Supplementary Material:

Bioactivity in Rhododendron: A systemic analysis of antimicrobial and cytotoxic activities and their phylogenetic and phytochemical origins

Anne Grimbs 1, Abhinandan Shrestha 1, Ahmed Sayed Deyab Rezk 1, Sergio Grimbs 1, Inamullah Hakeem Said 1, Hartwig Schepker 2, Marc-Thorsten Hütt 1, Dirk Carl Albach 3,

Klaudia Brix 1, Nikolai Kuhnert 1 and Matthias S. Ullrich 1,*

1Department for Life Sciences & Chemistry, Jacobs University Bremen, Bremen, Germany
2Stiftung Bremer Rhododendronpark, Bremen, Germany
3Institute for Biology and Environmental Sciences, Carl von Ossietzky University Oldenburg, Oldenburg, Germany

List of Figures

S1 PCA scores plots colored with respect to cytotoxicity classifications .................................. 2
S2 PCA loadings plot ...................................................................................................................... 2
S3 Chemical diversity across the subgenera of Rhododendron .................................................. 3
S4 Box plots of 282 identified polyphenolics by antimicrobial activity classification ................. 12
S5 Box plots of all 292 identified polyphenolics by cytotoxicity classification towards HaCaT cells .......................................................... 22
S6 Box plots of all 292 identified polyphenolics by cytotoxicity classification towards IEC-6 cells .......................................................... 31
S7 Distribution of the detected LC-MS peaks regarding m/z ratios and retention times ............... 32
S8 The 25 most- and least-predictive LC-MS peaks regarding Cohen’s κ for cytotoxicity towards IEC-6 cells.......................................................... 33
S9 The 25 most- and least-predictive LC-MS peaks regarding Cohen’s κ for cytotoxicity towards HaCaT cells .......................................................... 34
S10 The 27 most-predictive LC-MS peaks regarding Cohen’s κ for antimicrobial activity and non-cytotoxicity towards both cell lines .................................................................................. 36
S11 The 50 most-predictive LC-MS peak combinations with additive effects regarding Cohen’s κ for antimicrobial activity ......................................................................... 38
S12 The 50 most-predictive LC-MS peak combinations with alternative effects regarding Cohen’s κ for antimicrobial activity ........................................................................ 39
S13 Heatmap of the most-predictive LC-MS peak combinations with alternative effects regarding Cohen’s κ for antimicrobial activity .......................................................... 40
S14 The 50 most-predictive LC-MS peak combinations with additive effects regarding Cohen’s κ for cytotoxicity towards IEC-6 cells ........................................................................ 41
S15 The 50 most-predictive LC-MS peak combinations with alternative effects regarding Cohen’s κ for cytotoxicity towards IEC-6 cells ........................................................................ 42
S16 Heatmap of the most-predictive LC-MS peak combinations with additive and alternative effects, respectively, regarding Cohen’s κ for cytotoxicity towards IEC-6 cells ........................................................................ 43

List of Tables

S1 Average and standard deviation of m/z ratios and retention times for most-predictive LC-MS peaks .......................................................................................................................... 32
S2 The 23 most-predictive LC-MS peaks regarding Cohen’s κ for antimicrobial activity .......... 35
S3 The 27 most-predictive LC-MS peaks regarding Cohen’s κ for antimicrobial activity and non-cytotoxicity .................................................................................. 37
S4 The 12 out of 23 most-predictive individual LC-MS peaks involved in the most-predictive additive peak combinations regarding antimicrobial activity .......................................................................... 44
S5 The 23 most-predictive individual LC-MS peaks involved in the most-predictive alternative peak combinations regarding antimicrobial activity .......................................................... 45
S6 The 6 identified caffeoylquinic acids acting additive in LC-MS peak combinations .......................................................................................................................... 45
S7 The 15 identified regioisomeric chlorogenic acids acting as functional alternative structures in LC-MS peak combinations .................................................................................. 46
S8 Quercetin and the 16 identified quercetin-O-glycosides acting as functional alternative structures in LC-MS peak combinations .................................................................................. 46
S9 Sequence generation for phylogenetic analysis including DNA regions, primer sequences, and respective PCR protocols .................................................................................................. 47
**Figure S1 |** Principal component analysis (using Pareto scaling) of the phytochemical data for all 87 *Rhododendron* species. The scores of the principal components corresponding to the species are colored with respect to the cytotoxicity towards HaCaT cells (**A**) and IEC-6 cells (**B**), respectively. The item shape highlights the subgenus of each species (△:*Azaleastrum*, □:*Hymenanthes*, ○:*Pentanthera*, ◄:*Rhododendron*, ▽:*Tsutsusi*). The item size as well illustrates the cytotoxicity classification, toxic – large and turquoise, safe – small and gray.

