

Leveraging open cheminformatics tools for non-targeted metabolomics analysis of *C. elegans*: a workflow comparison and application to strains related to xenobiotic metabolism and neurodegeneration

Gianfranco Frigerio^{1,2*}, Yunjia Lai³, Emma L. Schymanski^{1*}, Gary W. Miller³.

¹ Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, 6, Avenue du Swing, L-4367 Belvaux, Luxembourg

² Center for Omics Sciences (COSR), IRCCS San Raffaele Scientific Institute, Milan, Italy

³ Department of Environmental Health Sciences, Mailman School of Public Health at Columbia University, New York, NY, USA

* Corresponding authors: GF: frigerio.gianfranco@hsr.it & ELS: emma.schymanski@uni.lu

ORCID: GF: [0000-0002-3538-1443](https://orcid.org/0000-0002-3538-1443), YL: [0000-0002-1081-0897](https://orcid.org/0000-0002-1081-0897), ELS: [0000-0001-6868-8145](https://orcid.org/0000-0001-6868-8145), GWM: [0000-0001-8984-1284](https://orcid.org/0000-0001-8984-1284)

Supplementary data 1

Supplementary figures S1: Summary of the number of features obtained from the elaboration of the different combinations of analyses performed in RPLC NEG and HILIC POS, using patRoön or MS-DIAL. The total number of features initially retrieved was reduced to include only those passing the pooled QC check, and further divided by sample preparation scheme. The lower section shows the number of statistically significant features (ANOVA) among the groups of strains.

Supplementary figures S2, S5, S8, S11: For each of the four combinations of analysis type and data elaboration tool used, two Eulero-venn diagrams are reported: the first one shows the features that passed the QC check that are in common among the considered extraction schemes (scheme 1 or scheme 2); the second one, analogously, shows features statistically significant among sample groups.

Supplementary figures S3-S4, S5-S7, S9-S10, S12-S13: For each of the four combinations of analysis type and data elaboration tool used, a table reporting the number of annotated features, according to the levels of reporting, is shown for the total number of features; a similar table is then reported for features that passed the QC criteria; in the second column a similar table is reported separately considering samples treated with scheme 1 and scheme 2; in the third column similar tables are reported for features that were significantly different among considered groups.

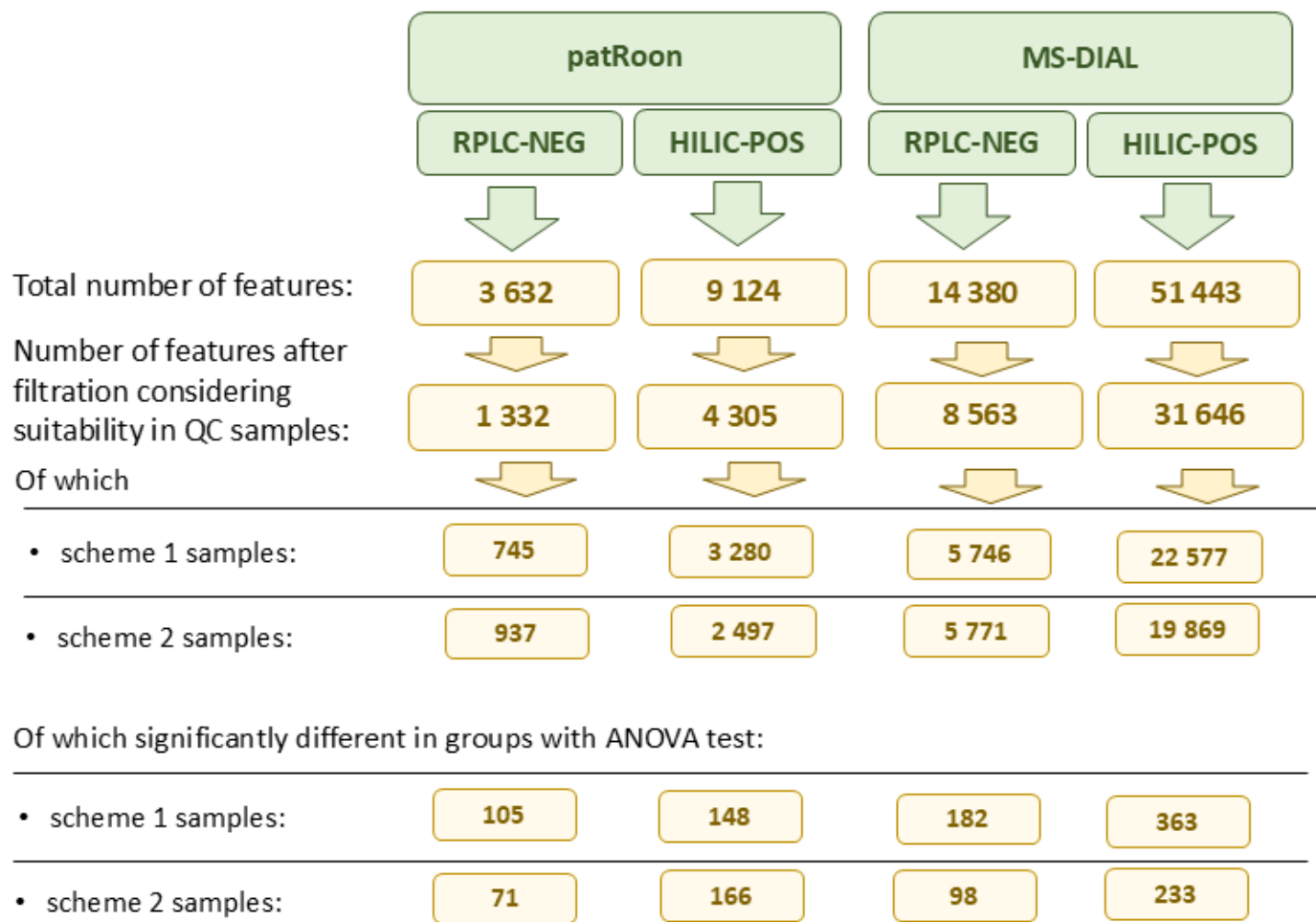
Supplementary figure S14: Upset plot showing the unique compounds annotated at level 3 or above, obtained considering all samples preparation schemes and chromatographic runs, grouped by the four annotation strategies implemented and showing the intersection among them.

Supplementary figure S15: SankeyNetwork graph illustrating an overview of the compound classes of annotated compounds at level 3 or above.

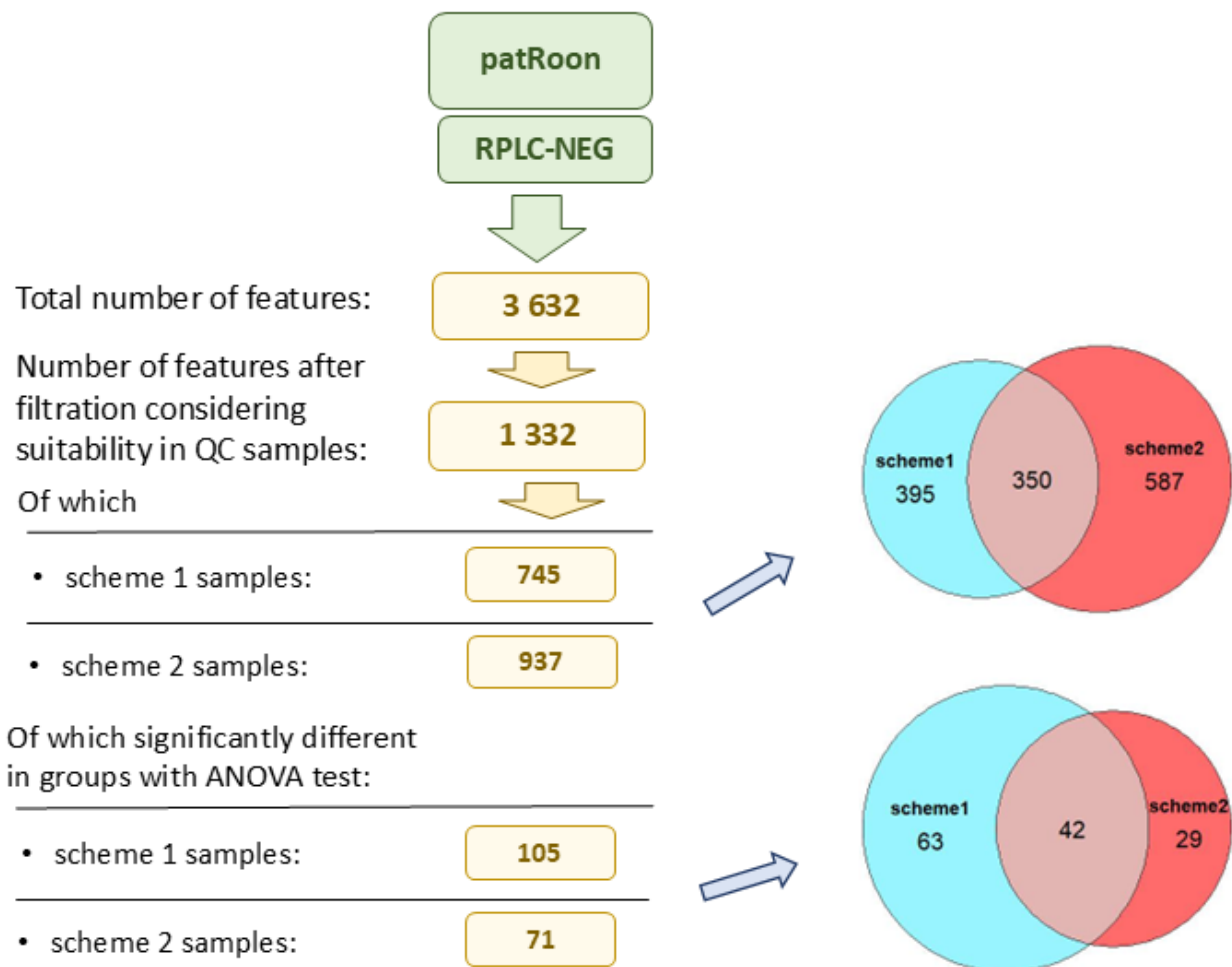
Supplementary figure S16: SankeyNetwork graph illustrating an overview of the compound classes of annotated compounds at level 3 or above with patRoön.

