

Supplement to "*Contrasting organic aerosol molecular composition between the urban and agricultural environment of the Po Valley*"

Luca D'Angelo et al.

Correspondence: Luca D'Angelo (l.dangelo@arpalombardia.it), Alexander L. Vogel (vogel@iau.uni-frankfurt.de)

Contents

	1	Calculations	1
	1.1	Normalized-Signal Intensity (nSI)	1
	1.2	Residuals contribution calculation	2
5	2	Supplement figures	3
	2.1	Brown carbon-related compounds	3
	2.2	Time serie of confirmed compounds	6
	2.3	Clustergrams	7
	2.4	Clusters molecular fingerprints	9
10	2.4.1	Clusters molecular fingerprints - MI site	9
	2.4.2	Clusters molecular fingerprints - SKI site	13
	2.5	Site-specific compound fingerprints	16
	2.6	Fingerprint of potential BrC features	18
	2.7	Attribution of potential BrC features at both sites	20
15	2.8	Aerosol Ångstrom Exponent for pair samples	21
	3	Compound Discoverer workflows	22

1 Calculations

1.1 Normalized-Signal Intensity (nSI)

In order to take into account the ionization efficiency and ion transmission variability between positive and negative ionization mode and to compare the time serie, we choose to scale the signal intensity of each individual compound detected in each sample by the maximum signal intensity detected over the whole time serie taking into account the ionization mode.

Let us call $SI_{c,pol,d}$ the signal intensity of compound c detected in polarization mode pol on a sample collected on day d and let us call $MAX(SI_{pol})$ the maximum SI within the entire dataset detected in pol . Then:

$$nSI_{c,d} = \frac{SI_{c,pol,d}}{MAX(SI_{pol})} \quad (1)$$

25

As a consequence, the normalized-Total Signal Intensity ($nTSI$) of the sample collected on day d is:

$$nTSI_d = \sum_{c=1}^C nSI \quad (2)$$

1.2 Residuals contribution calculation

Let us indicate $nSI_{MI,c,d}$ the normalized signal intensity (nSI) for compound c detected by the HPLC-HESI-HRMS system on the $PM_{2.5}$ sample collected at Milan (MI) on day d . Similarly, $nSI_{SKI,c,d}$ is the nSI for compound c detected on the sample collected on the same day d at Schivenoglia (SKI).

Then:

$$Residuals_{c,d} = nSI_{MI,c,d} - nSI_{SKI,c,d} \quad (3)$$

35 Positive *Residuals* means that $nSI_{c,d}$ is higher at MI, whereas negative *Residuals* indicates higher $nSI_{c,d}$ at SKI.

To evaluate the relative contribution amount of compound c detected with higher intensity at the site where it is higher (*site*) to the nTSI on day d , we calculate:

$$Contribution = \frac{Residuals_{c,d}}{nTSI_{site,d}} \times 100 \quad (4)$$

2 Supplement figures

40 2.1 Brown carbon-related compounds

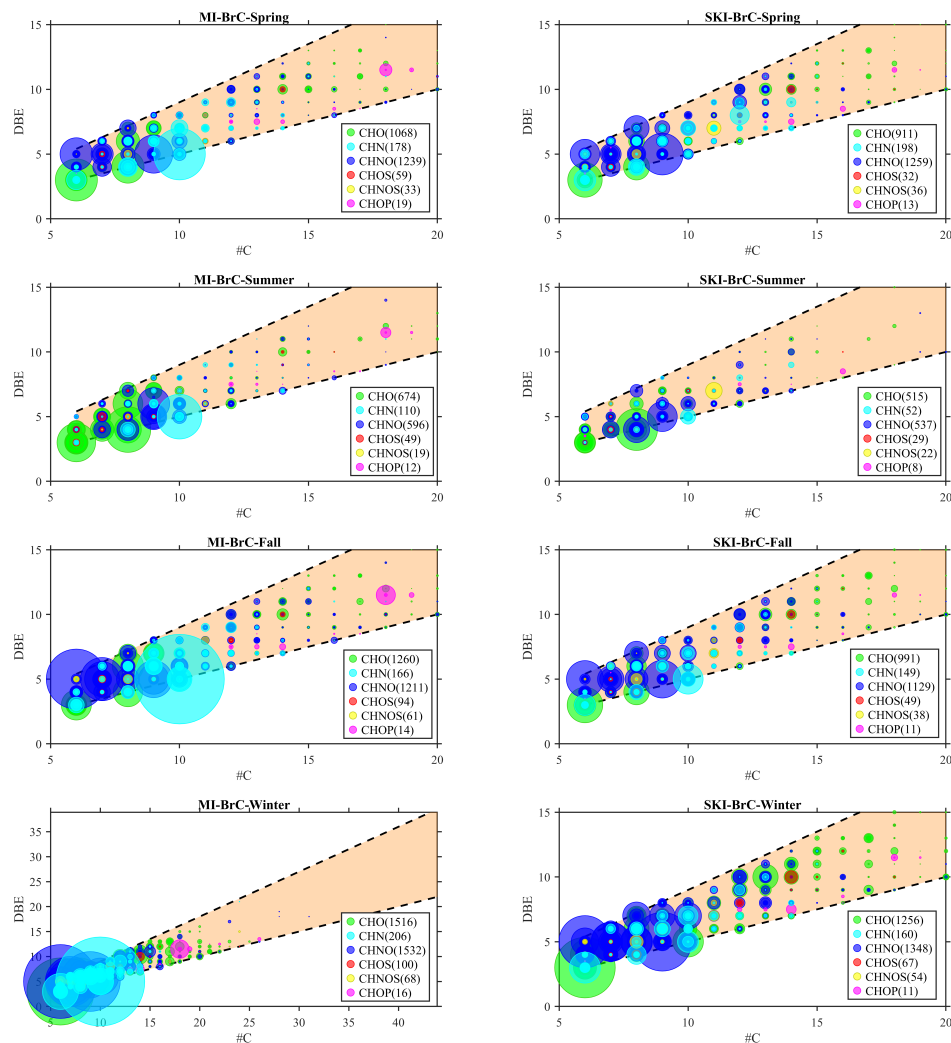


Figure S1. Seasonal fingerprints of BrC-related compounds for MI (left) and SKI (right) sites. The plots highlight the molecular variability between the two environments over the seasons.

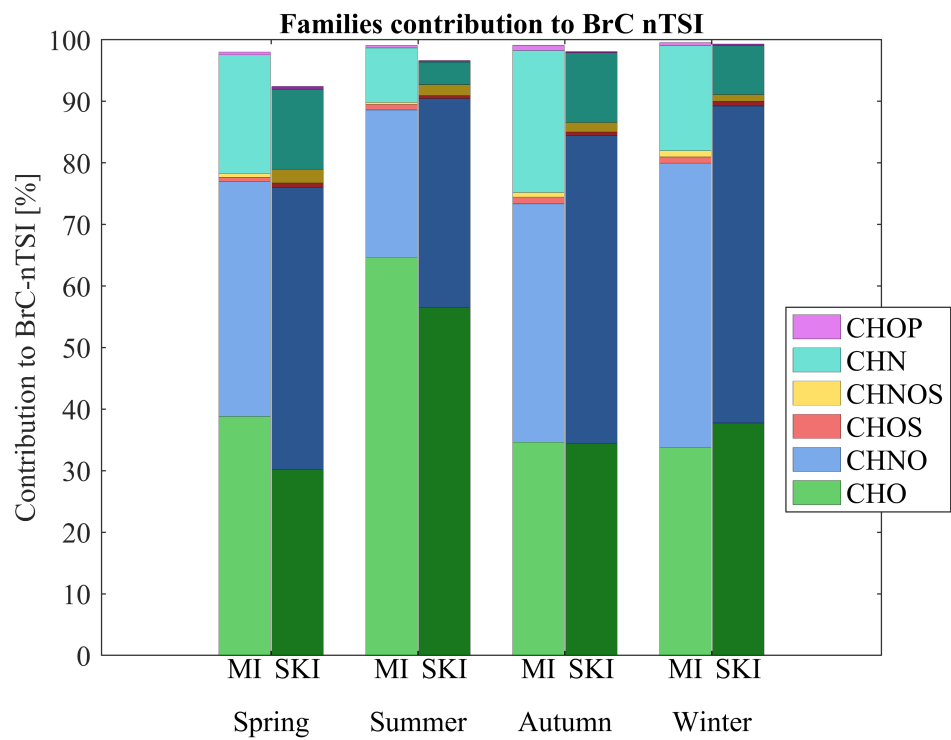


Figure S2. Contribution of the compound families related to BrC nTSl. The residual contribution (not shown) is due to compounds belonging to the "Other" group.

