J. Biophotonics **9**, No. 1–2, 181–189 (2016) / **DOI** 10.1002/jbio.201500176

Journal of BIOPHOTONICS

FULL ARTICLE

Identification of aqueous pollen extracts using surface enhanced Raman scattering (SERS) and pattern recognition methods

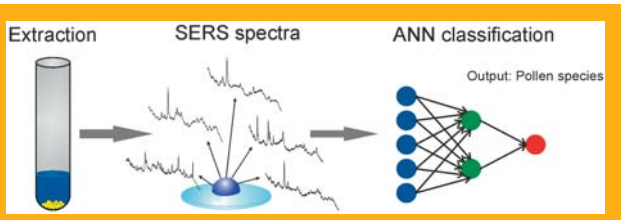
Stephan Seifert[1, 2], *Virginia Merk*[1], *and Janina Kneipp**[, 1, 2]

[1] Humboldt-Universität zu Berlin, Department of Chemistry, Brook-Taylor-Straße 2, 12489 Berlin, Germany
[2] BAM Federal Institute for Materials Research and Testing, Richard-Willstätter-Straße 11, 12489 Berlin, Germany

Received 8 June 2015, Revised 9 July 2015, Accepted 9 July 2015
Published online 7 August 2015

Key words: Surface enhanced Raman scattering (SERS), artificial neural networks (ANN), multivariate statistics, pollen, pattern recognition

Aqueous pollen extracts of varying taxonomic relations were analyzed with surface enhanced Raman scattering (SERS) by using gold nanoparticles in aqueous suspensions as SERS substrate. This enables a selective vibrational characterization of the pollen water soluble fraction (mostly cellular components) devoid of the spectral contributions from the insoluble sporopollenin outer layer. The spectra of the pollen extracts are species-specific, and the chemical fingerprints can be exploited to achieve a classification that can distinguish between different species of the same genus. In the simple experimental procedure, several thousands of spectra per species are generated. Using an artificial neural network (ANN), it is demonstrated that analysis of the intrinsic biochemical information of the pollen cells in the SERS data enables the identification of pollen from different plant species at high accuracy. The ANN extracts the taxonomically-relevant information from the data in spite of high intra-species spectral variation caused by signal fluctuations and preparation specifics. The results show that SERS can be used for the reliable characterization and identification of pollen samples. They have implications for improved investigation of pollen physiology and for allergy warning.



1. Introduction

The composition, morphology and physiology of pollen grains has been investigated repeatedly by spectroscopists, using Raman and infrared spectroscopies [1–6] not only with the goal of developing faster identification and classification tools, e.g., for allergy warning but also for better understanding of physiological processes [7] and the influence of environmental factors [8]. The pollen grain's outer layer consists of sporopollenin, a stable biopolymer of mainly unknown composition [9] that is dominating the Raman spectral fingerprint of pollen grains [2]. The pollen shell also contains, among others, flavonoid [10], carotenoid [11], and lipid [12] components. The pollen interior comprises typical cellular bioorganic constituents such as nucleic acids, proteins, lipids, and sugars [2], which make up much less biomolecular material than the polymer coat, but contain important determinants of, e.g., plant reproduction, metabolism, and allergenic potential [13].

* Corresponding author: e-mail: janina.kneipp@chemie.hu-berlin.de, Phone: +49 30 20 93 71 71, Fax: +49 30 2093 5559

With regard to the high fluorescence background that has been observed in the Raman spectra of some pollen species [6, 14, 15], obtaining surface-enhanced Raman scattering (SERS) rather than normal Raman spectra can provide significant improvements in the vibrational characterization of pollen, since the noble metal nanostructures that are used as SERS substrates provide efficient quenching of fluorescence together with high signal enhancement [16, 17]. Nevertheless, SERS spectral information from the outer coat highly depends on the portion of the sporopollenin ultrastructure that interacts with the metal nanoparticles and requires their integration in the structure, for example by *in situ* synthesis [16]. So far, in automated species identification, Raman-based classification relies mainly on pollen outer shell biochemistry [2]. It would be desirable to aim at species-specific differences in the pollen phenome that display in the pollen cells for Raman-spectroscopic classification. Here, due to the low sensitivity of normal, non-resonant Raman scattering, we demonstrate that SERS enables short acquisition times and works for small amounts of sample obtained from a limited number of cells that are present in a few pollen grains.

As opposed to attaining high reproducibility SERS spectra by using immobilized nanoparticles [17], we show now that species-specific classification of pollen extract SERS data is possible with high numbers of spectra that are obtained very quickly with simple nanoparticle solutions. To make use of this high amount of hyperspectral data, automated pattern recognition tools, in particular artificial neural networks (ANN), are applied. ANN are powerful for classification of large sets of spectra, and have attracted significant interest in analytical and biomedical spectroscopy [18–22].

As suggested in recent studies on using support vector machines for SERS-based cancer [23] and virus [24] detection, the application of supervised learning based on pattern recognition to SERS data could be very promising. The high potential of combining artificial neural networks with SERS data has convincingly been demonstrated by multiplex investigation of biomolecular mixtures, specifically to identify molecular species [25]. Here, rather than screening for specific components, we aim at a fingerprint analysis of multiple biomolecules in pollen extracts for taxonomic classification. For this purpose we apply to the SERS data both, frequently-used multivariate statistics as well as analysis by ANN. Thereby, we demonstrate that the identification of pollen from different plant species based on SERS spectra of their respective cellular fraction is feasible at high accuracy and may be useful for improved pollen investigation and allergy warning.