Supplementary figure S17: SankeyNetwork graph illustrating an overview of the compound classes of annotated compounds at level 3 or above that were significantly different among the considered groups of strains.

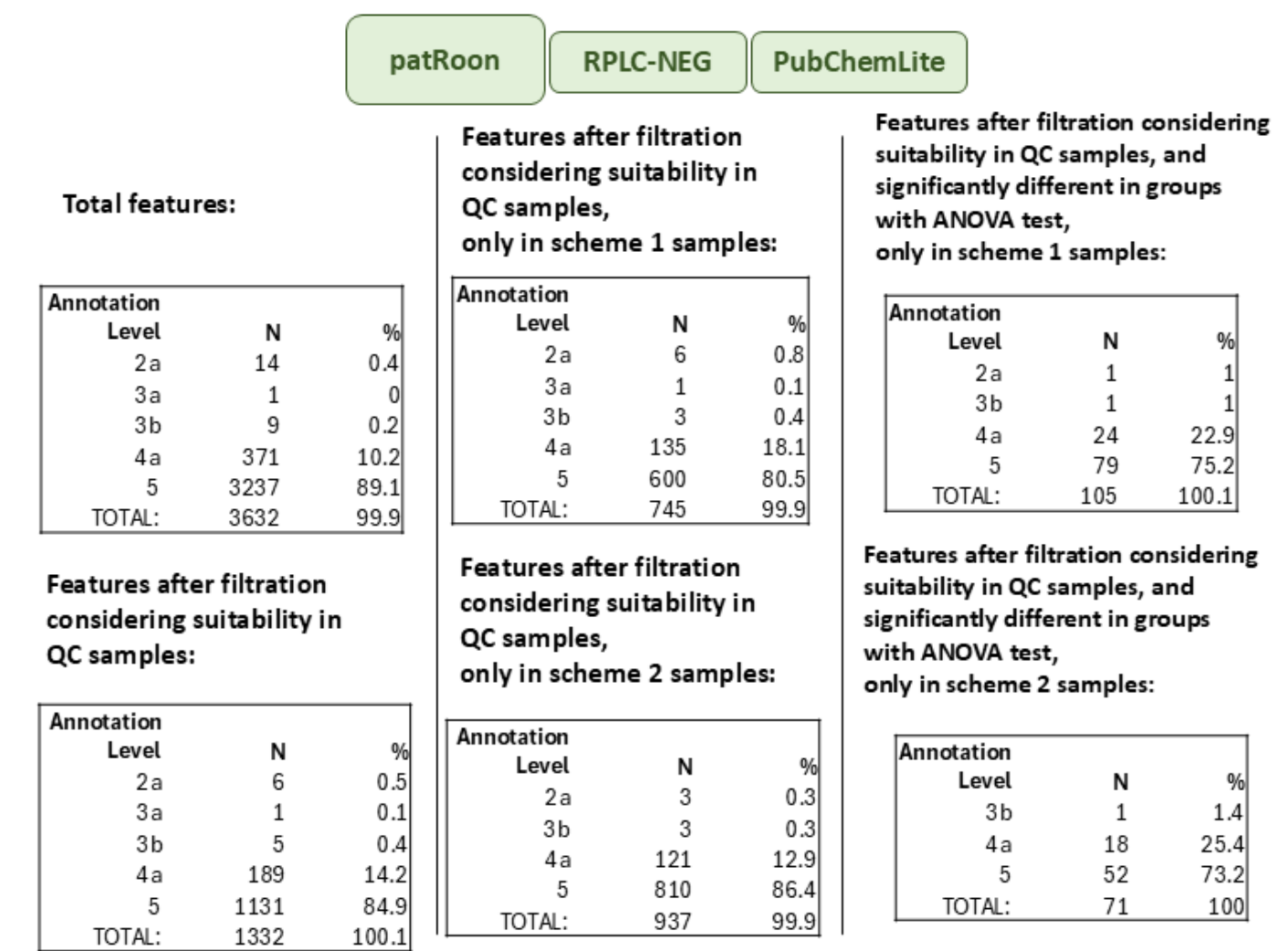
Supplementary figure S18: Visual representation of the enrichment analyses performed with the FELLA package considering all the annotated compounds that were significantly different among the considered groups of strains.



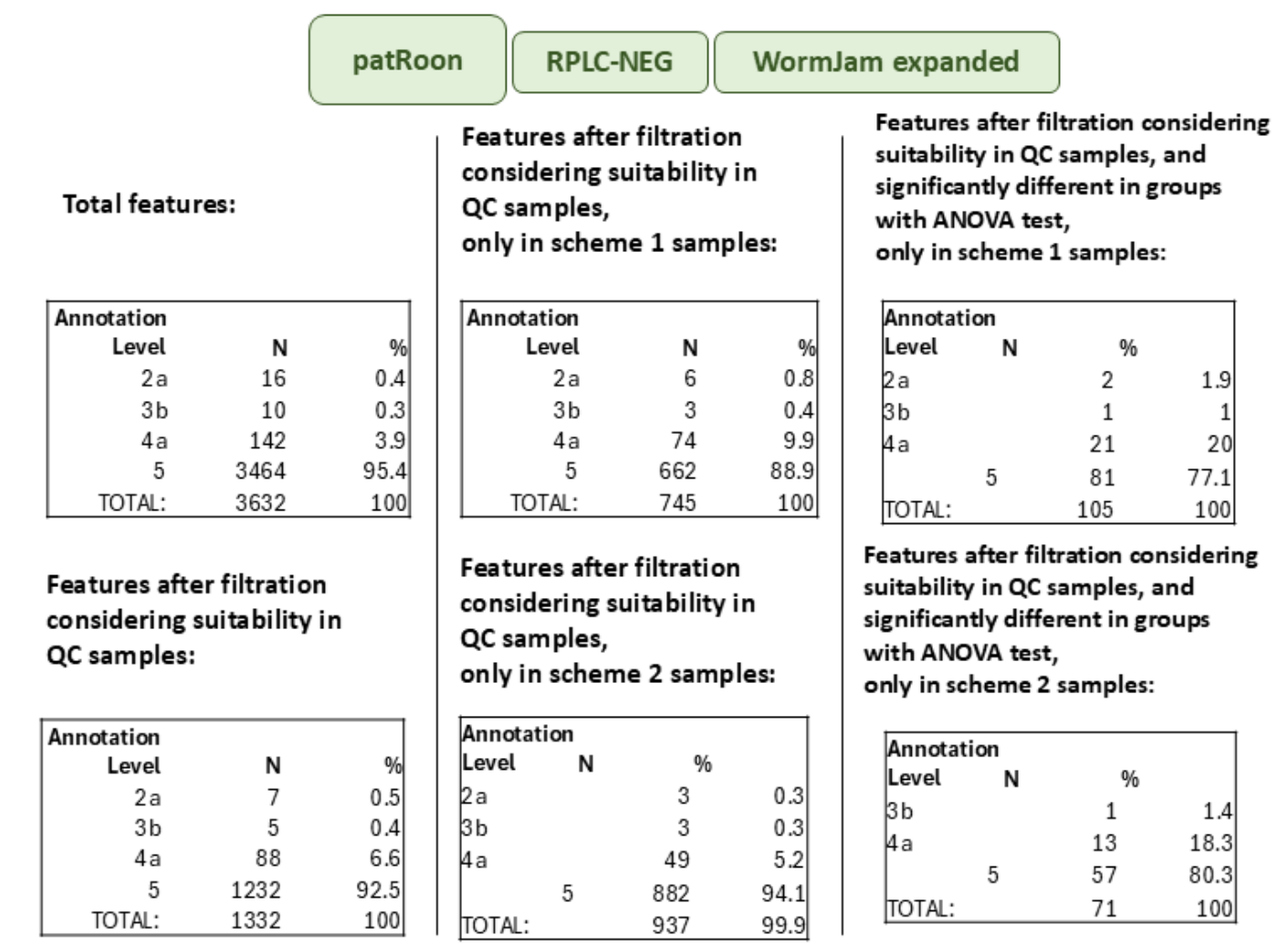
Supplementary figures S1: Summary of the number of features obtained from the elaboration of the different combinations of analyses performed in RPLC NEG and HILIC POS, using patRoön or MS-DIAL. The total number of features initially retrieved was reduced to include only those passing the pooled QC check, and further divided by sample preparation scheme. The lower section shows the number of statistically significant features (ANOVA) among the groups of strains.