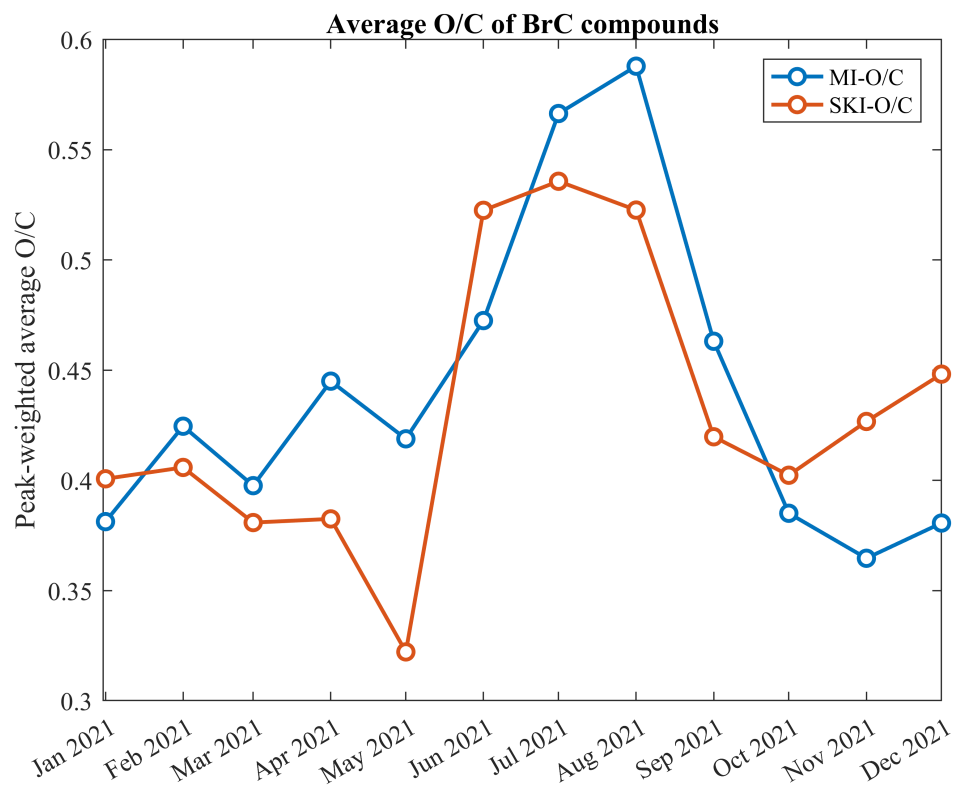


Figure S3. Peak-weighted average of O/C for BrC-related compounds. The monthly average suggests a stronger oxidation regime during summer.

2.2 Time serie of confirmed compounds

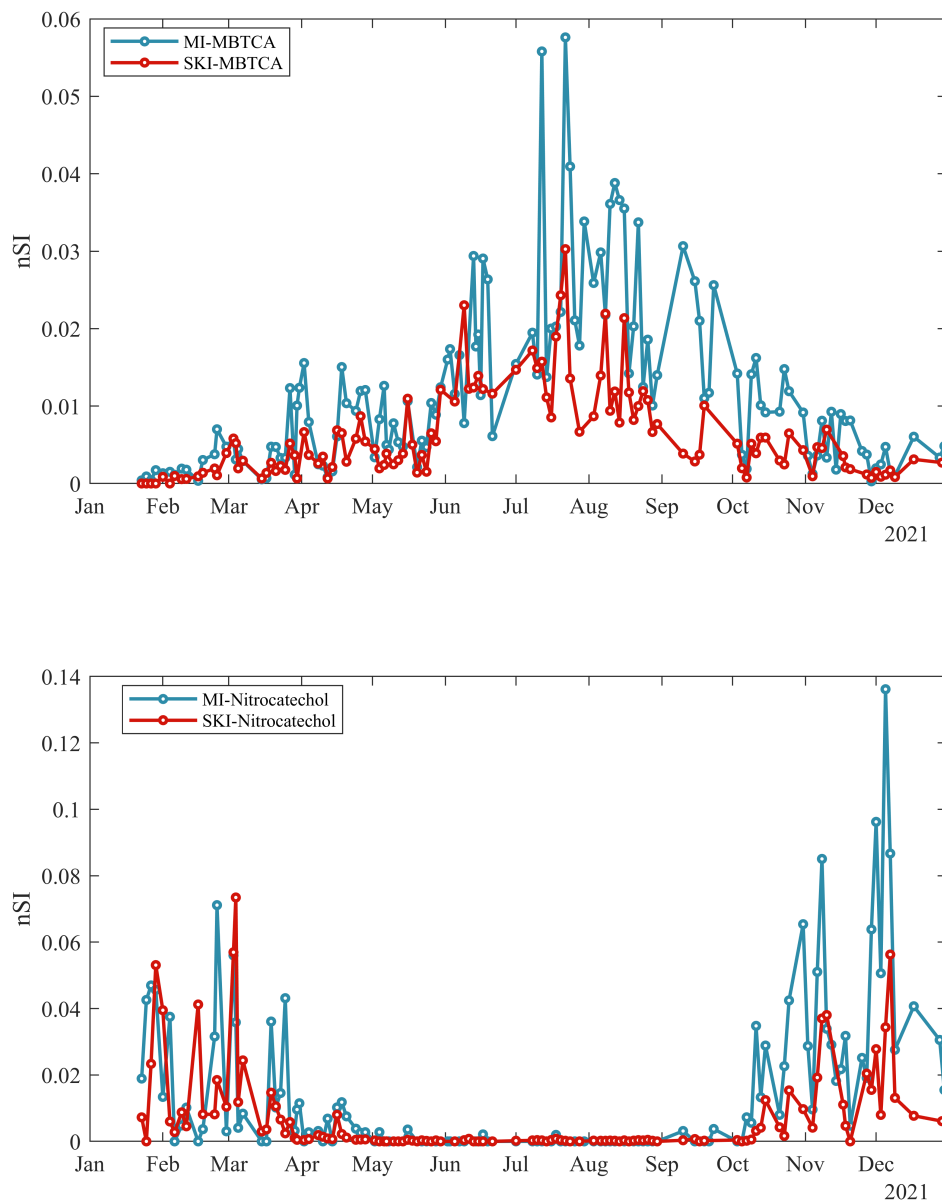


Figure S4. Time series of L1 identification confidence compounds for biogenic (MBTCA, upper plot) and anthropogenic (nitrocatechol, lower plot) SOA at MI and SKI sites.

2.3 Clustergrams

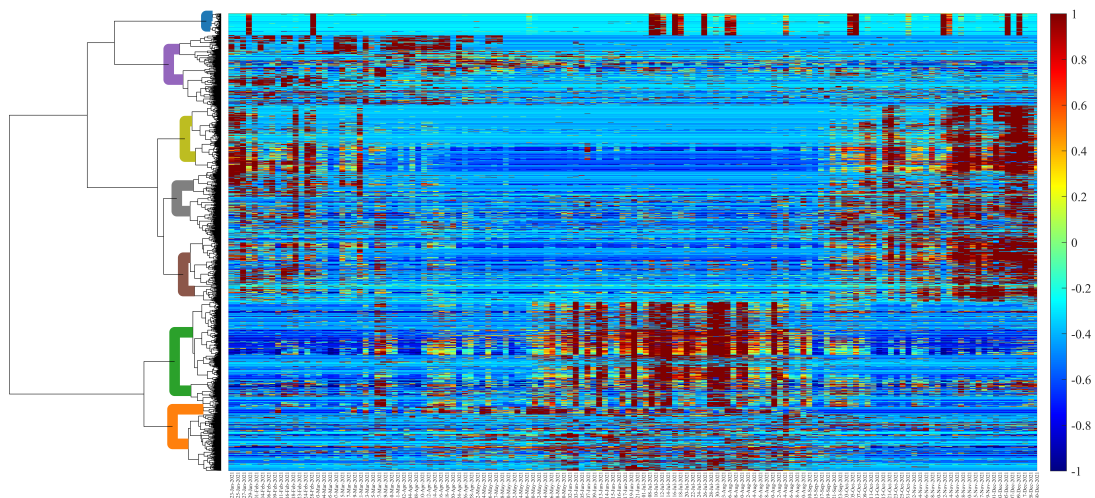


Figure S5. Graphic result of the hierarchical clustering performed on MI dataset. Warmer colors indicate higher values after standardization over the time series of each individual compound. The assigned clusters are marked with the same colors showed in Figure 3 in the main text.

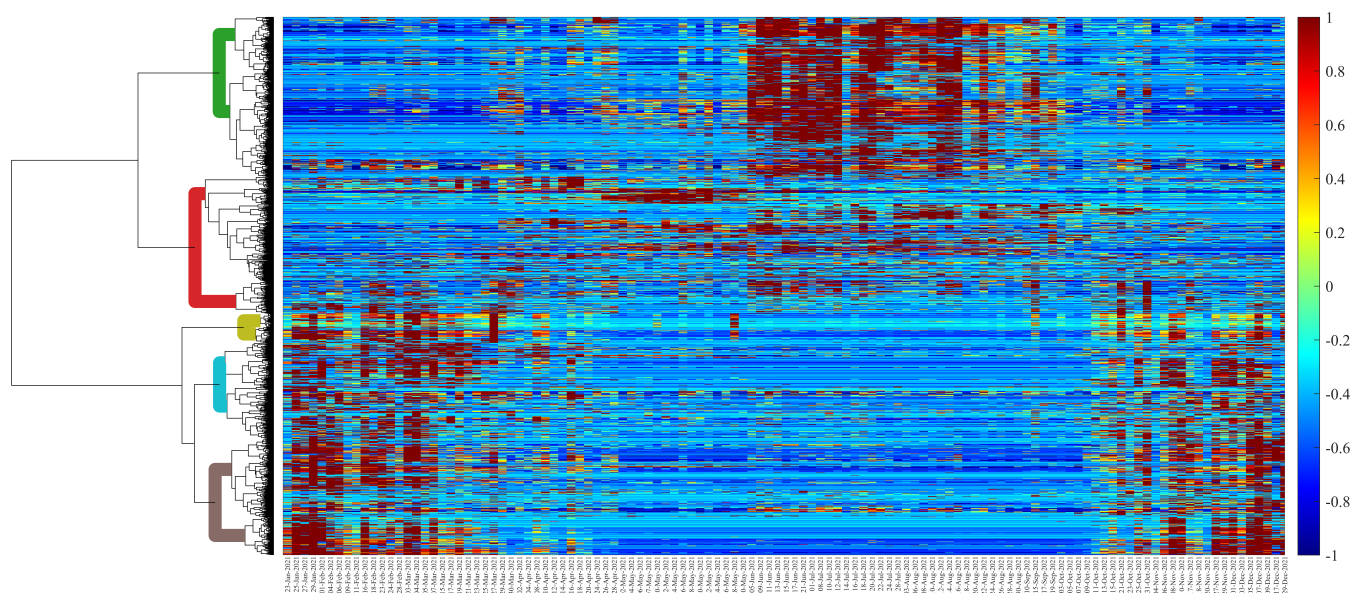


Figure S6. Graphic result of the hierarchical clustering performed on SKI dataset. Warmer colors indicate higher values after standardization over the time series of each individual compound. The assigned clusters are marked with the same colors showed in Figure 3 in the main text.

