Supporting Information for:

Automated compound speciation, cluster analysis, and quantification of organic vapors and aerosols using comprehensive two-dimensional gas chromatography and mass spectrometry.

Xiao He[1], Xuan Zheng[1*], Shuwen Guo[1], Lewei Zeng[1], Ting Chen[1], Bohan Yang[1], Shupei Xiao[1], Qiongqiong Wang[2], Zhiyuan Li[3], Yan You[4], Shaojun Zhang[5,6,7,8], and Ye Wu[5,6,7,8]

[1]College of Chemistry and Environmental Engineering, Shenzhen University, Shenzhen 518060, China

[2]Department of Atmospheric Science, School of Environmental Studies, China University of Geosciences, Wuhan 430074, China

[3]School of Public Health (Shenzhen), Sun Yat-sen University, Guangzhou 510275, China

[4]National Observation and Research Station of Coastal Ecological Environments in Macao, Macao Environmental Research Institute, Macau University of Science and Technology, Macao SAR 999078, China

[5]School of Environment, State Key Joint Laboratory of Environment Simulation and Pollution Control, Tsinghua University, Beijing 100084, China

[6]State Environmental Protection Key Laboratory of Sources and Control of Air Pollution Complex, Beijing 100084, China

[7]Beijing Laboratory of Environmental Frontier Technologies, School of Environment, Tsinghua University, Beijing 100084, China

[8]Laboratory of Transport Pollution Control and Monitoring Technology, Transport Planning and Research Institute, Ministry of Transport, Beijing 100028, China

Summary of the Supporting Information:

Pages: 14

Supplementary Table: 1

Supplementary Figures: 17

S1. Indicative reaction schemes that are included in the model development

Compounds containing hydrocarbon chains give rise to a series of ions distant from each other by 14 Da (-CH_2-) and the masses where they appear depend on the chemical structures linked to them. For example, aliphatic alkanes including *n-*/*i*-alkanes, dissociate to the formula C_nH_{2n+1} regularly, with the most abundant ion fragments being mass-to-charge ratio (m/z) 43, 57, 71, and 85, corresponding to C_3H_7, C_4H_9, C_5H_{11}, and C_6H_{13}, respectively, and the M^+ being weakly observed (Lavanchy, Houriet and Gäumann 1979). Aliphatic alkenes exhibit the most abundant ion fragments at m/z 41, 55, 69, and 83. A short alkyl chain attached to the benzene ring yields an abundant peak at m/z 91, with the base peak (ion with the highest abundance relative to other ions in the mass spectrum) shifting to larger ions gradually as the alkyl chain extends. Saturated rings undergo the loss of one methyl group and yields an abundant $[M - 15]^+$ signal. The appearance of the m/z 60 peak for carboxylic acids is a characteristic feature of the McLafferty rearrangement and provides valuable information about the identification of the compounds (Figure S3b). Aliphatic alcohols frequently cleave at the β bond to give a resonance-stabilized ion at m/z 31 (Figure S3c). Dissimilarly, aromatic alcohols generate a prominent M^+ ion, as the case for phenols. Esters consist of an alcohol and an acid component. When the acid portion is the major component, the fragmentation is partially characterized by typical acid peaks, where the loss of an alkyl radical through α-cleavage occurs predominantly, producing a discernable peak at m/z 74 (Figure S3f). Conversely, when the alcohol portion of the ester dominates, fragmentation similar to that of an alcohol is observed (Figure S3g). Trimethylsilyl derivatization is commonly used in GC analysis, and the fragments of m/z 73 and 147 are systematically observed in high abundance (He et al. 2018). Water elimination, indicated by an 18 Da loss, is common for hydroxyl groups (Harrison 2012). The McLafferty rearrangement dominates the mass spectrum via the transfer of a γ-hydrogen atom to a keto-group and the β-cleavage through a six-atom ring where appropriate (McLafferty 1959). Besides, the heteroatoms (e.g., oxygen) prevent the appearance of M^+, whereas the conjugated aromatic rings intensity the M^+. Aliphatic amines often undergo cleavage at the α-C−C bond to produce relatively stable ions: $CH_2NH_2^+$ (m/z 30), $C_2H_4NH_2^+$ (m/z 44), and $C_3H_6NH_2^+$ (m/z 58) for amine groups attached to the primary carbon, secondary carbon, and tertiary carbon, respectively (Figure S4a) (Mikaia 2022). M^+ is frequently detected for amides, following the nitrogen rule. α-cleavage or McLafferty rearrangement produces dominant ions in the mass spectrum, as illustrated in Figure S4b. Neutral loss of HC≡N and McLafferty rearrangement produce abundant $[M - 27]^+$ and m/z 41 ions for nitriles. C-N and N-O bonds are weak for aliphatic nitro compounds, and M^+ is seldomly observed in the EI spectrum. Aromatic nitro compounds exhibit additional fragmentation pattens, including ring arrangement and ring-opening reactions. The dominant ions originate from the loss of NO^+ (m/z 30) or NO_2^+ (m/z 46).

S2. The description of unresolved peaks

The unidentified peaks accounted for 15%, 18%, and less than 1% for HDDV vapor, HDDV aerosol, and ambient aerosol samples, respectively. Those identified peaks could be divided into three sub-categories:

The first category is column bleedings that appear in the GC×GC plot regularly underneath alkane groups. The majority of column bleedings come up at the end of GC thermal program when the temperature is over 240 °C. These column bleedings were isolated and removed for further analysis at the very beginning of data treatment processes. Some other column bleedings with aliphatic siloxane skeleton appeared regularly from the beginning to the end of the GC×GC tests. The number of silicon atoms within the skeleton varied much, as a consequence of which their characteristic ions differed, and these bleeding peaks mixed with the chromatographic peaks. Their mass spectrum and example chemical structures are shown in Figure S5.