**Figure S2 |** Loadings of the principal component analysis of the phytochemical data of all 87 *Rhododendron* species. Labeled are the 12 peaks (defined by mass-to-charge ratio and retention time) which are most related to the antimicrobial active *Rhododendron* samples based on the underlying separation.
Figure S3 | Chemical diversity with respect to the subgenus classification of the 87 *Rhododendron* species in terms of total number of identified polyphenolics (A) and of detected LC-MS peaks (B).
Intensity (m/z 345.08, rt 9.2 min)
Methyl gallate-O-hexoside, 9.2

Intensity (m/z 581.12, rt 43.7 min)
Myricetin-O-dipentoside (same side)

p-hydroxyphenethyl alcohol-1-O-
Myricetin-O-galloyl-rhamnoside, 39.9
Myricetin-7-O-galloyl-3-O-hexoside
p-Coumaric acid-O-hexoside, 27.3
p-Coumaric acid-O-hexoside, 24.9

Intensity (m/z 463.09, rt 34.1 min)
Myricetin-O-rhamnoside

Intensity (m/z 493.06, rt 37.8 min)
Myricetin-O-glucuronide

Intensity (m/z 615.10, rt 39.9 min)
Myricetin-O-galloyl-rhamnoside, 39.9
Methyl gallate-O-hexoside, 9.2
Methyl gallate-O-hexoside, 12.3
Myricetin-O-galloyl-hexoside
Myricetin-O-pentoside, 33.5
Myricetin-O-pentoside, 38.9
Procyanidin tetramer, 31.4
(Procyanidin Trimer C, 19.8
Procyanidin Trimer C, 16.9
Procyanidin Trimer C, 24.4
Procyanidin Trimer C, 5.9
(Procyanidin Trimer C, 631.10, rt 31.2 min)
487.15, rt 14.7 min)
615.10, rt 39.9 min)
457.14, rt 19.4 min)
345.08, rt 7.9 min)
479.08, rt 31.5 min)
449.07, rt 33.5 min)
576.13, rt 22.7 min)
576.13, rt 31.4 min)
493.06, rt 37.8 min)
449.07, rt 38.9 min)
576.13, rt 27.1 min)
345.08, rt 9.2 min)
865.19, rt 24.4 min)
865.20, rt 16.9 min)
317.03, rt 40.3 min)
865.20, rt 19.8 min)
(Procyanidin Trimer C, 345.08, rt 12.3 min)
(Myricetin-D-(6"-O-galloyl)-hexoside
p-Coumaric acid-O-hexoside, 24.9
p-Coumaric acid-O-hexoside-O-pentoside , 19.4
p-Coumaric acid-O-dihexoside, 21.3
p-Coumaric acid-O-dihexoside, 14.7
Procyanidin Trimer C, 24.4
Procyanidin tetramer, 31.4
Procyanidin tetramer, 22.7

9
Figure S4 | Box plots of the 282 identified polyphenolics with no significant difference in LC-MS intensity with respect to antimicrobial activity classification of all 87 *Rhododendron* species. Those 10 compounds showing significant differences are given in Fig 4. 17 of the 87 *Rhododendron* species are denoted as antimicrobial active (orange), i.e., the radius of the agar diffusion assay is \( \geq 0.6 \) cm, and, thus, 70 species are characterized as antimicrobial inactive (violet).
| m/z  | rt (min) | Peak Intensity |
|------|--------|----------------|
| 593.13 | 14.8 | (Epi)gallocatechin-(epi)catechin |
| 593.13 | 13.3 | (Epi)gallocatechin-(epi)catechin |
| 895.17 | 16.9 | (Epi)gallocatechin-(4,8'/2,7')-(epi)gallocatechin-(4',8")-(epi)catechin |
| 745.14 | 21.3 | (Epi)gallocatechin-(4,8')-3'-O-galloyl-(epi)catechin |
| 761.13 | 26.5 | (Epi)gallocatechin-(epi)gallocatechin, 26.5 |
| 609.13 | 7.0 | (Epi)gallocatechin-(epi)gallocatechin, 7.0 |
| 593.13 | 17.7 | (Epi)gallocatechin-(epi)catechin, 17.7 |
| 609.13 | 7.0 | (Epi)gallocatechin-(epi)gallocatechin, 7.0 |
| 593.13 | 10.3 | (Epi)gallocatechin-(epi)catechin, 10.3 |
| 593.13 | 11.2 | (Epi)gallocatechin-(epi)catechin, 11.2 |
| 593.13 | 12.5 | (Epi)gallocatechin-(epi)catechin, 12.5 |
| 593.13 | 13.5 | (Epi)gallocatechin-(epi)catechin, 13.5 |
| 593.13 | 14.8 | (Epi)gallocatechin-(epi)catechin, 14.8 |
| 593.13 | 16.6 | (Epi)gallocatechin-(epi)catechin, 16.6 |