2. Materials and methods

2.1 Sample preparation

Pollen samples of 14 different species listed in Table 1 were purchased from Sigma. Furthermore, 180 samples of 10 genera and 71 species of the orders *Fagales* (85 samples) and *Coniferales* (95 samples) were collected in parks and in the Botanic Garden Berlin-Dahlem. (Table 2) For the commercially available samples two aliquots and for the collected pollen samples one aliquot of 0.2 mg of pollen grains were extracted with 100 µl of deionized water and left for sedimentation for 5 minutes. Thus, extraction can be carried out much faster than for Raman investigations where several hours can be necessary [5]. After centrifugation, the liquid supernatant was separated from the sedimented pollen grains. For SERS measurements, 2 µl of the supernatant were

Table 1 Commercially available pollen samples investigated in this study and their respective plant family.

species	family	label
Bassia scoparia (firebrush)	*Amaranthaceae*	a
Artemisia absinthium (wormwood)	*Asteraceae*	b
Artemisia tridentata (sagebrush)	*Asteraceae*	c
Iva xantifolia (giant poverty)	*Asteraceae*	d
Betula occidentalis (water birch)	*Betulaceae*	e
Juniperus scopulorum (Rocky Mountain juniper)	*Cupressaceae*	f
Juglans nigra (black walnut)	*Juglandaceae*	g
Secale cereale (rye)	*Poaceae*	h
Sorghum halepense (johnson grass)	*Poaceae*	i
Populus deltoides (cottonwood poplar)	*Salicaceae*	j
Populus nigra (black poplar)	*Salicaceae*	k
Populus tremuloides (Quaking aspen)	*Salicaceae*	l
Populus trichocarpa (black cottonwood)	*Salicaceae*	m
Ulmus pumila (Chinese elm)	*Ulmaceae*	n

Table 2 Pollen samples collected in the field and their respective plant genus, family and order.

genus	family	order	no. of different species	no. of samples
Juniperus	*Cupressaceae*	*Coniferales*	2	7
Abies	*Pinaceae*	*Coniferales*	3	6
Picea	*Pinaceae*	*Coniferales*	12	25
Pinus	*Pinaceae*	*Coniferales*	13	40
Taxus	*Taxaceae*	*Coniferales*	3	17
Alnus	*Betulaceae*	*Fagales*	9	19
Betula	*Betulaceae*	*Fagales*	10	21
Corylus	*Betulaceae*	*Fagales*	6	19
Myrica	*Myricaceae*	*Fagales*	2	6
Quercus	*Fagaceae*	*Fagales*	11	20
			total number of samples	180

mixed with 2 µl of sodium chloride solution (100 mM) and 20 µl of gold nanoparticle suspension (~30 nm in size as determined by electron microscopy) synthesized by the method reported in ref [26]. From each of the two supernatant samples that were prepared with the commercially available pollen, four mixtures with nanoparticles were generated and measured separately, yielding eight sets of Raman spectra per pollen species. From every pollen extract from the samples collected in the field, two mixtures with nanoparticles were prepared. Drops of 20 µl of the mixture were transferred onto a CaF_2 slide, and Raman spectra were acquired using a water immersion objective. Effectively, in each SERS experiment, supernatant from an amount of 4 µg of pollen was probed. To obtain the spectra shown in Figure S1 in the Supporting information for comparison, one SERS experiment was also conducted using effective supernatant samples corresponding to ~0.2 mg of rye pollen with the same amount of gold nanoparticles as in all other experiments.

2.2 Raman experiment

For each of the eight preparations from each species of commercially available pollen and for each of the two preparations of pollen collected in the field, 500 SERS spectra were obtained using an integration time of 1 second per spectrum. The spectra were acquired using a LabRam HR800 (Horiba Jobin Yvon, Bensheim, Germany) coupled to a BX41 microscope (Olympus, Hamburg, Deutschland) with a 60× water immersion objective. For excitation, a diode laser (Toptica, Germany) at 785 nm was used. The Raman scattered light was detected by a liquid nitrogen-cooled CCD detector (1024 × 256 pixels, Horiba). The laser intensity was $1.4 \cdot 10^6$ W/cm^2.

2.3 Data analysis

A schematic overview of the data processing and the algorithms applied to the SERS data is shown in the Supporting information Figure S2.

The spectra were frequency-calibrated using a spectrum of 4-acetamidophenol. Elimination of spikes, interpolation in the spectral range between 400 cm^{-1} and 1700 cm^{-1}, vector-normalization, calculation of first derivatives, averaging of the spectra and transformation to standard normal distribution was done using Matlab (The Mathworks, Inc.). For multivariate statistics, the Statistics toolbox of Matlab was used. For hierarchical clustering (HCA), the distance matrix was calculated using Euclidean distances, and Ward's algorithm was applied for clustering.