The second category is ester compounds scattering in the GC×GC plot. We have observed some aliphatic ester peaks in the middle of the GC×GC plot, one example given in Figure S6. Their mass spectra were compared to spectra from National Institute of Standards and Technology (NIST20) library and these peaks were identified with high match scores. There were only a few of these peaks, contributing to less than 1% of the total signal and they were not frequently detected in every sample. In this regard, the ester compounds were not included for further discussion and classified as unidentified.

The third category refers to the unresolved peaks literally. The mass spectra of these peaks were disorganized and did not follow distinct patterns. Some of the fragment ions stem from their own

89 peaks and others were interferences from the matrix. These peaks could not be classified into any
90 given groups in the model and were labeled unidentified.

91 Table 1. The list and linear correlation of the authentic standards.

Group	Name	1st RT (min)	2nd RT (s)	R^2
Acid	Isobutyric Acid	4.67	1.25	0.99
Acid	2-Methyl butyric acid	7.07	1.36	0.99
Alcohol	1-Decanol	19.67	1.42	0.98
Alcohol	1-Hexadecanol	34.14	1.38	0.96
Alcohol	n-Nonadecanol	39.94	1.39	0.95
Alcohol	1-Docosanol	45.00	1.43	0.99
Aldehyde	Octanal	11.60	1.48	1.00
Aldehyde	Decanal	17.80	1.46	0.99
Aldehyde	Pentanal	3.73	1.17	0.97
Aldehyde	Hexanal	5.67	1.41	1.00
Aldehyde	Decanal	17.80	1.55	0.99
Alkane	Heptane	3.74	0.74	0.97
Alkane	Octane	5.60	0.87	0.98
Alkane	Nonane	8.34	0.94	1.00
Alkane	Decane	11.47	0.97	1.00
Alkane	Undecane	14.67	0.98	0.99
Alkane	Dodecane	17.67	0.97	0.97
Alkane	Tridecane	20.47	0.99	0.97
Alkane	Tetradecane	23.20	0.98	0.99
Alkane	Pentadecane	25.67	1.00	0.99
Alkane	Hexadecane	28.07	0.98	0.99
Alkane	Heptadecane	30.34	1.00	0.97
Alkane	Octadecane	32.54	1.04	0.97
Alkane	Nonadecane	34.60	1.03	0.99
Alkane	Eicosane	36.54	1.07	0.96
Alkane	Heneicosane	38.40	1.08	0.96
Alkane	Docosane	40.20	1.11	0.98
Alkane	Tricosane	41.94	1.12	0.97
Alkane	Tetracosane	43.60	1.12	0.97
Alkane	Pentacosane	45.20	1.17	0.99
Alkane	Hexacosane	46.74	1.22	0.99
Alkane	Heptacosane	48.20	1.20	0.98
Alkane	Octacosane	49.60	1.25	0.98
Alkane	Nonacosane	51.00	1.27	0.97
Alkane	Triacontane	52.34	1.33	0.96
Alkane	Hentriacontane	53.60	1.41	0.97
Alkane	Dotriacontane	54.87	1.53	0.96
Alkane	Tritriacontane	56.14	1.77	0.96
Alkane	Tetratriacontane	57.60	2.09	0.96
Alkane	Pentatriacontane	59.27	2.46	0.99
Alkene	1-Octene	5.39	0.90	0.99

Alkene	1-Octadecene	32.39	1.07	0.99
Alkene	1-Docosene	40.13	1.13	0.99
Alkene	1-Decene	11.19	0.99	0.98
Alkene	1-Dodecene	17.39	1.03	0.96
Alkene	1-Tetradecene	22.99	1.02	0.98
Alkyl-benzene	Ethylbenzene	7.27	1.46	1.00
Alkyl-benzene	p-Xylene	7.54	1.42	0.99
Alkyl-benzene	o-Xylene	8.14	1.57	0.99
Alkyl-benzene	Isopropylbenzene	9.07	1.51	1.00
Alkyl-benzene	1,3,5-Trimethylbenzene	10.54	1.55	0.99
Alkyl-benzene	Butyl-benzene	13.27	1.55	0.99
Alkyl-benzene	Pentyl-benzene	16.40	1.54	0.97
Alkyl-benzene	Hexyl-benzene	19.40	1.50	0.97
Alkyl-benzene	4-Ethyltoluene	10.40	1.61	0.98
Alkyl-benzene	p-Cymene	12.33	1.61	0.97
Alkyl-PAHs	1-Methylnaphthalene	20.87	2.44	1.00
Alkyl-PAHs	1-Ethylnaphthalene	23.14	2.41	0.97
Amide	Acetamide	5.07	2.51	1.00
Amide	Propanamide	7.00	2.79	0.96
Amide	N,N-Dibutylformamide	20.54	1.95	0.95
Amine	Triethylamine	3.60	0.86	0.99
Amine	Dibutylamine	10.27	1.18	0.99
Amine	Aniline	10.74	2.58	0.98
Amine	1,2-Benzenediamine	17.80	3.50	1.00
Amine	1-Naphthalenamine	26.60	3.57	0.96
Amine	4-Biphenylamine	31.54	3.56	0.99
Cycloalkane	Ethyl-cyclohexane	6.59	0.99	0.95
Cycloalkane	Butyl-cyclohexane	12.59	1.12	0.99
Cycloalkane	Hexyl-cyclohexane	18.86	1.16	0.97
Cycloalkane	Octyl-cyclohexane	24.46	1.13	0.95
Cycloalkane	Decyl-cyclohexane	29.46	1.15	0.95
Cycloalkane	Dodeyl-cyclohexane	33.93	1.21	0.97
Cycloalkane	Tetradecyl-cyclohexane	37.99	1.24	0.97
Cycloalkane	Hexadecyl-cyclohexane	41.66	1.29	0.98
Ester	Butyl acetate	5.93	1.39	0.98
Ester	Amyl Acetate	8.73	1.51	0.97
Ester	Isoamyl Acetate	7.66	1.39	0.99
Furan	Furan	12.46	1.40	0.96
Ketone	2-Nonanone	14.34	1.50	0.98
Ketone	2-Dodecanone	23.00	1.43	0.97
Ketone	2-Pentadecanone	30.34	1.38	0.99
Ketone	2-Pentanone	3.53	1.09	0.99
Ketone	3-Heptanone	7.93	1.49	0.99
Oxy-PAH	1,4-Naphthalenedione	23.60	3.36	0.99