- **(Epi)gallocatechin-(epi)catechin**
- **(Epi)gallocatechin-(epi)gallocatechin**
- **(Epi)gallocatechin-(epi)gallocatechin, 26.5**
- **(Epi)gallocatechin-(epi)catechin, 7.0**
- **(Epi)gallocatechin-(epi)catechin, 17.7**
- **(Epi)gallocatechin-(epi)catechin, 10.3**
- **(Epi)gallocatechin-(epi)catechin, 11.2**
- **(Epi)gallocatechin-(epi)catechin, 12.5**
- **(Epi)gallocatechin-(epi)catechin, 13.5**
- **(Epi)gallocatechin-(epi)catechin, 14.8**
- **(Epi)gallocatechin-(epi)catechin, 16.6**

**Intensities:**
- Safe
- Toxic
### Intensity

#### (m/z 625.14, rt 36.4 min)
- Myricetin-3-O-hexoside-7-O-rhamnoside

#### (m/z 345.08, rt 7.9 min)
- Methyl gallate-O-hexoside, 7.9

#### (m/z 507.14, rt 8.6 min)
- Methyl gallate-O-dihexoside (diff. sides)

#### (m/z 491.12, rt 43.2 min)
- Methyl gallate-O-coumaroyl-hexoside

#### (m/z 449.07, rt 38.9 min)
- Myricetin-O-pentoside, 38.9

#### (m/z 615.10, rt 39.9 min)
- Myricetin-O-galloyl-rhamnoside, 39.9

#### (m/z 581.12, rt 43.7 min)
- Myricetin-O-dipentoside (same side)

#### (m/z 565.16, rt 38.9 min)
- Methyl gallate-O-hexoside, 9.2

#### (m/z 433.11, rt 39.9 min)
- Methyl gallate-O-hexoside, 12.3

#### (m/z 507.14, rt 8.6 min)
- Myricetin-O-pentoside, 40.5

#### (m/z 449.07, rt 40.5 min)
- Myricetin-O-pentoside, 40.5

#### (m/z 493.06, rt 37.8 min)
- Procyanidin tetramer, 31.4

#### (m/z 487.15, rt 21.3 min)
- Myricetin-O-hexoside, 31.5

#### (m/z 467.12, rt 23.0 min)
- Myricetin-O-hexoside, 36.3

#### (m/z 583.11, rt 47.5 min)
- Procyanidin tetramer, 31.4

#### (m/z 631.09, rt 33.6 min)
- Myricetin-O-hexoside, 31.5

#### (m/z 625.14, rt 36.4 min)
- Myricetin-O-hexoside, 36.3

#### (m/z 581.12, rt 43.7 min)
- Myricetin-O-dipentoside (same side)

#### (m/z 449.07, rt 40.5 min)
- Myricetin-O-pentoside, 40.5

#### (m/z 576.13, rt 31.4 min)
- Procyanidin tetramer, 31.4

#### (m/z 463.09, rt 34.1 min)
- p-hydroxyphenethyl alcohol-1-O-β-D-(6”-O-galloyl)-hexoside

#### (m/z 487.15, rt 21.3 min)
- p-Coumaric acid-O-dihexoside, 21.3

#### (m/z 463.09, rt 34.1 min)
- Naringenin-O-hexoside-O-pentoside

#### (m/z 449.07, rt 38.9 min)
- Naringenin-O-hexoside, 39.9

#### (m/z 317.03, rt 40.3 min)
- Naringenin-O-hexoside, 36.3

#### (m/z 479.08, rt 36.3 min)
- Procyanidin tetramer, 31.4

#### (m/z 463.09, rt 34.1 min)
- Myricetin-O-rhamnoside

#### (m/z 433.11, rt 39.9 min)
- Myricetin-O-rhamnoside

#### (m/z 487.15, rt 21.3 min)
- p-Coumaric acid-O-dihexoside, 21.3

#### (m/z 576.13, rt 31.4 min)
- Procyanidin tetramer, 31.4

#### (m/z 463.09, rt 34.1 min)
- Naringenin-O-hexoside-O-pentoside

#### (m/z 449.07, rt 40.5 min)
- Naringenin-O-hexoside, 39.9

#### (m/z 576.13, rt 31.4 min)
- Procyanidin tetramer, 31.4

#### (m/z 463.09, rt 34.1 min)
- Naringenin-O-hexoside-O-pentoside
Figure S5 | Box plots of all 292 identified polyphenolics showing no significant difference in LC-MS intensity with respect to cytotoxicity classification towards HaCaT cells of all 87 *Rhododendron* species divided in cytotoxic (1 species) and non-cytotoxic (86 species, gray).
Figure S6 | Box plots of all 292 identified polyphenolics showing no significant difference in LC-MS intensity with respect to cytotoxicity classification towards IEC-6 cells of all 87 *Rhododendron* species divided in cytotoxic (22 species, green) and non-cytotoxic (65 species, gray).
Table S1 | Average and standard deviation of mass-to-charge (m/z) ratios and retention times (rt) for most-predictive LC-MS peaks regarding antimicrobial activity as well as cytotoxicity towards HaCaT and IEC-6 cells.