2.4 Clusters molecular fingerprints

In this section we show the molecular fingerprint of each cluster. We chose to graphically described the clusters' characteristics based on their retention time (a proxy for the polarity of the molecules) against their calculated molecular weight (left plots), their H/C ratio against O/C ratio with the Van Krevelen diagram (center) and their average OS_C against the number of carbon in the Kroll plot (right). Within these plots, each feature is colored based on the molecular class and the size of the bullet is proportional to its nSI considering as null value all SI below the threshold after blank correction.

2.4.1 Clusters molecular fingerprints - MI site

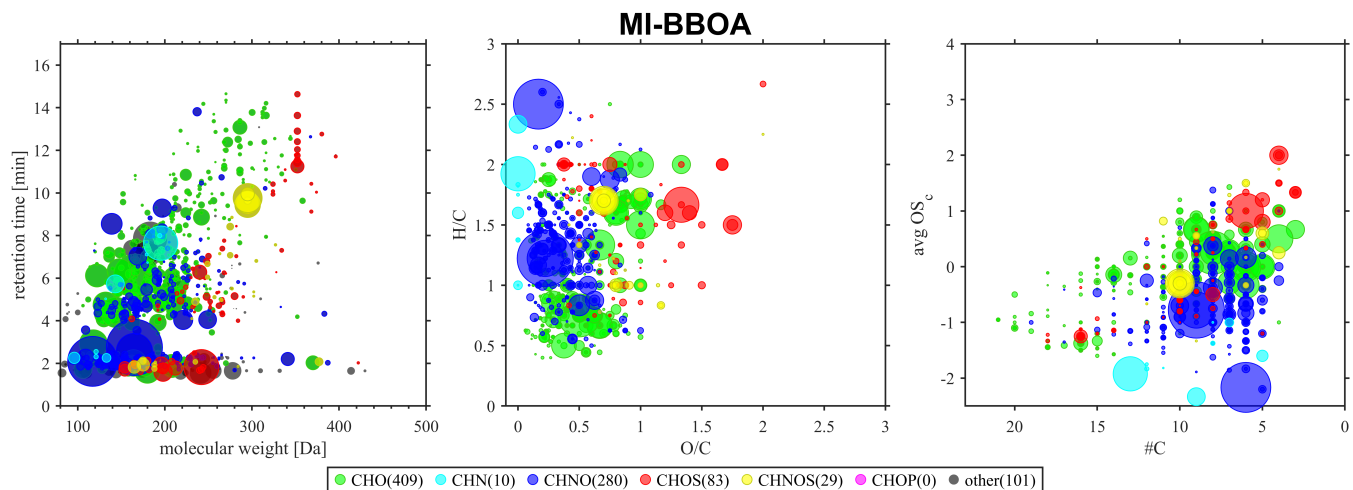


Figure S7. Molecular fingerprint of biomass burning organic aerosols at MI (MI-BBOA).

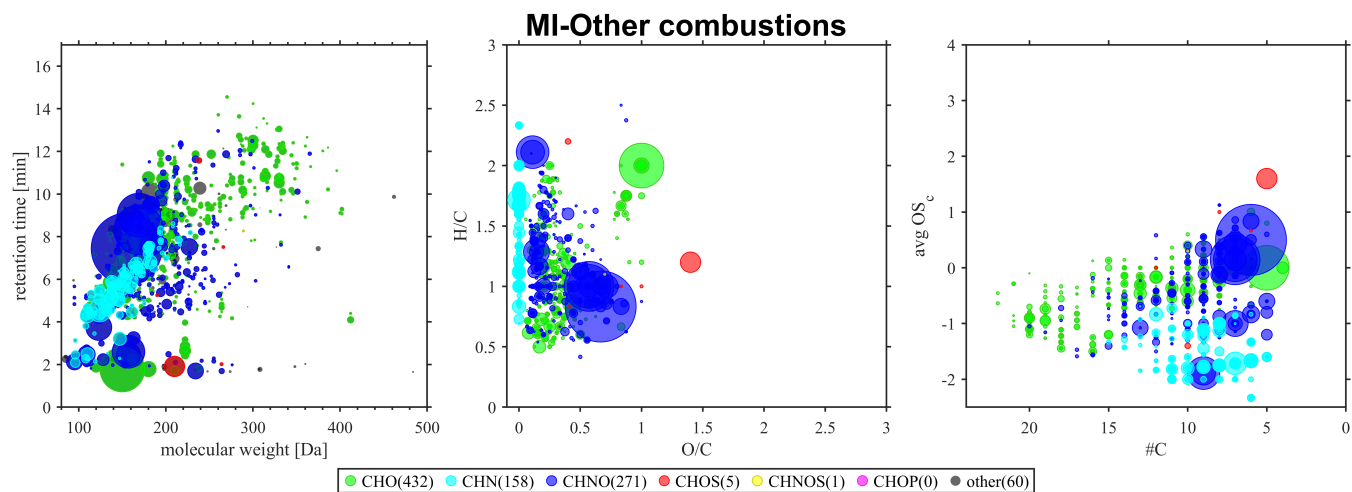


Figure S8. Molecular fingerprint of other combustion related features at MI (MI-Other combustion).

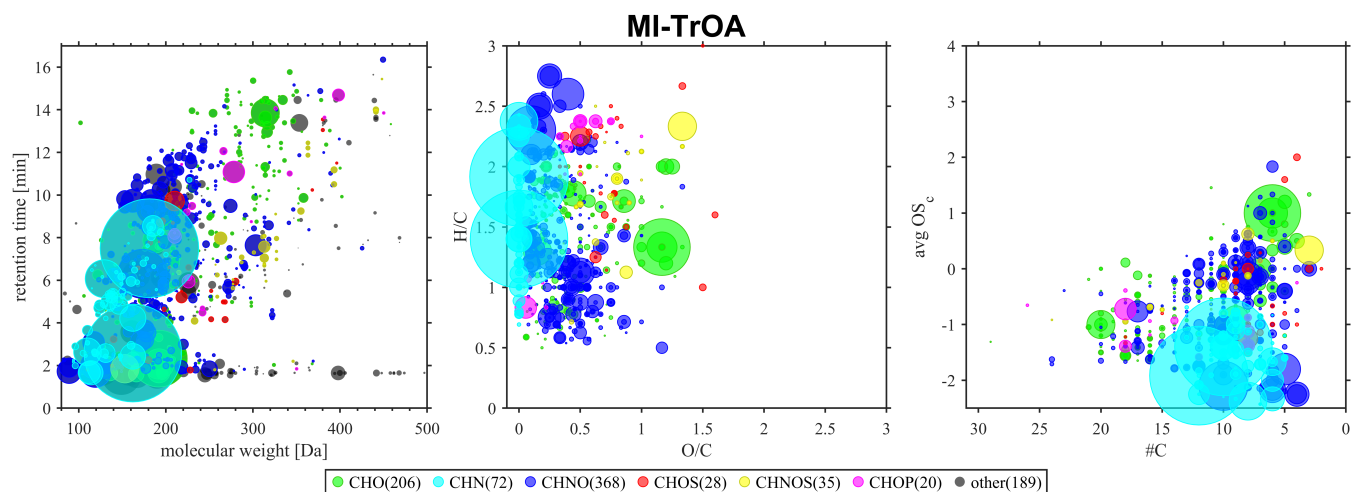


Figure S9. Molecular fingerprint of traffic organic aerosols at MI (MI-TrOA).

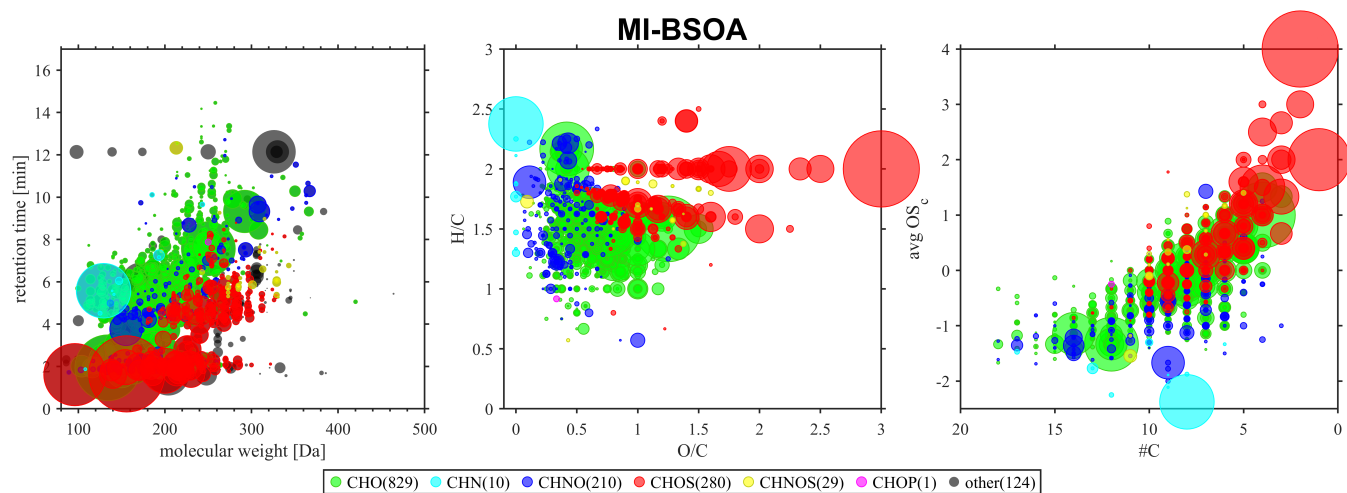


Figure S10. Molecular fingerprint of biogenic organic aerosols at MI (MI-BSOA).