Oxy-PAH	9,10-Anthracenedione	36.07	3.85	0.99
Oxy-PAH	Benz(a)anthracene-7,12-dione	46.74	4.62	0.98
PAHs	Indene	12.94	2.12	0.99
PAHs	Naphthalene	17.26	2.95	1.00
PAHs	Acenaphthylene	24.53	3.38	1.00
PAHs	Acenaphthene	25.39	3.27	0.99
PAHs	Fluorene	27.79	3.30	1.00
PAHs	Phenanthrene	32.26	3.73	0.99
PAHs	Anthracene	32.53	3.70	0.99
PAHs	Fluoranthene	37.00	3.04	0.99
PAHs	Pyrene	38.86	4.38	0.99
PAHs	Benz[a]anthracene	44.59	4.58	0.97
PAHs	Chrysene	44.73	4.76	0.97
PAHs	Benzo[b]fluoranthene	49.26	5.15	0.96
PAHs	Benzo[k]fluoranthene	49.39	5.15	0.98
PAHs	Benzo[a]pyrene	50.53	5.61	0.97
PAHs	Indeno[1,2,3-cd]fluoranthene	55.46	3.02	0.99
Pinene	α-Pinene	9.46	1.19	0.95
Pinene	β-Pinene	10.86	1.35	0.99
Polyphenyl	Biphenyl	22.74	2.44	0.95

92

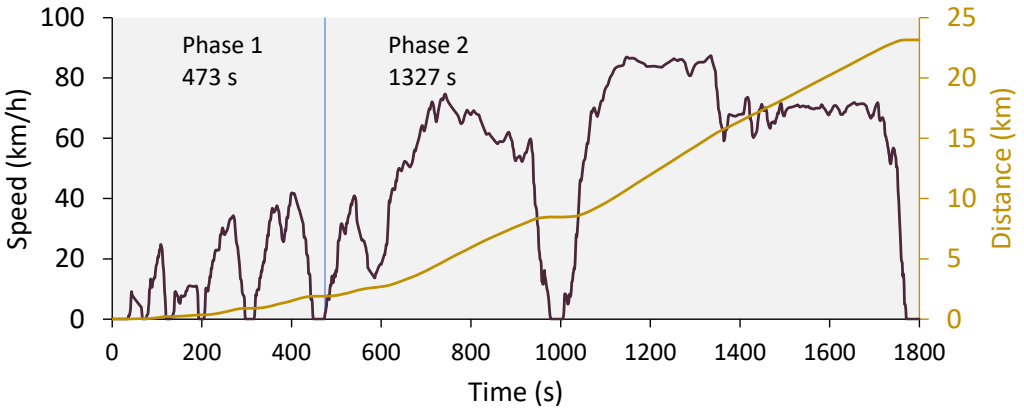


Figure S1. Speed trace for China heavy-duty commercial vehicle test cycle for tractor trailers (CHTC-HT) cycle. The total duration is 1800 s and divided into two phases of 344 s and 1330 s. The average speed is 46.5 km/h, and the idle ratio is 8.6%.

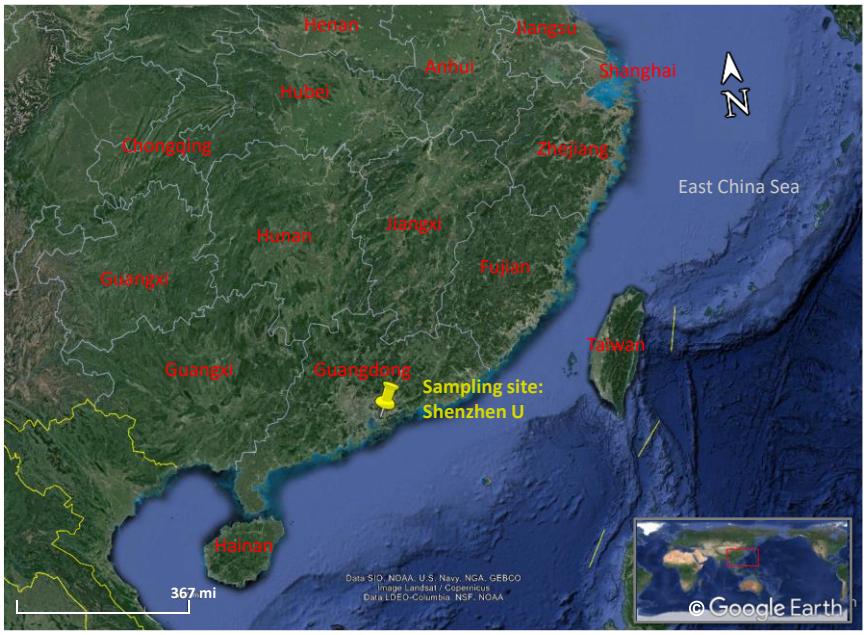
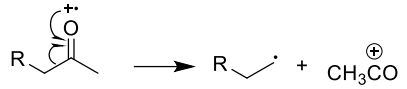


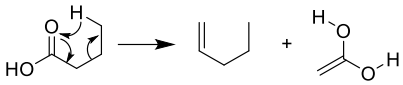
Figure S2. Sampling site location on Shenzhen University campus. $PM_{2.5}$ high-volume sampler was installed on the rooftop of a 5-story building surrounded by university campus, residential areas, greenbelts, and a golf park.

