Received: 27 October 2017 | Revised: 15 March 2018 | Accepted: 29 March 2018

DOI: 10.1002/ece3.4138

ORIGINAL RESEARCH

WILEY Ecology and Evolution Open Access

An optimized protocol for large-scale *in situ* sampling and analysis of volatile organic compounds

Jyothi V. Nair | Pragadheesh V. Shanmugam | Snehal D. Karpe | Uma Ramakrishnan | Shannon Olsson

National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore, India

Correspondence
Jyothi V. Nair, National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore, India.
Email: jyothivn@ncbs.res.in

Funding information
Stiftelsen Olle Engkvist Byggmästare, Sweden, Grant/Award Number: 2014/254 and 2016/348; NCBS-TIFR, Bangalore; CSIR, New Delhi; SPM-CSIR, New Delhi; Science and Engineering Board, India

Abstract

Chemical ecology is an ever-expanding field with a growing interest in population- and community-level studies. Many such studies are hindered due to lack of an efficient and accelerated protocol for large-scale sampling and analysis of chemical compounds. Here, we present an optimized protocol for such large-scale study of volatiles. A large-scale *in situ* study to understand role of semiochemicals in variation in mating success of lekking blackbuck was conducted. Suitable methods for sampling and statistical analysis were identified by testing and comparing the efficiencies of available techniques to reduce analysis time while retaining sensitivity and comprehensiveness. Solid-phase extraction using polydimethylsiloxane, analysis using a semiautomated detection of retention time and base peak, and statistical analysis using random forest algorithm were identified as the most efficient methods for large-scale *in situ* sampling and analysis of volatiles. The protocol for large-scale volatile analysis can facilitate evolutionary and metaecological studies of volatiles *in situ* from all types of biological samples. The protocol has potential for wider application with the analysis and interpretation methods being suitable for all kinds of semiochemicals, including nonvolatile chemicals.

KEYWORDS

chemical ecology, gas chromatography, lek mating, mass spectrometry, pheromones, polydimethylsiloxane, random forests, regression, volatiles

1 | INTRODUCTION

Chemical communication is the oldest and most ubiquitous mode of communication in the living world (Wyatt, 2003). From chemotaxis in bacteria (Adler, 1975) to foraging trails in ants (Billen & Morgan, 1998), examples of organisms using chemical cues to find resources (Leonhardt, Menzel, Nehring, & Schmitt, 2016; Wilson, 1965), conspecifics (Brennan & Kendrick, 2006; Dweck et al., 2015) or to avoid danger (Holopainen & Blande, 2012; Mathis & Smith, 1993) can be found across the tree of life. The sheer diversity and abundance of chemicals in the environment provide an opportunity for organisms to utilize them for survival and reproduction. Organisms often employ multiple chemical compounds as blends in specific ratios, from two to six compounds in moth pheromones (Roelofs, 1995) to mixtures of several compounds for individual recognition in mammals (Brennan & Kendrick, 2006). These chemicals are also released into a complex natural environment with thousands of chemicals emanating from every microbe, plant, and animal. This cacophony of information can make the collection, separation, and interpretation of chemical signals a daunting task, particularly for unknown analytes

embedded in a natural matrix. While organisms such as plants and insects have been extensively studied and constitute a large proportion of our existing knowledge of chemical cues, systems such as mammals (Burger, 2005) and marine organisms (Hay, 2014) have been relatively less frequently explored and are largely restricted to zoo or captive individuals, due to difficulty in large-scale sampling under natural conditions.

Large-scale sampling and analysis of infochemicals have become particularly relevant in recent years due to a remarkable shift in chemical ecology from an individual-centric (Vet, 1999) to a more population- and community-centric approach (Dicke, 2006). Chemical ecologists are now interested not only in the chemical cues produced by organisms and the behavioral responses elicited, but also at the individual variations in these infochemicals and responses (Vet, 1999), their impact on populations and interspecies interactions (Hay, 2014), and the role of chemical communication in shaping communities and ecosystems (Dicke, 2006). Nevertheless, large-scale studies of complex chemical matrices pose a considerable challenge for sampling, analysis, and interpretation of volatiles. In this study, we have attempted to address the major concerns associated with large-scale *in situ* sampling and analysis of volatile organic compounds (VOCs) and have developed an optimized protocol for such studies.

A wide variety of sampling techniques is available to extract and retain the volatile profile of a biological sample (Millar & Haynes, 1998). Large-scale volatile sampling requires a good balance of sensitivity, chemical retention and preservation, and sampling comprehensiveness. Relatively small amounts of volatiles are released from biological samples and diffuse in a large volume of air, which requires volatile sampling techniques to be highly sensitive to capture them. Often, large-scale sampling requires collection over long duration to obtain the required sample size, necessitating the storage of samples for lengthy periods before analysis and increasing the risk of microbial contamination and loss or degradation of sample (e.g., Birkemeyer et al., 2016). In addition, the complete volatile profile of a biological sample is often difficult to obtain due to differences in adsorption/absorption efficiencies of the volatile constituents. Different sampling techniques might thus capture different subsets of the same original volatile composition. When studying a novel organism with a large repertoire of unknown volatiles, it is therefore important to choose the most comprehensive sampling technique. *In situ* techniques can eliminate the need to collect and store samples but must also be optimized for sensitivity and comprehensiveness.

Identification is a major rate-limiting step in large-scale analysis of semiochemicals. Analysis of VOCs is commonly performed using gas chromatography for separation (Dewulf, Van Langenhove, & Wittmann, 2002) and mass spectrometry for identification and quantification. Mass spectrum for each extracted volatile is visually compared to reference chromatograms in mass spectral libraries to initiate the process of identification. This is a highly laborious process and becomes especially difficult when sample sizes are large, and samples are rich in volatiles, such as studies that look at individual recognition in populations (O'Dwyer & Nevitt, 2009). Analysis is particularly arduous for novel biological systems as existing libraries (Wiley MS library [http://www.palisade.com], National Institute of Standards and Technology MS library [http://www.nist.gov], and binbase library; Skogerson, Wohlgemuth, Barupal, & Fiehn, 2011) are often ill-equipped for natural product identification, particularly rare compounds.

Finally, statistical analysis of volatiles should be relevant to the study system as well as adhere to the nature of volatile data. Typically, volatile profiles of samples belonging to two or more distinct sample sets are compared using clustering or classification tools to identify distinguishing volatiles from each set. However, classifying a study system into explicit sets may not always be possible and/or justifiable. Researchers may, in some cases, be interested in the variation in volatile composition across the range of a particular factor or may be unsure about the classification categories. In such scenarios, regression and clustering analysis can be used, respectively, for functional inferences. A major limiting factor for statistical analysis of VOCs is high dimensionality, that is, relatively larger number of variables/volatiles (p) than samples (n) (Johnstone & Titterington, 2009). Another important aspect of analysis of volatiles which is often ignored is that volatile data are represented as relative proportions (instead of absolute concentrations) of extracted volatiles. Such "compositional" data require transformations rendering them suitable for standard statistical tools (Aitchison, 1982) or tools that conform to the nonindependent nature of such data (Ranganathan & Borges, 2011).

We have addressed concerns particular to large-scale field studies of volatiles at each stage: sampling, analysis, and interpretation of VOCs. In an attempt to alleviate as many of these concerns as possible, we have developed a pipeline for large-scale *in situ* studies of VOCs. To develop our pipeline, as a case study, we have explored the role of chemical communication in lek mating behavior of an antelope, Indian blackbuck, *Antilope cervicapra*.

