient, and fast alternative. [17, 18]. Such e-Noses have been successfully used for discriminating different medicinal plants on the basis of the different emitted VOCs [19, 20]. Quartz crystal microbalance (QCM)-based sensor arrays have also been previously used to differentiate between plants from the Lamiaceae family [21, 22]. For instance, a multichannel QCM (MQCM) with molecularly imprinted polystyrene membranes has been used for discrimination of terpenes emanated from freshly

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2021 doi:10.20944/preprints202106.0278.v1

dried species belonging to the Lamiaceae, such as Rosemary (*Rosmarinus officinalis* L.), Sweet Basil (*Ocimum basilicum*), and Common Sage (*Salvia officinalis*) [21].

A crucial point when fabricating QCM-based sensors is the detector material used for coating the QCM substrate. In this context, reticular compounds like metal-organic frameworks, or MOFs, carry a huge potential. These porous materials can be easily modified to yield different responses to VOCs, thus allowing to fabricate sensor arrays with each component showing different sensitivities.

In our previous work we have used an e-Nose to differentiate between different species of Mints or VOCs isolated from them [23]. In the present study we focus on QCM sensors coated with MOF thin films. Six different MOFs were investigated, including HKUST-1, Cu(BDC), Cu(BPDC), $Cu_2(DCam)_2$(dabco), $Cu_2(DCam)_2$(BiPy), and $Cu_2(DCam)_2$(BiPyB) [24]. MOF thin films were deposited using layer-by-layer methods, yielding co-called SURMOFs (surface anchored MOFs) [25]. To validate the performance of these SURMOF-based QCM arrays, we challenged them by testing the ability of these sensors to discriminate different chemotypes of closely related species (the two mints *Mentha aquatica* and *Mentha suaveolens*, and the closely related Korean Mint, *Agastache rugosa*), against the more distant Lemon Balm (*Melissa officinalis*), and the unrelated, but similarly scented Lemon Grass (*Cymbopogon citratus*). In addition, we were using three accessions of Basil (*Ocimum campechianum*, Amazonian Basil versus two accessions of *O. tenuiflorum*, Holy Basil, Tulsi). In contrast to Sweet Basil, which had been addressed by e-Noses previously [22], we wanted to test, to what extent it would be possible to discern true O. tenuiflorum from closely related sister species, since O. campechianum is a member of the closest known haplotype known for the genus Ocimum [12]. All accessions were selected from the authenticated reference plant collection of the Botanic Garden of the Karlsruhe Institute of Technology. The response times of the QCM sensors upon exposure to and removal of a particular scent were determined using Nonlinear Least-Square (NLS) fits to an exponential rise (or fall) function and were found to amount to less than 1 min. [26, 27] The exposure and purging datasets (4 cycles) obtained from a low-cost homemade portable e-Nose were analyzed using machine learning approaches, employing three different classification methods, i.e. PCA, LDA, and k-NN techniques. [22] The first and second cycles of the data sets were used for training and the following repeated cycles were used as unknown data for prediction. A statistical analysis revealed that more than 90% classification accuracy could be achieved within 8 different scent classes from 3 different plant leaves in very short time (less 6 min). Prediction accuracies with repeating test measurements reached up to 90% for LDA and k-NN from unknown data sets.

2. Materials and Methods

2.1 Plant material

The present study included freshly collected 3 different species of Basil, 4 different species of Mints, and a control sample of Lemon Grass grown in the Botanical Institute of Karlsruhe Institute of Technology (KIT), Germany (Table 1). The scents were collected from 3 g weight of fresh leaves from Ocimum campechianum Mill., Cymbopogon citratus, Ocimum tenuiflorum L., Melissa officinalis L, Mentha aquatica, Agastache rugosa,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2021 doi:10.20944/preprints202106.0278.v1

Mentha suaveolens. The abbreviations used throughout the text were defined in the Table 1.

Table 1: Accessions used in this study. The voucher number gives the code, under which the plants are available in the Botanical Garden of the KIT. The abbreviations used in the text are also given. The number of leaves harvested to reach 3 g is indicated as well.

Plant	Common name	Abbreviatio	KIT voucher
Ocimum campechianum Mill.	Amazon Basil	Bas7564	7564
Ocimum tenuiflorum L.	Tulsi	Bas5751	5751
Ocimum tenuiflorum L.	Krishna Tulsi	Bas8257	8257
Cymbopogon citratus	Lemon Grass	LemGra	5722
Melissa officinalis L	Common Balm	MeliOfL	4643
Mentha aquatica L.	Water Mint	MintAQ	8680
Agastache rugosa	Korean Mint	MintAR	7576
Mentha suaveolens	Apple Mint	MintSU	3638

2.2 Chemicals and fabrication of QCM sensors

Commercially available AT-cut HC-49U type 200 nm silver coated 10 MHz quartz crystals with 5 mm electrode radius (J. Walter Thompson, Shanghai, China) were used as QCM electrodes. Six QCM sensors were coated with six different SURMOFs, HKUST-1, Cu(BDC), Cu(BPDC), $Cu_2(DCam)_2$ (dabco), $Cu_2(DCam)_2$ (BiPy) $Cu_2(DCam)_2$(BiPyB) to establish the sensor array comprising the e-Nose system as shown in Figure 1.

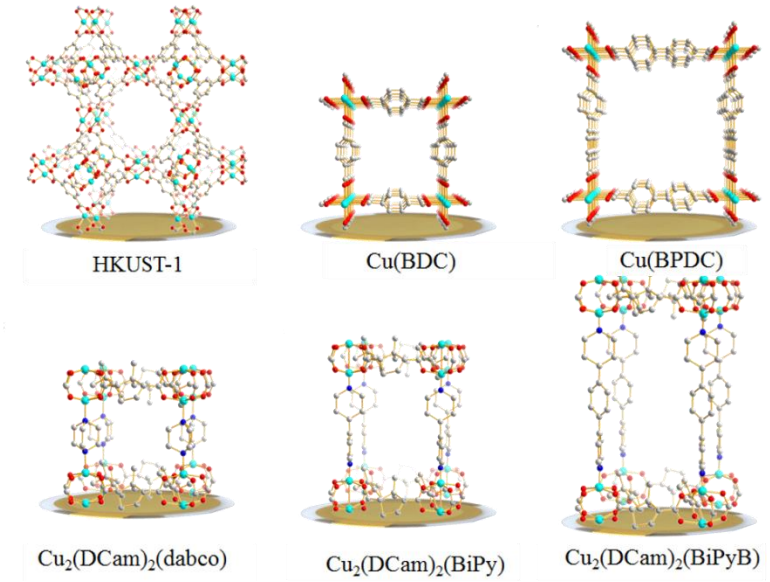


Figure 1. The six different SURMOFs structures of the sensor array used in the e-Nose system, namely HKUST-1, Cu(BDC), Cu(BPDC), $Cu_2(DCam)_2$(dabco), $Cu_2(DCam)_2$(BiPy), $Cu_2(DCam)_2$(BiPyB).