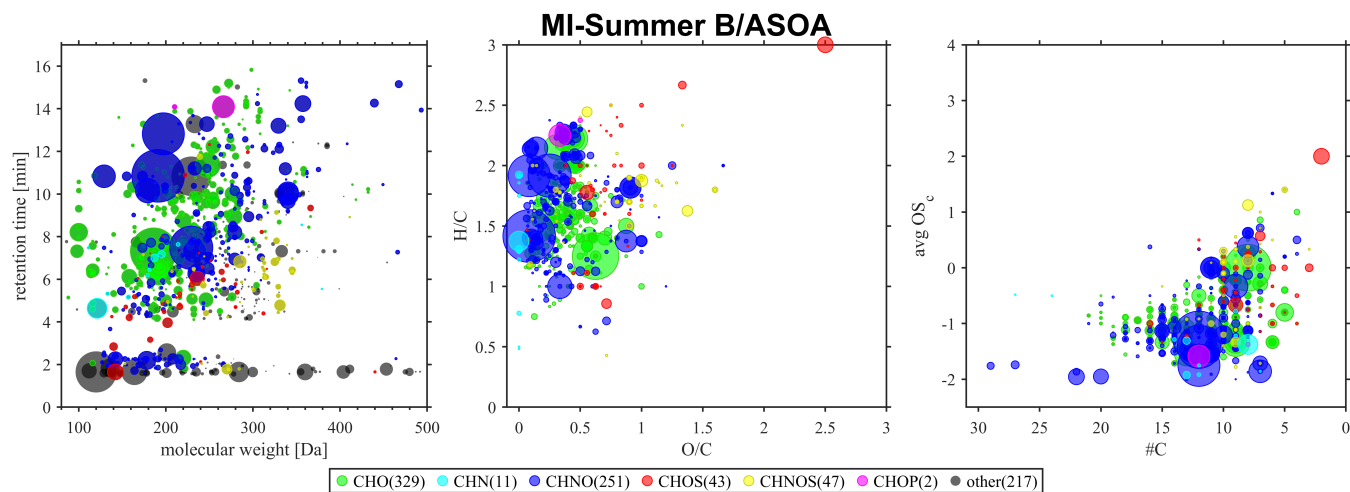


Figure S11. Molecular fingerprint of summertime mixed biogenic and anthropogenic aerosols at MI (MI-Summer B/ASOA).

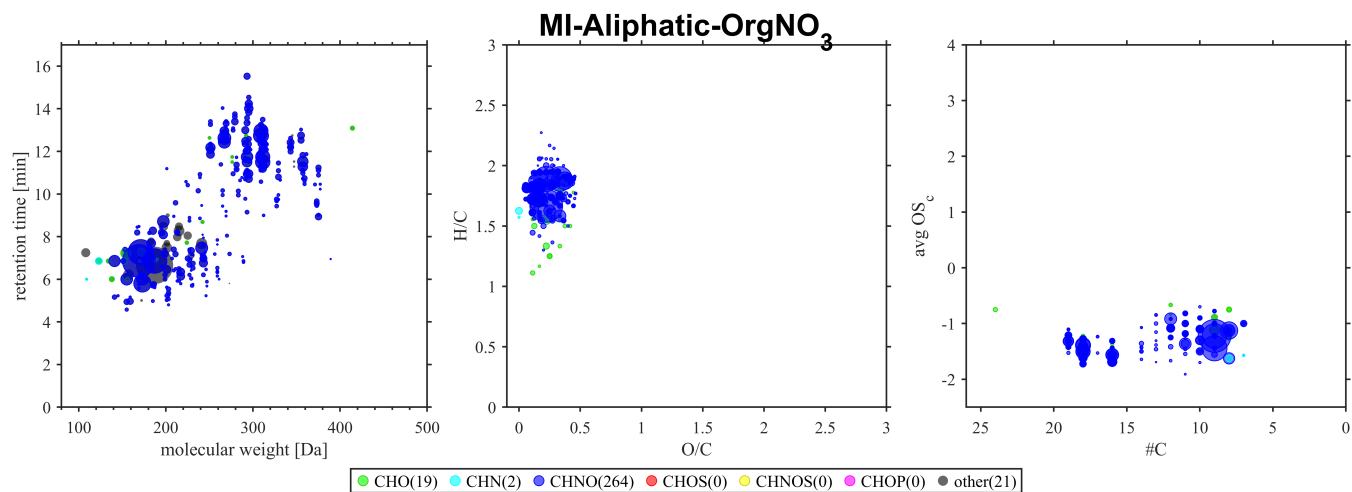


Figure S12. Molecular fingerprint of aliphatic organo-nitrate aerosols at MI (MI-OrgNO₃).

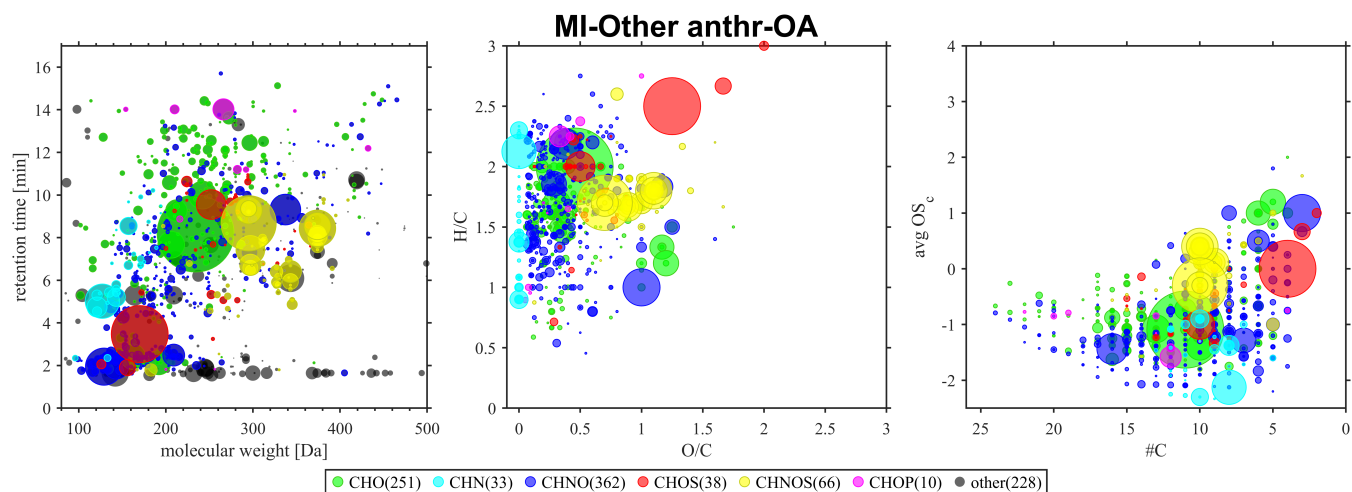


Figure S13. Molecular fingerprint of other anthropogenic organic aerosols at MI (MI-Other anthr-OA).

2.4.2 Clusters molecular fingerprints - SKI site

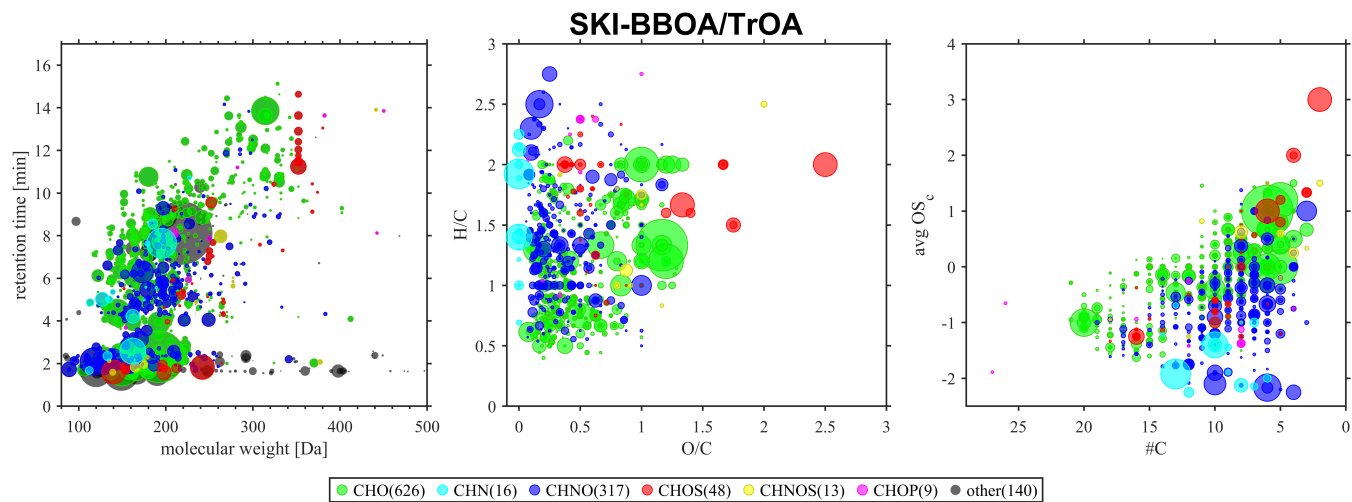


Figure S14. Molecular fingerprint of mixed biomass burning and traffic related organic aerosols at SKI (SKI-BBOA/TrOA).