2 | MATERIALS AND METHODS

2.1 | Case study species

The blackbuck is an antelope endemic to the Indian subcontinent. It is a near-threatened species, primarily found in grasslands and open woodlands in India (IUCN). Blackbuck are known to breed throughout the year with two annual mating peaks, March–April and August–October (Ranjitsinh, 1989). They have a wide variety of mating systems, including the rare lek mating, in which males aggregate and display to females on small, clustered, resource-less territories (Isvaran, 2003). There is a strong spatial skew in mating success of males in a blackbuck lek (aggregation of males), with 90% of the matings occurring in central territories (Isvaran & Jhala, 1999). All territories are repeatedly marked by males with dung and urine that accumulate to form dung piles (Figure 1a). The dung piles are periodically evaluated by potential mates as well as competitors, possibly for olfactory cues about age (like in white rhino, Marneweck, Jürgens, & Shrader, 2017), strength, virility, genetic compatibility, or

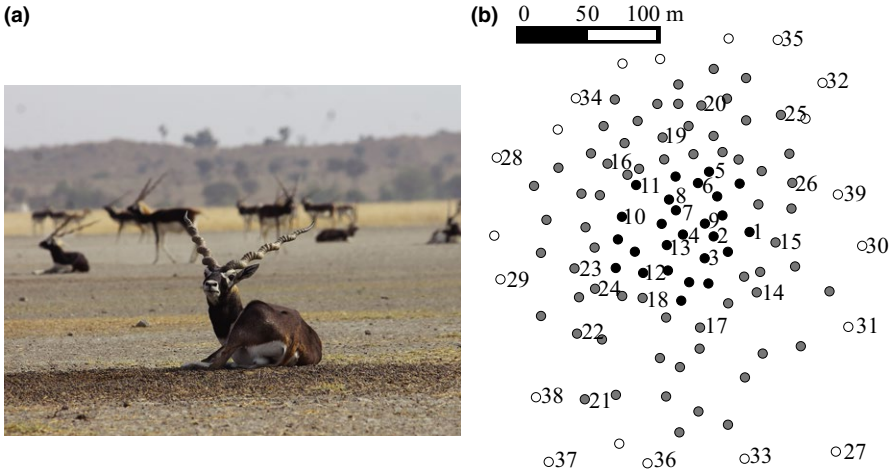

FIGURE 1 Large-scale chemical analysis case study system, the blackbuck, *Antilope cervicapra*. (a) Male blackbuck on territorial dung pile on lek. (b) Topography of dung piles on lek in February 2016 with principal dung piles (filled circles, central—dark gray, middle—light gray, and peripheral—white) and sampling dung piles (center—1 to 13, middle—14 to 26, and periphery—27 to 39)

identity (e.g., MHC and MUPS in mice, Cotton, 2007) of the defecator. A study on captive blackbuck (Rajagopal, Archunan, Geraldine, & Balasundaram, 2010) highlights variation in volatile profile of urine of males corresponding to dominance hierarchy. Similarly, in a lek, olfactory cues from dung piles are likely contributors to spatial variation in mating success of males and, in that case, are hypothesized to have a spatial variation corresponding to mating success.

2.2 | Case study site

Volatile sampling for the study was conducted in Tal Chhapar wildlife sanctuary, Rajasthan, India (27°47′53″N 74°26′06″E), during March–April 2015 and 2016. Tal Chhapar is a densely populated blackbuck sanctuary (>4,000 blackbuck in 800 ha area of grassland) hosting a single lek occupied by 150(±20) males during the mating peak. Before the onset of the study, all principal (large) dung piles on the lek were marked using GPS (Figure 1b) and the mean of their latitudes and longitudes was designated as lek center (Isvaran & Jhala, 1999). All territories within 65 m from lek center were defined as central territories (C), territories on the edge of the lek (120–250 m from lek center) were defined as peripheral territories (P), and the rest (between 65 and 120 m) were defined as middle territories (M). To determine spatial variation in volatile profiles of dung piles, 13 territories from each zone were selected for volatile sampling (Figure 1b).

2.3 | Sample collection

Freshly defecated fecal pellets (10–15 pellets) were collected, whenever available, from each of the 39 selected territories in sterilized glass vials using sterilized gloves. Owing to dry (15% humidity), hot (temp 35–43°C), and windy climate at sampling site during summer, pellets become very dry and brittle within 3–4 hr of defecation; hence, moisture and softness of pellet were used to detect freshness. Samples were collected on alternate days during least active period on the lek (12:00 noon–2:00 p.m.) to cause minimum disturbance to the animal. We collected 150 samples in the first season of sampling (March 2015) and 178 samples in second season (March 2016). Volatile extraction was optimized using season 1 samples in laboratory within 4 months of sampling. Optimized sampling protocol was used to extract season 2 samples *in situ*. Samples (season 1) were stored in −80°C till extraction, and extracts (seasons 1 and 2) were stored at −20°C till analysis to minimize microbial growth and loss of volatiles.

2.4 | Volatile extraction

Volatiles were extracted from fecal samples by three different techniques—solvent extraction, solid-phase extraction, and thermal desorption and their relative yields (number of volatiles and contaminants) were assessed using gas chromatography–mass spectrometry (GC-MS).

2.4.1 | Solvent extraction

Solvent extraction was performed using two most widely used solvents for volatile extraction (Millar & Haynes, 1998)—hexane (nonpolar) and dichloromethane (DCM, midpolar). Three different exposure times (3, 6, and 9 hr) were used to extract volatiles from five samples of March 2015. Three to four blackbuck pellets from each sample collection vial were ground together to obtain six replicates of 2 g each and immersed in approximately 5 ml solvent for each of the solvents and exposure times. The extracts were filtered using Whatman filter paper. Traces of water in the filtrate were removed using anhydrous sodium sulfate, and the filtrate was concentrated by evaporating the solvent using a slow stream of ultra-high-purity nitrogen gas. The concentrate was directly subjected to GC-MS analysis.

2.4.2 | Solid-phase extraction

Solid-phase extraction (SPE) was performed using preconditioned polydimethylsiloxane (PDMS) tubes procured from Carl Roth (Rotilabo®–silicone tube). PDMS tubes of 1.5 mm inner diameter and 3.5 mm outer diameter were cut into 5 mm pieces and soaked

for 4 hr in 1:1 mixture of acetonitrile and methanol. They were then dried using ultra-high-purity nitrogen gas and conditioned in a Gerstel Tube Conditioner by heating over the stream of nitrogen gas at 4 bar constant pressure. The entire process was repeated twice before using for extraction. For the extraction, two PDMS tubes were exposed to fecal pellets collected in each glass vial for 4 hr. On each day of sampling, an empty glass vial used as environmental control and volatiles were extracted using two tubes from this vial as well. The tubes were then removed and stored in labeled, sterilized 0.5-ml amber glass vials.

2.4.3 | Thermal desorption

Thermal desorption was performed using a Gerstel Thermal Desorption Unit (TDU) and Cooled Injection System (CIS 4) controlled by Gerstel Modular Analytical Systems Controller C506 and Gerstel Maestro 1 software which extracts volatiles and directly introduces them into a GC-MS for analysis. One fecal pellet from each collection vial was crushed using a sterilized spatula, and 2 g of the powder was added into the TDU liner by covering both the end of the liner with glass wool. The samples were introduced into the TDU at the initial temperature of 30°C using a Gerstel MultiPurpose Sampler (MPS). After a delay time and initial temperature of 1 min each at 30°C, the TDU temperature was increased to 200°C at the rate of 100°C/min and retained in 200°C for 10 min. Volatiles were desorbed from TDU and transferred to CIS at 210°C, trapped in the silanized glass wool liner of the CIS, and maintained at −50°C using liquid nitrogen. After the equilibration time of 0.20 min, the CIS was ramped to 220°C at the rate of 12°C/s and held at constant temperature for 5 min for optimal transfer of volatiles to GC. TDU-CIS was also used to desorb and introduce volatiles from solvent extract/PDMS tube into GC-MS.