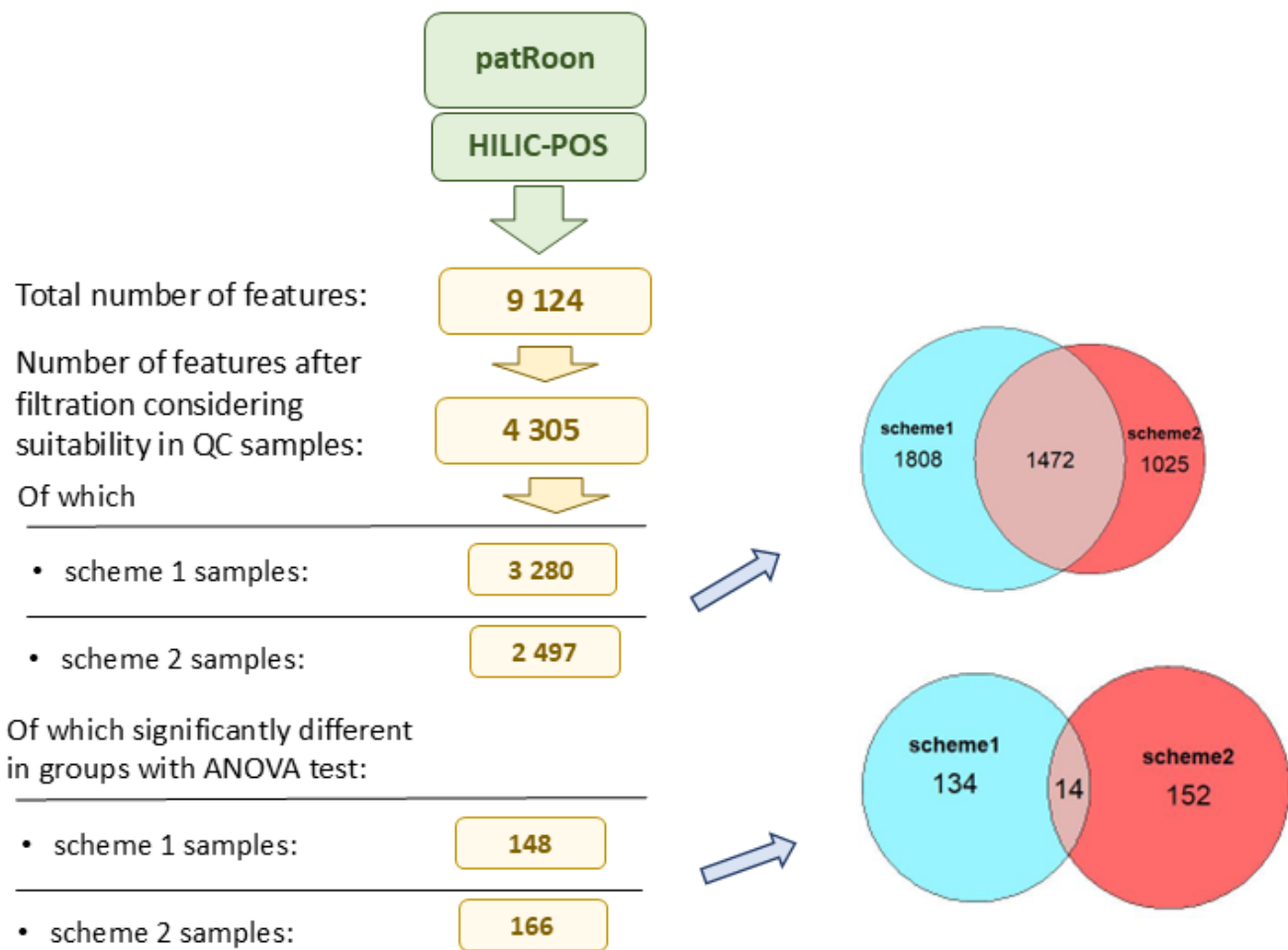
Supplementary figures S2: For RPLC NEG analyses elaborated with patRoön, two Eulero-venn diagrams are reported: the first one shows the features that passed the QC check that are in common among the considered extraction schemes (scheme 1 or scheme 2); the second one, analogously, shows features statistically significant among sample groups.



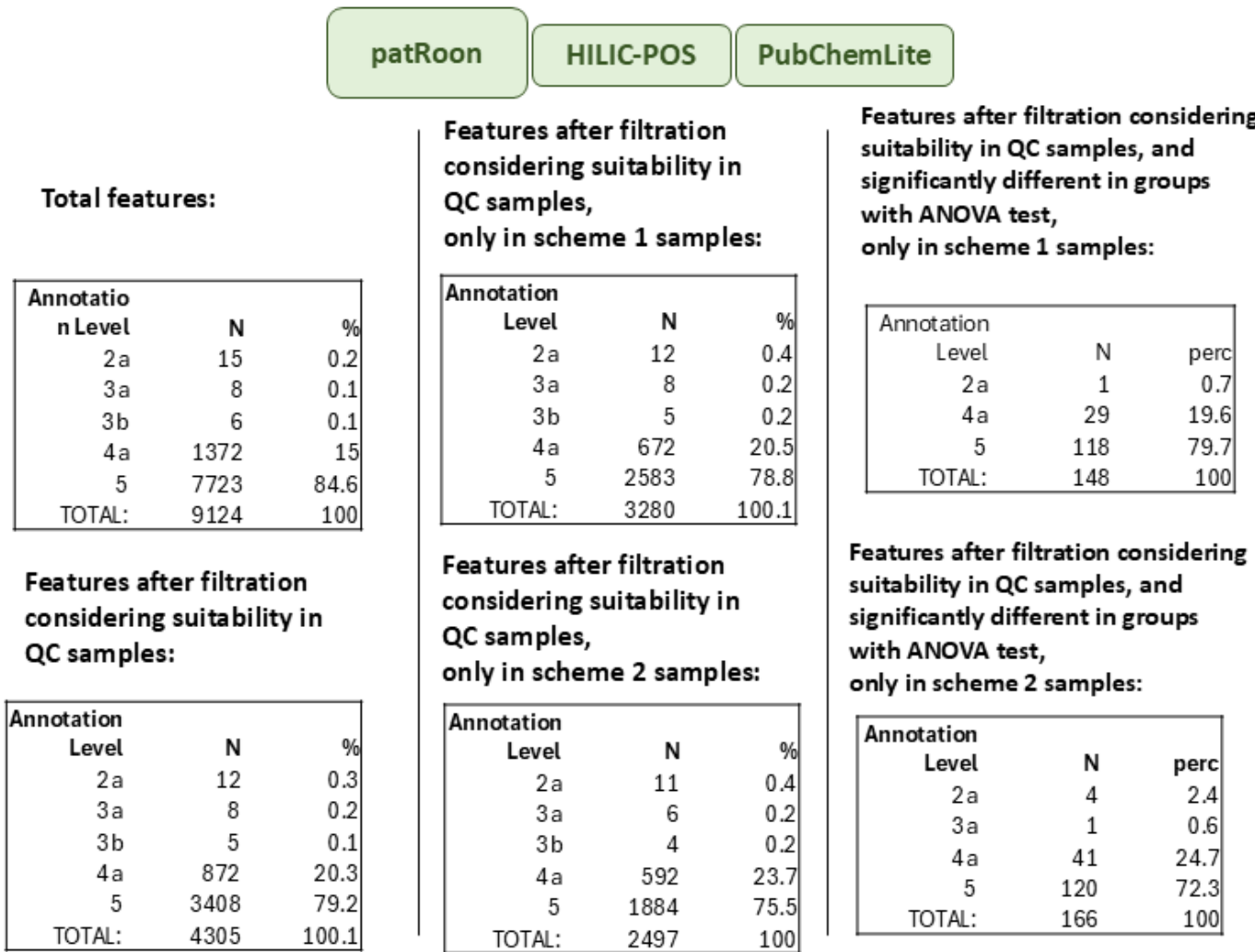
Supplementary figures S3: For RPLC NEG analyses elaborated with patRoön with the PubChemLite database, a table reporting the number of annotated features, according to the levels of reporting, is shown for the total number of features; a similar table is then reported for features that passed the QC criteria; in the second column a similar table is reported separately considering samples treated with scheme 1 and scheme 2; in the third column similar tables are reported for features that were significantly different among considered groups.