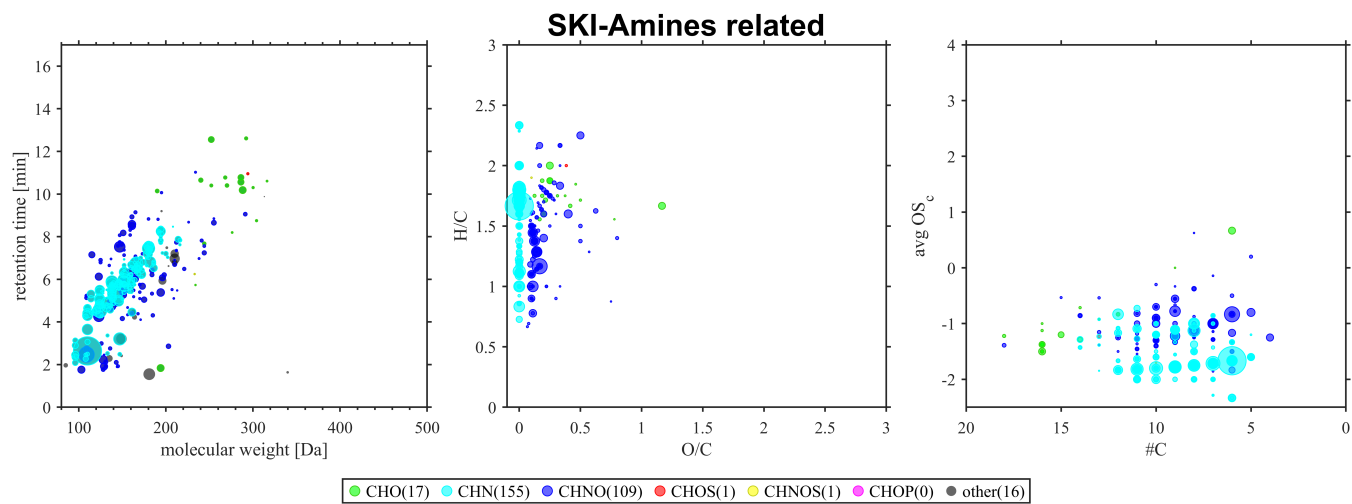


Figure S15. Molecular fingerprint of amines related organic aerosols at SKI (SKI-Amines related).

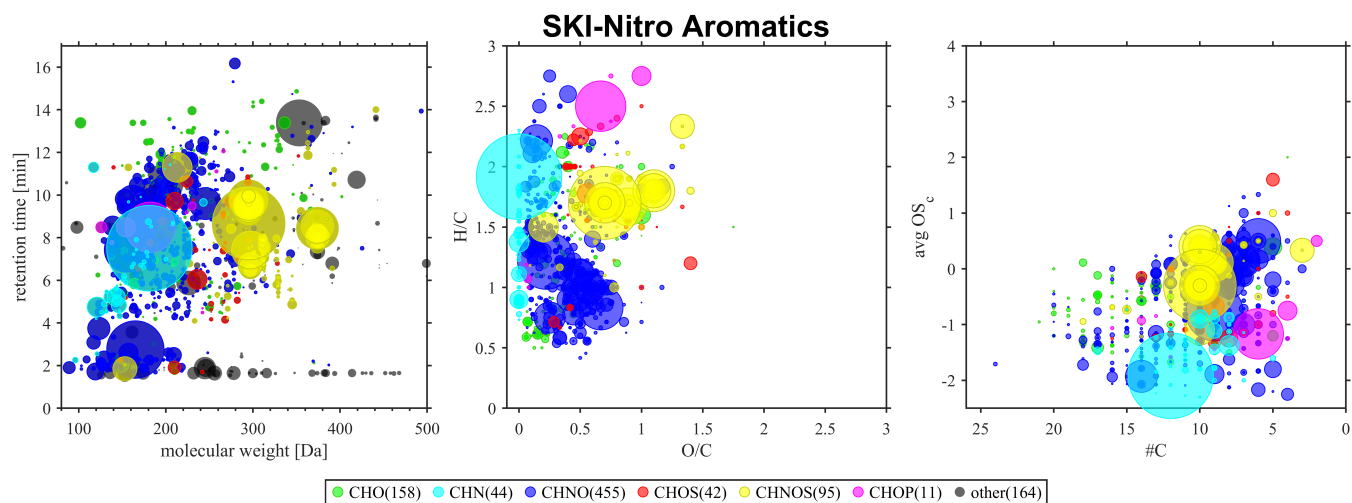


Figure S16. Molecular fingerprint of nitro-aromatic related organic aerosols at SKI (SKI-Nitro Aromatics).

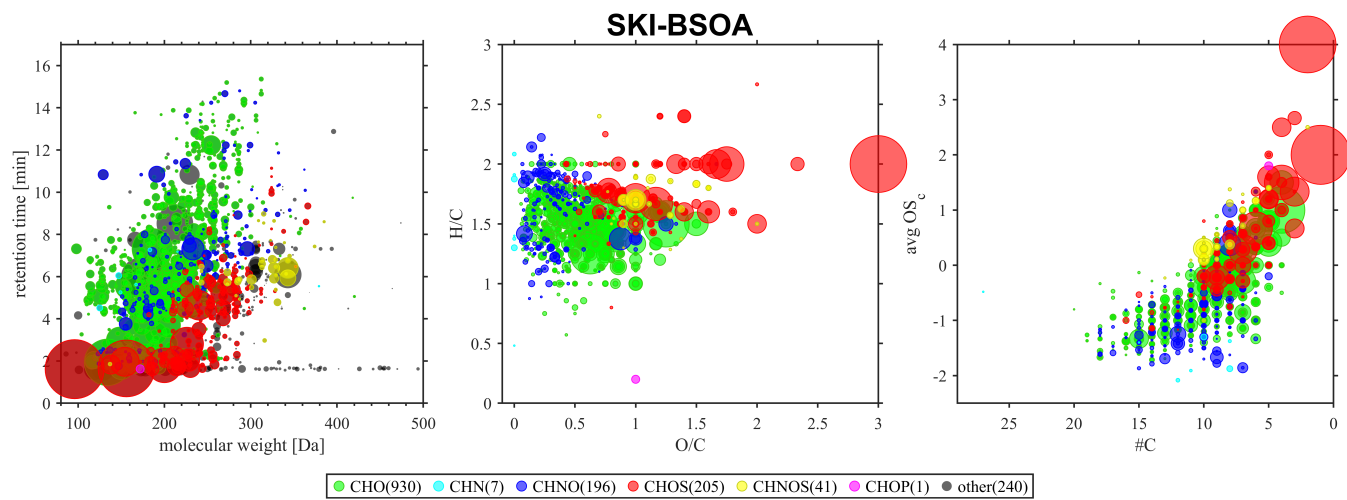


Figure S17. Molecular fingerprint of biogenic organic aerosols at SKI (SKI-BSOA).

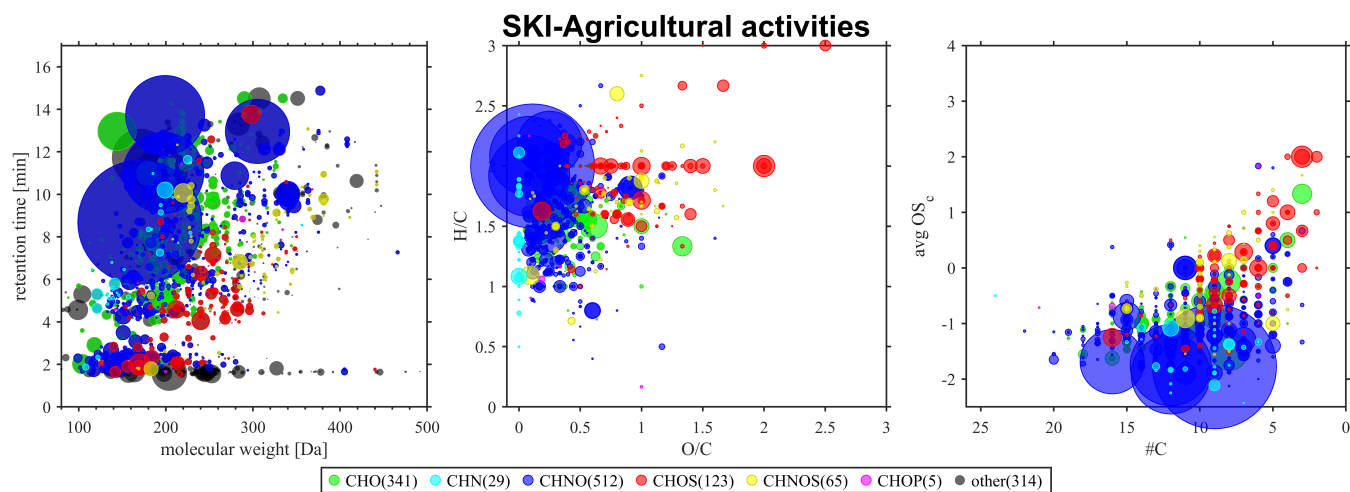


Figure S18. Molecular fingerprint of features linked to agricultural activities at SKI (SKI-Agricultural activities).

2.5 Site-specific compound fingerprints

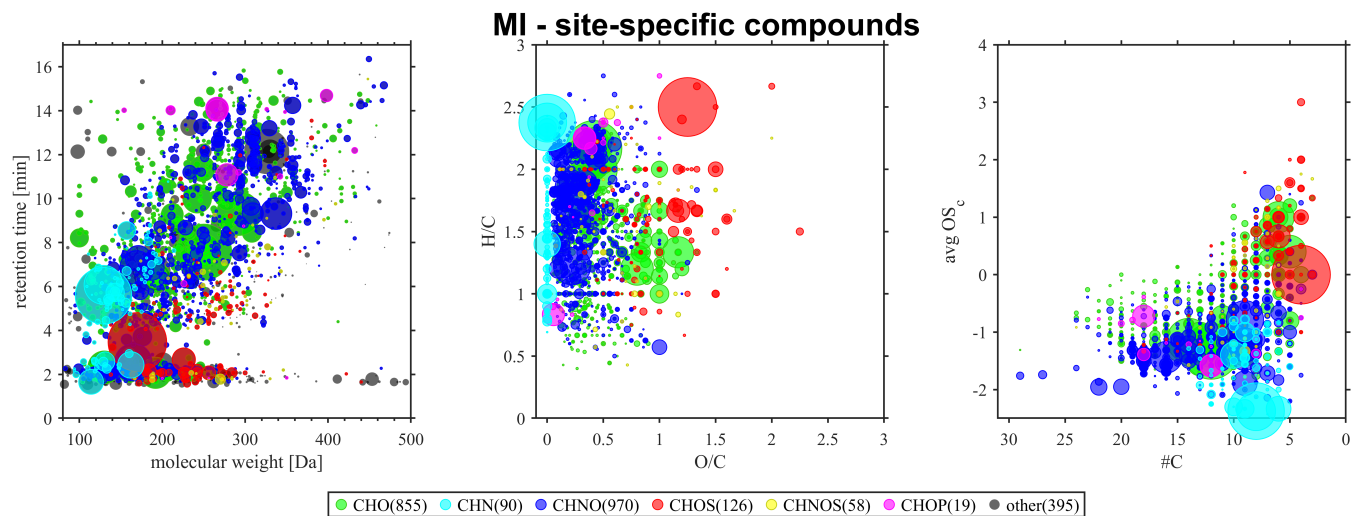


Figure S19. Molecular fingerprint of site-specific compounds detected at MI site and classified with the HCA.