2.5 | GC-MS analysis

Volatiles extracted from samples were separated and identified using an Agilent 7890B gas chromatograph coupled with a 5977A MSD mass spectrometer. An HP-5 MS column (30 m × 0.25 mm id, 0.25 μm film thickness) was used with helium as the carrier gas at a flow rate of 1 ml/min. The column oven was kept at 40°C for 1 min, increased to 180°C at a rate of 5°C/min, and finally increased to 270°C (with a 5 min holding temperature) in the second ramp at 25°C/min. Samples were introduced to the GC in solvent vent mode, directly or through TDU, with a purge flow to split vent of 30 ml/min at 1.5 min and vent flow of 70 ml/min; vent pressure was held at 7.07 psi for 0.01 min. The transfer line between the GC and MS was maintained at 250°C, whereas the source and quadrupole temperatures were 230 and 150°C, respectively. Ionization was performed in electron impact mode with ionization energy of 70 eV. GC-MS acquisition was performed using Agilent MassHunter Workstation software B.07.02.1938, and qualitative analysis was assessed by MassHunter Qualitative Analysis version B.07.00.

GC chromatograms of the different volatile extracts of replicate samples were compared to assess best volatile sampling method. Each GC chromatogram was also compared to a corresponding blank control to check for contaminants. Blank controls were empty sterilized glass vials exposed to similar environmental conditions and extraction procedures as the samples. The optimized laboratory-based volatile sampling protocol was then tested for feasibility for large-scale *in situ* sampling in season 2. Further optimization of downstream analysis was carried out using these samples.

Volatile analytes were identified by matching the mass spectral data of the peak with library spectra (NIST and personal libraries created from authentic standards), by comparing their relative retention index using a homologous series of n-alkanes (C_6–C_{30} hydrocarbons, Sigma-Aldrich), by comparing their elution order and/or by comparing their retention time with standards. Quantity of volatiles extracted was approximated by the area under each peak. A known peak (octamethylcyclotetrasiloxane) was taken as an internal standard, and all peak areas were normalized by dividing with the peak areas of internal standard for SPE extracts (Kallenbach et al., 2014). Contaminants are removed by comparing the spectra of samples with corresponding controls.

2.6 | Statistical analysis

In this study, we were interested in understanding the spatial variation in volatiles from dung piles across the lek. This variation can either be discrete with dung piles in the center of the lek being remarkably different from all other dung piles (like the mating behavior) or gradual with a stepwise change in volatile composition from center to periphery. As a result, it was necessary to assess different statistical approaches for proper biological inference of the chemical data. It also gave us the opportunity to explore different statistical approaches used in chemical studies and arrive at the most suitable tools for large-scale studies in general. Three common statistical approaches are—(1) clustering, (2) classification, and (3) regression. Clustering is an unsupervised statistical approach and was used to look for natural clusters of territories with similar chemical composition. Classification is a supervised statistical approach for which we predefined our study system into zones (center, middle, and periphery, as described in sampling methods) and analyzed chemical variation between these zones. To observe gradual variation, we used regression approaches and distance of each territory from the center of lek was used as the parameter to note variation. To optimize our protocol, we used two alternate statistical tools from each approach and compared their efficiencies using relevant statistics. The tools used were principal component analysis (PCA) and hierarchical clustering for clustering, linear discriminate analysis (LDA) and random forest classification (RF) for classification, and principal component reduction (PCR) and random forest regression for regression analysis.

Random forest is a machine learning algorithm which can be used to assess the importance of variables (volatiles) in classification and regression analysis (Breiman, 2001; Ehrlinger, 2015). It

builds decision trees using bootstrapping from samples and creates a ranked variable importance list by running permutations of decision trees. Variables to be considered can be assessed based on corresponding variable importance scores—mean decrease in accuracy (for classification) and increase in mean square error (for regression). PCR (Jolliffe, 1982) is a dimensionality reduction approach based on PCA used to circumvent problems of multidimensionality in data such as GC-MS data. PCA is performed on the observed data matrix for the explanatory variables to obtain the principal components (PCs), and then a subset of the PCs is selected, based on some appropriate criteria (e.g., variability explained), for the intended multivariate analysis (e.g., multivariate regression in PCR).

Principal component analysis, hierarchical clustering, and LDA were performed using PAST3 software (Ryan, Hammer, Harper, & Paul Ryan, 2001), an open source software for statistical analysis. Random forest classification and regression models and PCR were optimized, cross-validated (10 CV), and compared using "caret" package (Kuhn, 2015) in R. Classification models were compared based on classification accuracy, and regression models were compared based on root-mean-square error (RMSE).

3 | RESULTS AND DISCUSSION

3.1 | Volatile sampling

Thermal desorption was identified as the most comprehensive method with the highest number of volatiles and few contaminants (Figure 2a). It was assessed as an ideal choice for large-scale volatile sampling if field-collected samples can be immediately transferred to laboratories for ex situ extraction. It was also determined to be devoid of issues of volatile preservation post extraction as coupling with GC enabled immediate analysis of volatiles extracted by TDU. However, it was not suitable for the chosen case study as the crushed blackbuck pellets obstructed the liner of CIS risking malfunction of