|                          | # peaks | m/z ratio       | rt [min]       |
|--------------------------|---------|-----------------|----------------|
| Antimicrobial activity   | 23      | 399.69 ± 98.97  | 63.0 ± 6.7     |
| Cytotoxicity towards HaCaT cells | 26      | 719.60 ± 202.91 | 45.5 ± 15.2    |
| Cytotoxicity towards IEC-6 cells | 13      | 557.92 ± 237.94 | 34.9 ± 14.9    |

Figure S7 | Distribution of the detected LC-MS peaks (grey) regarding m/z ratios (left) and retention times (right). In addition, the m/z ratio and retention time of the 23 most-predictive peaks as well as top 1%, 2% and 5% peaks (dark orange to bright orange) with respect to Cohen’s $\kappa$ correlation to antimicrobial activity across all 87 *Rhododendron* species are shown. The arrows denote the average m/z ratio and retention time, respectively.
Figure S8 | The 25 most- and least-predictive LC-MS peaks with respect to Cohen’s $\kappa$ correlation for cytotoxicity towards IEC-6 cells across all 87 *Rhododendron* species. A sample is denoted as cytotoxic towards IEC-6 cells (green) if the MTT assay is significantly dropped. A compound, defined by an m/z ratio and retention time tuple, is denoted as present in a sample (gray) if its intensity is $\geq 10000$. The * represents the significance of the p-values according to multiple testing correction by Benjamini-Hochberg (0.05).
Figure S9 | The 25 most- and least-predictive LC-MS peaks with respect to Cohen’s $\kappa$ correlation for cytotoxicity towards HaCaT cells across all 87 Rhododendron species. A sample is denoted as cytotoxic towards HaCaT cells (green) if the MTT assay is significantly dropped. A compound, defined by an m/z ratio and retention time tuple, is denoted as present in a sample (gray) if its intensity is $\geq$ 10000. The * represents the significance of the p-values according to multiple testing correction by Benjamini-Hochberg (0.05).
Table S2 | The 23 most-predictive LC-MS peaks with respect to Cohen’s $\kappa$ correlation for antimicrobial activity across all 87 *Rhododendron* species. The peaks are uniquely determined by m/z ratio and retention time (rt) and have attributed the rank and correlation coefficient, $\kappa$, for antimicrobial activity, cytotoxicity towards HaCaT and IEC-6 cells. The highlighted rows depict the seven most-predictive peaks for antimicrobial active but non-cytotoxic compounds.

| Peak  | Antimicrobial activity | Cytotoxicity HaCaT | Cytotoxicity IEC-6 |
|-------|------------------------|-------------------|-------------------|
| m/z ratio | rt [min] | rank | $\kappa$ | rank | $\kappa$ | rank | $\kappa$ |
| 333.19 | 64.8 | 1 | 0.7704 | 31117 | −0.0220 | 7956 | 0.0821 |
| 455.20 | 66.0 | 1 | 0.7704 | 31117 | −0.0220 | 7956 | 0.0821 |
| 523.19 | 67.7 | 3 | 0.7601 | 34746 | −0.0223 | 25643 | −0.0036 |
| 333.22 | 67.3 | 4 | 0.7592 | 29299 | −0.0218 | 3987 | 0.1205 |
| 257.16 | 66.5 | 5 | 0.7468 | 27352 | −0.0215 | 1812 | 0.1617 |
| 257.20 | 66.5 | 5 | 0.7468 | 27352 | −0.0215 | 1812 | 0.1617 |
| 387.22 | 56.8 | 7 | 0.7383 | 31907 | −0.0221 | 3205 | 0.1307 |
| 334.17 | 68.3 | 8 | 0.7258 | 30246 | −0.0219 | 5717 | 0.1010 |
| 304.16 | 67.3 | 9 | 0.7200 | 34075 | −0.0223 | 21733 | 0.0124 |
| 375.22 | 59.5 | 10 | 0.7121 | 28363 | −0.0217 | 9472 | 0.0691 |
| 387.16 | 57.0 | 11 | 0.7076 | 32668 | −0.0222 | 4652 | 0.1119 |
| 499.17 | 68.4 | 11 | 0.7076 | 32668 | −0.0222 | 42896 | −0.0854 |
| 285.09 | 50.5 | 13 | 0.6969 | 26230 | −0.0213 | 4851 | 0.1092 |
| 333.16 | 66.4 | 13 | 0.6969 | 26230 | −0.0213 | 1103 | 0.1834 |
| 333.20 | 66.2 | 13 | 0.6969 | 26230 | −0.0213 | 1103 | 0.1834 |
| 384.95 | 56.9 | 13 | 0.6969 | 26230 | −0.0213 | 4851 | 0.1092 |
| 384.99 | 68.3 | 13 | 0.6969 | 26230 | −0.0213 | 4851 | 0.1092 |
| 559.15 | 65.0 | 13 | 0.6969 | 26230 | −0.0213 | 1103 | 0.1834 |
| 607.16 | 66.2 | 13 | 0.6969 | 26230 | −0.0213 | 4851 | 0.1092 |
| 387.23 | 55.5 | 21 | 0.6939 | 31117 | −0.0220 | 2267 | 0.1501 |
| 563.17 | 41.9 | 21 | 0.6939 | 31117 | −0.0220 | 21419 | 0.0141 |
| 523.21 | 68.3 | 23 | 0.6916 | 1777 | 0.0750 | 46192 | −0.1291 |
Figure S10 | The 27 most-predictive LC-MS peaks with a strong combined relation signal (≥ 0.68) for antimicrobial activity and non-cytotoxicity towards both cell lines across all 87 Rhododendron species (see Material and Methods). A sample is denoted as antimicrobial active (orange, left lane) if the radius of the agar diffusion assay is ≥ 0.6 cm and as cytotoxic towards HaCaT or IEC-6 cells (green, right lane) if the MTT assay is significantly dropped. A compound, defined by an m/z ratio and retention time tuple, is denoted as present in a sample (gray) if its intensity is ≥ 10,000. The seven most-predictive peaks regarding antimicrobial activity and non-cytotoxicity at once are highlighted in bright orange.
Table S3 | 27 most-predictive LC-MS peaks with respect to the combined relation signal for antimicrobial activity and non-cytotoxicity towards both cell lines across all 87 Rhododendron species. The peaks are uniquely determined by m/z ratio and retention time (rt) and have attributed the combined correlation coefficient, $\kappa$, and the individual ranks for antimicrobial activity and the cytotoxicity towards HaCaT and IEC-6 cells. The highlighted rows depict the seven most-predictive peaks for antimicrobial active but non-cytotoxic compounds.