DCam is a layer linker to produce pillared-layer MOF structures. The pillar linkers are diazabicyclo[2.2.2]octane (dabco), 4,4'-bipyridyl (BiPy), and 1,4-bis(4-pyridyl)benzene (BiPyB), respectively. BDC is benzene-1,4-dicarboxylate and BPDC is biphenyl-4,4'-dicarboxylate for the Cu(BDC), and Cu(BPDC) MOF structures [23, 24].

Prior to SURMOF deposition, the QCM substrates were functionalized by an O2 plasma treatment for 30 min. All films were prepared using 30 synthesis cycles. The SURMOF synthesis details were provided in the Supporting Information part of our previous work [24]. X-ray diffraction (XRD) was used to characterize the SURMOFs thin films grown on the QCM sensors, the diffractograms are shown in Figure SI-1. The XRD data reveals the presence of crystalline, oriented MOF thin films with the targeted structure.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2021 doi:10.20944/preprints202106.0278.v1

2.3 Data acquisition with the e-Nose

Fig. 2 shows a schematic view of the working principle of the 6-channel low-cost homemade portable e-Nose system used for discrimination of scents of Basil/Mint leaves. The sensor array and a humidity/temperature sensor were placed inside a 3D-printed head space in a cylindrical form. For the QCM data acquisition, 5 V/16 MHz ATMega32U4 microcontrollers and open-source Pierce oscillator circuits designed by openQCM have been used [28] to read the frequency change. Temperature and humidity were measured with Adafruit HHTU21D-F temperature & humidity sensor breakout board. The temperature of the chamber was kept constant at 25±0.5 °C. The software package MATLAB has been used to record and analyze the data.

3g of freshly collected leaves from each species of Basil and Mints were inserted separately into a 100 mL glass vial. The emanating VOCs emitted from the fresh plant leaves inside the bottle were circulated through the sensor array with a 3W small diaphragm pump with a small flow rate 0.1 L/m while the inlet and outlet of N_2 gas closed. The surface of the sensing thin films inside the head space was activated by purging with N2. This process led to a removal of residual compounds within the SURMOF pores. For each Basil/Mint scent accession, the change in resonance frequency was recorded 6 min for each cycle with 2 min exposure for adsorption, and subsequently, 4 min of purging during cleaning with dry N2 gas. The exposure and purging cycles were repeated 4 times.

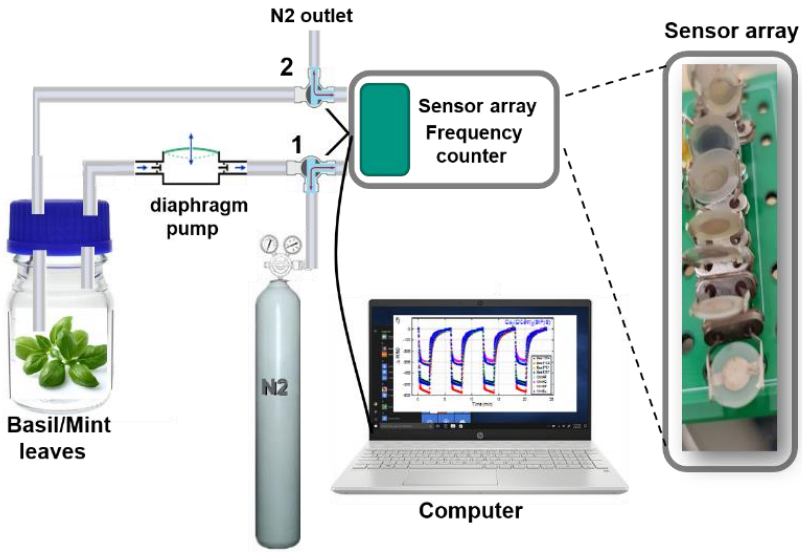


Figure 2. A schematic view of the working principle of the 6-channel low-cost homemade portable E-nose system used for discrimination of scents of Basil/Mint leaves.

MOFs are highly porous materials with huge specific surfaces [29]. SURMOFs coated on a QCM will adsorb the VOCs on the outer surface, as well as inside the pores (see Fig. 3)[24]. Of course, for the latter it is required that the pores and channels inside the MOF are sufficiently large to accommodate diffusion of the VOC into the pore system. A quantitative determination of the total amount of a particular VOC loaded into a MOF thin film can be carried out using a QCM. In the present case, the scent emitted from a plant consist of a large variety of different compounds, their number typically exceeding 20[30-32].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2021 doi:10.20944/preprints202106.0278.v1

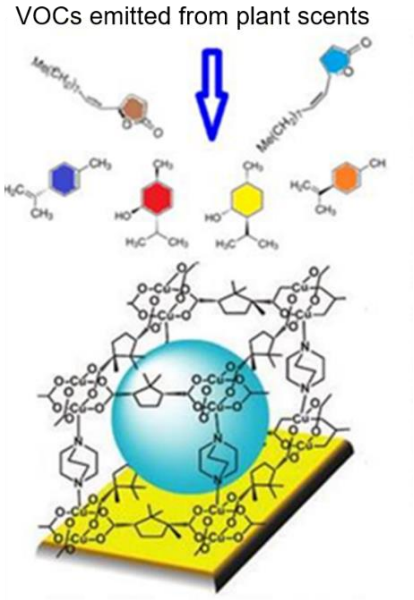


Figure 3: MOFs. Highly porous SURMOFs with huge specific surfaces coated on a QCM will adsorb the VOCs on the outer surface, as well as inside the pores [33].

2.4. Data Analysis and Classification

The QCM-response after exposure to the plant scent and after purging with dry nitrogen is shown in Fig. 4. Is was found that single-component rise and fall functions well described the QCM data for times up to 60 a after start of exposure/purging. At later times there is a linear behavior, indicating diffusion into and out of the pores [34-36]. The frequency shift of the QCM sensors is directly proportional to the change of the absorbed mass according to the Sauerbrey relation [37], e.g. $\Delta F(t) = -C\Delta m(t)$, where C is the QCM mass sensitivity constant, which is related to structural and physical properties of the piezo electrical quartz sensor material. The frequency response times were calculated from Non-linear Least-Square (NLS) fits of the QCM response to an exponential rise function [26, 27] in the time interval between 5s and 60s.