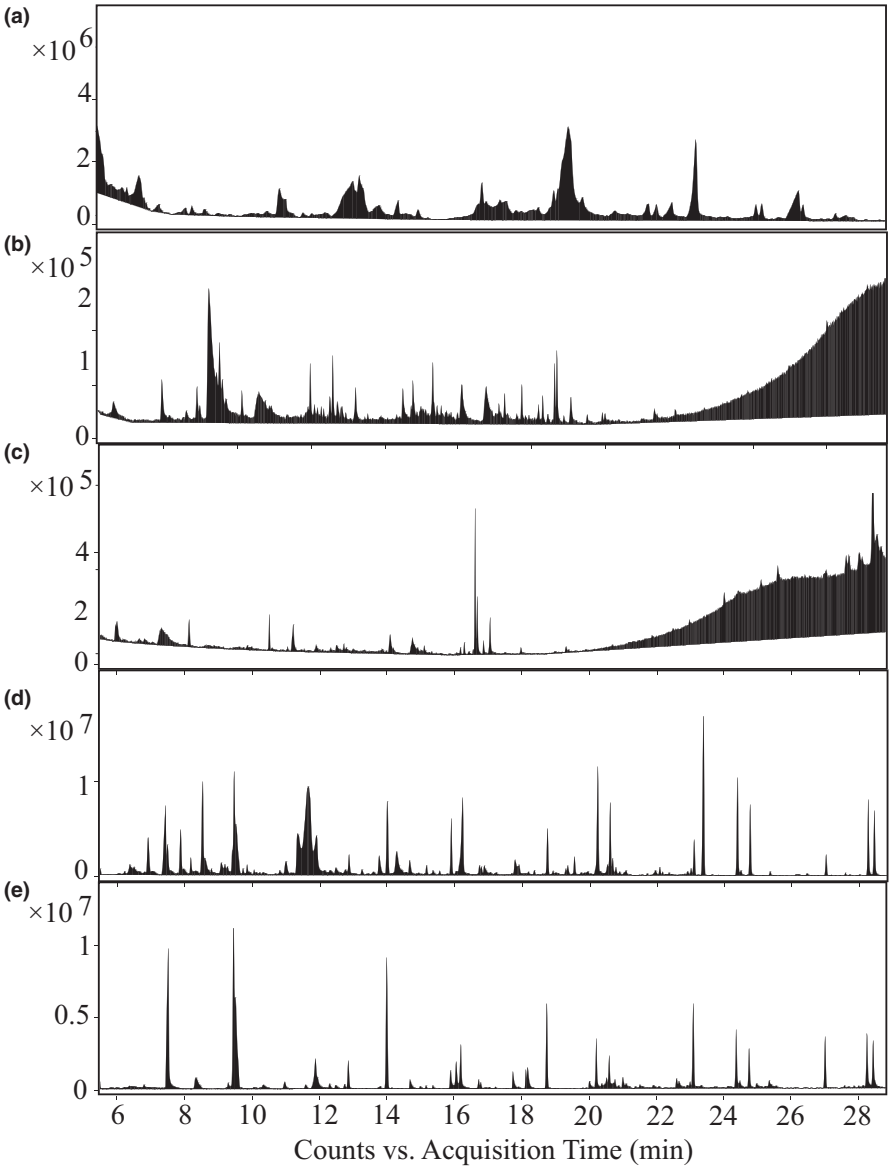

FIGURE 2 Chromatograms of volatiles extracted from blackbuck fecal pellets using different volatile sampling techniques. (a) Thermal desorption. (b) Solvent extraction using dichloromethane (9 hr). (c) Solvent extraction using Hexane (9 hr). (d) Solid-phase extraction using polydimethylsiloxane (PDMS). (e) Volatiles extracted from blank using PDMS

the instrument. Besides, TDU is a laboratory-based technique and therefore requires removal of biological samples from the field. This was not ideal for our study as we were unable to immediately analyze the samples and fecal matter is subject to rapid microbial conversion altering the VOC profile from that which is present *in situ*.

Solvent extraction was the least comprehensive method of extraction in this study. Both hexane and DCM extracted very few analytes despite extended periods of exposure (9 hr) with blackbuck fecal samples (Figure 2b,c). Even though DCM was successfully used to extract volatiles from urine in a previous study on captive blackbuck (Rajagopal et al., 2010), it was found to be not appropriate for fecal samples in the current study. Solvent extraction was recognized as a highly sensitive technique that requires little equipment and sample, albeit with the risk of degradation of the source sample as well as volatiles by the solvent during preparation of extracts. Maceration of the tissue for solvent extraction can also lead to release of chemicals not generally present in the headspace of the system *in situ* and lead to misinformation. This method was thus concluded to be not suitable for large-scale *in situ* sampling of volatiles, especially for novel study systems where volatiles are previously unknown.

Solid-phase extraction using PDMS followed by thermal desorption extracted many volatiles in our samples (Figure 2d), but the sampling PDMS tube itself produced many contaminants (Figure 2e), unlike TDU or solvent extraction (Figure S1). However, the contaminants from the PDMS tubes could be recognized by their mass spectral peaks and comparison with blank samples and could therefore be removed from analysis. PDMS was assessed as convenient for large-scale *in situ* volatile sampling tool due to low cost of the material, compactness, and no requirement of equipment in the field collection stage. The solid phase of PDMS does not provide an active surface to catalyze the artefactual reaction of compounds, which may happen during thermal desorption directly from samples or when using other solid-phase extraction materials such as TENAX or PORAPAK. Solid-phase extraction under ambient conditions is therefore more likely to reflect compounds that are volatilized from the sample, whereas both solvent extracts and direct thermal desorption can include compounds which would not be volatilized under ambient conditions. Solid-phase extraction using PDMS also retains the analytes unaltered for a longer time when stored in −20°C (Kallenbach et al., 2014).

Our chosen method was successfully used for concurrent large-scale *in situ* volatile sampling of 39 blackbuck territories on lek (Figure 1b) over a period of 15 days (season 2 of sampling) and 178 volatile samples (62 central, 65 middle, and 52 peripheral) were collected. For volatile preservation postextraction for long durations (3–4 months in this study), PDMS tubes were stored in −20°C. This technique is fairly sensitive and convenient for replicate sampling, as multiple PDMS tubes can be used to sample volatiles from same biological samples which can be stored for backup. The only potential drawback is that PDMS has biased affinity to volatiles of low molecular weight (Kallenbach et al., 2014) and thus might underrepresent the volatile profile of a sample.

A thorough experimental comparison of three commonly used volatile sampling techniques in terms of comprehensiveness, sensitivity, volatile preservation, and suitability for large-scale *in situ* sampling concludes that solid-phase extraction using PDMS along with thermal desorption is the most efficient and practical sampling technique for large-scale study of volatiles. While other field-based sampling and analysis methods are available, including solid-phase microextraction and portable GC-MS (Kücklich et al., 2017; Marneweck et al., 2017), these methods are both costly and cannot be applied for very large-scale concurrent sampling as required in this study. In particular, our method is effective for initial studies that can reveal major analytes within and between samples. However, other methods could be more suitable for detailed and/or small-scale analyses requiring a comprehensive overview of the total VOC profile.

3.2 | Analysis of volatiles

Fast and efficient analysis of large number of volatiles from the SPE samples was performed by avoiding recursive analysis of same volatile in different samples using an automation protocol to group them. Groups were made based on retention time (RT) and base peak (BP) allowing a margin of error of 0.1 s in RT to account for peak shifts. RT, BP, and peak area of each peak in a chromatogram were obtained in a tabular form using Agilent MassHunter Workstation software. Such tables were acquired for all samples and corresponding controls. A PERL code (see Appendix S1) was used to generate a metatable of all RT-BP combinations of all peaks in samples and controls which was then used for RT-BP grouping, contamination removal, as well as for normalization of the peak areas with internal standards (allowed margin of error in RT, BP value for known contaminants, and RT-BP for internal standard are user provided, see Appendix S1). RT-BP combinations missing in a sample or control were denoted as zeros in the metatable. Three representatives from each RT-BP group (volatiles with same RT [±0.1 s as defined for this study]—BP from three different samples) were selected, and their mass spectra were compared with each other for validation of the PERL code and verified for the removal of the PDMS contaminants. Each RT-BP group ideally corresponds to a single analyte. Comparison of mass spectra revealed two kinds of errors—(1) different volatiles grouped in same RT-BP group (0.01%) and (2) same volatile grouped in different RT-BP groups (4.2%). These errors were low and asserted the use of RT-BP grouping as a useful automation tool for analysis. All contaminants were also successfully removed. Contaminant removal included removing analytes (RT-BP groups) from samples which were present in controls as well as contaminants with known BP values, like silicates (see Appendix S1). For further analysis, the number of samples in which each analyte was present was calculated and analytes that were found in less than 5% of the total samples were eliminated to reduce noise. Mass spectrometer chromatogram and relative retention index (RRI) (Kováts, 1958) of one representative from each RT-BP group was used to identify the corresponding compound.