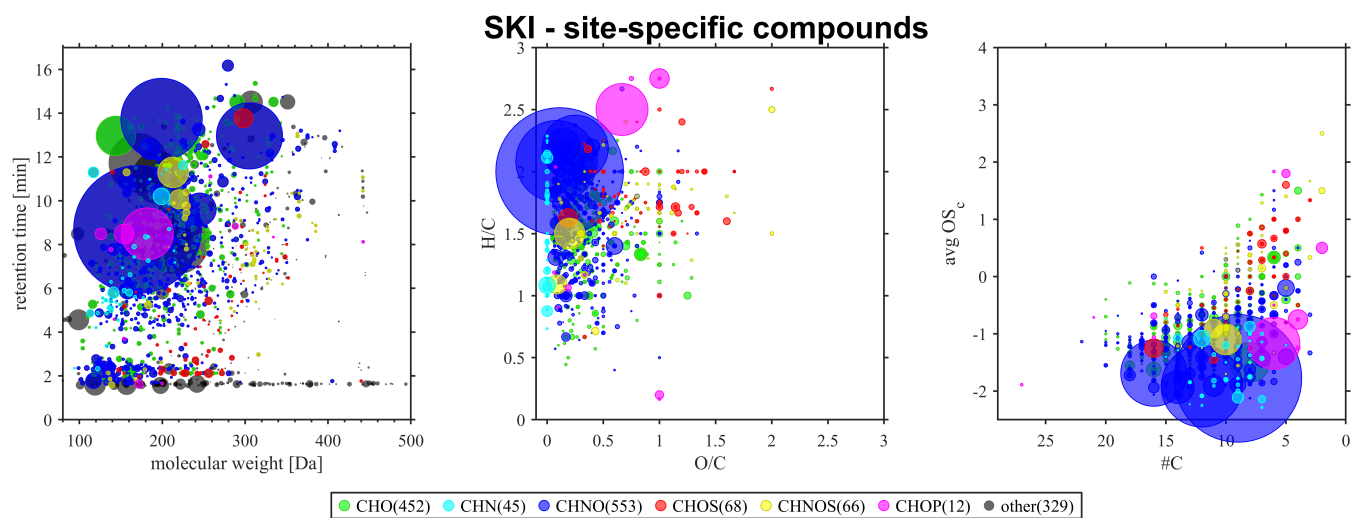


Figure S20. Molecular fingerprint of site-specific compounds detected at SKI site and classified with the HCA.

2.6 Fingerprint of potential BrC features

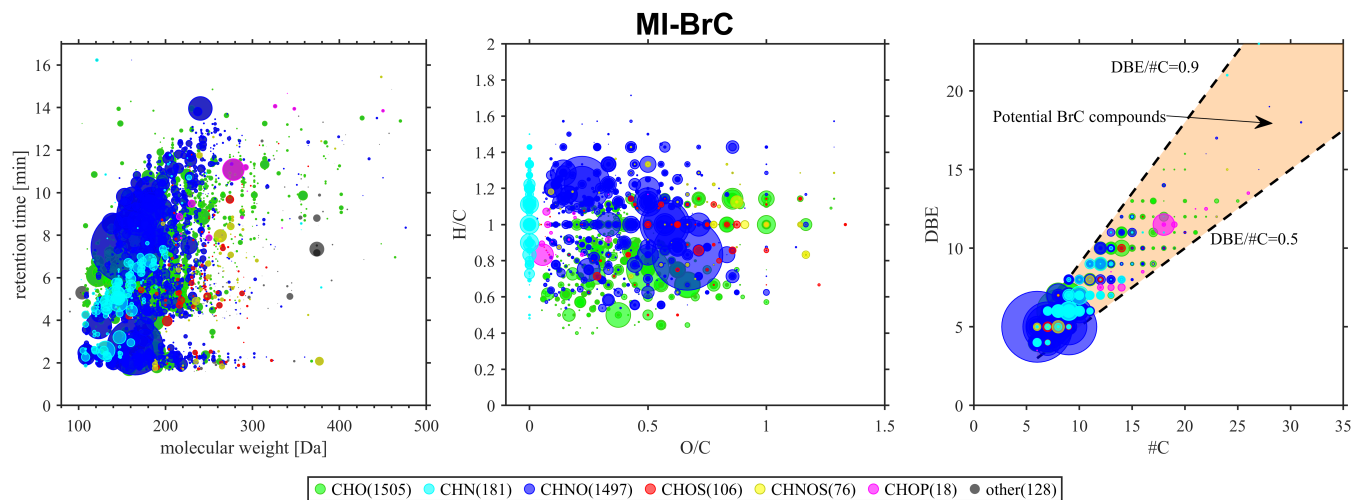


Figure S21. Molecular fingerprints of potential BrC-related compounds detected and classified at MI site.

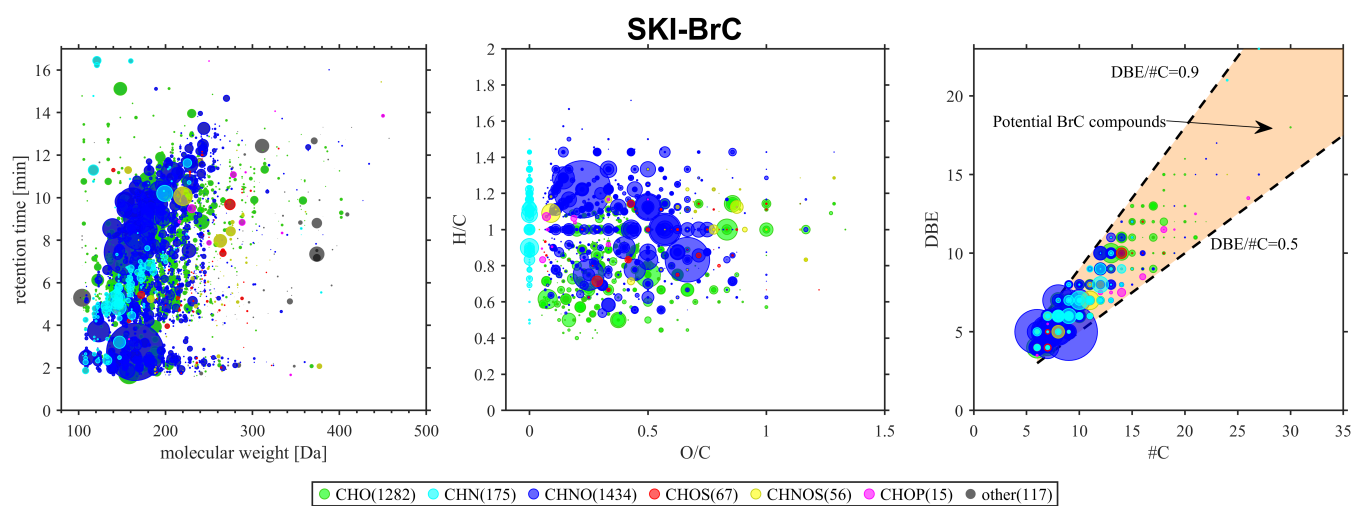


Figure S22. Molecular fingerprints of potential BrC-related compounds detected and classified at SKI site.