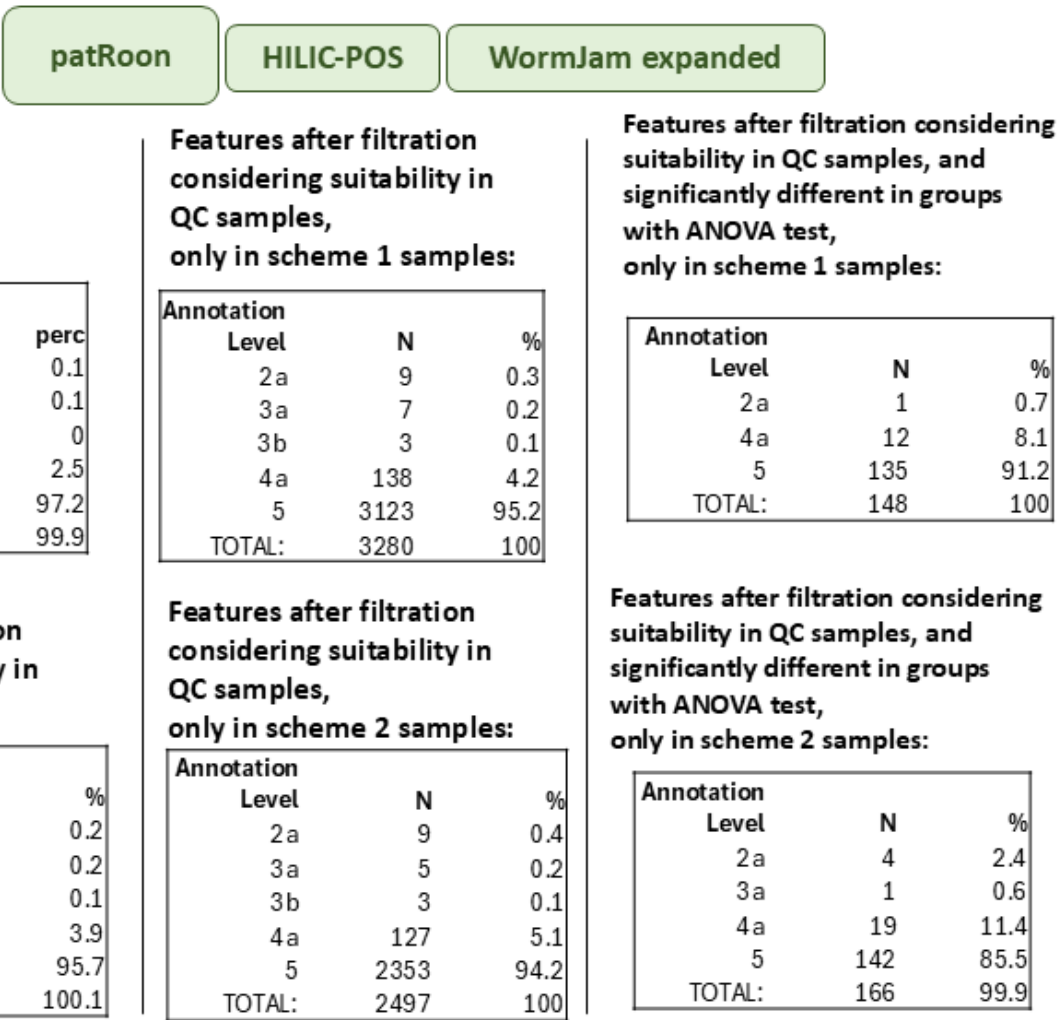
Supplementary figures S4: For RPLC NEG analyses elaborated with patRoom with the WormJam expanded database, a table reporting the number of annotated features, according to the levels of reporting, is shown for the total number of features; a similar table is then reported for features that passed the QC criteria; in the second column a similar table is reported separately considering samples treated with scheme 1 and scheme 2; in the third column similar tables are reported for features that were significantly different among considered groups.



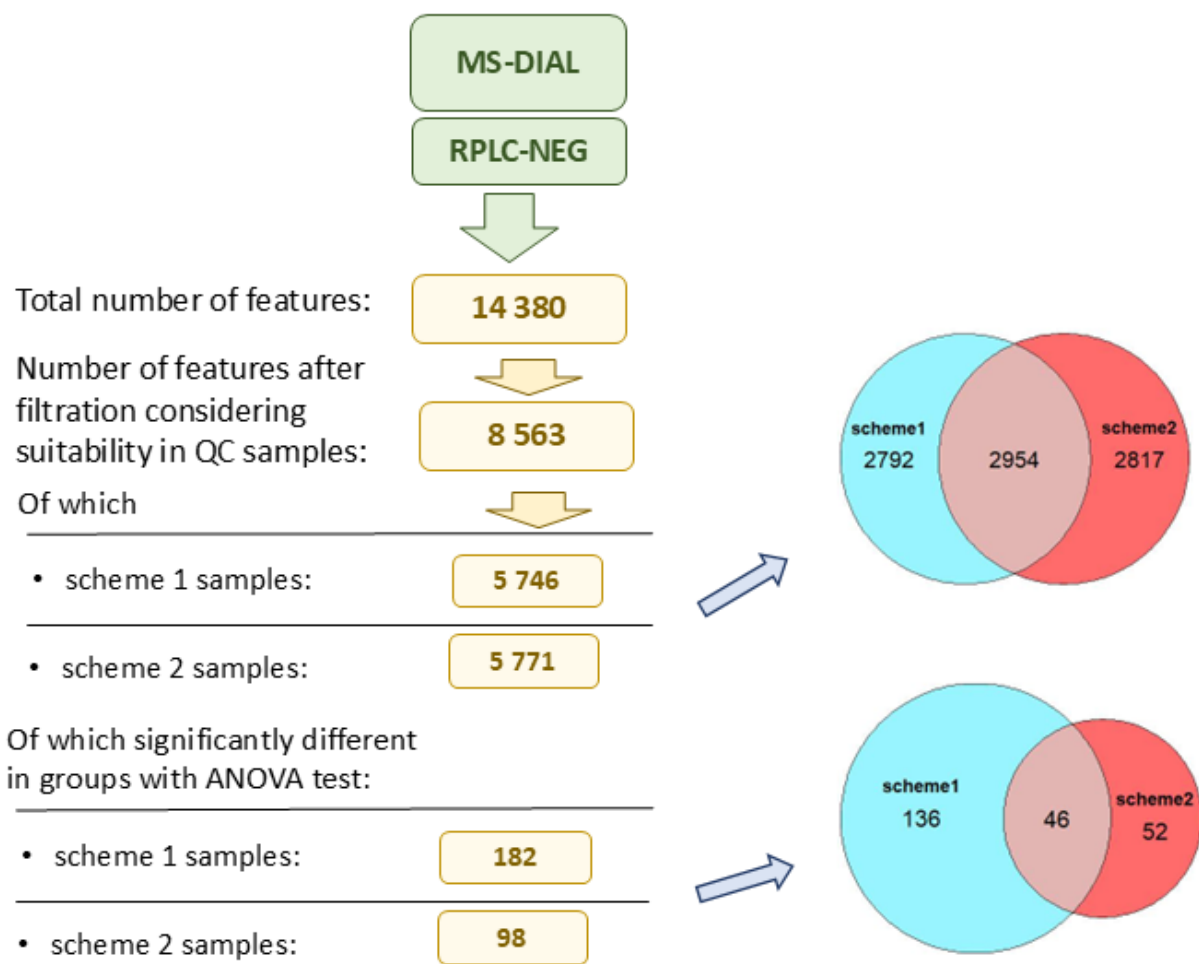
Supplementary figures S5: For HILIC POS analyses elaborated with patRoön, two Eulero-venn diagrams are reported: the first one shows the features that passed the QC check that are in common among the considered extraction schemes (scheme 1 or scheme 2); the second one, analogously, shows features statistically significant among sample groups.