TABLE 1 Volatiles from blackbuck lek dung piles

Compound no.	Compound name	Mode of identification	RRI
1	Farnesene type compound	MS	–
2	6-Methyl-2-heptanone	MS and RRI	958
3	Diterpene 1	MS	–
4	β-Citronellene	MS and RRI	948
5	2-Octanone	MS and RRI	997
6	2,6,6-Trimethyl-2-cyclohexene-1,4-dione	MS and RRI	1,145
7	1-Pentanol	MS	–
8	*m*-Cresol	MS and RRI	1,077
9	4-Heptanone	MS and RRI	876
10	6-Methyl-5-heptene-2-one	MS and RRI	958
11	Unidentified 1	–	–
12	2-Decanone	MS and RRI	1,202
13	β-Dihydroterpeneol	MS and RRI	1,140
14	(*E*)-β-ocimene	MS and RRI	1,054
15	4-Propylphenol	MS and RRI	1,273
16	2,6-Lutidine	MS	–
17	Unidentified 2	–	–
18	β-Cyclocitral	MS and RRI	1,223
19	3,7-Dimethyl-2-octene	MS and RRI	966
20	2-Nonanone	MS and RRI	1,101
21	2,2,6-Trimethylcyclohexanone	MS and RRI	1,033
22	2-Undecanone	MS and RRI	1,296
23	Unidentified 3	–	–
24	(*Z*)-β-Ocimene	MS and RRI	1,045
25	Unidentified 4	–	–
26	β-Caryophyllene	MS and RRI	1,430
27	Hexadecane	MS and RRI	1,601
28	Nonane	MS and RRI	900
29	γ-Terpinene	MS and RRI	1,065
30	Indole	MS and RRI	1,296
31	*Cis-m*-menth-8-ene	MS and RRI	1,030
32	2,4,7,9-Tetramethyl-5-decyne-4,7-diol	MS and RRI	1,430
33	*p*-Menth-4(8)-ene	MS and RRI	1,021
34	(2*E*)-3,7,11,15-Tetramethyl-2-hexadecene	MS	–
35	2-Methyl-1-heptene-6-one	MS and RRI	972
36	Isophorone	MS and RRI	1,130
37	Unidentified 5	–	–
38	1a,2,5,5-Tetramethyl-trans-1a,4a,5,6,7,8-hexahydro-gamma-chromene	MS and RRI	1,350
39	Unidentified 6	–	–
40	Unidentified 7	–	–
41	*p*-Menth-2-ene	MS and RRI	1,009
42	Unidentified 8	–	–
43	*Trans*-pinane	MS and RRI	976
44	Modephene	MS and RRI	1,385

(Continues)

TABLE 1 (Continued)

Compound no.	Compound name	Mode of identification	RRI
45	Unidentified 9	–	–
46	Unidentified 10	–	–
47	Unidentified 11	–	–
48	Pinocarvone	MS and RRI	1,164
49	Menthone	MS and RRI	1,156
50	2(4H)-Benzofuranone,5,6,7,7a-tetrahydro-4,4,7a-trimethyl-	MS and RRI	1,539
51	Butyl cyclohexane	MS and RRI	1,037
52	Unidentified 12	–	–
53	1H-Cycloprop[e]azulene, decahydro-1,1,4,7-tetramethyl-, [1aR-(1aα,4β,4aβ,7β,7aβ,7bα)]	MS and RRI	1,371
54	2-Phenylethyl alcohol	MS and RRI	1,119
55	Unidentified 13	–	–
56	Unidentified 14	–	–
57	Unidentified 15	–	–
58	1-Cyclohexyl ethanone	MS and RRI	991
59	3-Heptadecene, (Z)-	MS and RRI	1,691
60	Unidentified 16	–	–
61	Frontalin	MS and RRI	938
62	Humulene	MS and RRI	1,457
63	Unidentified 17	–	–
64	Unidentified 18	–	–
65	2,4-Dimethyl phenethyl alcohol	MS	–
66	Unidentified 19	–	–
67	*Trans*-pinocarveol	MS and RRI	1,139
68	Unidentified 20	–	–
69	*p*-Cresol	MS and RRI	1,078
70	β-Selinene	MS and RRI	1,490
71	Unidentified 21	–	–
72	Unidentified 22	–	–
73	Unidentified 23	–	–
74	2-Methyl thiophene	MS and RRI	775
75	Furan-2,5-dihydro-3,4-dimethyl	MS	–
76	Hexyl-(*E*)-2-methylbut-2-enoate	MS and RRI	1,278
77	2-Pentadecanone-6,10,14-trimethyl	MS and RRI	1,853
78	Unidentified 24	–	–
79	Acenaphthene	MS and RRI	1,485
80	Copaene	MS and RRI	1,373
81	2-Heptanone-4,6-dimethyl-	MS and RRI	1047
82	Phenyl acetaldehyde	MS and RRI	1,047
83	1,2,4,4-Tetramethylcyclopentene	MS and RRI	867
84	Isomenthone	MS and RRI	1,170
85	Dihydroeudesmol	MS and RRI	1,666
86	Unidentified 25	–	–
87	1-Hexanol	MS and RRI	866
88	Cyclosativene	MS and RRI	1,369

(Continues)

TABLE 1 (Continued)

Compound no.	Compound name	Mode of identification	RRI
89	Isopropyl myristate	MS and RRI	1,836
90	2-Methyl decalin	MS and RRI	1,139
91	Unidentified 26	—	—
92	2-Pentyl furan	MS and RRI	991
93	Unidentified 27	—	—
94	3,5-di-*tert*-Butyl-4-hydroxybenzaldehyde	MS and RRI	1,772
95	Phenylethyl ketone	MS and RRI	1,741
96	Unidentified 28	—	—
97	Unidentified 29	—	—
98	Unidentified 30	—	—
99	Unidentified 31	—	—
100	Unidentified 32	—	—

MS, mass spectral match; RRI, relative retention index match.
Unidentified—compounds could not be identified using mass spectra and relative retention index.

In this study, more than 200 unique analytes were detected and quantified out of which 100 (Table 1) were found in at least 5% of the samples. These analytes were sorted in decreasing order of their abundance, and their ranks in the sorted list were used as their identification number. Chemical names of 68 of these analytes were identified (Table 1). RT-BP-based grouping and rank-based nomenclature of compounds ensured inclusion of all detected analytes for downstream comparative analysis, which could have otherwise been overlooked due to difficulty in identification of their chemical name. This is particularly useful for studies of novel organisms such as blackbuck which potentially have several compounds not listed in existing mass spectrum libraries. The developed semiautomation analysis method helped in reducing the interpretation time of 178 blackbuck samples (from season 2), each of which had more than 400 total analytes (~200 sample volatiles and >200 PDMS-induced contaminants per sample) to roughly 170 hr (2 weeks at 12 hr per day, including chromatogram integration, acquiring RT-BP data, running the code, and cross-validation). By hand, taking even just 5 min per compound ID, these numbers of samples would take over 6,000 hr. This efficiency of RT-BP-based large-scale analysis of volatiles makes it extremely beneficial for large-scale population-level studies that investigate individual or genetic as well as spatial, temporal, or environment-induced variations in chemical signatures. The Perl script used in this study is applicable to data obtained from other GC-MS software apart from Agilent MassHunter. The input file