The analysis by three-layer, feedforward artificial neural networks was performed using nprtool of Matlab Neural Networks toolbox. A schematic overview of the structure of the artificial neural networks is shown in the Supporting information Figure S3. For the investigation of commercially available pollen, the data of one extract of each species were used to build the ANN (28 000 spectra), and the data of the other extracts were used as pure test set (28000 spectra). The data used to build the net were further divided into training data (70%), validation data (25%) and test data (5%). A network consisting of 894 input and 14 output neurons, using 50 hidden neurons was used. Conjugate gradient backpropagation algorithm was applied. The training of the ANN was stopped at the minimum validation mean square error, which was reached after 356 steps. Subsequently, the performance of the ANN was assessed using the test data. The assignment of the 500 spectra of each data set was averaged, and classification was obtained using the winner-takes-all (WTA) method [27]. For the investigation of samples collected in the field 46 000 spectra of 46 samples of *Coniferales* and 47 000 spectra of 47 samples of *Fagales* divided into training (70%) and validation

(30%) data were used to build the ANN. The network consisting of 894 input and 2 output neurons, using 50 hidden neurons was trained with conjugate gradient backpropagation algorithm until the validation mean square error was reached after 302 steps. Subsequently, the net was tested with 49 000 spectra of 49 different samples of *Coniferales* and 38 000 spectra of 38 different samples of *Fagales*. The assignment of the 500 spectra of each data set was averaged, and classification was obtained using the winner-takes-all (WTA) method. The spectra with an assignment higher than 0.9 to the respective taxonomic order were defined as 'taxonomically relevant'. They were extracted from the data sets, averaged and used for further multivariate analyses.

3. Results and discussion

3.1 SERS spectra

We have investigated pollen species from varying taxonomic relations (Table 1). From 14 species that were commercially available, 4000 spectra per species were obtained (two samples, eight preparations per species in total), by using gold nanoparticles in aqueous suspensions as SERS substrate. (for a description of the single SERS spectra see Section 1 and Figure S4 in the Supporting information) Figure 1 shows the averaged spectra (averages of 500) of 14 different species. Most bands are assigned to signals from nucleobases and amino acids (Table 3 and reference spectra in the Supporting information Figure S5). Specifically, there occurs a strong variation in the contribution of the ring breathing vibration at 734 cm^{-1} depending on the species. (see also Supporting Information Figure S1 and its discussion). Furthermore, several bands are highly specific for a particular plant species. For example, the spectra of *Juniperus scopulorum* pollen extract (Figure 1f) show bands at 1134 cm^{-1} and 1230 cm^{-1} that can be assigned to the NH_3^+ deformation vibration of amino acids and the ring breathing vibration of tryptophan, respectively [28]. These bands are either not present, or by far less intense in the spectra of the other investigated species (Figure 1). The observation of amino acid and nucleic acid components in the spectra is in accord with the known composition of the plant cells that make up the interior of the pollen grains.

In order to evaluate SERS for classification approaches using the inner pollen grain spectral fingerprint that comes from these cells, the SERS spectra of the pollen extracts were subjected to hierarchical cluster analysis (HCA), principal component analysis (PCA) and an artificial neural network (ANN). The

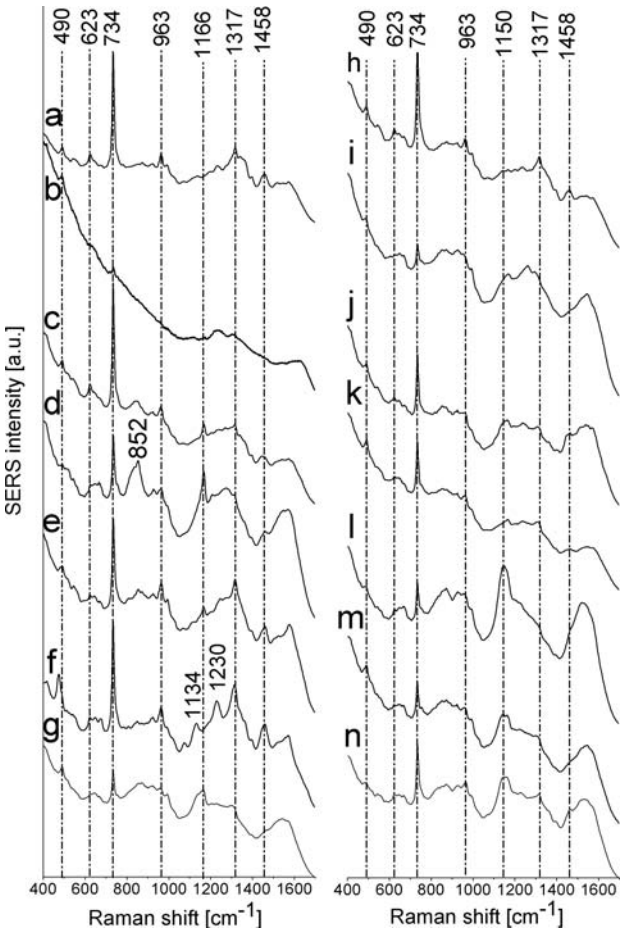


Figure 1 Mean spectra of 500 vector-normalized SERS spectra (accumulation time per individual spectrum: 1 second), of the aqueous supernatant of pollen from the species labeled in Table 1. Excitation wavelength: 785 nm (intensity $1.4 \cdot 10^6$ W/cm^2).

spectral information contained in each individual SERS spectrum depends on the particular biomolecule(s) that are in the vicinity of the SERS-active gold nanoaggregates. Since different nanoparticles or nanoaggregates are residing in the focal volume at the time each spectrum is acquired, the variance within the spectra from one sample is high (see also Supporting information Section 1 and Figure S4). In order to obtain spectra that represent as many molecules contained in a liquid extract as possible, mean spectra of each of the 8 data sets (á 500 spectra) per species obtained from the 14 commercially available pollen species were used for both PCA and HCA.