| Peak m/z ratio | Combined $\kappa$ | Individual ranks | AM | HaCaT | IEC-6 |
|---------------|------------------|-----------------|---|--------|-------|
| 499.17        | 0.8152           | 11              | 32668 | 42896  |
| 523.19        | 0.7861           | 3               | 34746 | 25643  |
| 417.17        | 0.7556           | 33              | 31907 | 40552  |
| 523.21        | 0.7457           | 23              | 1777  | 46192  |
| 365.19        | 0.7362           | 27              | 33392 | 33332  |
| 523.19        | 0.7362           | 27              | 33392 | 33332  |
| 304.16        | 0.7299           | 9               | 34075 | 21733  |
| 359.16        | 0.7210           | 24              | 29299 | 28761  |
| 404.22        | 0.7210           | 24              | 29299 | 28761  |
| 433.22        | 0.7111           | 560             | 34075 | 47583  |
| 333.19        | 0.7104           | 1               | 31117 | 7956   |
| 455.20        | 0.7104           | 1               | 31117 | 7956   |
| 358.21        | 0.7067           | 46              | 30246 | 33644  |
| 563.17        | 0.7019           | 21              | 31117 | 21419  |
| 451.17        | 0.7006           | 185             | 35392 | 42037  |
| 517.11        | 0.6949           | 64              | 36005 | 33108  |
| 384.25        | 0.6949           | 42              | 39555 | 28111  |
| 357.22        | 0.6934           | 104             | 31117 | 37580  |
| 373.20        | 0.6934           | 104             | 31117 | 37580  |
| 829.40        | 0.6934           | 104             | 31117 | 37580  |
| 427.23        | 0.6900           | 265             | 36005 | 43695  |
| 359.22        | 0.6893           | 429             | 22316 | 46572  |
| 375.16        | 0.6887           | 33              | 31907 | 25424  |
| 503.10        | 0.6882           | 204             | 23675 | 41567  |
| 242.95        | 0.6808           | 192             | 31907 | 40552  |
| 242.99        | 0.6808           | 192             | 31907 | 40552  |
| 433.22        | 0.6808           | 192             | 31907 | 40552  |
Figure S11 | The 50 most-predictive peak combinations with additive effects with respect to Cohen’s $\kappa$ correlation for antimicrobial activity across all 87 Rhododendron species. A sample is denoted as antimicrobial active (orange) if the radius of the agar diffusion assay is $\geq 0.6$ cm. A compound, defined by an m/z ratio and retention time tuple, is denoted as present in a sample (gray) if its intensity is $\geq 10000$. The * represents the significance of the p-values according to multiple testing correction by Benjamini-Hochberg (0.05). The combinations comprising one of the seven most-predictive peaks regarding antimicrobial activity and non-cytotoxicity at once are highlighted in bright orange.
Figure S12 | The 50 most-predictive peak combinations with alternative effects with respect to Cohen’s $\kappa$ correlation for antimicrobial activity across all 87 *Rhododendron* species.