(a) Radical initiated α-cleavage



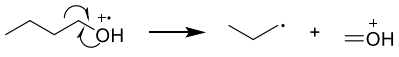
m/z 43

(b) McLafferty rearrangement



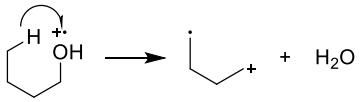
m/z 88 m/z 60

(c) Loss of CH_2OH



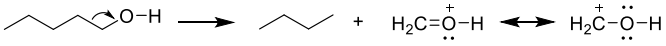
m/z 74 m/z 31

(d) Loss of H_2O



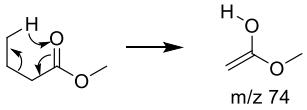
m/z 56

(e) β-cleavage



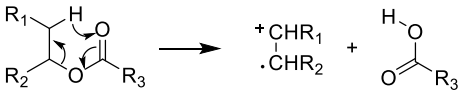
m/z 31

(f) Ester with acid portion dominant



m/z 74

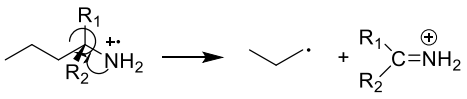
(g) Ester with alcohol portion dominant



R3 ≪ R1 or R2

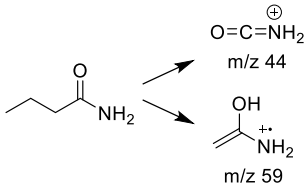
Figure S3. Interpretation of characteristic ions in the EI spectrum for O-containing species, including (a) aliphatic ketones, (b-d) aliphatic alcohols, (e) carboxylic acids, (f) esters with acid portion dominant, and (g) esters with alcohol portion dominant.

(a) Radical initiated α-cleavage



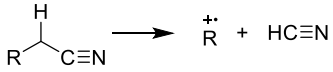
R_1, R_2 = H, m/z 30
R_1 = CH_3, R_2 = H, m/z 44
R_1 = CH_3, R_2 = CH_3, m/z 58

(b) McLafferty rearrangement

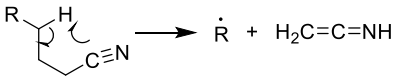


m/z 44
m/z 59

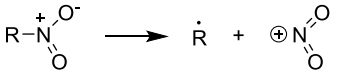
(c) Loss of HCN



(d) McLafferty rearrangement

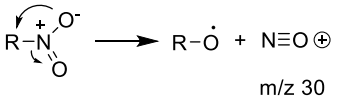


(e) Loss of NO_2



m/z 46

(f) Loss of NO



m/z 30

Figure S4. Interpretation of the characteristic ions in the EI spectrum for N-containing species, including (a) aliphatic amines, (b) amides, (c) and (d) nitriles, and (e) organic nitrates, and (e) organic nitrites.

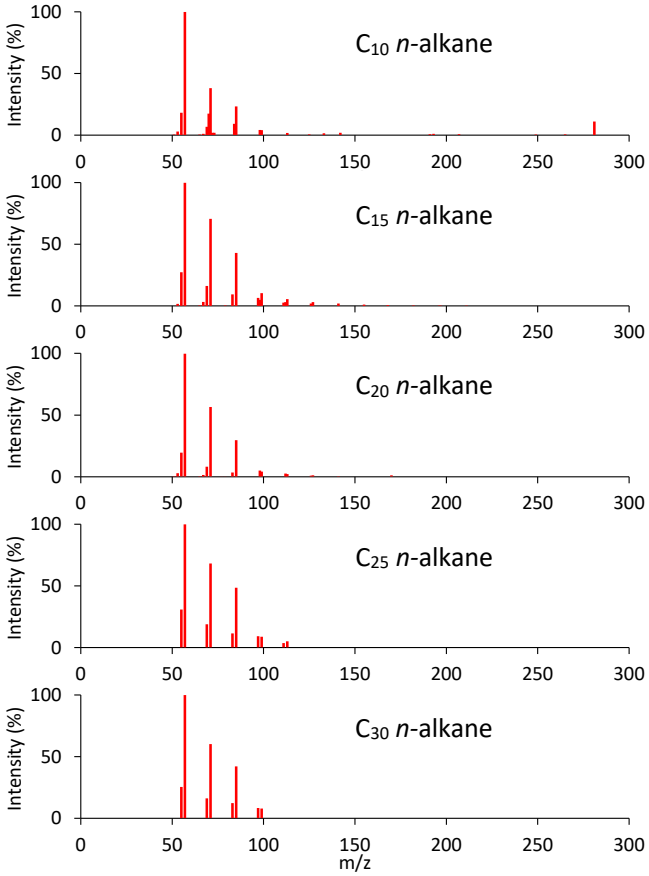


Figure S5. The mass spectra of homologous *n*-alkane standards from C_{10} to C_{30}, which exhibit high similarity. In such cases, the similarity score could be elevated incorrectly and mislead compound identification results.