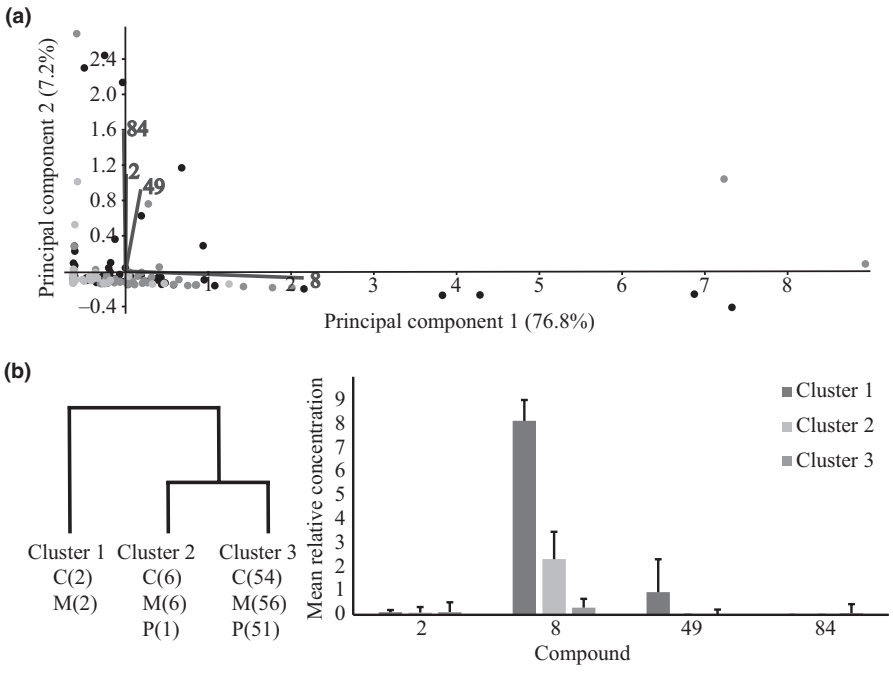

FIGURE 3 Statistical analysis using different clustering algorithms. (a) Principal component analysis (using SVD algorithm, 1,000 bootstraps), filled circles indicate samples from central (black), middle (dark gray), and peripheral (light gray) territories. Principal component analysis biplots (thick gray lines) for four most variable compounds (dark gray numbers) are shown (b) Hierarchical clustering (using Ward's method. Euclidian distance. 1,000 bootstraps). Numbers in brackets in dendrogram represent number of samples from each zone. Bar plot shows intercluster variation in concentrations of the four most variable compounds

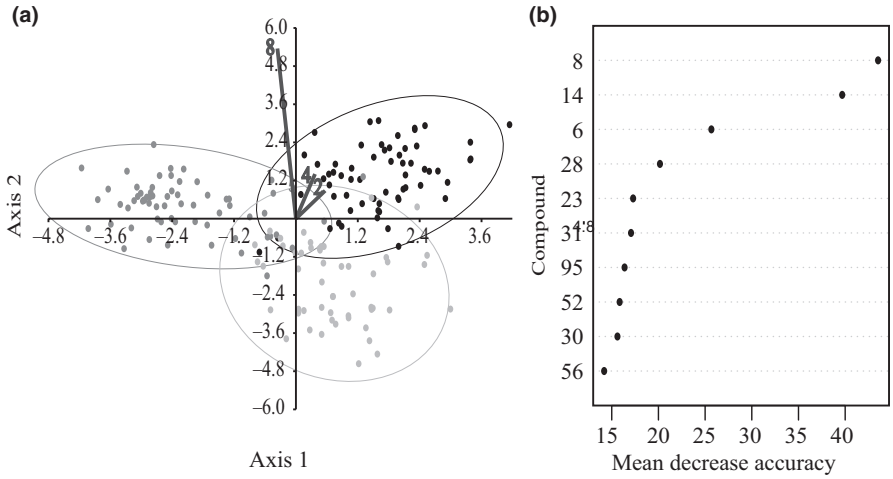

FIGURE 4 Statistical analysis using different classification algorithms. (a) Linear discriminant analysis. (b) Random forests—top 10 important volatiles distinguishing center, middle, and periphery based on mean decrease in accuracy. Filled circles in linear discriminate analysis scatter plot indicate samples from central territories (black), middle (dark gray), and periphery (light gray). Biplots (thick gray lines) for three most variable compounds (dark gray numbers) are shown

needed for the script is a simple comma-separated file (.csv) consisting of peak area, base peak, and retention time (see Appendix S1). This method is less sophisticated than the several existing freely available and open source data preprocessing programmes such as XCMS (Smith, Want, O'Maille, Abagyan, & Siuzdak, 2006) and CAMERA (Kuhl, Tautenhahn, Böttcher, Larson, & Neumann, 2012) but has an inherent advantage in terms of simultaneous removal of many known contaminants (like PDMS-derived volatiles in this study). While silica-based compounds have signature base peaks that can be fed into the Perl script for easy removal, other contaminants with unique retention time and base peak can also be easily removed with user-provided information (Appendix S1).

3.3 | Statistical analysis

Principal component analysis did not produce strong clusters with unscaled (Figure 3a) as well as scaled and log-transformed data (Figure S2), while hierarchical clustering produced three clusters, varying predominantly in concentration of compound 8 "*meta*-cresol" (Figure 3b). High levels of *meta*-cresol were detected in a few samples from central and middle territories, whereas almost all samples from peripheral territories showed low concentrations of *meta*-cresol (Figure 3a,b). Overall, there was more variation between volatile profiles of territories within a zone (C/M/P) than between the zones in chemical composition. Clustering tools alone were thus not sufficient to convey meaningful information about spatial variation in chemical signature of blackbuck dung piles, probably due to individual variation being much higher than spatial variations.

Classification by LDA had high accuracy of 92.8% in sorting samples to corresponding zones (Figure 4a). Classification by random forest classification model had a lower efficiency (65.10%). Among the zones, classification accuracy was highest for central territories (Table 2). Compound 8 (*meta*-cresol) was the most important volatile distinguishing between the zones (Figure 4b). Both methods arrived at similar results, and in first glance, LDA analysis appeared to be better than random forests in terms of classification accuracy. Accuracy checks on a test dataset using the 10 cross-validation

TABLE 2 Classification error in optimized random forest model

Classification zone	Classification error
C	0.2096774
M	0.4218750
P	0.4038462

C, center; M, middle; P, periphery.

method revealed that the random forest model produced consistent results, and the model was optimized to use 1,000 trees (ntree) and 15 variables per try (mtry) to produce 65.10% accuracy. LDA, however, failed at 10 cross-validation possibly due to large number of zero values in the data, so the model could not be optimized. Jackknifing reduced the classification efficiency of LDA drastically to 17.42%. In this study, among these two methods, random forest was thus determined as a better statistical approach in terms of accuracy and consistency to assess discrete spatial variation in chemical composition of lek. Considerable difference between the three zones in chemical composition corroborated the hypothesis of spatial variation in mating success of males being correlated with spatial variation in chemical signature of the lek. *meta*-cresol was again determined as the most important driver of this variation.

Among the two regression models, optimized random forest regression model (ntree = 1,000, ntry = 5, 10 CV) explained 26.17% of variation in volatile composition across territories (Figure 5a). The margin of root-mean-square error (RMSE = 55.512 m) is very high (>1/5th of the distance between center and outermost territory on the lek) indicating a weak trend along distance from center. For principal component regression, PCA was used for dimensionality reduction and model optimization. The optimized model with four principal components had higher error (RMSE = 60.82273 m) than random forests (Figure 5b). Regression analyses detected a small but consistent variation in chemical composition of lek from center to periphery. *meta*-cresol was again predicted as the most important compound varying with distance from center (Figure 5c). Random forest regression method was determined as a better tool to assess gradual spatial variation in chemical composition of lek.