A sample is denoted as antimicrobial active (orange) if the radius of the agar diffusion assay is $\geq 0.6$ cm. A compound, defined by an $m/z$ ratio and retention time tuple, is denoted as present in a sample (gray) if its intensity is $\geq 10000$. The * represents the significance of the p-values according to multiple testing correction by Benjamini-Hochberg (0.05). The combinations comprising one of the seven most-predictive peaks regarding antimicrobial activity and non-cytotoxicity at once are highlighted in bright orange.
Figure S13 | Heatmap of the top 250 LC-MS peaks involved in the most-predictive alternative peak combinations regarding Cohen’s $\kappa$ correlation for antimicrobial activity across all 87 Rhododendron species (main panel). The upper right panel provides the overview of all 5,414 peaks included in the outperforming alternative peak combinations, namely $\kappa \geq 0.77$. The combinations highlighted in red involve at least one of the 23 most-predictive peaks regarding the individual peak analysis, $\kappa \geq 0.68$. The corresponding individual peak correlation coefficients are depicted in the thinner horizontal and vertical panels. Orange labels emphasize the two out of seven most-predictive peaks regarding antimicrobial activity and non-cytotoxicity.
Figure S14 | The 50 most-predictive LC-MS peak combinations with additive effects with respect to Cohen’s $\kappa$ correlation for cytotoxicity towards IEC-6 cells across all 87 *Rhododendron* species. A sample is denoted as cytotoxic towards IEC-6 cells (green) if the MTT assay is significantly dropped. A compound, defined by an $m/z$ ratio and retention time tuple, is denoted as present in a sample (gray) if its intensity is $\geq 10000$. The * represents the significance of the p-values according to multiple correction by Benjamini-Hochberg (0.05).
Figure S15 | The 50 most-predictive LC-MS peak combinations with alternative effects with respect to Cohen’s κ correlation for cytotoxicity towards IEC-6 cells across all 87 Rhododendron species. A sample is denoted as cytotoxic towards IEC-6 cells (green) if the MTT assay is significantly dropped. A compound, defined by an m/z ratio and retention time tuple, is denoted as present in a sample (gray) if its intensity is ≥ 10000. The * represents the significance of the p-values according to multiple testing correction by Benjamini-Hochberg (0.05).
Figure S16 | Heatmap of the top 250 LC-MS peaks involved in the most-predictive (A) additive and (B) alternative peak combinations regarding Cohen’s $\kappa$ correlation for cytotoxicity towards IEC-6 cells across all 87 Rhododendron species (main panel). The respective upper right panel provides the overview of all 7,763 and 8,665 peaks included in the outperforming additive and alternative peak combinations, namely $\kappa \geq 0.38$. The combinations highlighted in green involve at least one of the 13 most-predictive peaks regarding the individual peak analysis, $\kappa \geq 0.35$. The corresponding individual peak correlation coefficients are depicted in the thinner horizontal and vertical panels.
Table S4 | The 12 out of 23 most-predictive individual LC-MS peaks involved in the most-predictive additive peak combinations regarding antimicrobial activity across all 87 *Rhododendron* species. The peaks are uniquely determined by $m/z$ ratio and retention time (rt) and have attributed the respective maximum Cohen’s $\kappa$ correlation coefficient with respect to ‘AND’ operation, $\kappa$, and the (minimum) ranks for antimicrobial activity of ‘AND’ combination and individual (single) occurrence. The highlighted rows depict the seven most-predictive peaks for antimicrobial active but non-cytotoxic compounds.

| $m/z$ ratio | rt [min] | $\kappa$ | min $rk_{AND}$ | $rk_{single}$ |
|------------|----------|---------|----------------|--------------|
| 523.19     | 67.7     | 0.8538  | 9              | 3            |
| 304.16     | 67.3     | 0.8470  | 16             | 9            |
| 523.21     | 68.3     | 0.8470  | 16             | 23           |
| 333.19     | 64.8     | 0.8395  | 32             | 1            |
| 455.20     | 66.0     | 0.8395  | 32             | 1            |
| 387.16     | 57.0     | 0.8395  | 32             | 11           |
| 387.22     | 56.8     | 0.8395  | 32             | 7            |
| 499.17     | 68.4     | 0.8041  | 159            | 11           |
| 333.22     | 67.3     | 0.7943  | 317            | 4            |
| 334.17     | 68.3     | 0.7943  | 317            | 8            |
| 563.17     | 41.9     | 0.7943  | 317            | 21           |
| 387.23     | 55.5     | 0.7943  | 317            | 21           |
Table S5 | The 23 most-predictive individual LC-MS peaks involved in the most-predictive alternative peak combinations regarding antimicrobial activity across all 87 Rhododendron species. The peaks are uniquely determined by m/z ratio and retention time (rt) and have attributed the respective maximum Cohen’s κ correlation coefficient with respect to ‘OR’ operation, κ, and the (minimum) ranks for antimicrobial activity of ‘OR’ combination and individual (single) occurrence. The highlighted rows depict the seven most-predictive peaks for antimicrobial active but non-cytotoxic compounds.