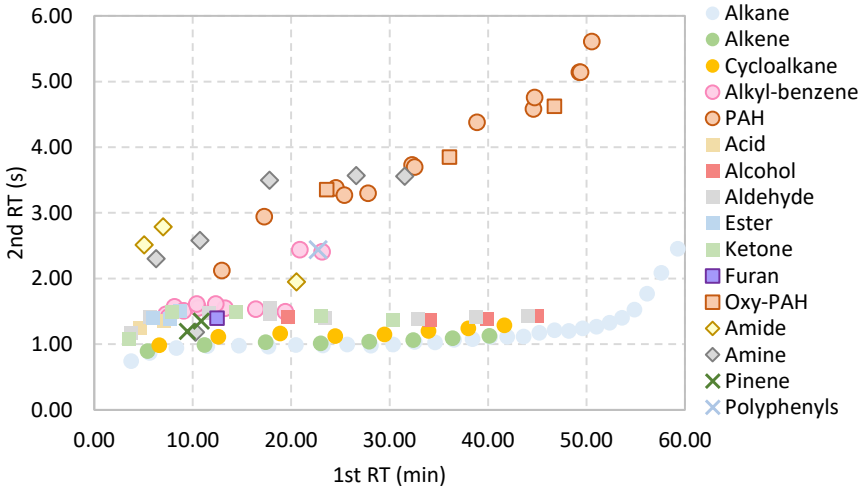


Figure S6. The distribution of all authentic standards, coded by colors and shapes.

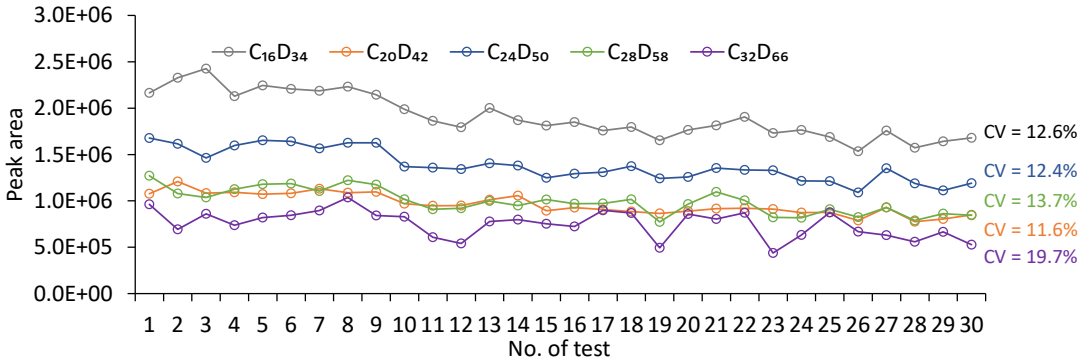


Figure S7. The variation of peak areas of internal standards tested throughout the whole measurement.

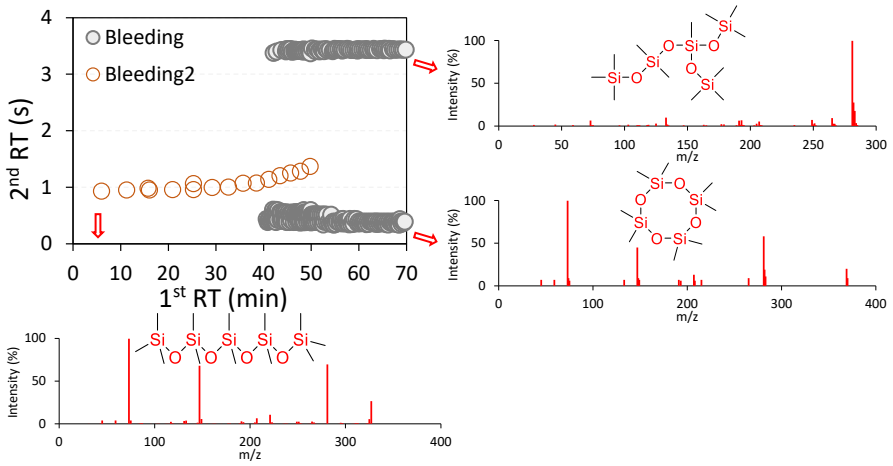


Figure S8. The distribution of the column bleedings in an example GC×GC plot and their respective mass spectra and chemical structures.

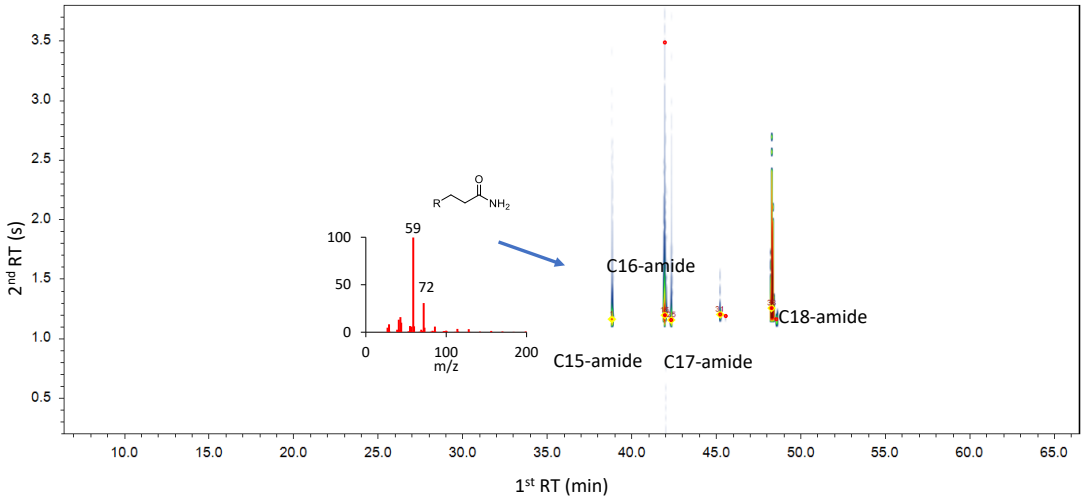


Figure S9. The chromatogram and mass spectra of tentatively identified amide representatives.