3.2 Hierarchical cluster analysis (HCA)

The omnipresence of the band at 734 cm^{-1} (see Section 1 and Figure S4 in the Supporting information),

Table 3 Raman shifts of characteristic bands in the SERS spectra of pollen extracts and their tentative assignments to vibrational modes of reference spectra.

Raman shift [cm^{-1}]	tentative assignment[a]
1458	G, A: δ,ν[b]
1363	Trp: ring[b]
1317	C: ν[b]
1230	Trp: ring breathe[b]
1205	His: δ ring[c]
1194	C: ν[b]
1166	His: δ[c]
1150	Met: NH_3^+: δ[c]
1134	Amino acids NH_3^+: δ[c]
1014	Phe: ring breathe[b]
994	Trp: asym indole breathe[b]
963	A, G: δ[b]
922	T: δ[b]
890	Indole[c]
854	Ser: ν[c]
827	Val: ν[c]
800	C: ring breathe[b]
758	Trp: sym Indole breathe[b]
734	A: ring breathe[b]
662	G: ring breathe[b]
623	A: δ[b]
542	Indole[c]
490	C, G, T: δ[b]

[a] Abbreviations: ν, stretching; δ, deformation; A, adenine; C, cytosine; T, thymine; G, guanine; Met, Methionine; His, Histidine; Phe, Phenylalanine; Trp, Tryptophan; V, Valine
[b] according to reference spectra (see Figure S5 in the Supplementary information)
[c] according to Refs. [28, 29]

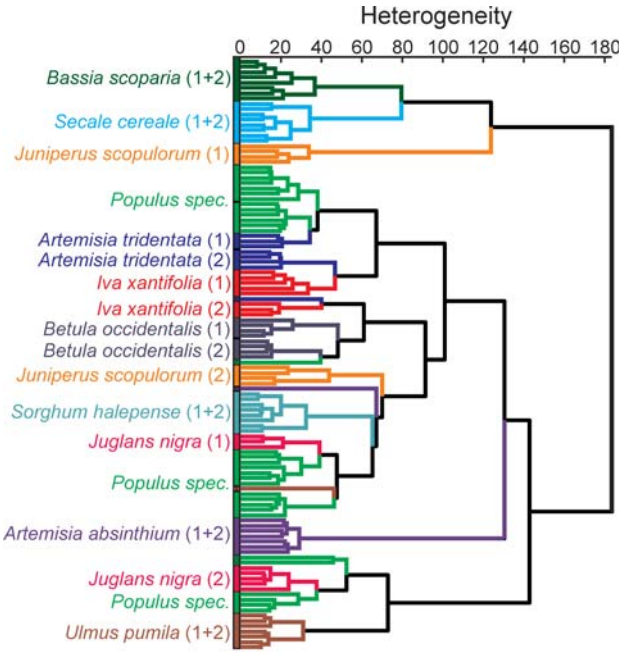


Figure 2 Result of a hierarchical cluster analysis of 112 average spectra (8 average spectra per species corresponding to 8 separate SERS experiments from 14 different pollen species). Cluster analysis using the spectral range of 400–1700 cm^{-1} was done on standardized first derivatives of the mean spectra of 500 vector-normalized spectra. Euclidean distances were used and Ward's algorithm was applied for clustering. For clarity, spectra from one respective species are shown in the same color but with different enumeration. Labels (1) and (2) correspond to different clusters of the same species. In most cases those labels correspond to the two respective extracts that were prepared from each sample.

as well as similarities in some other spectral features strongly suggests that the whole spectral range needs to be used for the distinction of different species on the basis of their SERS spectra. In order to avoid over-weighting of very dominant spectral features, the data were transformed to standard normal distribution before the multivariate analysis. The results of an HCA, performed on first derivatives of 8 average spectra per species are shown in Figure 2. The spectra obtained with the pollen extracts of *Bassia scoparia*, *Secale cereale*, *Sorghum halepense*, *Artemisia absinthium,* and *Ulmus pumila* form sub-clusters each exclusively representing one respective species (Figure 2). This indicates a very low variance within the spectra of each of these species even though they result from different preparations. The spectra of *Populus nigra*, *Populus deltoides*, *Populus trichocarpa* and *Populus tremuloides* form three clusters containing data from all the different extracts of all species (Figure 2). The species of the *Populus* genus cannot be separated by HCA (cluster *Populus* spec. in Figure 2). The spectra of *Juniperus scopulorum*, *Betula occidentalis*, *Artemisia tridentata*, *Iva xantifolia* and *Juglans nigra*, display two different patterns, leading to two clusters, respectively, for each of these species. (Figure 2) Each of the sub-clusters (labelled with (1) and (2) respectively) within these species classes can mainly be assigned to the spectra of one single extract. This indicates that there is a high intra-class variance, caused by differences in the chemical composition of the extract. In order to achieve a species-specific classification, an analysis is needed that is robust enough to neglect these preparation-based differences.