The QCM signal drop observed after removing a particular scent has been determined by a NLS fit to a exponential decay function in the time interval between 125s and 180s using the following expression:

$$\Delta F(t) = \Delta F_{max}\,(1 - e^{t/\tau_{ads}}) \qquad \text{(Adsorption)}, \qquad (1)$$

$$\Delta F(t) = \Delta F_{max}\, e^{-t/\tau_{des}} \qquad \text{(Desorption)} \qquad (2)$$

where τ_{ads} and τ_{des} are the relaxation time related to the association constant of the adsorption process and desorption process respectively.

During the discrimination analysis of the scents, the first cycle of the loading/purging curve was used for training while the other 3 repeated cycles were used for testing and prediction the eight different class of scents emitted from the plant leaves as a source. The exposure data at highest response between 1 min - 2 min just before beginning of the purging process were cut to be used as training dataset for the discrimination accuracy calculations. Similarly, for prediction tests, one minute of exposure data at highest response were cut for the other cycles e.g. the data between 7-8 min for 2nd. cycle and 13-14 min for 3rd. cycle, 19-20 min for 4th. cycle. Three different classification algorithms were

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2021 doi:10.20944/preprints202106.0278.v1

tested: Principle Component Analysis (PCA), Linear Discrimination Analysis (LDA), and Nearest Neighbors (k-NN) using the scripts written in MATLAB.

PCA is an unsupervised machine learning method that uses dimension reduction, and data visualization [38, 39]. This algorithm transforms the original dataset into a new set of so-called Principal Components (PC). Hence, a large number of datasets is effectively compressed in a smaller set of PC variables. The 3D-PCA image classification can be obtained by projecting principal component scores in x, y and z axes. This makes it possible to visualize the separation of classes or clusters.

On the other hand, LDA is a supervised machine learning method that maximizes the discrimination among known categories by creating a new linear axis and projecting the data points on that axis. LDA and PCA are similar classification techniques. [40] They both compute the linear combinations of variables which best explain the data. LDA give a difference model between the classes of the data. PCA, in contrast, does not consider any difference in class. LDA implements the data with independent variables of continuous categorical observation. The objective of LDA is to find the projection hyperplane that minimizes the interclass variance and maximizes the separation distance between the projected classes. LDA has also been used in the literature due to its relatively fast model computation. Therefore, in this paper we evaluate the effectiveness of the model, in classifying the scents of two different plant species with 8 different classes.

In k-NN discrimination analysis, k nearest neighbors is a simple algorithm that classifies new cases by scanning the distances of the classified elements of the nearest neighbors by comparing all the stored known cases. k-NN has been used for statistical estimation and pattern recognition. k is a parameter that defines the number of nearest neighbors before rendering a classification decision. In this paper we also evaluate the effect of the number of nearest neighbors on the classification and prediction accuracy of the scents of two different plant genera with 8 different classes.

3. Results and discussions

3.1 Sensor array responses

Figure 4 shows the response of a QCM sensor coated with a SURMOF of the type Cu2(DCam)2(BiPyB) after exposure to the scents emitted from different plants, and after purging with nitrogen gas. For all scents, the sensor reached saturation frequency (of 99.3% of ⌾Fmax) in average 29±8 s after start of exposure, and - after purging - recovered back (to 0.7% of ⌾Fmax) in average 54±4 s as shown in Figure SI-3 and Table SI-1 in Supporting Information Appendix. The frequency response times were calculated from Nonlinear Least square fits to an exponential rise or decrease function. Among all SURMOF-based QCM sensors, the frequency response time was fastest for HKUST-1, 27.5±10.8 s for adsorption and 44.0±1.1 s for desorption. The longest frequency response time was observed for Cu2(Dcam)2 (BiPy), with a value of 69.3±7.3 s for adsorption and 66.7±11.0 s for desorption. In general, adsorption process is faster than desorption showing a strong affinity between the sensing MOFs and the scent molecules emitted from both Basil and Mint leaves.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2021 doi:10.20944/preprints202106.0278.v1

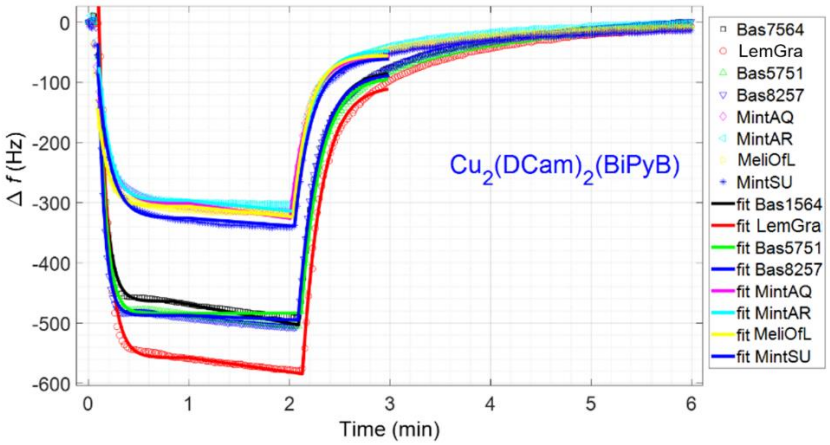


Figure 4. An example illustration of the one cycle response of the sensor array, the change in resonance frequency of the sensing film Cu2(DCam)2(BiPyB) due to adsorption with 2 min exposure, and subsequently, 4 min purging during cleaning with dry N2 gas.

The resonance frequency shifts of the sensor array consisting of 7 QCM sensors coated with all the 7 different sensing materials (see Table 1 for abbreviations) during 4 cycles of exposure to the individual Basil/Mint leaves are shown in Figure SI-2 in Supporting Information Appendix.

For the comparison of the effect of SURMOF modification on bare Ag coated QCM sensors, the maximum scale of the plots were kept constant at -600 Hz. For all MOF-materials, the QCM sensors showed the highest response to LemGra and the lowest response to MintAQ. Interestingly, the responses of all sensors to both Basil and Mint species can be separated into two categories. The red circle with the highest response belongs to the scent of the control sample LemGra.