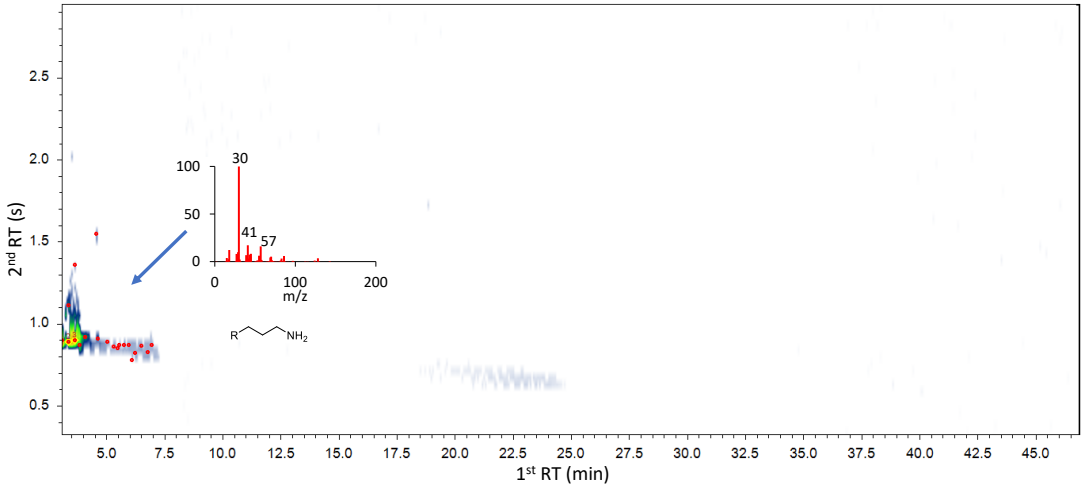


Figure S10. The chromatogram and mass spectra of tentatively identified amine representatives.

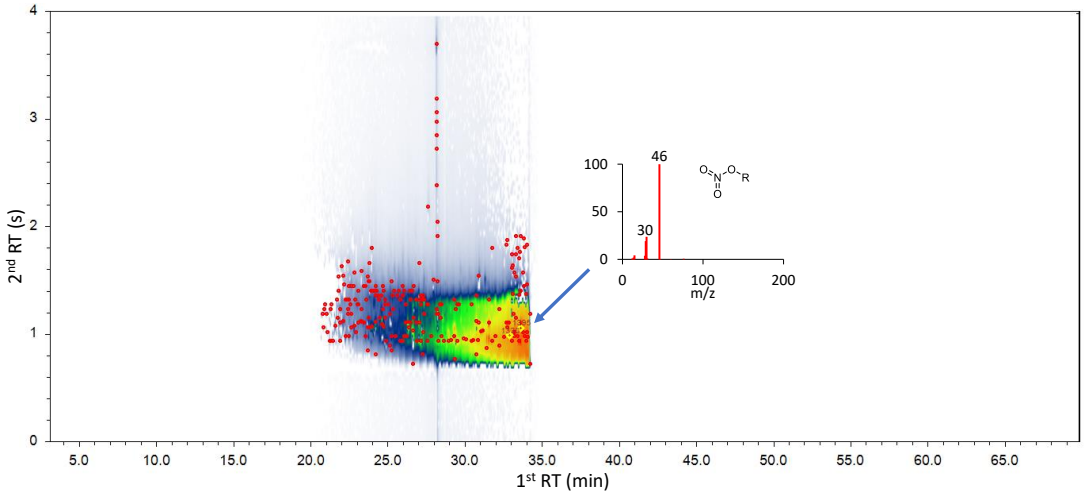


Figure S11. The chromatogram and mass spectra of tentatively identified nitro representatives.

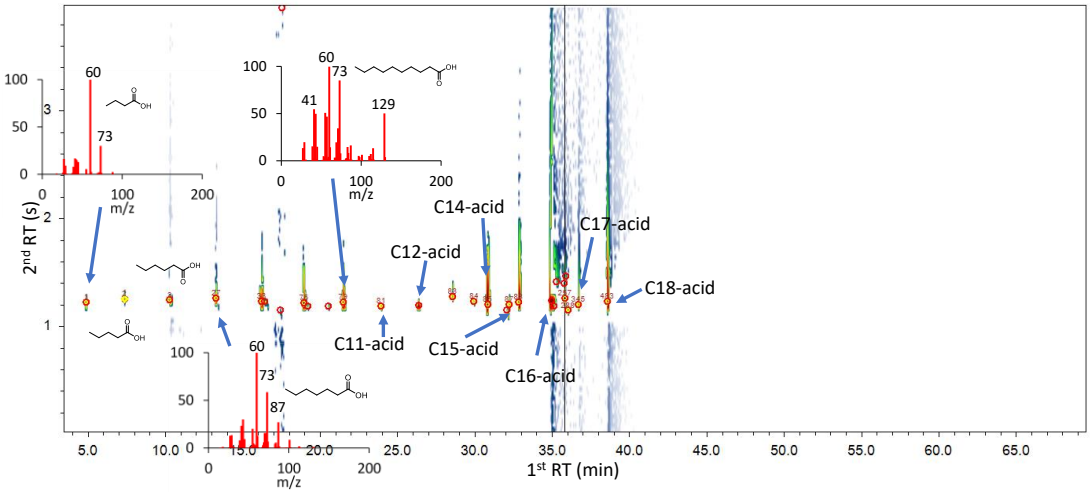


Figure S12. The chromatogram and mass spectra of tentatively identified aliphatic acid representatives.