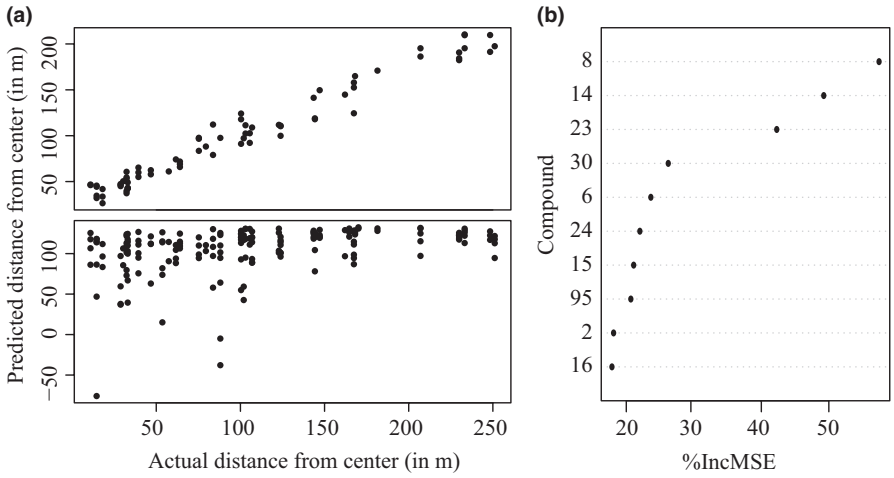

FIGURE 5 Statistical analysis using different regression models. (a) Regression plot between actual and predicted distance of samples from lek center—random forest regression (top), principal component regression (bottom). (b) Top 10 important volatiles varying with distance from center based on increasing mean square error (IncMSE) using random forest regression

All statistical tools explored in this study, in consensus, indicated a strong spatial variation in concentration of *m*-cresol that corresponded to spatial variation of mating success of male blackbuck in a lek. However, random forest algorithms that operate by construction of several decision trees were more efficient (as measured by accuracy and RMSE) than the classical ordination-based multivariate analysis tools used in this study. Standard multivariate statistics, like PCA, operate with the assumption of independence of data points, which does not hold true for compositional and interdependent data like VOCs (Ranganathan & Borges, 2011), as volatiles are usually scored in terms of relative proportions Commonly used corrective measures such as square root transformations or log transformations with the addition of a constant (ranging from 0.01 to 0.00001) to accommodate zero data points could greatly alter the dataset and should be used with caution. Most multivariate tools also falter when the dataset consists of more variables than samples, that is, when total number of analytes is higher than total number of samples. Although dimensionality reduction approaches such as PCR can be used (like in this study), this approach proves ineffective if the chosen subset does not help in significantly reducing the number of variables. The random forest approach is ideal for large-scale analysis of volatiles due to its inherent ability to deal with correlation between variables, high classification and regression efficiency, and prediction of variables of importance. It allows for more variables than samples and does not overfit the data (Breiman, 2001; Ranganathan & Borges, 2011). Large-scale studies can be affected by inherent noise in the system as well as manual and instrumental variations introduced during elucidation. Replicate sampling and analysis can help to track and counter such variations to a certain extent. Other common issues associated with large-scale studies such as repeated samples and nested data can be dealt with tools such as GLMM and random slopes (Jamil, Ozinga, Kleyer, & ter Braak, 2013; Weiß et al., 2017). A combination of multiple statistical approaches catering to the study system can also help to improve efficiency of analysis.

4 | CONCLUSIONS

Here, we have developed a pipeline for large-scale *in situ* studies of VOCs employing solid-phase *in situ* extraction, thermal desorption coupled with GC-MS, semiautomated analysis using retention time and base peak, and statistical analysis using random forests. Each of the selected methods exhibited both advantages and drawbacks, but each step was selected to maximize efficiency, sensitivity, chemical retention and preservation, and comprehensiveness of analysis. Our methodology helped in analyzing about 200 samples in two weeks (not including the GC-MS runtime) as opposed to 7–8 months (details explained in Section 2) that would have been required to do this work manually. While the TDU-GC-MS instrument is specialized and costly, the other techniques explored are inexpensive and require relatively little expertise to perform. In addition, our pipeline is not restricted to mammals as used for the case study but could be employed with nearly any living system in land or sea (PDMS can also absorb chemicals in aqueous solutions). Our method was also specifically developed for large-scale, population-level studies where neither compounds nor sample groups may be previously known. As such, it can be employed in new systems that have little prerequisite knowledge of the natural products involved. Further studies that employ this method may incorporate additional steps that increase the effectiveness of our novel pipeline methodology.

ACKNOWLEDGMENTS

Authors are thankful to the Rajasthan Forest Department, Jaipur, India, and Tal Chhapar wildlife sanctuary, Rajasthan, for permits to collect samples. Authors are grateful to Bhagwana Ram, Rupsy

Khurana, Anushree Suryavanshi, Pankhuri Chaudhary, Umadevi P, Irfan, Dilawar Khan, and Yash Ketkar for assistance in collection of samples and Mr. Srinivas Rao for assistance with GC-MS analysis of samples. We are thankful to Rupsy Khurana for providing blackbuck pictures used in figure 1a and graphical abstract. JVN is grateful to CSIR, New Delhi, for fellowship. SK is grateful to SPM-CSIR, New Delhi, for fellowship. VSP and SO were funded by Stiftelsen Olle Engkvist Byggmästare, Sweden (under grant nos. 2014/254 and 2016/348). SO was also funded by Science and Engineering Board, India (Ramanujan Fellowship). This project was completely funded by NCBS-TIFR, Bangalore.

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

JVN, SO, URK, and VSP conceived the ideas and designed methodology; JVN collected the data; VSP identified the compounds; SDK wrote the code for semiautomation analysis; JVN analyzed the data; JVN and SO led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

DATA ACCESSIBILITY

The data used in this study has been provided in Dryad Digital Repository: https://doi.org/10.5061/dryad.kp18283.

REFERENCES

Adler, J. (1975). Chemotaxis in bacteria. *Annual Review of Biochemistry*, *44*, 341–356. https://doi.org/10.1146/annurev.bi.44.070175.002013

Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society. Series B (Methodological)*, *44*, 139–177. https://doi.org/10.2307/2345821

Billen, J., & Morgan, E. D. (1998). Pheromone communication in social insects: Sources and secretions. In R. K. Vander Meer, M. D. Breed, M. L. Winston & K. E. Espelie (Eds.), *Pheromone communication in social insects: Ants*, *wasps*, *bees*, *and termites* (pp. 3–33). Boulder, CO: Westview Press.

Birkemeyer, C., Thomsen, R., Jänig, S., Kücklich, M., Slama, A., Weiß, B. M., & Widdig, A. (2016). The sampling of body odour of primates: Cotton swabs sample semi-volatiles rather than volatiles. *Chemical Senses*, *41*(6), 525–535. https://doi.org/10.1093/chemse/bjw056

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5–32. https://doi.org/10.1023/A:1010933404324

Brennan, P. A., & Kendrick, K. M. (2006). Mammalian social odours: Attraction and individual recognition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*, 2061–2078. https://doi.org/10.1098/rstb.2006.1931

Burger, B. V. (2005). Mammalian semiochemicals. *Topics in Current Chemistry*, *240*, 231–278. https://doi.org/10.1007/b98318

Cotton, S. (2007). Individual recognition: Mice, MUPS and the MHC. *Current Biology*, *17*, R971–R973. https://doi.org/10.1016/j.cub.2007.09.024

Dewulf, J., Van Langenhove, H., & Wittmann, G. (2002). Analysis of volatile organic compounds using gas chromatography. *TrAC Trends in Analytical Chemistry*, *21*, 637–646. https://doi.org/10.1016/j.cub.2007.09.024

Dicke, M. (2006). Chemical ecology from genes to communities : Integrating 'omics' with community ecology. In M. Dicke & W. Takken (Eds.), *Chemical ecology from genes to communities* (pp. 175–189). Dordrecht, the Netherlands: Springer.