A radar plot (Fig. 5) of the maximum frequency shift response for the sensor arrays shows that each sensor responds differently. In case of the latter, for all scents the response was very small, less than -10 Hz, as expected. The maximum frequency shift response values of the different SURMOF-based QCM sensors for the different scents shown in the radar plot are also listed in Table 2. The highest response of around -600 Hz comes from the sensor coated with a Cu2(Dcam)2 (BiPyB) SURMOF thin film. The lowest response was obtained from the sensor coated with Cu2(Dcam)2(dabco). The large difference in response between different MOFs for the same scent results from the different chemical structure of the various SURMOFs. In addition to the chemical structure, also the different pores-sizes can have an influence. Since each scent contains many different VOCs, a precise identification of the underlying mechanisms is beyond the scope of this article.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2021 doi:10.20944/preprints202106.0278.v1

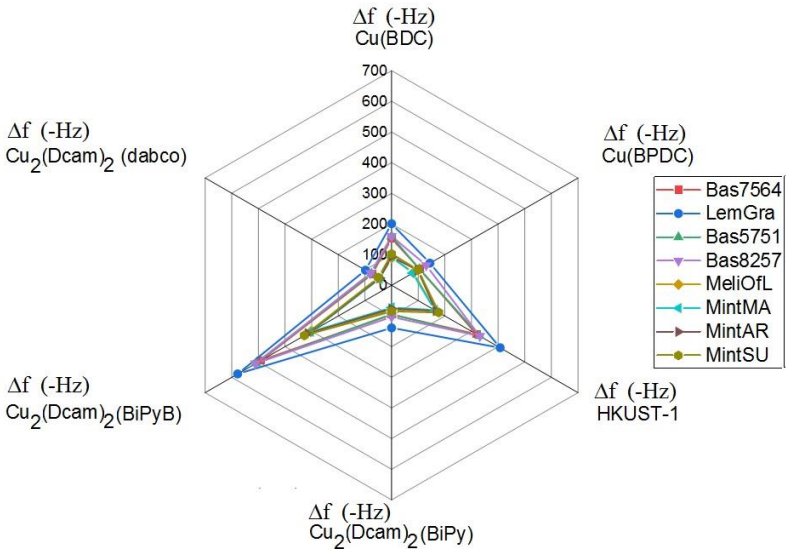


Figure 5. Radar plot of the maximum frequency shifts responses of sensor array.

Table 2. The maximum frequency shifts response values of the sensor arrays shown in the radar plot in Figure 5.

	Bas7564	LemGra	Bas5751	Bas8257	MeliOfL	MintMA	MintAR	MintSU
	⌾f(-Hz)	⌾f(-Hz)	⌾f(-Hz)	⌾f(-Hz)	⌾f(-Hz)	⌾f(-Hz)	⌾f(-Hz)	⌾f(-Hz)
Cu(BDC)	153,1	200,9	159,4	160,0	89,5	94,4	93,1	102,4
Cu(BPDC)	105,4	144,0	104,8	130,1	82,6	78,9	98,2	100,1
HKUST-1	319,1	407,6	330,6	331,3	163,3	162,1	166,7	177,3
$Cu_2(Dcam)_2$ (BiPy)	95,9	139,0	95,0	103,0	80,4	72,5	76,1	84,8
$Cu_2(Dcam)_2$ (BiPyB)	492,6	578,3	509,0	510,9	302,8	303,4	314,4	327,6
$Cu_2(Dcam)_2$ (dabco)	72,9	98,2	75,4	79,0	44,7	44,5	45,9	49,3

3.2. Principal Component Analysis (PCA)

Fig. 6 shows a 3D projection of the principal component scores in x, y and z axes calculated using Principal Component Analysis from 451 measurements for the eight different scents. These components group into clearly separated clusters. Interestingly, the two accessions from Mentha clustered with Agastache rugosa (Korean Mint, AR, belonging to a neighbouring clade), but also with the more distantly related Melissa officinalis (MeliOfL). The three Basil scents were clearly separated – here, the two accessions for O. tenuiflorum (Tulsi) were close to each other, but unequivocally resolved from the closely related O. campechianum. This is astonishing, because the latter species belongs to the sister clade closest with O. tenuiflorum within the entire genus. The other surprise comes from the complete separation of Lemon Grass (LemGra) from Melissa officinalis (MeliOfL), since both species have a very similar lemon-like scent and are often used for mutual surrogation in commercial samples. The clear separation indicates that the e-Nose can pick up even subtle differences in the VOC profile that would go unnoticed for most human noses.

The sum of the three scores of the total variance explained by each principal component in 3D plot given in Fig. 6 is equal to 96.3%. By introducing the 4th. and 5th. PCA components, the visual PCA discrimination accuracy reaches to 99,8%.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2021 doi:10.20944/preprints202106.0278.v1