3.3 Principal component analysis (PCA)

In order to separate the different types of variance that occur in the data set, PCA was applied to the pollen extract SERS data. The scores, loadings and variances of the first four principal components, which represent 64% of the total variance in the data set, are shown in Figure 3. Before PCA, the

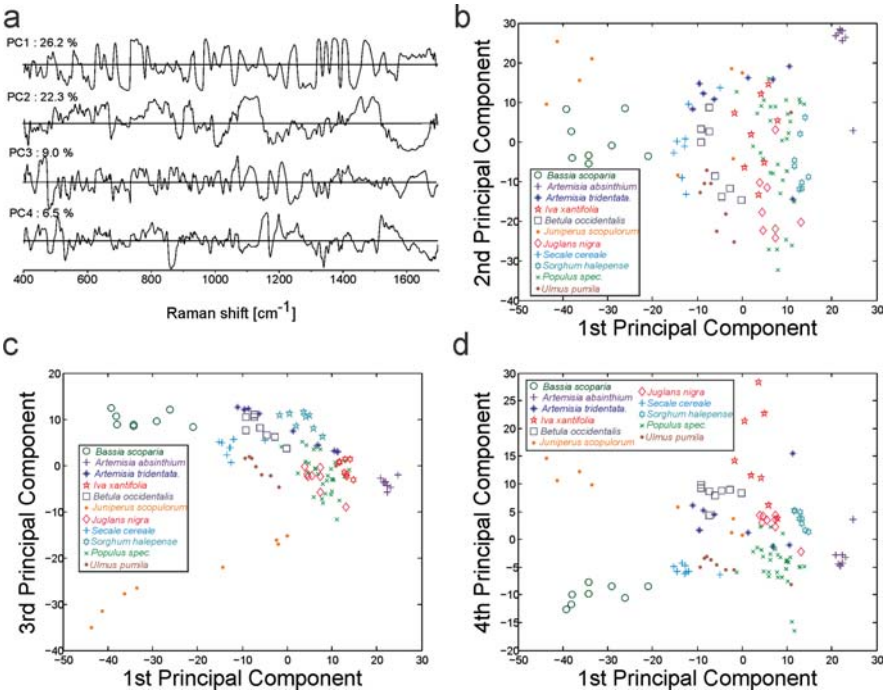


Figure 3 Results of the principal component analysis using the spectral range of 400–1700 cm^{-1} of 112 average spectra from 14 different commercially available pollen species. Loadings of the first four PCs (**a**), as well as the scores of the first and second (**b**), first and third (**c**) and first and fourth PC (**d**) are shown. PCA was done with standardized first derivatives of the mean spectra of 500 vector-normalized spectra.

data were transformed to standard normal distribution. The loadings (Figure 3a) show that the normalization leads to an equal weighting of the different spectral regions. This is not the case if the data are analyzed without normalization, where the variation in the strong band at 734 cm^{-1} dominates the loadings of several principal components (compare Figure S6 in the Supporting information). Combination of the scores of the first four principal components (Figure 3b–d) makes it possible to distinguish between the spectra of the species that could also be separated in HCA (compare with Figure 2). As an advantage over the HCA result, the variance weighting capabilities of the PCA lead to grouping of the spectra of *Juniperus scopulorum*, *Betula occidentalis* and *Iva xantifolia*, specifically by combining the information contained in the first, third and fourth PCs (Figure 3c and d). Still, grouping of the spectra of *Artemisia tridentata, Juglans nigra*, and the different *Populus* species is not achieved by PCA (Figure 3b–d). As illustrated by the group of spectra from *Sorghum halepense* (Figure 3), the scores of the first principal component are essential for the separation of different species. Although not shown in the scores plots of Figure 3 for clarity, the variance represented by the first principal component was also found to be responsible for discrimination between different extracts of the same species, as seen, e.g., in the data from *Juniperus scopulorum* (compare also Figure 2). This illustrates that the two different contributions to spectral variance cannot be completely separated by PCA.

3.4 Artifical neural network (ANN) for species identification

The major challenge in achieving correct classification of the data using the above-described experimental procedure is the separation of characteristic differences between spectra that are species-specific and the spectral variances due to different preparations of pollen from the same species. Since the variance of spectra from pollen grains of different species cannot be extracted by the approaches discussed before, a pattern recognition method that can emphasize the relevant spectral information for species separation has to be applied. As supervised pattern recognition tool that can cope with large amounts of data, ANN were trained. ANN, compared to support vector machines, another tool for supervised pattern recognition that has been used in the classification of SERS spectra [23, 24], provide advantages for the identification of several groups instead of simple differentiation of two classes. 894 spectral data points were used as Input Neurons (for a schematic overview of the structure of the ANN see Figure S3 in the Supporting information). 2000 spectra of the four preparations of one of the two pollen extracts for each species were applied to build and validate the neural network, while the 2000 spectra of the four preparations of the respective other extracts were only used to test the network. The results of the testing (Figure 4) show that a correct identification of the data was achieved for 97% of the sam-