Dweck, H. K. M., Ebrahim, S. A. M., Thoma, M., Mohamed, A. A. M., Keesey, I. W., Trona, F., … Hansson, B. S. (2015). Pheromones mediating copulation and attraction in *Drosophila*. *Proceedings of the National Academy of Sciences*, *112*, E2829–E2835. https://doi.org/10.1073/pnas.1504527112

Ehrlinger, J. (2015). ggRandomForests: Visually exploring a random forest for regression. *arXiv*, 1–28. https://doi.org/10.1073/pnas.1504527112

Hay, M. E. (2014). Challenges and opportunities in marine chemical ecology. *Journal of Chemical Ecology*, *40*, 216–217. https://doi.org/10.1007/s10886-014-0393-5

Holopainen, J. K., & Blande, J. D. (2012). Molecular plant volatile communication. *Advances in Experimental Medicine and Biology*, *739*, 17–31. https://doi.org/10.1007/s10886-014-0393-5

Isvaran, K. (2003). When should males lek? Insights from a dynamic state variable model. *Behavioral Ecology*, *14*, 876–886. https://doi.org/10.1093/beheco/arg066

Isvaran, K., & Jhala, Y. (1999). Variation in lekking costs in blackbuck (*Antilope cervicapra*): Relationship to lek-territory location and female mating patterns. *Behaviour*, *137*, 547–563. https://doi.org/10.2307/4535722

Jamil, T., Ozinga, W. A., Kleyer, M., & ter Braak, C. J. F. (2013). Selecting traits that explain species–environment relationships: A generalized linear mixed model approach. *Journal of Vegetation Science*, *24*(6), 988–1000. https://doi.org/10.1111/j.1654-1103.2012.12036.x

Johnstone, I. M., & Titterington, D. M. (2009). Statistical challenges of high-dimensional data. *Philosophical Transactions. Series A*, *Mathematical*, *Physical*, *and Engineering Sciences*, *367*, 4237–4253. https://doi.org/10.1098/rsta.2009.0159

Jolliffe, I. T. (1982). A note on the use of principal components in regression. *Applied Statistics*, *31*, 300. https://doi.org/10.2307/2348005

Kallenbach, M., Oh, Y., Eilers, E. J., Veit, D., Baldwin, I. T., & Schuman, M. C. (2014). A robust, simple, high-throughput technique for time-resolved plant volatile analysis in field experiments. *The Plant Journal*, *78*, 1060–1072. https://doi.org/10.1111/tpj.12523

Kováts, E. (1958). Gas-chromatographische charakterisierung organischer verbindungen. Teil 1: Retentionsindices aliphatischer halogenide, alkohole, aldehyde und ketone. *Helvetica Chimica Acta*, *41*, 1915–1932. https://doi.org/10.1002/hlca.19580410703

Kücklich, M., Möller, M., Marcillo, A., Einspanier, A., Weiß, B. M., Birkemeyer, C., & Widdig, A. (2017). Different methods for volatile sampling in mammals. *PLoS One*, *12*(8), e0183440. https://doi.org/10.1371/journal.pone.0183440

Kuhl, C., Tautenhahn, R., Böttcher, C., Larson, T. R., & Neumann, S. (2012). CAMERA: An integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Analytical Chemistry*, *84*(1), 283–289. https://doi.org/10.1021/ac202450g

Kuhn, M. (2015). A short introduction to the caret package. *R Foundation for Statistical Computing*, 1–10.

Leonhardt, S. D., Menzel, F., Nehring, V., & Schmitt, T. (2016). Ecology and evolution of communication in social insects. *Cell*, *164*, 1277–1287. https://doi.org/10.1186/1471-2105-12-321

Marneweck, C., Jürgens, A., & Shrader, A. M. (2017). Dung odours signal sex, age, territorial and oestrous state in white rhinos. *Proceedings of the Royal Society of London B: Biological Sciences*, *284*, https://doi.org/10.1186/1471-2105-12-321

Mathis, A., & Smith, R. J. F. (1993). Intraspecific and cross-superorder responses to chemical alarm signals by brook stickleback. *Ecology*, *74*, 2395–2404. https://doi.org/10.2307/1939590

Millar, J. G., & Haynes, K. F. (1998). *Methods in chemical ecology*. New York, NY: Chapman & Hall.

O'Dwyer, T. W., & Nevitt, G. A. (2009). Individual odor recognition in procellariiform chicks. *Annals of the New York Academy of Sciences*, *1170*, 442–446. https://doi.org/10.1111/j.1749-6632.2009.03887.x

Rajagopal, T., Archunan, G., Geraldine, P., & Balasundaram, C. (2010). Assessment of dominance hierarchy through urine scent marking and its chemical constituents in male blackbuck *Antelope cervicapra*, a critically endangered species. *Behavioural Processes*, *85*, 58–67. https://doi.org/10.1016/j.beproc.2010.06.007

Ranganathan, Y., & Borges, R. M. (2011). To transform or not to transform. *Plant Signaling & Behavior*, *6*, 113–116. https://doi.org/10.4161/psb.6.1.14191

Ranjitsinh, M. K. (1989). *The Indian blackbuck*. Dehradun, India: Natraj Publishers.

Roelofs, W. L. (1995). *Colloquium Paper: Roelofs*. *92*, 44–49.

Ryan, P. D., Hammer, Ø., Harper, D. A., & Paul Ryan, D. D. (2001). PAST: Paleontological statistics software package for education and data analysis. *Palaeontologia Electronica 178 kb. T. Harper. Geological Museum*, *4*, 5–7.

Skogerson, K., Wohlgemuth, G., Barupal, D. K., & Fiehn, O. (2011). The volatile compound BinBase mass spectral database. *BMC Bioinformatics*, *12*, 321. https://doi.org/10.1186/1471-2105-12-321

Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., & Siuzdak, G. (2006). XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry*, *78*(3), 779–787. https://doi.org/10.1021/ac051437y

Vet, L. E. M. (1999). From chemical to population ecology: Infochemical use in an evolutionary context. *Journal of Chemical Ecology*, *25*, 31–49. https://doi.org/10.1023/a:1020833015559

Weiß, B. M., Marcillo, A., Manser, M., Holland, R., Birkemeyer, C., & Widdig, A. (2017). A non-invasive method for sampling the body odour of mammals. *Methods in Ecology and Evolution*, *9*, 420–429. https://doi.org/10.1111/2041-210X.12888

Wilson, E. O. (1965). Chemical communication in the social insects. *Science (New York, N.Y.)*, *149*, 1064–1071. https://doi.org/10.1126/science.149.3688.1064

Wyatt, T. D. (2003). *Pheromones and animal behaviour*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511615061

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Nair JV, Shanmugham PV, Karpe SD, Ramakrishnan U, Olsson S. An optimized protocol for large-scale *in situ* sampling and analysis of volatile organic compounds. *Ecol Evol*. 2018;8:5924–5936. https://doi.org/10.1002/ece3.4138