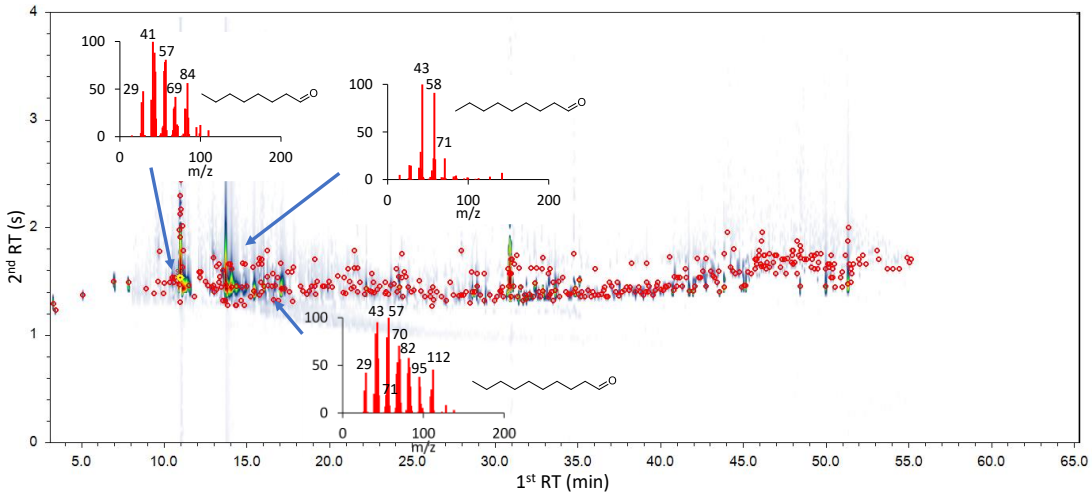


Figure S13. The chromatogram and mass spectra of tentatively identified aliphatic aldehyde and ketone representatives.

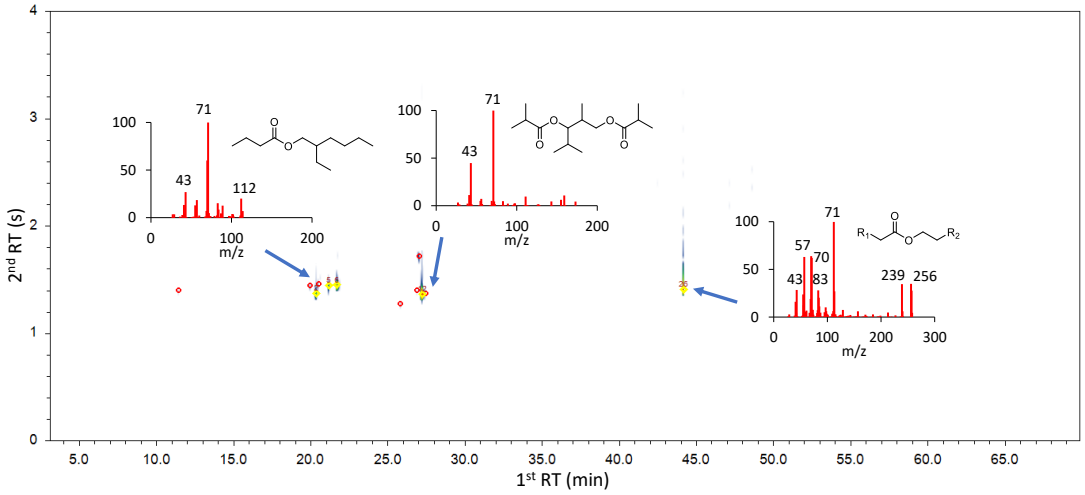


Figure S14. The chromatogram and mass spectra of tentatively identified aliphatic ester representatives.

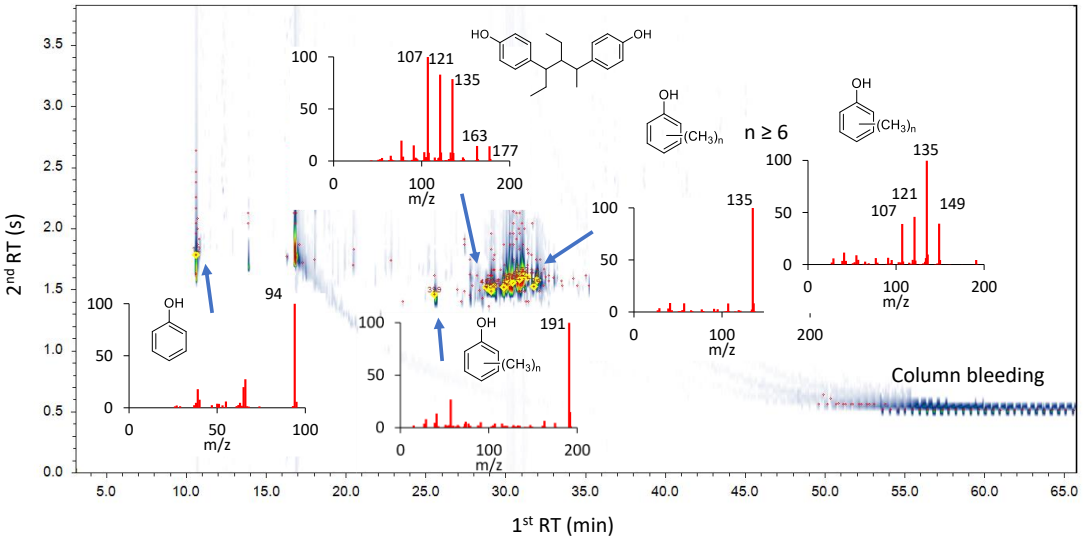


Figure S15. The chromatogram and mass spectra of tentatively identified phenol and alkyl-substituted phenol representatives in HDDV vapor samples.