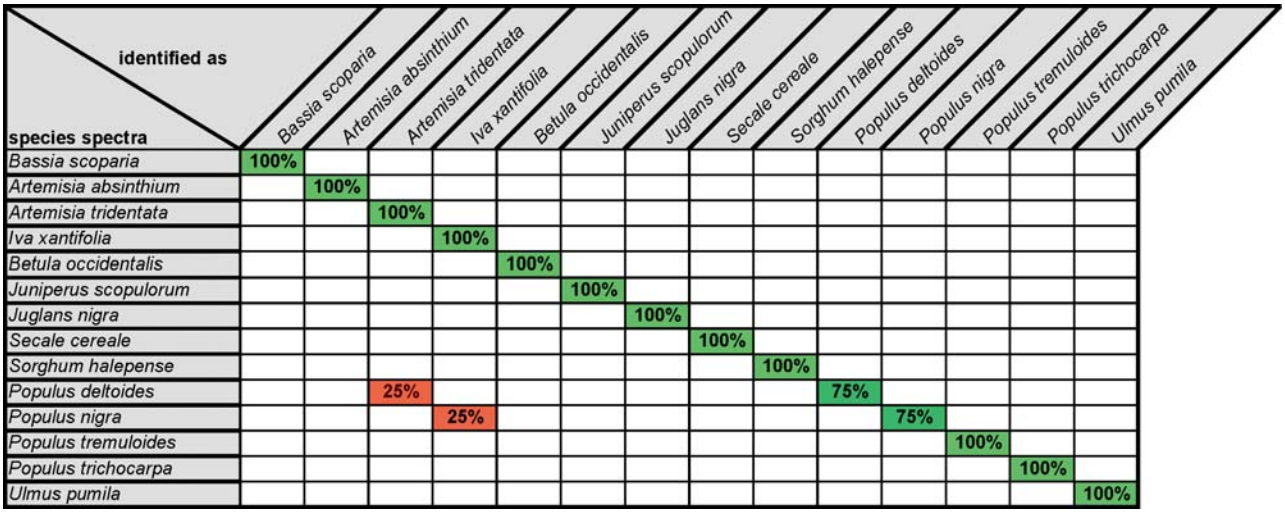


Figure 4 Classification results for 14 commercially available pollen species after the analysis of their SERS spectra with an artificial neural network. The assignment of 500 spectra each was averaged and the classification was done using the winner-takes-all (WTA) method.

ples. Only one data set of *Populus nigra* and one of *Populus deltoides* were assigned to the wrong class. Most of the spectra of different species of the *Populus* genus that could not be separated by PCA or HCA (Figures 2 and 3) were distinguished by this pattern recognition method. As spectra of different extracts were used for training and testing, respectively, identification must be based on species-specific chemical make-up rather than preparation/sampling specifics.

3.5 Combination of ANN with PCA

In order to validate whether classification by ANN can also be applied to other taxonomic categories, classification was applied to SERS spectra of pollen grains of different genera and species of *Coniferales* (95 samples) and *Fagales* (85 samples) orders collected in the field. (Table 2) The scores plot of the PCA (Figure 5a) on mean spectra shows that separation of the spectra of pollen from the two plant orders is not possible, because order-specific variance is superimposed by other variances in the data set. Subsequently, 93 000 individual spectra were used for training and validation of an ANN, and 87 000 spectra were used as test set. The results show that 92% of the *Coniferales* and 91% of the *Fagales* samples were assigned to the correct taxonomic order (Figure 5b). This illustrates that the ability of taxonomic identification can be extended from the level of species (Figure 4) to other, more heterogeneous taxonomic groups. For each sample the assignment to the respective order for each of the 500 spectra was used for separating SERS spectra that were characteristic of the respective plant order from those containing information that is less relevant for taxonomic separation: Spectra that provided an assignment lower than 0.9 were neglected in the further analysis and not included for the calculation of average spectra. Again, a PCA was executed on the average spectra (Figure 5c). The scores of the first PC show that the spectra that were responsible for an insufficient separation result in Figure 5a could effectively be eliminated from the full data. This leads to different loadings of this PC compared to the respective component of the PCA executed without pre-selection by the ANN (see Figure S7 in the Supporting information). The mean spectra of the pre-selected data can also be combined with HCA for plant order classification (Figure S8).

The potential of ANN for fingerprint analyses of complex mixtures in biomaterials based on vibrational spectra has impressively been demonstrated by applications to bacteria [27], the identification of Malaria parasite cell cycles [30], as well as tissue imaging [31] in previous work. The data shown here demonstrate that, applications with similar objective of ANN to SERS spectra, which display several specifics compared to normal Raman and infrared data, is feasible as well. As illustrated by this work, and comparing the data with our previous results obtained with normal Raman scattering [2, 3] the SERS data of the pollen extract represent a selected phenotypic information. This is illustrated by the results of the HCA and PCA: Even in clusters that are related to species, the hierarchy of classification in HCA does not represent phylogenetic connections (Figure 2). In contrast, in the case of, e.g., normal Raman spectra [2] or infrared spectra [4], the pollen coat – an important determinant when offspring is to