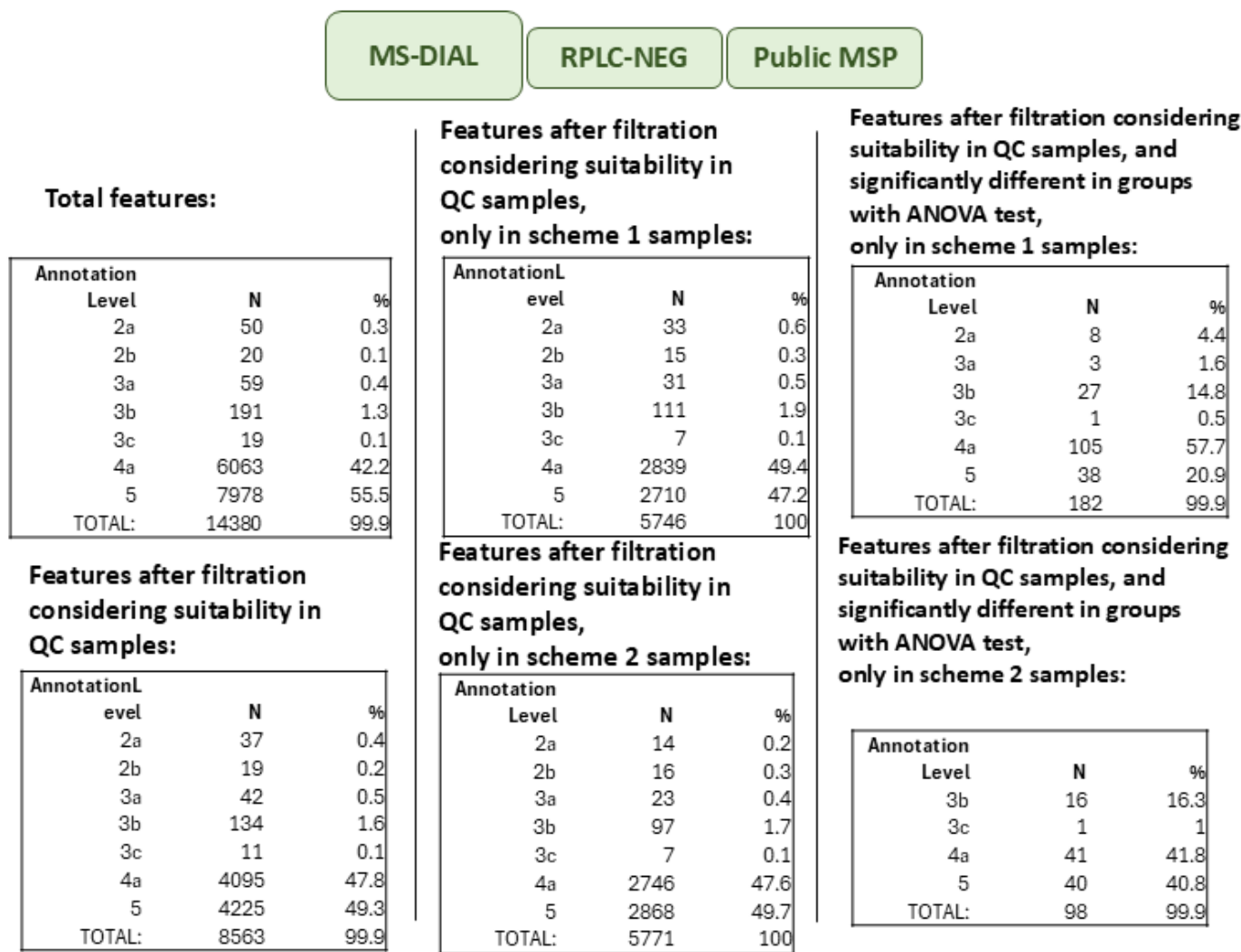
Supplementary figures S6: For HILIC POS analyses elaborated with patRoön with the PubChemLite database, a table reporting the number of annotated features, according to the levels of reporting, is shown for the total number of features; a similar table is then reported for features that passed the QC criteria; in the second column a similar table is reported separately considering samples treated with scheme 1 and scheme 2; in the third column similar tables are reported for features that were significantly different among considered groups.



Supplementary figures S7: For HILIC POS analyses elaborated with patRoön with the WormJam expanded database, a table reporting the number of annotated features, according to the levels of reporting, is shown for the total number of features; a similar table is then reported for features that passed the QC criteria; in the second column a similar table is reported separately considering samples treated with scheme 1 and scheme 2; in the third column similar tables are reported for features that were significantly different among considered groups.



Supplementary figures S8: For RPLC NEG analyses elaborated with MS-DIAL, two Eulero-venn diagrams are reported: the first one shows the features that passed the QC check that are in common among the considered extraction schemes (scheme 1 or scheme 2); the second one, analogously, shows features statistically significant among sample groups.



Supplementary figures S9 For RPLC NEG analyses elaborated with MS-DIAL with the public MPS libraries, a table reporting the number of annotated features, according to the levels of reporting, is shown for the total number of features; a similar table is then reported for features that passed the QC criteria; in the second column a similar table is reported separately considering samples treated with scheme 1 and scheme 2; in the third column similar tables are reported for features that were significantly different among considered groups.

MS-DIAL

RPLC-NEG

WormJam expanded

Total features:

Annotation	Level	N	%
	2a	9	0.1
	3a	30	0.2
	3b	66	0.5
	4a	2431	17.1
	5	11679	82.2
	TOTAL:	14215	100.1

Features after filtration considering suitability in QC samples, only in scheme 1 samples:

Annotation	Level	N	%
	2a	7	0.1
	3a	17	0.3
	3b	31	0.5
	4a	1307	23
	5	4317	76
	TOTAL:	5679	99.9

Features after filtration considering suitability in QC samples, and significantly different in groups with ANOVA test, only in scheme 1 samples:

Annotation	Level	N	%
	2a	1	0.6
	3b	2	1.1
	4a	41	22.8
	5	136	75.6
	TOTAL:	180	100.1

Features after filtration considering suitability in QC samples:

Annotation	Level	N	%
	2a	8	0.1
	3a	21	0.2
	3b	44	0.5
	4a	1768	20.9
	5	6623	78.2
	TOTAL:	8464	99.9

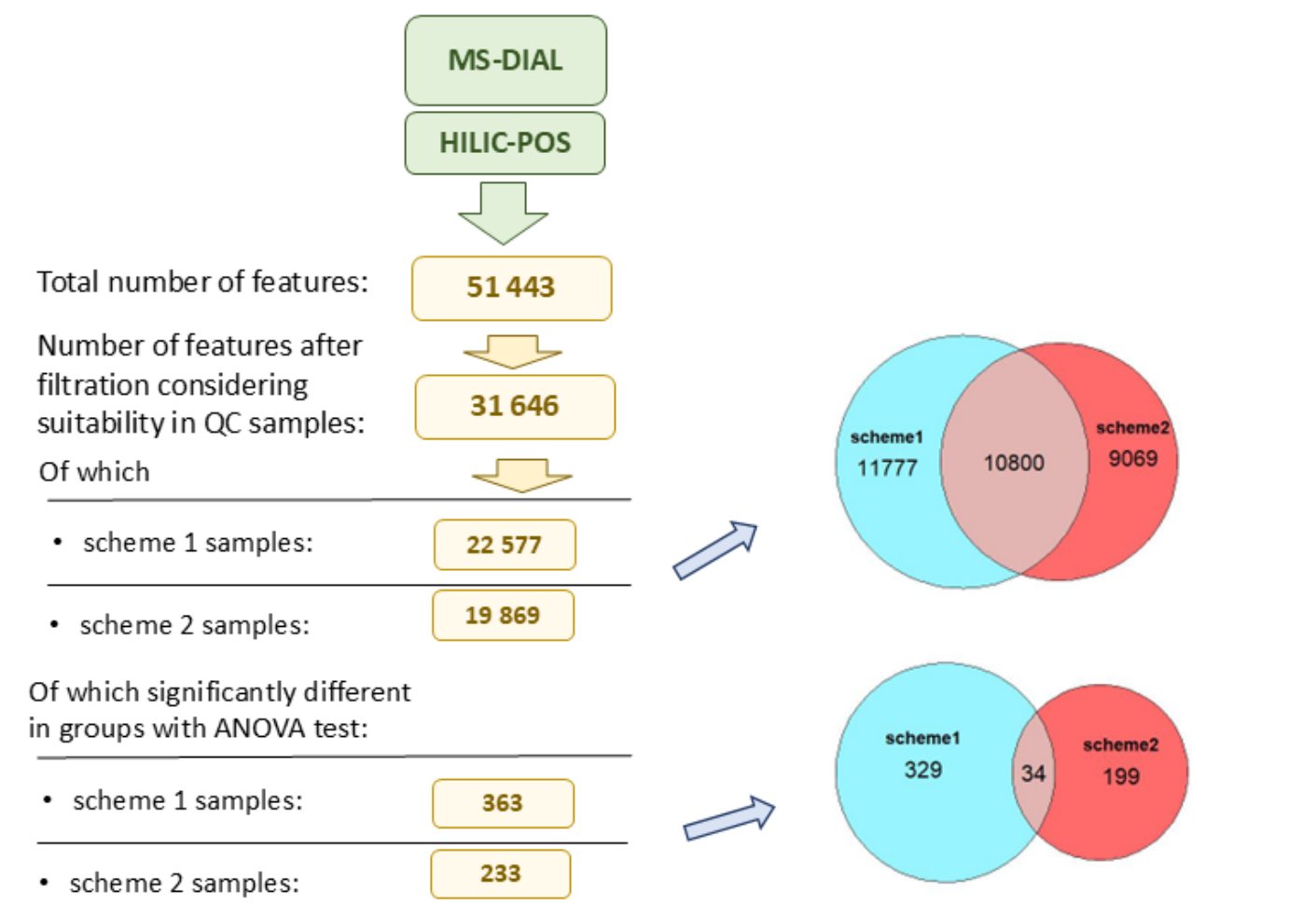
Features after filtration considering suitability in QC samples, only in scheme 2 samples:

Annotation	Level	N	%
	2a	3	0.1
	3a	12	0.2
	3b	22	0.4
	4a	1147	20.1
	5	4520	79.2
	TOTAL:	5704	100

Features after filtration considering suitability in QC samples, and significantly different in groups with ANOVA test, only in scheme 2 samples:

Annotation	Level	N	%
	2a	1	1.1
	4a	27	28.7
	5	66	70.2
	TOTAL:	94	100

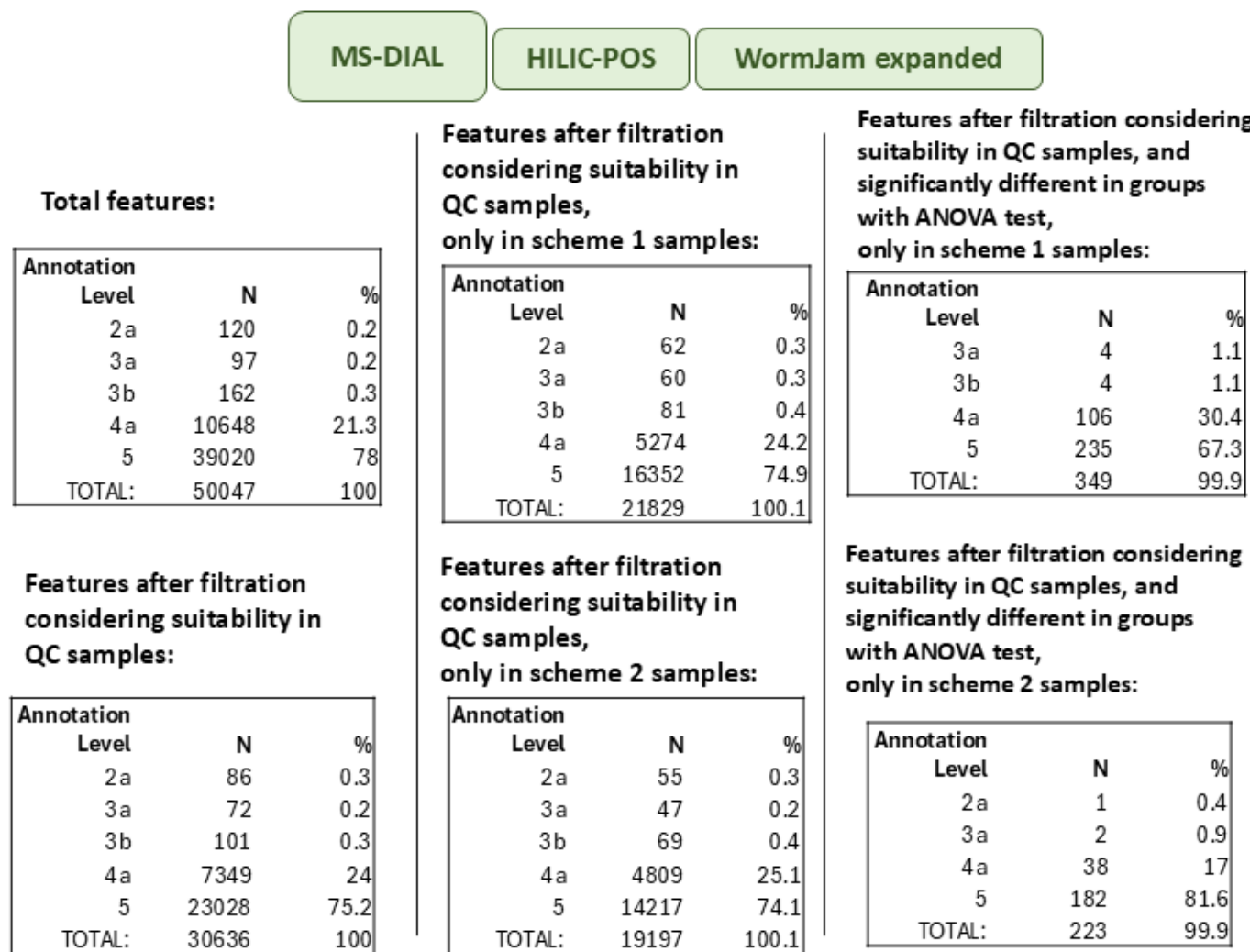
Supplementary figures S10: For RPLC NEG analyses elaborated with MS-DIAL with the WormJam expanded MSP libraries, a table reporting the number of annotated features, according to the levels of reporting, is shown for the total number of features; a similar table is then reported for features that passed the QC criteria; in the second column a similar table is reported separately considering samples treated with scheme 1 and scheme 2; in the third column similar tables are reported for features that were significantly different among considered groups.



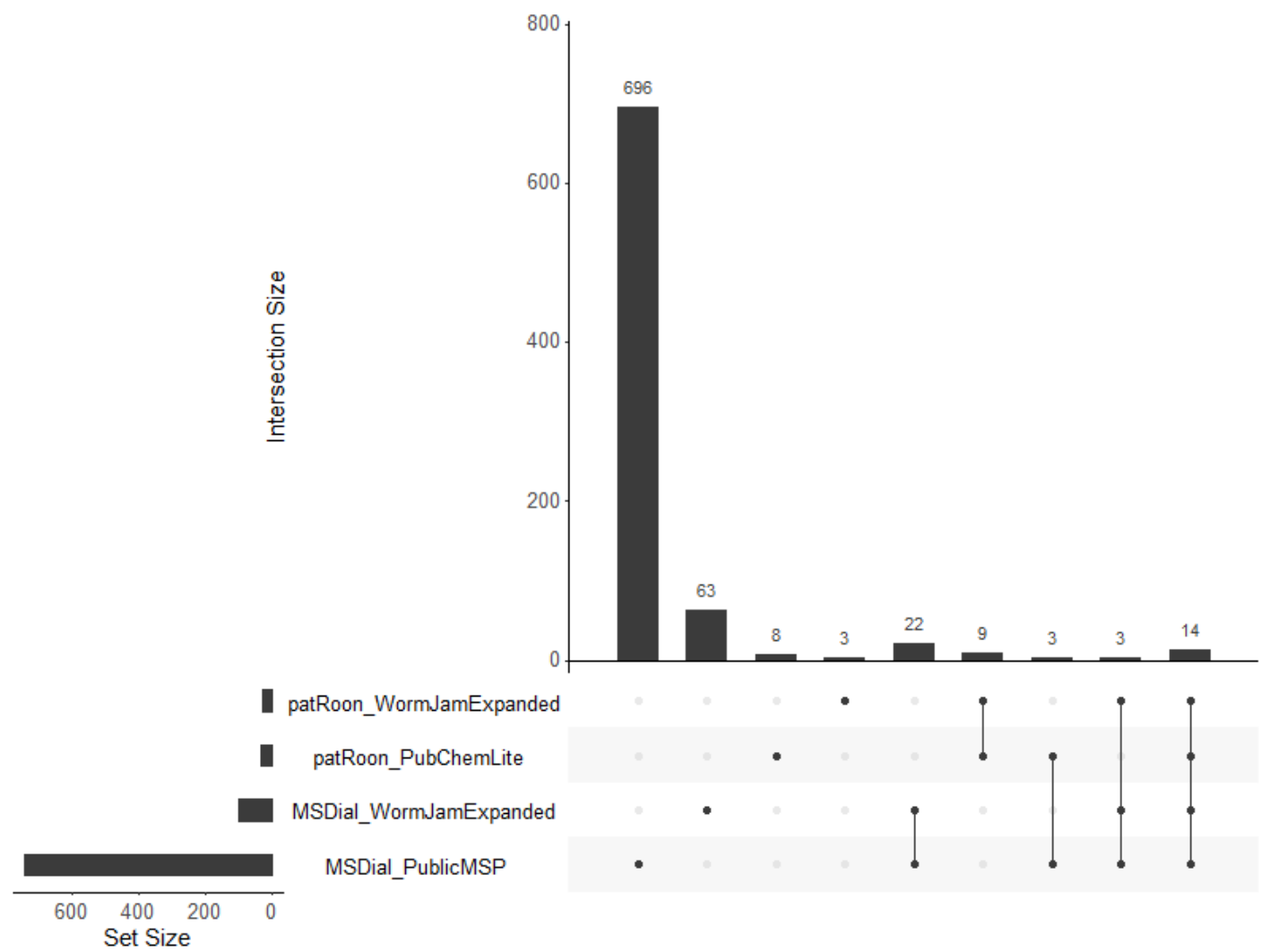
Supplementary figures S11: For HILIC POS analyses elaborated with MS-DIAL, two Eulero-venn diagrams are reported: the first one shows the features that passed the QC check that are in common among the considered extraction schemes (scheme 1 or scheme 2); the second one, analogously, shows features statistically significant among sample groups.