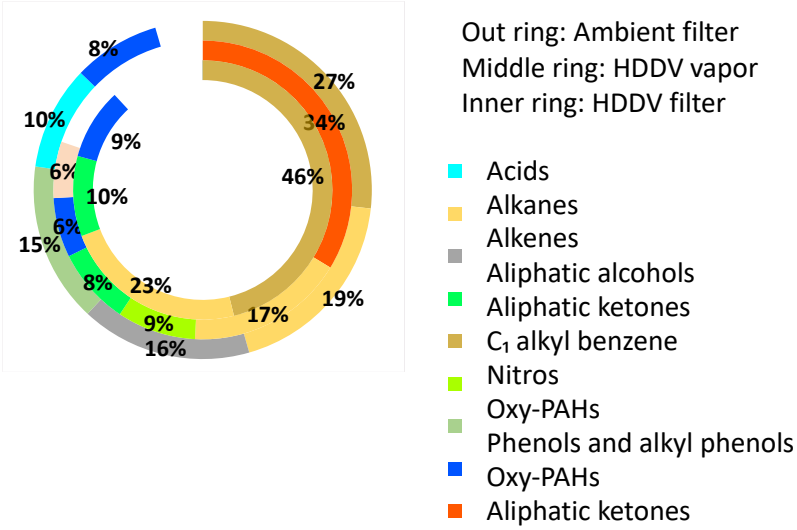


Figure S16. The top few species (contribution fraction exceeding 5%) distributions in different samples: the out ring refers to the ambient aerosol sample, the middle ring refers to the HDDV tailpipe vapor emission, and the inner ring refers to the HDDV tailpipe aerosol emission.

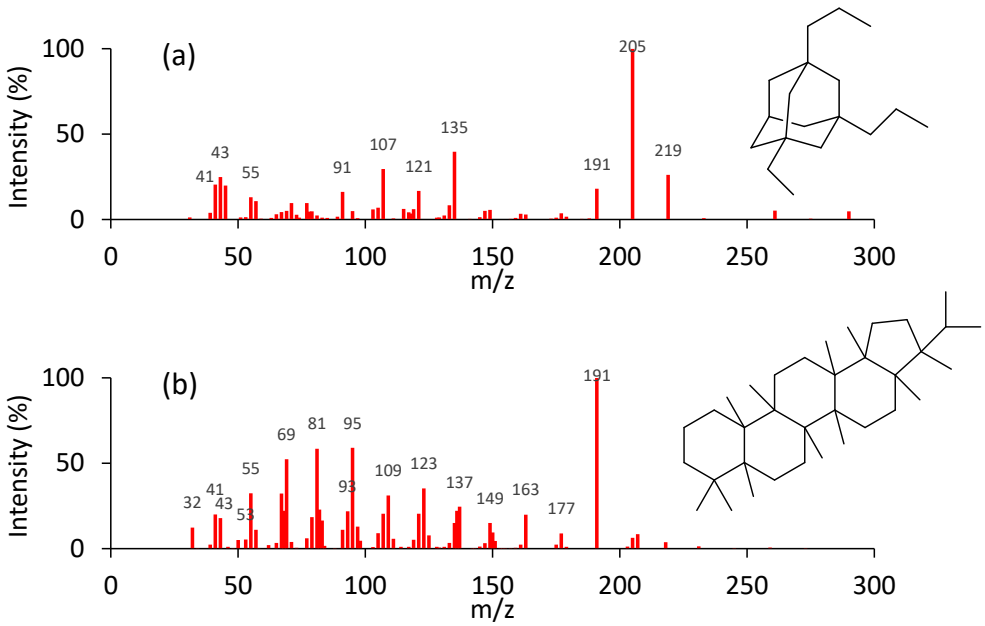


Figure S17. The chemical structures and mass spectra of (a) 1.1-Ethyl-3,5-dipropyladamantane adamantane and (b) 17α(H),21β(H)-hopane.

References:

Harrison, A. G. Pathways for water loss from doubly protonated peptides containing serine or threonine. *J Am Soc Mass Spectrom,* 23**,** 116-123, https://www.ncbi.nlm.nih.gov/pubmed/22065406, 2012.

He, X., X. H. H. Huang, K. S. Chow, Q. Wang, T. Zhang, D. Wu & J. Z. Yu Abundance and Sources of Phthalic Acids, Benzene-Tricarboxylic Acids, and Phenolic Acids in PM2.5 at Urban and Suburban Sites in Southern China. *ACS Earth and Space Chemistry,* 2**,** 147-158, 2018.

Lavanchy, A., R. Houriet & T. Gäumann The mass spectrometric fragmentation of n-alkanes. *Organic Mass Spectrometry,* 14**,** 79-85, https://doi.org/10.1002/oms.1210140205, 1979.

McLafferty, F. W. Mass Spectrometric Analysis. Molecular Rearrangements. *Anal Chem,* 31**,** 82-87, https://dx.doi.org/10.1021/ac60145a015, 1959.

Mikaia, A. Protocol for Structure Determination of Unknowns by EI Mass Spectrometry. I. Diagnostic Ions for Acyclic Compounds with up to One Functional Group. *Journal of Physical and Chemical Reference Data,* 51**,** 031501, 2022.