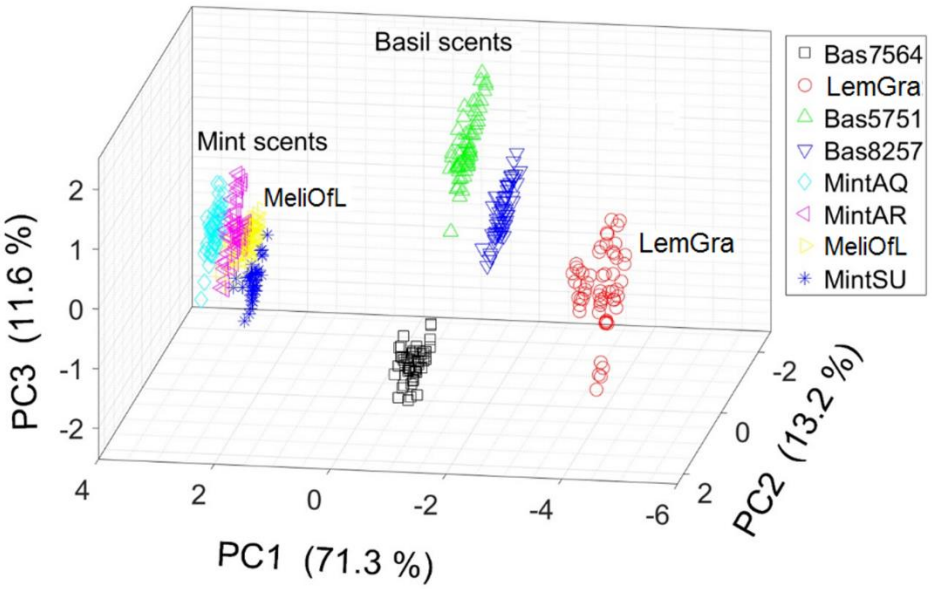


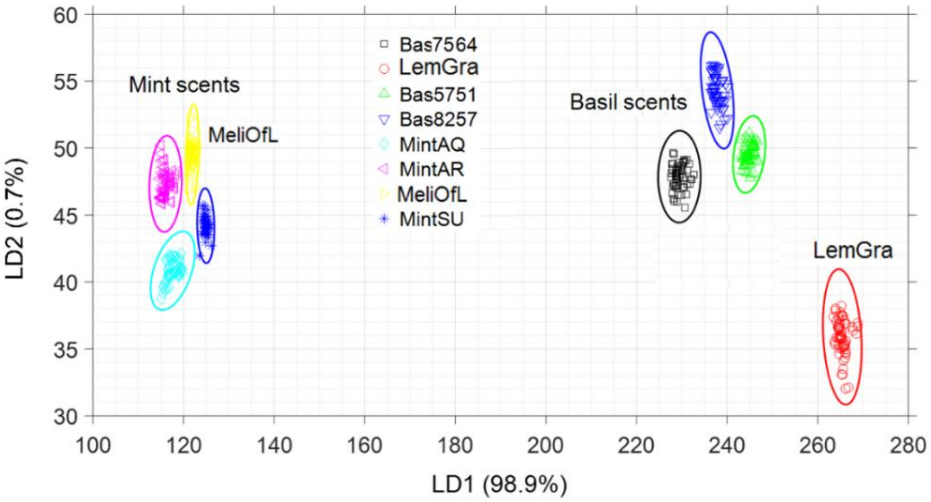
Figure 6. 3D plot of the Principal Component coefficients from 451 observations.

3.3. Linear Discrimination Analysis (LDA)

The 2D plot of the Linear Discriminant Analysis for the 8 different species with 95% confidence ellipse is presented in Fig 7(a). The so-called confusion matrix was calculated from 10-fold LDA cross validation partition using 451 observations with 406 training size and 45 test size obtained from the first cycle of the e-Nose measurements as shown in Fig 7(b). The LDA plot again shows an obvious clustering. The sum of the first two LDA vector components is 99.6%, and the LDA discrimination accuracy reached 100%. A calculated confusion matrix chart given in Fig.7 also confirms that the categorized (raw) labels match 100% with the true labels (columns) given during the training. The diagonal cells show the correctly classified observations, while the off-diagonal values show the percentage of the misclassification.

Figure 7 show the Linear Discriminant Analysis of eight Basil/Mint species including a control sample. The 2D plot of the 10-fold Linear-Discriminant Analysis was obtained from the training data sets shown with the colored symbols (first cycle e-Nose measurements) in Fig 7a, and the prediction confusion matrix for the unknown data sets from the second cycle of the e-Nose measurements shown in Fig7b. As we can clearly see (Figure 8) that the Mentha group of plants form a cluster separated from the basil group of plants and the lemon grass which was used as an outlier forms a completely separate cluster. The prediction matrix (Figure 8b) shows 9.8% overlap in case of Bas8257 (Krishna Tulsi) and Bas5751 (Tulsi). This could be attributed to the fact that both samples belong to plants of the same species.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2021 doi:10.20944/preprints202106.0278.v1

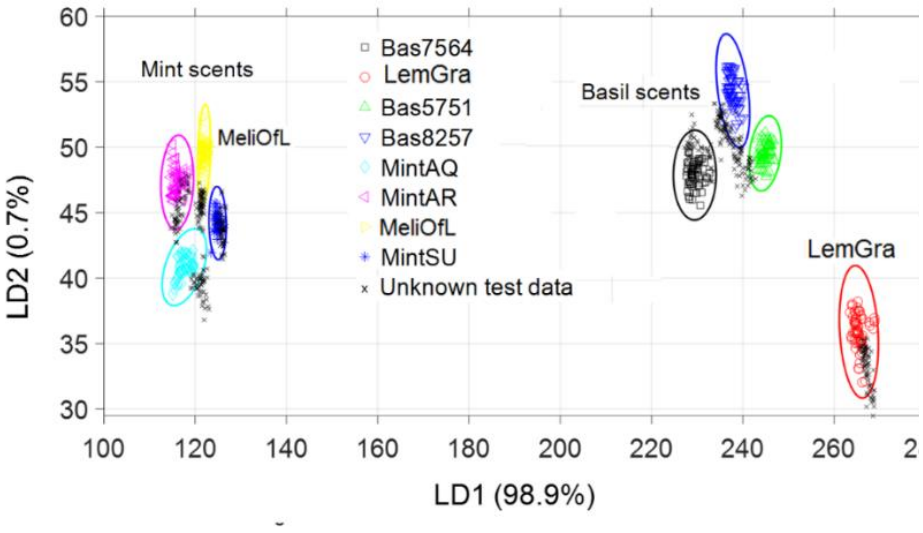


(a)

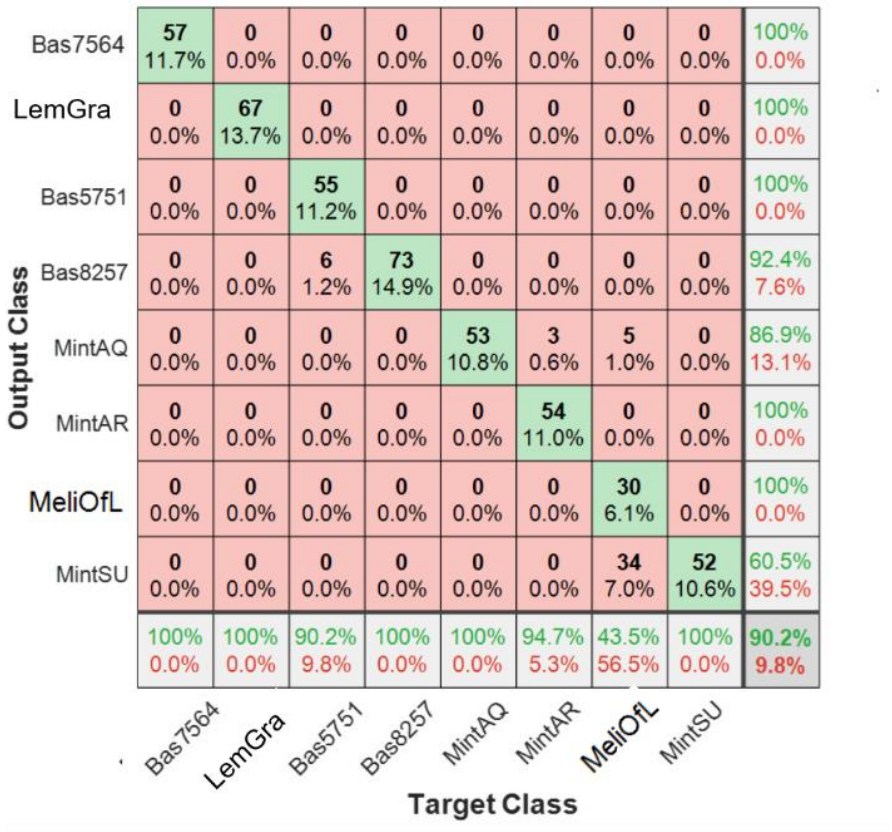
(b)