2.7 Attribution of potential BrC features at both sites

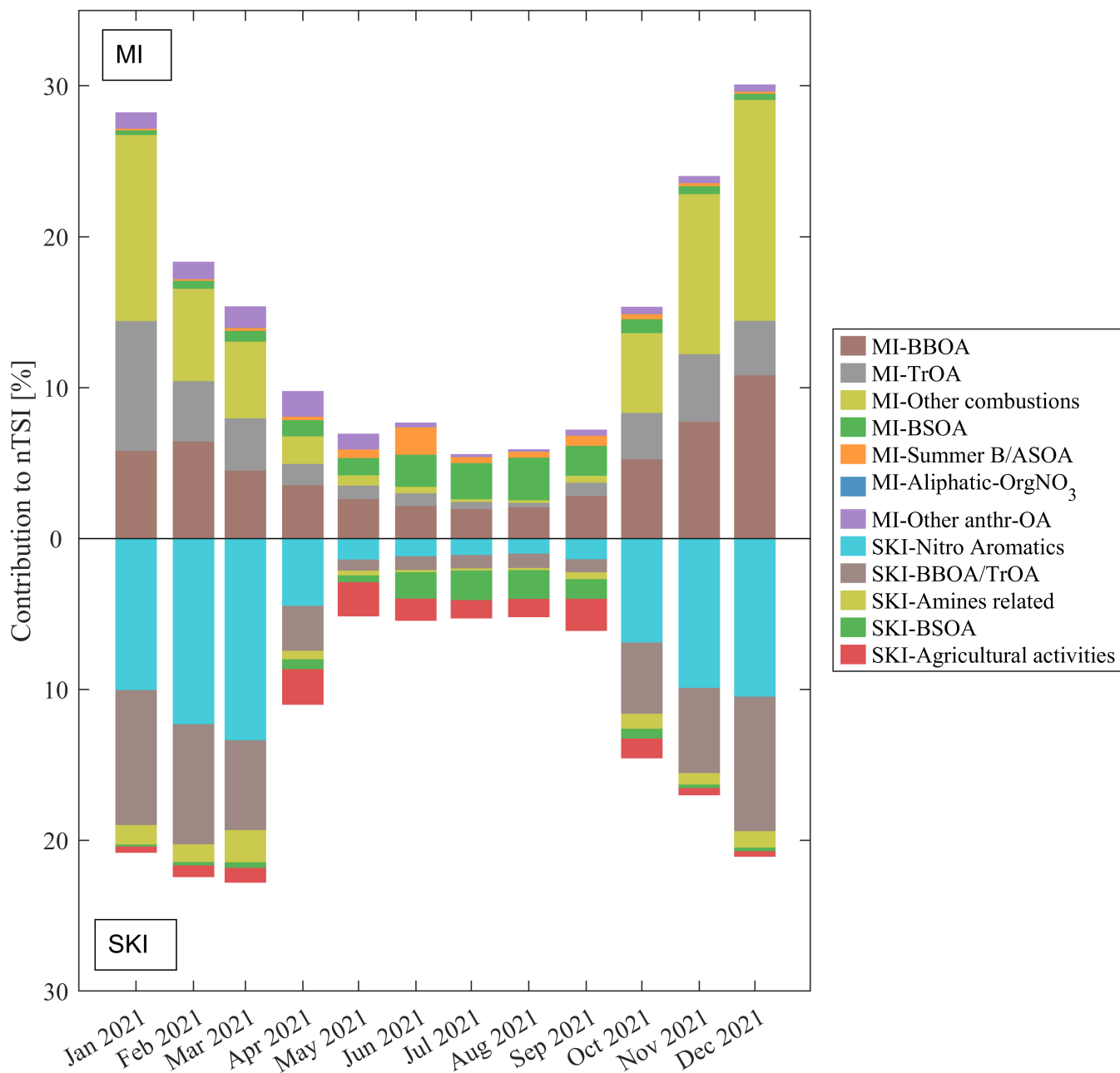


Figure S23. Contribution to nTSI due to features identified as potential BrC-related compounds.

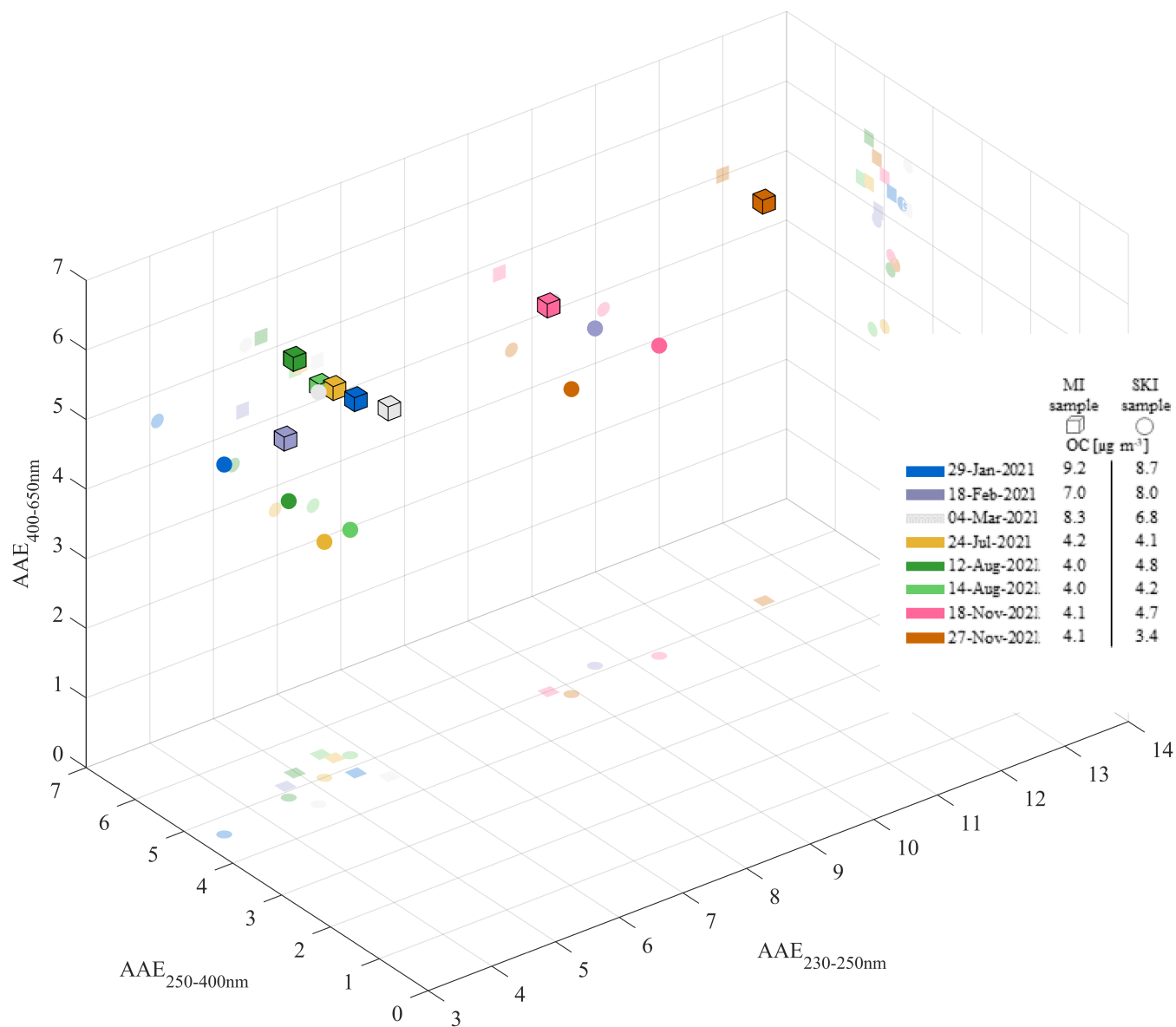


Figure S24. AAE for pairs of samples with similar OC concentrations and contribution to PM_{2.5}. OC concentrations are shown in the legend. Same colors indicates samples collected on the same days whereas the shape refers to the collection site: MI (cubes) or SKI (spheres). The marker projections make it easier to appreciate the distances between the points across the three planes, and they are oriented toward the reader.

3 Compound Discoverer workflows

Table S1. Workflow of the non-target software Compound Discoverer 3.3.2.31 - negative ionization mode data

Processing node	Parameter	Settings
Select Spectra	1. Spectrum Properties Filter:	Lower RT Limit: 0 Upper RT Limit: 16.5 First Scan: 0 Last Scan: 0 Ignore Specified Scans: (not specified) Lowest Charge State: 0 Highest Charge State: 0 Min. Precursor Mass: 50 Da Max. Precursor Mass: 5000 Da Total Intensity Threshold: 10000 Minimum Peak Count: 1
	2. Scan Event Filters:	Mass Analyzer: Is FTMS MS Order: Is MS2, MS1 Activation Type: Is HCD Min. Collision Energy: 0 Max. Collision Energy: 1000 Scan Type: Any Polarity Mode: Is -
	3. Peak Filters:	S/N Threshold (FT-only): 5
	4. Replacements for Unrecognized Properties:	Unrecognized Charge Replacements: 1 Unrecognized Mass Analyzer Replacements: ITMS Unrecognized MS Order Replacements: MS2 Unrecognized Activation Type Replacements: HCD Unrecognized Polarity Replacements: -

continued on next page

Processing node	Parameter	Settings
Align Retention Times	5. General Settings:	Unrecognized MS Resolution@200 Replacements: 60000 Unrecognized MSn Resolution@200 Replacements: 30000
		Precursor Selection: Use MS1 Precursor Use Isotope Pattern in Precursor Reevaluation: True Provide Profile Spectra: Automatic Store Chromatograms: False
Detect Compounds	1. General Settings:	Alignment Model: Adaptive curve Alignment Fallback: Use Linear Model Maximum Shift [min]: 0.2 Shift Reference File: True Mass Tolerance: 5 ppm Remove Outlier: True
		Mass Tolerance [ppm]: 2 ppm Min. Peak Intensity: 10000 Min. [#] Scans per Peak: 5 Use Most Intense Isotope Only: False
	2. Trace Detection:	Max. Number of Gaps to Correct: 2 Min. Number of Adjacent Non-Zeros: 2
	3. Peak Detection:	Chromatographic S/N Threshold: 1.5 Remove Baseline: False Gap Ratio Threshold: 0.35 Max. Peak Width [min]: 2 Min. Relative Valley Depth: 0.2
	4. Isotope Pattern Detection:	
continued on next page		