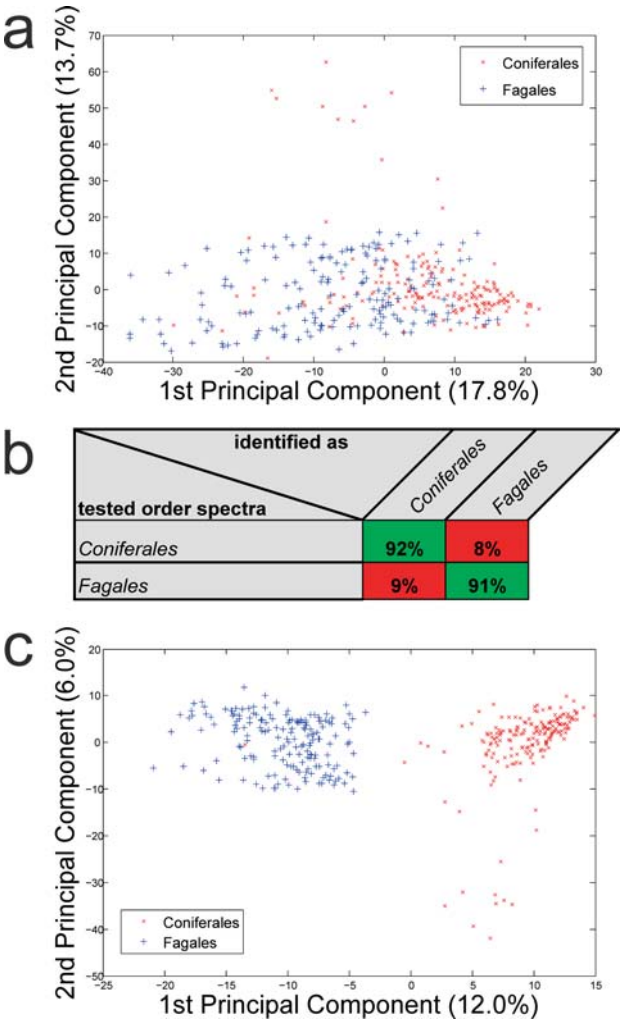


Figure 5 (**a**) Scores of the first and second PC of a PCA executed on 360 average spectra (each calculated from 500 individual spectra) from *Coniferales* and *Fagales* pollen grains. (**b**) Results of the classification by ANN using the winner-takes-all (WTA) method. (**c**) PCA on the same data set as in (**a**) after elimination of those individual spectra that were not assigned to the respective order by ANN before calculation of the average spectra. For PCA, standardized first derivatives in the spectral range of 400–1700 cm^{-1} were used.

be produced and hence very characteristic of a separate, even closely related species – contributes strongly to the spectra, resulting in a reconstruction of the phylogenetic hierarchy by these spectroscopies. Different from the methods that generate mainly information from the pollen coat, the SERS data from the pollen interior as obtained here will be useful to compare different physiological states in the same type of pollen and to understand subtle physiological changes in the pollen cells. Considering potential analytical applications, it should be pointed out that the strength of the ANN-based pattern recognition of pollen SERS data lies in the identification of pollen species and of specific biochemical changes, e.g., caused by environmental influences on the composition of the pollen interior.

The general ability of the ANN to separate information that is relevant for identification of taxonomic relationships from information that represents a specific composition due to sampling/preparation suggests ANN also for pre-selection of SERS data in a hierarchical classification/identification approach. Similar strategies have been proposed for the classification of bacteria [27] and moulds [32].

4. Conclusions

In this paper, we have shown the identification of aqueous pollen extracts with SERS. As demonstrated, SERS probes the chemical composition of the water soluble fraction of the pollen grains. Thereby, new and complementary information on the complex composition of pollen grains is gained in addition to the results from other spectroscopic methods that are mainly focused on the pollen outer shell. Multivariate classification using HCA and PCA of the SERS data to achieve separation of spectra from different plant species is possible to some extent, but intra-species spectral variation due to the nature of the SERS experiment is in some cases greater than variation between species. This finding is in accordance with other recent works on the application of multivariate analyses to SERS data in bioanalytics, e.g., in diagnostics [33, 34], microorganisms classification [35, 36], and plant research [37]. Here, analysis of the SERS data with an artificial neural network showed that intrinsic biochemical information of the pollen cells can be utilized in spite of the great intra-species spectral variance that is caused by variations in the SERS signals and by preparation specifics. We find that the SERS data can be used for classification and identification of different pollen species. Therefore, future applications aim for a large ANN database for SERS-based pollen classification. Furthermore, we show that ANN cannot only be exploited to identify single components in complex mixtures as shown recently [25], but also for the classification of complex biological extracts. In the work presented here, ANN were used to pre-select spectra that contain information relevant for classification, enabling additional analyses, e.g. by PCA or HCA. This provides a means to use SERS data for bioanalytical identification purposes with other complex samples as well.

We conclude that SERS in combination with ANN shows great potential for improved pollen detection and warning. The combination of the SERS data with other vibrational spectroscopic information

will contribute to a comprehensive characterization of pollen biochemistry. Unlike other spectroscopic methods for pollen investigation [3, 7] and classification [2, 4], SERS probes the chemistry of the cellular fraction of the pollen grains. Therefore, it will be very useful to investigate pollen physiology, and the impact of environmental factors.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website.

Acknowledgements We acknowledge funding by European Research Council (ERC) Grant No. 259432 (MULTIBIOPHOT). J.K. also acknowledges support by Einstein Foundation Berlin (grant A-2011-77).

The authors thank Thomas Dürbye of the Botanic Garden and Botanical Museum Berlin-Dahlem and Carsten Später of Bezirksamt Berlin Tempelhof-Schöneberg for their support during sample collection.

Author biographies Please see Supporting Information online.