Figure 7. Linear Discriminant Analysis of eight species. (a) 2D plot of the Linear Discriminant Analysis with 95% confidence ellipse, (b) confusion matrix obtained from 10-fold LDA calculations using as the training data set obtained from the first cycle of the e-Nose measurements.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2021 doi:10.20944/preprints202106.0278.v1



(a)



(b)

Figure 8. Linear Discriminant Analysis of six Mint species. (a) 2D plot of the 10-fold Linear-Discriminant Analysis obtained from the training data sets shown with the colored symbols (first cycle e-Nose measurements), and from the prediction data sets shown with the black symbols (second cycle of the e-Nose measurements), (b) prediction confusion matrix for the unknown data sets from the second cycle of the e-Nose measurements.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2021 doi:10.20944/preprints202106.0278.v1

Table 3 shows a summary of LDA prediction results for unknown data sets obtained from different cycles of measurement after training with data sets from the 1st cycle, 2nd cycle, and 3rd cycle. The discrimination accuracy for each cycle are 100%. Nevertheless, the crosscheck prediction accuracies are obtained as between 73.5% and 90.2% with the average of 79.2%. The prediction accuracies in Table 3 show similar overlap between Bas8257 (Krishna Tulsi) and Bas5751 (Tulsi) confirming that both samples originating from the same plant species.

Table 3. LDA prediction results for unknown data sets obtained from different cycles of measurement after training with data sets from 1st cycle, 2nd cycle, 3rd cycle.

Training data	Unknown test data	Discrimination Accuracy	Prediction Accuracy
1st cycle	2nd cycle	100	90.2%
1st cycle	3rd cycle	100	75.3%
1st cycle	4th cycle	100	77.7%
2nd cycle	3rd cycle	100	79.3%
2nd cycle	4th cycle	100	79.0%
3rd cycle	4th cycle	100	73.5%

3.4. Nearest Neighbour Analysis (k-NN)

As a third non-parametric classification scheme, we applied the k-NN analysis with a 10-fold (k=10) calculation of the unknown data sets from the second cycle as compared to the true assignment from the training data set collected during the first cycle. The data sets from the second cycle of the e-Nose measurement were used for the k-NN calculation to determine prediction accuracy for the unknown observations. The k-NN discrimination accuracy was obtained as 94.2% with 5.8% misclassification (See Fig 9a). The overall prediction accuracy for the unknown data sets was smaller than in case of LDA with 82.3% corresponding to 17.7% misclassification.

The change in the k-NN discrimination and prediction accuracies with increasing number of nearest neighbors between 2 and 50 is given in Figure SI-4 in Supporting information. The k-NN discrimination accuracy drops from 100% to 90.2% with increasing nearest neighbor due to overlapping of classified data. Similarly, the k-NN prediction accuracy drops from 85.1% to 72.3% with increasing numbers of nearest neighbors.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 June 2021 doi:10.20944/preprints202106.0278.v1

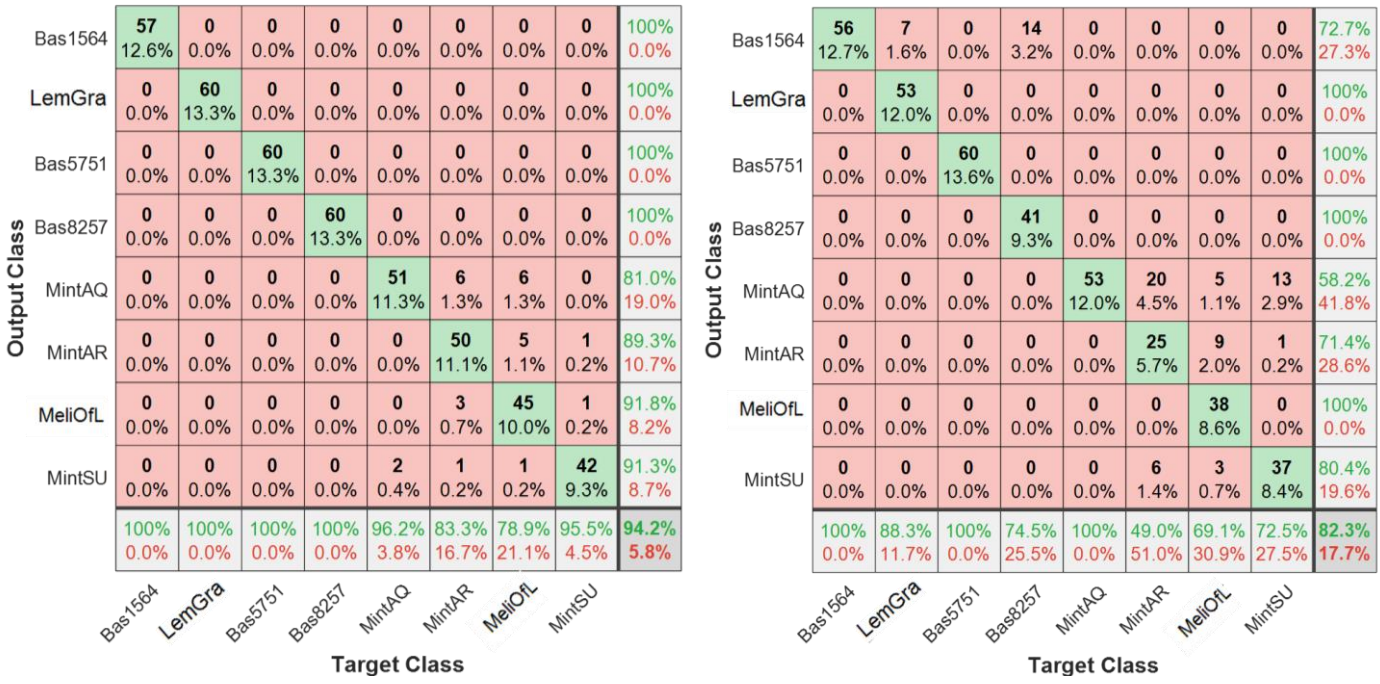


Figure 9. The discrimination (a) and prediction (b) confusion matrixes obtained from the k-NN analysis with10-fold (k=10) calculation with the unknown data sets from the second cycle to compare with the training data set (true labels) obtained from the first cycle.