| m/z ratio | rt [min] | κ  | min rk('OR') | rk('single') |
|-----------|----------|----|--------------|--------------|
| 257.16    | 66.5     | 0.9626 | 1 | 5          |
| 257.20    | 66.5     | 0.9626 | 1 | 5          |
| 285.09    | 50.5     | 0.9300 | 3 | 13         |
| 333.22    | 67.3     | 0.9300 | 3 | 4          |
| 384.95    | 56.9     | 0.9300 | 3 | 13         |
| 384.99    | 66.6     | 0.9300 | 3 | 13         |
| 384.99    | 68.2     | 0.9300 | 3 | 13         |
| 455.20    | 66.0     | 0.9300 | 3 | 1          |
| 607.16    | 66.2     | 0.9235 | 64 | 13        |
| 559.15    | 65.0     | 0.9235 | 64 | 13        |
| 333.16    | 66.4     | 0.9235 | 64 | 13        |
| 333.20    | 66.2     | 0.9235 | 64 | 13        |
| 333.19    | 64.8     | 0.8972 | 160 | 1         |
| 334.17    | 68.3     | 0.8972 | 160 | 8          |
| 387.22    | 56.8     | 0.8972 | 160 | 7          |
| 375.22    | 59.5     | 0.8927 | 224 | 10         |
| 499.17    | 68.4     | 0.8657 | 4025 | 11        |
| 387.16    | 57.0     | 0.8657 | 4025 | 11        |
| 387.23    | 55.5     | 0.8657 | 4025 | 21        |
| 563.17    | 41.9     | 0.8657 | 4025 | 21        |
| 523.19    | 67.7     | 0.8355 | 15781 | 3           |
| 523.21    | 58.3     | 0.8065 | 32005 | 23        |
| 304.16    | 67.3     | 0.8065 | 32005 | 9          |

Table S6 | The 6 identified caffeoylquinic acids (CQA) acting additive in LC-MS peak combinations (AND).

| CQA m/z ratio | Peak in combination m/z ratio | Cohen’s κ |
|--------------|-------------------------------|-----------|
| m/z ratio   | rt [min]                      |          |
| 353.08      | 24.6                          | -0.0640  |
| 353.09      | 13.9                          | 0.0175   |
| 353.09      | 15.4                          | 0.0106   |
| 353.09      | 19.0                          | -0.0497  |
| 353.09      | 21.6                          | -0.0086  |
| 353.09      | 24.2                          | 0.0235   |
Table S7 | The 15 identified regioisomeric chlorogenic acids (CGA) acting as functional alternative structures in LC-MS peak combinations (OR).

| CGA m/z ratio rt [min] | CGA in combination m/z ratio rt [min] | Cohen’s κ | κ'CGA' | κ'CGA | κ'OR' |
|------------------------|----------------------------------------|-----------|--------|--------|--------|
| 337.09 19.0            | 367.10 28.7                            | −0.0200   | 0.0433 | 0.0386 |
| 337.09 20.3            | 353.09 13.9                            | −0.0783   | 0.0175 | 0.0110 |
| 337.09 21.9            | 367.10 32.7                            | 0.0372    | −0.1932| −0.0760|
| 337.09 25.7            | 353.09 21.6                            | −0.0086   | −0.0086| 0.0173 |
| 337.09 27.6            | 353.09 13.9                            | −0.0792   | 0.0175 | 0.0112 |
| 337.09 29.6            | 353.09 24.2                            | −0.1233   | 0.0235 | 0.0052 |
| 353.08 24.6            | 337.09 25.7                            | −0.0640   | −0.0086| −0.0007|
| 353.09 13.9            | 337.09 19.0                            | −0.0175   | 0.0200 | 0.0374 |
| 353.09 15.4            | 337.09 25.7                            | 0.0106    | −0.0086| 0.0052 |
| 353.09 19.0            | 337.09 25.7                            | −0.0497   | −0.0086| −0.0007|
| 353.09 21.6            | 337.09 25.7                            | −0.0086   | −0.0086| 0.0173 |
| 353.09 24.2            | 367.10 22.5                            | 0.0235    | −0.0097| 0.0173 |
| 367.10 22.5            | 337.09 19.0                            | 0.0177    | −0.0097| −0.0200|
| 367.10 28.7            | 337.09 19.0                            | 0.0386    | 0.0433 | −0.0200|
| 367.10 32.7            | 353.09 24.2                            | 0.0052    | −0.1932| 0.0235 |

Table S8 | Quercetin (Q) and the 16 identified quercetin-O-glycosides (QG) acting as functional alternative structures in LC-MS peak combinations (OR).