References

[1] C. S. Pappas, P. A. Tarantilis, and M. G. Polissiou, Appl. Spectrosc. **57**, 23–27 (2003).
[2] F. Schulte, J. Lingott, U. Panne, and J. Kneipp, Anal. Chem. **80**, 9551–9556 (2008).
[3] F. Schulte, J. Mader, L. W. Kroh, U. Panne, and J. Kneipp, Anal. Chem. **81**, 8426–8433 (2009).
[4] B. Zimmermann, Appl. Spectrosc. **64**, 1364–1373 (2010).
[5] B. G. Pummer, H. Bauer, J. Bernardi, B. Chazallon, S. Facq, B. Lendl, K. Whitmore, and H. J. Grothe, Raman Spectrosc. **44**, 1654–1658 (2013).
[6] N. Ivleva, R. Niessner, and U. Panne, Anal. Bioanal. Chem. **381**, 261–267 (2005).
[7] F. Schulte, U. Panne, and J. Kneipp, J. Biophotonics **3**, 542–547 (2010).
[8] B. Zimmermann and A. Kohler, PLoS One **9**, e95417 (2014).
[9] J. Brooks and G. Shaw, Trans. Bose Res. Inst. (Calcutta) **40**, 19–38 (1977).
[10] K. S. Gould and C. Lister, Flavonoids: Chemistry, Biochemistry and Applications (CRC Press Taylor & Francis Group, Boca Raton, FL, USA, 2006), pp. 414–415.
[11] J. Brooks and H. Shaw, Nature **219**, 532–533 (1968).
[12] R. W. Scott and M. J. Strohl, Phytochemistry **1**, 189–193 (1962).
[13] H. Behrendt and W. M. Becker, Curr. Opin. Immunol. **13**, 709–715 (2001).
[14] M. L. Laucks, G. Roll, G. Schweiger, and E. J. Davis, J. Aerosol Sci. **31**, 307–319 (2000).
[15] A. R. Boyain-Goitia, D. C. S. Beddows, B. C. Griffiths, and H. H. Telle, Appl. Opt. **42**, 6119–6132 (2003).
[16] V. Joseph, F. Schulte, H. Rooch, I. Feldmann, I. Dorfel, W. Osterle, U. Panne, and J. Kneipp, J. Chem. Commun. **47**, 3236–3238 (2011).
[17] V. Joseph, M. Gensler, S. Seifert, U. Gernert, J. P. Rabe, and J. Kneipp, J. Phys. Chem. C **116**, 6859–6865 (2012).
[18] Y. Liu, B. R. Upadhyaya, and M. Naghedolfeizi, Appl. Spectrosc. **47**, 12–23 (1993).
[19] H. G. Schulze, M. W. Blades, A.V. Bree, B. B. Gorzalka, L. S. Greek, and R. F. B. Turner, Appl. Spectrosc. **48**, 50–57 (1994).
[20] V. O. Santos, F. C. C. Oliveira, D. G. Lima, A. C. Petry, E. Garcia, P. A. Z. Suarez, and J. C. Rubim, Anal. Chim. Acta **547**, 188–196 (2005).
[21] L. A. Reisner, A. Cao, and A. K. Pandya, Chemom. Intell. Lab. Syst. **105**, 83–90 (2011).
[22] W. Schumacher, M. Kuehnert, P. Roesch, and J. Popp, J. Raman Spectrosc. **42**, 383–392 (2010).
[23] S. X. Li, Q. Y. Zeng, L. F. Li, Y. J. Zhang, M. M. Wan, Z. M. Liu, H. L. Xiong, Z. Y. Guo, and S. H. Liu, Biomed. Opt., DOI: 10.1117/1.JBO.18.2.027008.
[24] P. Negri, J. Y. Choi, C. Jones, S.M. Tompkins, R. A. Tripp, and R. A. Dluhy, Anal. Chem. **86**, 6911–6917 (2014).
[25] S. Kasera, L. O. Herrmann, J. del Barrio, J. J. Baumberg, and O. A. Scherman, Sci. Rep. **4**, DOI: 10.1038/srep06785.
[26] P. C. Lee and D. Meisel, J. Phys. Chem. **86**, 3391–3395 (1982).
[27] T. Udelhoven, D. Naumann, and J. Schmitt, Appl. Spectrosc. **54**, 1471–1479 (2000).
[28] S. Stewart and P. M. Fredericks, Spectrochimica Acta A **55**, 1641–1660 (1999).
[29] J. Bukowska and K. Jackowska, J. Electroanal. Chem. **322**, 347–356 (1992).
[30] G. T. Webster, K. A. De Villiers, T. J. Egan, S. Deed, L. Tilley, M. J. Tobin, K. R. Bambery, D. McNaughton, and B. R. Wood, Anal. Chem. **81**, 2516–2524 (2009).
[31] R. Goodacre, Vib. Spectrosc. **32**, 33–45 (2003).
[32] K. H. Liland, A. Kohler, and V. Shapaval, Chemometr. Intell. Lab. Syst. **138**, 41–47 (2014).
[33] D. Lin, S. Feng, J. Pan, Y. Chen, J. Lin, G. Chen, S. Xie, H. Zeng, and R. Chen, Optics Express **19**, 13565–13577 (2011).
[34] S. Feng, D. Lin, J. Lin, B. Li, Z. Huang, G. Chen, W. Zhang, L. Wang, J. Pan, R. Chen, and H. Zeng, Analyst **138**, 3967–3974 (2013).
[35] O. E. Rivera-Betancourt, R. Karls, B. Grosse-Siestrup, S. Helms, F. Quinn, and R. A. Dluhy, Analyst **138**, 6774–6785 (2013).
[36] H. Zhou, D. Yang, N. P. Ivleva, N. E. Mirescu, R. Niessner, and C. Haisch, Anal. Chem. **86**, 1525–1533 (2014).
[37] L. Qiu, P. Liu, L. Zhao, M. Q. Wen, H. Y. Yang, S. G. Fan, and L. Z. Zhou, Vib. Spectrosc. **72**, 134–141 (2014).