5. Conclusions

In this work, sensor arrays based on six different SURMOFs have been used successfully for the discrimination of eight aromatic plants, seven of which belonged to the taxonomically challenging family of the Lamiaceae. The exposure and purging datasets (4 cycles) obtained from a low-cost custom-made portable e-Nose were analyzed with a Linear Discriminant Analysis (LDA) classification model. The first and second cycles of the data sets were used for training and the following repeated cycles were used as unknown data for prediction. More than 90% classification accuracy has been obtained within 8 different scent classes. Prediction accuracies with repeating test measurements reached up to 90% for LDA from unknown data sets. We can show that it is not only possible to discern and identify plants on the genus level (Mentha, Agastache, Melissa, all belonging to the Mentheae tribe within the Lamiaceae), but even to discriminate closely related sister clades within a genus (Basil). In addition, we were able to unequivocally separate Lemon Grass (Cymbopogon citratus) from Common Balm (Melissa officinalis L), although these species share an intense lemon-like scent and are often used for mutual surrogation and adulteration, demonstrating that the e-Nose excels most human noses that can be easily tricked by replacements of these two species. This study paves way for potential use of the sensors in detection of food adulterants. The portability and quick response of the sensor arrays demonstrates a huge potential for a future fabrication of cheap monitoring devices for use in food industry and food surveillance.

Author Contributions: "Conceptualization, S.O., M.S., A.K, S.V., S.V., L.F., F.G., P.N., and U.L.; methodology, S.O., M.S., A.K., P.N., C.L., Z.Z., and U.L. ; software, S.O., R.H. L.F., F.G.,; validation, S.O., M.S., R.H, A.K., S.V., P.N., and U.L.; formal analysis, S.O., R.H., A.K., L.H., S.V., P.N., and U.L. ; investigation, S.O.; resources, S.O., C.W., and P.N., and U.L.; data curation, S.O., and L.H.; writing—original draft preparation, S.O., M.S., R.H., C.L., Z.Z., S.V., P.N.; writing—review and editing, R.H., A.K., C.L., Z.Z., L.H., C.W., S.V., L.F., F.G., P.N., and U.L. ; visualization, S.O.. ; supervision, C.W., P.N., and U.L. ; project administration, S.O., C.W., P.N., and U.L. ; funding acquisition, C.W., P.N., and U.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—2082/1—390761711, Philipp Schwartz Fellowship (a program run by the Alexander von Humboldt Foundation) provided by the Karlsruhe Institute of Technology, and Iraqi Ministry of Education and Science.

Institutional Review Board Statement: "Not applicable" for studies not involving humans or animals.

Acknowledgments: The corresponding authors as well as co-authors A. K. and S. O. obtained support through a Philipp Schwartz Fellowship (a program run by the Alexander von Humboldt Foundation) provided by the Karlsruhe Institute of Technology. Co-author M. S. was supported by a PhD fellowship from the Iraqi Ministry of Education and Science; U.L. and C.W. acknowledge support from Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—2082/1—390761711. The authors would like to thank Mr. Joachim Daumann for competent cultivation and maintenance of the plant material used in this study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available for now.

Conflicts of Interest: No conflicts of interest.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Figure S1: title, Table S1: title, Video S1: title.

References

1. Choesin, D. N.; Boerner, R. E., Allyl isothiocyanate release and the allelopathic potential of Brassica napus (Brassicaceae). *American journal of botany* **1991,** 78, (8), 1083-1090.
2. Bruneton, J., *Pharmacognosy, phytochemistry, medicinal plants*. Lavoisier publishing: 1995.
3. Kubitzki, K.; Rohwer, J.; Bittrich, V., *The families and genera of vascular plants*. Springer: 1990; Vol. 1.
4. Butterworth, M.; Davis, G.; Bishop, K.; Reyna, L.; Rhodes, A., What Is a Superfood Anyway? Six Key Ingredients for Making a Food "Super". *Gastronomica* **2020,** 20, (1), 46-58.
5. Zoller, A.; Nordwig, H., *Heilpflanzen der ayurvedischen Medizin: ein praktisches Handbuch; mit 404 Tabellen*. Haug: 1997.
6. Murthy, K. S. R.; Reddy, M. C.; Rani, S. S.; Pullaiah, T., Bioactive principles and biological properties of essential oils of Burseraceae: A review. *Journal of Pharmacognosy and Phytochemistry* **2016,** 5, (2), 247.
7. Malladi, S.; Ratnakaram, V. N.; Pullaiah, T., Phytochemical screening of Caralluma lasiantha: Isolation of C21 pregnane steroid. *Oriental Journal of Chemistry* **2017,** 33, (2), 963-967.
8. Mediratta, P.; Sharma, K., Effect of essential oil of the leaves and fixed oil of the seeds of Ocimum sanctum on immune responses. *Journal of Medicinal and Aromatic Plant Sciences* **2000,** 22, (1B), 694-700.
9. Jürges, G.; Beyerle, K.; Tossenberger, M.; Häser, A.; Nick, P., Development and validation of microscopical diagnostics for 'Tulsi'(Ocimum tenuiflorum L.) in ayurvedic preparations. *European Food Research and Technology* **2009,** 229, (1), 99-106.
10. Cohen, Y.; Ben Naim, Y.; Falach, L.; Rubin, A. E., Epidemiology of basil downy mildew. *Phytopathology* **2017,** 107, (10), 1149-1160.
11. Agrawal, P.; Rai, V.; Singh, R., Randomized placebo-controlled, single blind trial of holy basil leaves in patients with noninsulin-dependent diabetes mellitus. *International journal of clinical pharmacology and therapeutics* **1996,** 34, (9), 406.