| Q(G) m/z ratio rt [min] | QG in combination m/z ratio rt [min] | κ’Q(G) | κ’CG | κ’OR |
|------------------------|-------------------------------------|--------|------|------|
| 301.05 45.5            | 433.08 42.5                         | −0.1105| 0.0056| 0.0056|
| 433.08 35.1            | 433.08 42.5                         | −0.0249| 0.0056| 0.0056|
| 433.08 38.3            | 433.08 42.5                         | −0.0177| 0.0056| 0.0056|
| 433.08 40.8            | 609.15 31.7                         | −0.0121| 0.0089| 0.0114|
| 433.08 42.5            | 301.05 45.5                         | 0.0056 | −0.1105| 0.0056|
| 447.10 40.0            | 433.08 42.5                         | −0.0903| 0.0056| 0.0056|
| 463.09 35.5            | 433.08 42.5                         | −0.0903| 0.0056| 0.0056|
| 463.09 36.4            | 433.08 42.5                         | −0.0903| 0.0056| 0.0056|
| 463.09 41.9            | 433.08 42.5                         | −0.1154| 0.0056| 0.0|
| 609.12 47.5            | 433.08 42.5                         | −0.0601| 0.0056| 0.0056|
| 609.13 48.8            | 609.15 36.0                         | 0.0261 | −0.1333| −0.1008|
| 609.13 50.6            | 433.08 42.5                         | −0.1266| 0.0056| 0.0056|
| 609.15 26.5            | 433.08 42.5                         | −0.3533| 0.0056| 0.0056|
| 609.15 30.1            | 433.08 40.8                         | 0.0110 | −0.0121| 0.0056|
| 609.15 31.4            | 609.15 36.0                         | 0.0235 | −0.1333| 0.0172|
| 609.15 31.7            | 433.08 40.8                         | 0.0089 | −0.0121| 0.0114|
| 609.15 36.0            | 609.15 31.4                         | −0.1333| 0.0235| 0.0172|
## Table S9 | Sequence generation for phylogenetic analysis including DNA regions, primer sequences, and respective PCR protocols.

| Region | Primer sequence (starting from 5' end) | Initialization | Amplification cycles | Final elongation |
|--------|--------------------------------------|----------------|----------------------|-----------------|
| **matK** |  |  |  |  |
| trnK707F | ACT GTA TCG CAC TAT GTA TCA | Milne et al. (2010) | 94°C, 2min | 94°C, 30s; 61.5°C, 1min; 72°C, 1min |
| trnK2R | AAT TAG TCG GAT GGA GGA G | Johnson and Soltis (1995) | 30× | 72°C, 7min |
| MK1447F | GCC TCA ATA TCT TCT GAA ACC TT | Milne et al. (2010) | 94°C, 2min | 94°C, 30s; 61.5°C, 1min; 72°C, 1min |
| MK1645R | AGC CAA AAT GGC TTT TCC TT | Milne et al. (2010) | 94°C, 2min | 94°C, 30s; 61.5°C, 1min; 72°C, 1min |
| MK1538F | TAT GGG TGT TTA AAG AGC | Milne et al. (2010) | 94°C, 2min | 94°C, 30s; 61.5°C, 1min; 72°C, 1min |
| MK1785R | TCT ATC ATT TGA CTC GGT ACC A | Milne et al. (2010) | 94°C, 2min | 94°C, 30s; 61.5°C, 1min; 72°C, 1min |
| **trnL-F** |  |  |  |  |
| trnL-5(UAA)F | CGA AAT CGG TAG ACG CTA CG | Taberlet et al. (1991) | 94°C, 1min | 94°C, 30s; 54°C, 30s; 72°C, 1min |
| trnF(GAA) | ATT TGA ACT GGT GAC ACG AG | Taberlet et al. (1991) | 94°C, 1min | 94°C, 30s; 54°C, 30s; 72°C, 1min |
| **ITS** |  |  |  |  |
| ITS-4 | TCC TCC GCT TAT TGA TAT GC | White et al. (1990) | 94°C, 1min | 94°C, 18s; 54°C, 30s; 72°C, 1min |
| ITS-A | GGA AGG AGA AGT GGT AAC AAG G | Blattner (1999) | 94°C, 1min | 94°C, 18s; 54°C, 30s; 72°C, 1min |

*a* – Internal sequencing primer,  
*b* – modified internal sequencing primer
References

Blattner, F. R. (1999). Direct amplification of the entire ITS region from poorly preserved plant material using recombinant PCR. *BioTechniques* 27, 1180–1186

Johnson, L. A. and Soltis, D. E. (1995). Phylogenetic Inference in *Saxifragaceae Sensu Stricto* and *Gilia (Polemoniaceae)* Using *matK* Sequences. *Ann. Mo. Bot. Gard.* 82, 149–175. doi:10.2307/2399875

Milne, R. I., Davies, C., Prickett, R., Inns, L. H., and Chamberlain, D. F. (2010). Phylogeny of *Rhododendron* subgenus *Hymenanthes* based on chloroplast DNA markers: between-lineage hybridisation during adaptive radiation? *Plant. Syst. Evol.* 285, 233–244. doi:10.1007/s00606-010-0269-2

Taberlet, P., Gielly, L., Pautou, G., and Bouvet, J. (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Mol. Biol.* 17, 1105–1109. doi:10.1007/bf00037152

White, T., Bruns, T., Lee, S., and Taylor, J. (1990). Amplification and Direct Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics. In *PCR protocols: A Guide to Methods and Applications*, eds. M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White (Academic Press). 315–322. doi:10.1016/b978-0-12-372180-8.50042-1