	MS-DIAL	HILIC-POS	Public MSP																																																																																				
Total features:	Features after filtration considering suitability in QC samples, only in scheme 1 samples: <table> <tr><th>Annotation</th><th>Level</th><th>N</th><th>%</th></tr> <tr><td></td><td>2 a</td><td>1008</td><td>2</td></tr> <tr><td></td><td>3 a</td><td>517</td><td>1</td></tr> <tr><td></td><td>3 b</td><td>1830</td><td>3.6</td></tr> <tr><td></td><td>4 a</td><td>27355</td><td>53.2</td></tr> <tr><td></td><td>5</td><td>20733</td><td>40.3</td></tr> <tr><td></td><td>TOTAL:</td><td>51443</td><td>100.1</td></tr> </table>	Annotation	Level	N	%		2 a	1008	2		3 a	517	1		3 b	1830	3.6		4 a	27355	53.2		5	20733	40.3		TOTAL:	51443	100.1	Features after filtration considering suitability in QC samples, only in scheme 2 samples: <table> <tr><th>Annotation</th><th>Level</th><th>N</th><th>%</th></tr> <tr><td></td><td>2 a</td><td>468</td><td>2.4</td></tr> <tr><td></td><td>3 a</td><td>228</td><td>1.1</td></tr> <tr><td></td><td>3 b</td><td>773</td><td>3.9</td></tr> <tr><td></td><td>4 a</td><td>11449</td><td>57.6</td></tr> <tr><td></td><td>5</td><td>6951</td><td>35</td></tr> <tr><td></td><td>TOTAL:</td><td>19869</td><td>100</td></tr> </table>	Annotation	Level	N	%		2 a	468	2.4		3 a	228	1.1		3 b	773	3.9		4 a	11449	57.6		5	6951	35		TOTAL:	19869	100	Features after filtration considering suitability in QC samples, and significantly different in groups with ANOVA test, only in scheme 1 samples: <table> <tr><th>Annotation</th><th>Level</th><th>N</th><th>%</th></tr> <tr><td></td><td>2 a</td><td>13</td><td>3.6</td></tr> <tr><td></td><td>3 a</td><td>6</td><td>1.7</td></tr> <tr><td></td><td>3 b</td><td>15</td><td>4.1</td></tr> <tr><td></td><td>4 a</td><td>197</td><td>54.3</td></tr> <tr><td></td><td>5</td><td>132</td><td>36.4</td></tr> <tr><td></td><td>TOTAL:</td><td>363</td><td>100.1</td></tr> </table>	Annotation	Level	N	%		2 a	13	3.6		3 a	6	1.7		3 b	15	4.1		4 a	197	54.3		5	132	36.4		TOTAL:	363	100.1
Annotation	Level	N	%																																																																																				
	2 a	1008	2																																																																																				
	3 a	517	1																																																																																				
	3 b	1830	3.6																																																																																				
	4 a	27355	53.2																																																																																				
	5	20733	40.3																																																																																				
	TOTAL:	51443	100.1																																																																																				
Annotation	Level	N	%																																																																																				
	2 a	468	2.4																																																																																				
	3 a	228	1.1																																																																																				
	3 b	773	3.9																																																																																				
	4 a	11449	57.6																																																																																				
	5	6951	35																																																																																				
	TOTAL:	19869	100																																																																																				
Annotation	Level	N	%																																																																																				
	2 a	13	3.6																																																																																				
	3 a	6	1.7																																																																																				
	3 b	15	4.1																																																																																				
	4 a	197	54.3																																																																																				
	5	132	36.4																																																																																				
	TOTAL:	363	100.1																																																																																				
Features after filtration considering suitability in QC samples:	Features after filtration considering suitability in QC samples, only in scheme 1 samples: <table> <tr><th>Annotation</th><th>Level</th><th>N</th><th>%</th></tr> <tr><td></td><td>2 a</td><td>738</td><td>2.3</td></tr> <tr><td></td><td>3 a</td><td>369</td><td>1.2</td></tr> <tr><td></td><td>3 b</td><td>1256</td><td>4</td></tr> <tr><td></td><td>4 a</td><td>18091</td><td>57.2</td></tr> <tr><td></td><td>5</td><td>11192</td><td>35.4</td></tr> <tr><td></td><td>TOTAL:</td><td>31646</td><td>100.1</td></tr> </table>	Annotation	Level	N	%		2 a	738	2.3		3 a	369	1.2		3 b	1256	4		4 a	18091	57.2		5	11192	35.4		TOTAL:	31646	100.1	Features after filtration considering suitability in QC samples, only in scheme 2 samples: <table> <tr><th>Annotation</th><th>Level</th><th>N</th><th>%</th></tr> <tr><td></td><td>2 a</td><td>10</td><td>4.3</td></tr> <tr><td></td><td>3 a</td><td>5</td><td>2.1</td></tr> <tr><td></td><td>3 b</td><td>13</td><td>5.6</td></tr> <tr><td></td><td>4 a</td><td>94</td><td>40.3</td></tr> <tr><td></td><td>5</td><td>111</td><td>47.6</td></tr> <tr><td></td><td>TOTAL:</td><td>233</td><td>99.9</td></tr> </table>	Annotation	Level	N	%		2 a	10	4.3		3 a	5	2.1		3 b	13	5.6		4 a	94	40.3		5	111	47.6		TOTAL:	233	99.9	Features after filtration considering suitability in QC samples, and significantly different in groups with ANOVA test, only in scheme 2 samples: <table> <tr><th>Annotation</th><th>Level</th><th>N</th><th>%</th></tr> <tr><td></td><td>2 a</td><td>10</td><td>4.3</td></tr> <tr><td></td><td>3 a</td><td>5</td><td>2.1</td></tr> <tr><td></td><td>3 b</td><td>13</td><td>5.6</td></tr> <tr><td></td><td>4 a</td><td>94</td><td>40.3</td></tr> <tr><td></td><td>5</td><td>111</td><td>47.6</td></tr> <tr><td></td><td>TOTAL:</td><td>233</td><td>99.9</td></tr> </table>	Annotation	Level	N	%		2 a	10	4.3		3 a	5	2.1		3 b	13	5.6		4 a	94	40.3		5	111	47.6		TOTAL:	233	99.9
Annotation	Level	N	%																																																																																				
	2 a	738	2.3																																																																																				
	3 a	369	1.2																																																																																				
	3 b	1256	4																																																																																				
	4 a	18091	57.2																																																																																				
	5	11192	35.4																																																																																				
	TOTAL:	31646	100.1																																																																																				
Annotation	Level	N	%																																																																																				
	2 a	10	4.3																																																																																				
	3 a	5	2.1																																																																																				
	3 b	13	5.6																																																																																				
	4 a	94	40.3																																																																																				
	5	111	47.6																																																																																				
	TOTAL:	233	99.9																																																																																				
Annotation	Level	N	%																																																																																				
	2 a	10	4.3																																																																																				
	3 a	5	2.1																																																																																				
	3 b	13	5.6																																																																																				
	4 a	94	40.3																																																																																				
	5	111	47.6																																																																																				
	TOTAL:	233	99.9																																																																																				