12. Jürges, G.; Sahi, V.; Rios Rodriguez, D.; Reich, E.; Bhamra, S.; Howard, C.; Slater, A.; Nick, P., Product authenticity versus globalisation—The Tulsi case. *PloS one* **2018,** 13, (11), e0207763.
13. Brümmer, J. M., Klein-Rabe-Weiss - Textsammlung-Lebensmittelrecht. 51. Ergänzungslieferung. B. Behrs Verlag GmbH & Co., 2000 Hamburg 76, Febr. 1992. 120 Blatt - lose Blattausgabe im Streifband. Unverbindliche Preisempfehlung DM 131,61 + MwSt. *Starch - Stärke* **1992,** 44, (11), 442-443.
14. Ichim, M. C.; Häser, A.; Nick, P., Microscopic authentication of commercial herbal products in the globalized market: Potential and limitations. *Frontiers in pharmacology* **2020,** 11, 876.
15. Zaheer SH, P. B., Chopra RN, Santapau H, Krishnan MS, Deshaprabhu SB, *Wealth of India–A Dictionary of Raw Materials and Industrial Products*. CSIR New Delhi, 1966; Vol. VII., p 10.
16. Vani, S. R.; Cheng, S.; Chuah, C., Comparative study of volatile compounds from genus Ocimum. *American Journal of Applied Sciences* **2009,** 6, (3), 523.
17. Ganie, S. H.; Upadhyay, P.; Das, S.; Sharma, M. P., Authentication of medicinal plants by DNA markers. *Plant Gene* **2015,** 4, 83-99.
18. Zou, H.-Q.; Lu, G.; Liu, Y.; Bauer, R.; Tao, O.; Gong, J.-T.; Zhao, L.-Y.; Li, J.-H.; Ren, Z.-Y.; Yan, Y.-H., Is it possible to rapidly and noninvasively identify different plants from Asteraceae using electronic nose with multiple mathematical algorithms? *journal of food and drug analysis* **2015,** 23, (4), 788-794.
19. Kiani, S.; van Ruth, S. M.; Minaei, S.; Ghasemi-Varnamkhasti, M., Hyperspectral imaging, a non-destructive technique in medicinal and aromatic plant products industry: Current status and potential future applications. *Computers and electronics in agriculture* **2018,** 152, 9-18.
20. Kiani, S.; Minaei, S.; Ghasemi-Varnamkhasti, M., Real-time aroma monitoring of mint (Mentha spicata L.) leaves during the drying process using electronic nose system. *Measurement* **2018,** 124, 447-452.
21. Iqbal, N.; Mustafa, G.; Rehman, A.; Biedermann, A.; Najafi, B.; Lieberzeit, P. A.; Dickert, F. L., QCM-arrays for sensing terpenes in fresh and dried herbs via bio-mimetic MIP layers. *Sensors* **2010,** 10, (7), 6361-6376.
22. Okur, S.; Sarheed, M.; Huber, R.; Zhang, Z.; Heinke, L.; Kanbar, A.; Wöll, C.; Nick, P.; Lemmer, U., Identification of Mint Scents Using a QCM Based E-Nose. *Chemosensors* **2021,** 9, (2), 31.
23. Okur, S.; Zhang, Z.; Sarheed, M.; Nick, P.; Lemmer, U.; Heinke, L., Towards a MOF e-Nose: A SURMOF sensor array for detection and discrimination of plant oil scents and their mixtures. *Sensors and Actuators B: Chemical* **2020,** 306, 127502.
24. Okur, S.; Qin, P.; Chandresh, A.; Li, C.; Zhang, Z.; Lemmer, U.; Heinke, L., An enantioselective e-nose: An array of nanoporous homochiral MOF films for stereospecific sensing of chiral odors. *Angewandte Chemie International Edition* **2020,** 132, 1 – 7.
25. Shekhah, O.; Wang, H.; Kowarik, S.; Schreiber, F.; Paulus, M.; Tolan, M.; Sternemann, C.; Evers, F.; Zacher, D.; Fischer, R. A., Step-by-step route for the synthesis of metal− organic frameworks. *Journal of the American Chemical Society* **2007,** 129, (49), 15118-15119.
26. Okur, S.; Kuş, M.; Özel, F.; Yılmaz, M., Humidity adsorption kinetics of water soluble calix [4] arene derivatives measured using QCM technique. *Sensors and Actuators B: Chemical* **2010,** 145, (1), 93-97.
27. Horzum, N.; Taşçıoglu, D.; Okur, S.; Demir, M. M., Humidity sensing properties of ZnO-based fibers by electrospinning. *Talanta* **2011,** 85, (2), 1105-1111.
28. Muckley, E. S.; Anazagasty, C.; Jacobs, C. B.; Hianik, T.; Ivanov, I. N. In *Low-cost scalable quartz crystal microbalance array for environmental sensing*, Organic sensors and bioelectronics IX, 2016; International Society for Optics and Photonics: 2016; p 99440Y.
29. Heinke, L.; Wöll, C., Surface-Mounted Metal–Organic Frameworks: Crystalline and Porous Molecular Assemblies for Fundamental Insights and Advanced Applications. *Advanced Materials* **2019,** 31, (26), 1806324.

30. Sarheed, M. M. Allelopathic Compounds from Mint Target the Cytoskeleton from Cell Biology Towards Application as Bioherbicides. KIT-Bibliothek, 2019.
31. Sarheed, M. M.; Rajabi, F.; Kunert, M.; Boland, W.; Wetters, S.; Miadowitz, K.; Kaźmierczak, A.; Sahi, V. P.; Nick, P., Cellular base of mint allelopathy: Menthone affects plant microtubules. *Frontiers in plant science* **2020,** 11, 1320.
32. Sarheed, M. M.; Rajabi, F.; Kunert, M.; Boland, W.; Wetters, S.; Miadowitz, K.; Kaźmierczak, A.; Sahi, V. P.; Nick, P., Cellular Base of Mint Allelopathy: Menthone Affects Plant Microtubules. *Frontiers in Plant Science* **2020,** 11, (1320).
33. Gu, Z.-G.; Grosjean, S.; Bräse, S.; Wöll, C.; Heinke, L., Enantioselective adsorption in homochiral metal–organic frameworks: the pore size influence. *Chemical Communications* **2015,** 51, (43), 8998-9001.
34. Heinke, L.; Gu, Z.; Wöll, C., The surface barrier phenomenon at the loading of metal-organic frameworks. *Nature communications* **2014,** 5, (1), 1-6.
35. Zhou, W.; Wöll, C.; Heinke, L., Liquid-and gas-phase diffusion of ferrocene in thin films of metal-organic frameworks. *Materials* **2015,** 8, (6), 3767-3775.
36. Heinke, L., Diffusion and photoswitching in nanoporous thin films of metal-organic frameworks. *Journal of Physics D: Applied Physics* **2017,** 50, (19), 193004.
37. Sauerbrey, G.; Jung, G., Vibrational Modes of Planoconvex Quartz Plates. *Z Angew Physik* **1968,** 24, (2), 100-106.
38. Kambhatla, N.; Leen, T. K., Dimension reduction by local principal component analysis. *Neural computation* **1997,** 9, (7), 1493-1516.
39. Sharifzadeh, S.; Ghodsi, A.; Clemmensen, L. H.; Ersbøll, B. K., Sparse supervised principal component analysis (SSPCA) for dimension reduction and variable selection. *Engineering Applications of Artificial Intelligence* **2017,** 65, 168-177.
40. Shaoning, P.; Ozawa, S.; Kasabov, N., Incremental linear discriminant analysis for classification of data streams. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **2005,** 35, (5), 905-914.