Processing node	Parameter	Settings
Group Compounds		Group Isotopes for: Br, Cl
		Use Peak Quality for Isotope Grouping: True
		Filter out Features with Bad Peaks Only: True
		Zig-Zag Index Threshold: 0.2
		Jaggedness Threshold: 0.4
		Modality Threshold: 0.9
		Remove Potentially False Positive Isotopes: True
	5. Compound Detection:	Ions: [2M-H]-1; [M-H]-1
		Base Ions: [M-H]-1
		Remove Singlets: True
	6. AcquireX Settings:	Detect Persistent Background Ions: False
	1. General Settings:	Mass Tolerance: 2 ppm
		RT Tolerance [min]: 0.2
		Align Peaks: False
		Preferred Ions: [M-H]-1
		Area Integration: All Ions
	2. Peak Rating Contributions:	Area Contribution: 3
		CV Contribution: 1
		FWHM to Base Contribution: 5
		Jaggedness Contribution: 5
Fill Gaps		Modality Contribution: 5
		Zig-Zag Index Contribution: 5
	3. Peak Rating Filter:	Peak Rating Threshold: 4
		Number of Files: 4
	1. General Settings:	

continued on next page

Processing node	Parameter	Settings
Mark Background Compounds		Mass Tolerance: 2 ppm S/N Threshold: 1.5 Use Real Peak Detection: True Apply Restrictive Gap Filling: True
	1. General Settings:	Max. Sample/Blank: 5 Max. Blank/Sample: 0 Hide Background: False
Assign Compound Annotations		
	1. General Settings:	
60		Mass Tolerance: 2 ppm
	2. Data Sources:	
		Data Source #1: Predicted Compositions
		Data Source #2: mzVault Search
		Data Source #3: mzCloud Search
		Data Source #4: ChemSpider Search
	3. Scoring Rules:	
		Use mzLogic: True
		Use Spectral Distance: True
		SFit Threshold: 20
Search mzCloud		SFit Range: 20
	4. Reprocessing:	
		Clear Names: False
	1. General Settings:	
		Compound Classes: All
		Precursor Mass Tolerance: 5 ppm
		FT Fragment Mass Tolerance: 5 ppm
		IT Fragment Mass Tolerance: 0.4 Da
		Library: Autoprocessed; Reference

continued on next page

Processing node	Parameter	Settings
		Post Processing: Recalibrated
		Max. # Results: 10
		Annotate Matching Fragments: False
		Search MSn Tree: False
	2. DDA Search:	
		Identity Search: HighChem HighRes
		Match Activation Type: True
		Match Activation Energy: Match with Tolerance
		Activation Energy Tolerance: 20
		Apply Intensity Threshold: True
		Similarity Search: Similarity Reverse
		Match Factor Threshold: 60
	3. DIA Search:	
		Use DIA Scans for Search: False
		Max. Isolation Width [Da]: 500
		Match Activation Type: False
		Match Activation Energy: Any
		Activation Energy Tolerance: 100
		Apply Intensity Threshold: False
	Match Factor Threshold: 20	
Predict Compositions		
	1. Prediction Settings:	
		Mass Tolerance: 2 ppm
		Min. Element Counts: C H
		Max. Element Counts: C ₉₀ H ₁₉₀ Br ₃ Cl ₄ N ₄ O ₂₀ P S ₃
		Min. RDBE: 0
		Max. RDBE: 40
		Min. H/C: 0.1
		Max. H/C: 3.5
		Max. # Candidates: 10
		Max. # Internal Candidates: 200
	2. Pattern Matching:	
		Intensity Tolerance [%]: 10

continued on next page

Processing node	Parameter	Settings
Search mzVault		Intensity Threshold [%]: 0.1 S/N Threshold: 3 Min. Spectral Fit [%]: 30 Min. Pattern Cov. [%]: 90 Use Dynamic Recalibration: True
	3. Fragments Matching:	Use Fragments Matching: True Mass Tolerance: 10 ppm S/N Threshold: 5
	1. Search Settings:	mzVault Library: <i>add libraries created with mzVault</i> Max. # Results: 10 Match Factor Threshold: 50 Search Algorithm: HighChem HighRes Match Analyzer Type: True IT Fragment Mass Tolerance: 0.4 Da FT Fragment Mass Tolerance: 10 ppm Use Retention Time: True Precursor Mass Tolerance: 10 ppm Apply Intensity Threshold: True Match Ionization Method: True Ion Activation Energy Tolerance: 20 Match Ion Activation Energy: Match with Tolerance Match Ion Activation Type: True Compound Classes: All Remove Precursor Ion: True RT Tolerance [min]: 3

Table S2. Workflow of the non-target software Compound Discoverer 3.3.2.31 - positive ionization mode data

Processing node	Parameter	Settings
Select Spectra		
	1. Spectrum Properties	
	Filter:	Lower RT Limit: 0 Upper RT Limit: 16.5 First Scan: 0 Last Scan: 0 Ignore Specified Scans: (not specified) Lowest Charge State: 0 Highest Charge State: 0 Min. Precursor Mass: 50 Da Max. Precursor Mass: 5000 Da Total Intensity Threshold: 10000 Minimum Peak Count: 1
	2. Scan Event Filters:	Mass Analyzer: Is FTMS MS Order: Is MS2, MS1 Activation Type: (not specified) Min. Collision Energy: 0 Max. Collision Energy: 1000 Scan Type: Any Polarity Mode: Is + MS1 Mass Range: (not specified)
	FAIMS CV: (not specified)	3. Peak Filters:
		S/N Threshold (FT-only): 1.5
	4. Replacements for	
	Unrecognized Properties:	Unrecognized Charge Replacements: 1 Unrecognized Mass Analyzer Replacements: ITMS Unrecognized MS Order Replacements: MS2 Unrecognized Activation Type Replacements: CID Unrecognized Polarity Replacements: +

continued on next page

Processing node	Parameter	Settings
		Unrecognized MS Resolution@200 Replacements: 60000
		Unrecognized MSn Resolution@200 Replacements: 30000
	5. General Settings:	Precursor Selection: Use MS1 Precursor Use Isotope Pattern in Precursor Reevaluation: True Provide Profile Spectra: Automatic Store Chromatograms: False
Align Retention Times	1. General Settings:	Alignment Model: Adaptive curve Alignment Fallback: Use Linear Model Maximum Shift [min]: 0.2 Shift Reference File: True Mass Tolerance: 5 ppm Remove Outlier: True
Detect Compounds	1. General Settings:	Mass Tolerance [ppm]: 2 ppm Min. Peak Intensity: 100000 Min. [#] Scans per Peak: 5 Use Most Intense Isotope Only: True
	2. Trace Detection:	Max. Number of Gaps to Correct: 2 Min. Number of Adjacent Non-Zeros: 2
	3. Peak Detection:	Chromatographic S/N Threshold: 1.5 Remove Baseline: False Gap Ratio Threshold: 0.35 Max. Peak Width [min]: 1 Min. Relative Valley Depth: 0.1
	4. Isotope Pattern Detection:	
continued on next page		

Processing node	Parameter	Settings
Group Compounds		Group Isotopes for: Br, Cl
		Use Peak Quality for Isotope Grouping: True
		Filter out Features with Bad Peaks Only: True
		Zig-Zag Index Threshold: 0.2
		Jaggedness Threshold: 0.4
		Modality Threshold: 0.9
		Remove Potentially False Positive Isotopes: True
	5. Compound Detection:	
		Ions: [2M+H]+1; [M+H]+1; [2M+Na]+1; [M+Na]+1
		Base Ions: [M+H]+1
		Remove Singlets: True
	6. AcquireX Settings:	
		Detect Persistent Background Ions: False
	1. General Settings:	
		Mass Tolerance: 2 ppm
		RT Tolerance [min]: 0.2
		Align Peaks: False
		Preferred Ions: [M+H]+1; [M+Na]+1
		Area Integration: All Ions
2. Peak Rating		
Contributions:		
	Area Contribution: 3	
	CV Contribution: 1	
	FWHM to Base Contribution: 5	
	Jaggedness Contribution: 5	
	Modality Contribution: 5	
	Zig-Zag Index Contribution: 5	
3. Peak Rating Filter:		
	Peak Rating Threshold: 4	
	Number of Files: 4	
Fill Gaps		
	1. General Settings:	

continued on next page

Processing node	Parameter	Settings
Mark Background Compounds		Mass Tolerance: 2 ppm S/N Threshold: 1.5 Use Real Peak Detection: True Apply Restrictive Gap Filling: True
	1. General Settings:	Max. Sample/Blank: 5 Max. Blank/Sample: 0 Hide Background: False
Assign Compound Annotations	1. General Settings:	Mass Tolerance: 2 ppm
	2. Data Sources:	Data Source #1: Predicted Compositions Data Source #2: mzCloud Search
	3. Scoring Rules:	Use mzLogic: True Use Spectral Distance: True SFit Threshold: 20 SFit Range: 20
	4. Reprocessing:	Clear Names: False
Search mzCloud	1. General Settings:	Compound Classes: All Precursor Mass Tolerance: 5 ppm FT Fragment Mass Tolerance: 5 ppm IT Fragment Mass Tolerance: 0.4 Da Library: Autoprocessed; Reference Post Processing: Recalibrated Max. # Results: 10

continued on next page

Processing node	Parameter	Settings
		Annotate Matching Fragments: False
		Search MSn Tree: False
	2. DDA Search:	
		Identity Search: HighChem HighRes
		Match Activation Type: True
		Match Activation Energy: Match with Tolerance
		Activation Energy Tolerance: 20
		Apply Intensity Threshold: True
		Similarity Search: Similarity Forward
		Match Factor Threshold: 80
	3. DIA Search:	
		Use DIA Scans for Search: False
		Max. Isolation Width [Da]: 500
		Match Activation Type: False
		Match Activation Energy: Any
		Activation Energy Tolerance: 100
		Apply Intensity Threshold: False
		Match Factor Threshold: 20
	Predict Compositions	
1. Prediction Settings:		
	Mass Tolerance: 2 ppm	
	Min. Element Counts: C H	
	Max. Element Counts: C ₉₀ H ₁₉₀ Br ₃ Cl ₄ N ₄ O ₂₀ S ₃ P	
	Min. RDBE: 0	
	Max. RDBE: 40	
	Min. H/C: 0.1	
	Max. H/C: 3.5	
	Max. # Candidates: 10	
	Max. # Internal Candidates: 200	
2. Pattern Matching:		
	Intensity Tolerance [%]: 10	
	Intensity Threshold [%]: 0.1	
	S/N Threshold: 3	

continued on next page

Processing node	Parameter	Settings
		Min. Spectral Fit [%]: 30
		Min. Pattern Cov. [%]: 90
		Use Dynamic Recalibration: True
	3. Fragments Matching:	
		Use Fragments Matching: True
		Mass Tolerance: 10 ppm
		S/N Threshold: 5