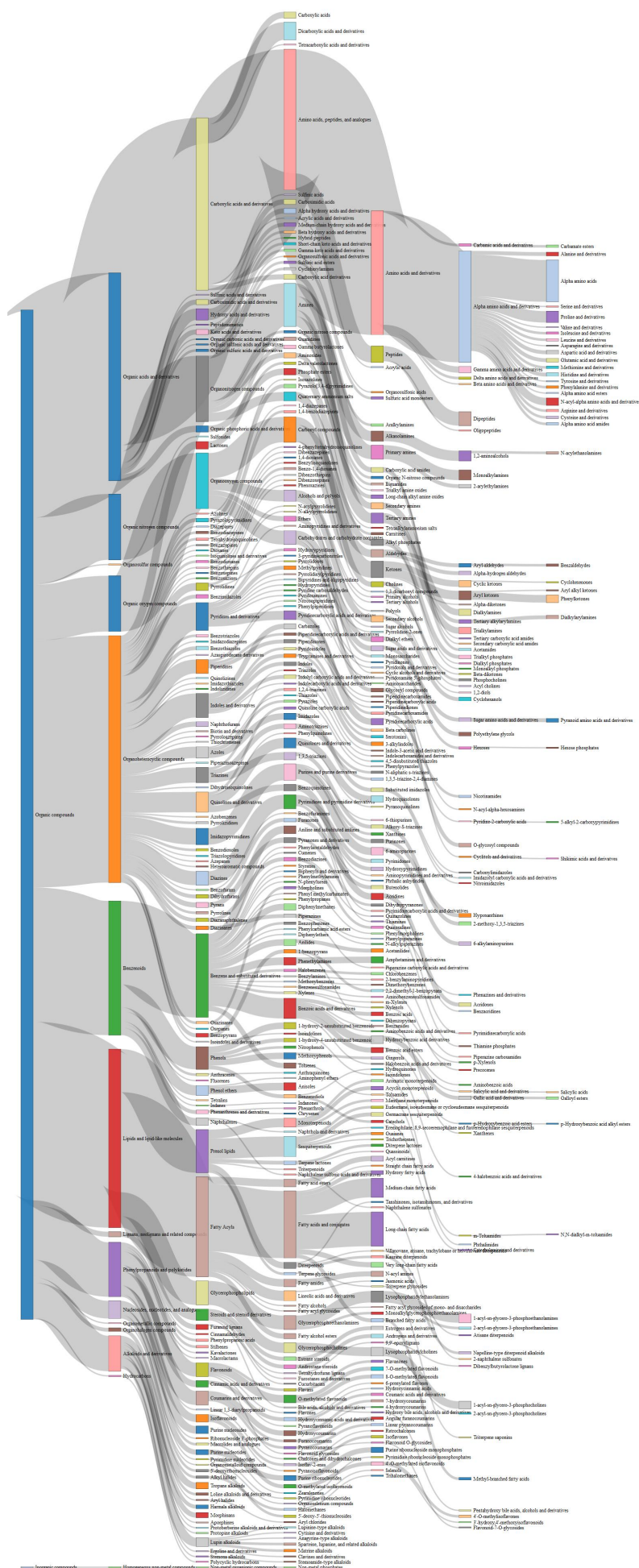
Supplementary figures S12: For HILIC POS analyses elaborated with MS-DIAL with the public MSP libraries, a table reporting the number of annotated features, according to the levels of reporting, is shown for the total number of features; a similar table is then reported for features that passed the QC criteria; in the second column a similar table is reported separately considering samples treated with scheme 1 and scheme 2; in the third column similar tables are reported for features that were significantly different among considered groups.



Supplementary figures S13: For HILIC POS analyses elaborated with MS-DIAL with the WormJam expanded MSP libraries, a table reporting the number of annotated features, according to the levels of reporting, is shown for the total number of features; a similar table is then reported for features that passed the QC criteria; in the second column a similar table is reported separately considering samples treated with scheme 1 and scheme 2; in the third column similar tables are reported for features that were significantly different among considered groups.

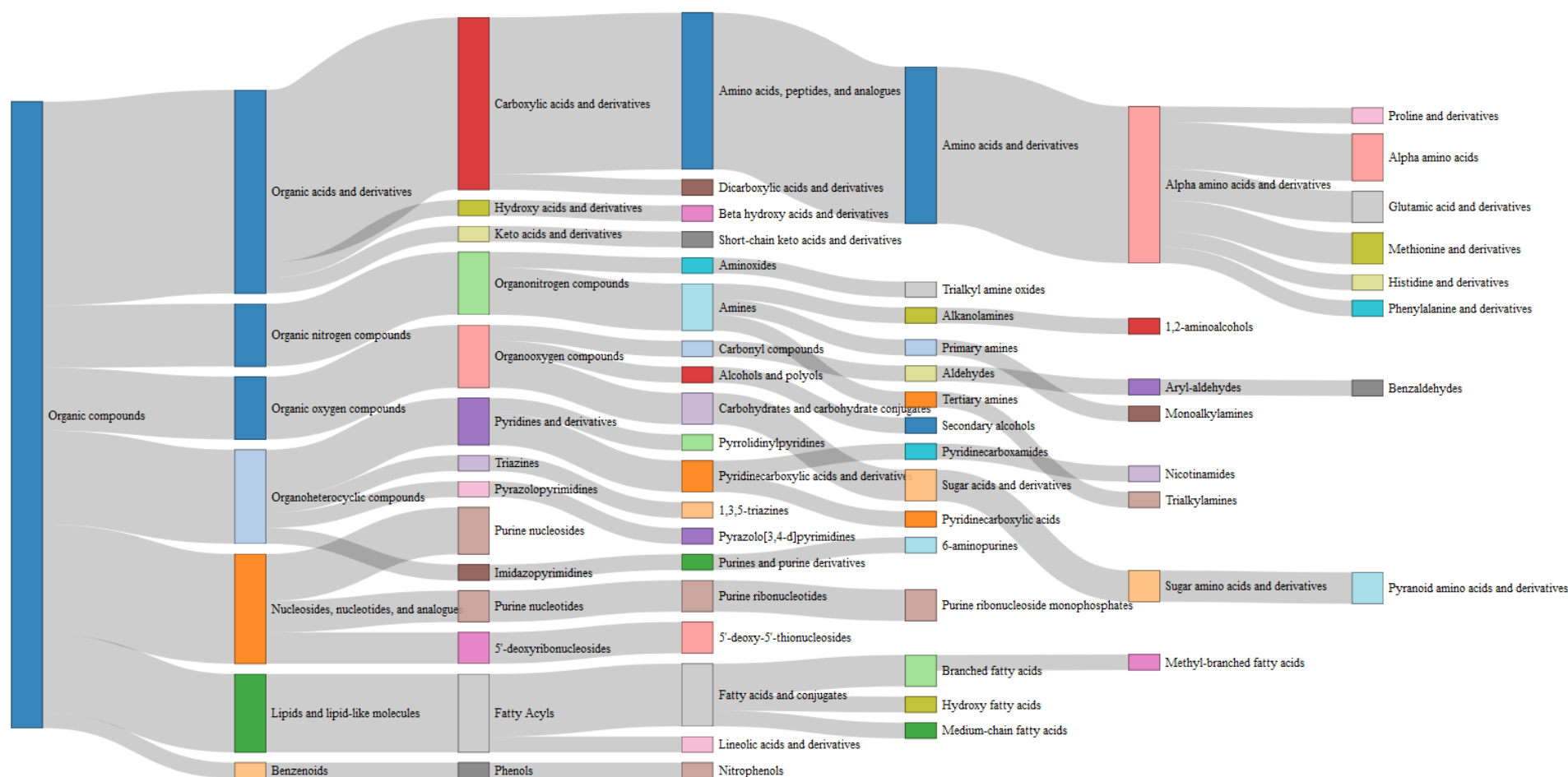


Supplementary figure S14: Upset plot showing the unique compounds annotated at level 3 or above, obtained considering all samples preparation schemes and chromatographic runs, grouped by the four annotation strategies implemented and showing the intersection among them.

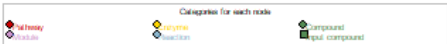


Supplementary figure S15: SankeyNetwork graph illustrating an overview of the compound classes of annotated compounds at level 3 or above.

Supplementary data 1



Supplementary figure S16: SankeyNetwork graph illustrating an overview of the compound classes of annotated compounds at level 3 or above with patRoom.



Supplementary figure S18: Visual representation of the enrichment analyses performed with the FELLA package considering all the annotated compounds that were significantly different among the considered groups of strains.