Bridging *In-Silico* and Experimental: Chemoinformatics Analysis of Mass Spectrometry-Based Metabolomics Study of Soybean

Divya Karade1,2, Swapnil Mundhe3,Narendra Kadoo2,3, Milind Ratnaparkhe4, Renu Vyas5, Muthukumarasamy Karthikeyan\*,1,2

1. Chemical Engineering and Process Development (CEPD) Division, CSIR-National Chemical Laboratory, Pune – 411008, India
2. Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India
3. Biochemical Sciences Division, CSIR-National Chemical Laboratory, Pune 411008, India
4. Indian Institute of Soybean Research, Indian Council of Agricultural Research (ICAR), Indore 452001, India
5. MIT School of Bioengineering, Sciences & Research, ADT University, Pune 412201, India

**Supplementary Tables:**

**Supplementary** **Table 1**: Morpho-physiological characteristics of the four soybean varieties used for the study

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sr No** | **Character** | **NRC119** | **JS335** | **JS7105** | **JS9305** |
| 1 | Growth type | Semi-determinate | Semi-determinate | Determinate | Semi-determinate |
| 2 | Flower color | White | Purple | Purple | Purple |
| 3 | Seed size | Bold | Small | Small | Small |
| 4 | Seed color | Yellow | Yellow | Yellow | Yellow |
| 5 | Hilum | Light black to Black | Black | Light black to Black | Black |
| 6 | Seed longevity | -N.A.- | High | Poor | High |
| 7 | Resistance to lodging | -N.A.- | No | Yes | No |
| 8 | Resistance to shattering | Yes | Yes | No | Yes |
| 9 | Tolerant to stem fly | Yes | Yes | Yes | Yes |
| 10 | Resistance to bacterial pustule | Yes | Yes | Yes | Yes |
| 11 | Resistant to YMV | Yes | No | No | No |
| 12 | Resistance to Myrothecium Leaf Spot | Highly resistant | Moderately resistant | Moderately susceptible | Highly resistant |

-N.A.-: Data not available

**Supplementary** **Table 2**: Small organic molecules (n=20) validated through tandem mass spectrometry by performing *in-silico* fragmentation approach (CFM-ID) using putatively annotated and identified molecules in UHPLC-MS experiments with up to 2 ppm error and highest probability score, i.e., 1 for soybean samples in positive and negative polar modes

| **Sr. no.** | **m/ z** | **Retention Time**  **(min.)** | **Molecular Formula** | **Adduct** | **ppm error** | **Probability score** | **Molecule name** | **Structure** | **Sample Tissue** | **Polarity** | **Reported/ Unreported** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 416.1499 | 7.45 | C22H24O8 | [M]-  [M+Cl]-  [M+HCOOK]- | 0.32 | 1 | 1-Acetoxypinoresinol |  | Leaf, seed | Negative | Non-reported for soybean  (Reported in *Olea europaea Linne*) - [Ref. 1](https://pubmed.ncbi.nlm.nih.gov/12729011/) |
| 2 | 430.2102 | 11.79 | C23H30N2O6 | [M]-  [M+HCOOH]- | 0.00021 | 1 | Cinegalline |  | Leaf, seed | Positive | Non-reported for soybean  (Reported in *Genista cinerea*) - [Ref. 2](https://pubmed.ncbi.nlm.nih.gov/4971408/) |
| 3 | 448.19 | 11.97 | C27H28O6 | [M]-  [M-H]- | 2.72 | 1 | Lonchocarpenin |  | Leaf, seed | Positive, Negative | Non-reported for soybean  (Reported in *Millettia richardiana*) - [Ref. 3](https://pubmed.ncbi.nlm.nih.gov/24079177/) |
| 4 | 570.1944 | 9.41, 9.45 | C26H34O14 | [M]-  [M+Cl]- | - | 1 | Decuroside III |  | Leaf, seed | Negative | Non-reported for soybean  (Reported in *Peucedanum decursivum Maxim*.) - [Ref. 4](https://www.thieme-connect.com/products/ejournals/abstract/10.1055/s-2007-969099) |
| 5 | 596.1386 | 7.94, 8.35 | C26H28O16 | [M]-  [M+Na+HCOONa]- | 0.19 | 1 | Quercetin 3-O-[beta-D-xylosyl-(1->2)-beta-D-glucoside] |  | Seed | Negative | Non-reported for soybean  (Reported in *Eucommia ulmoides*) - [Ref. 5](https://pubmed.ncbi.nlm.nih.gov/25039180/) |
| 6 | 772.2056 | 7.39, 7.54 | C33H40O21 | [M]- | - | 1 | Kaempferol 3-sophorotrioside |  | Leaf | Positive,Negative | Non-reported for soybean  (Reported in *Ficus carica* L.) - [Ref. 6](https://link.springer.com/article/10.1007/s12298-018-0550-3) |
| 7 | 824.4615 | 7.95, 8.09 | C43H68O15 | [M]- | - | 1 | Yiamoloside B |  | Leaf, seed | Positive, Negative | Non-reported for soybean  (Reported in *Phytolacca octandra*) - [Ref. 7](https://www.sciencedirect.com/science/article/abs/pii/0031942281800672?via%3Dihub) |
| 8 | 926.52 | 12.44, 12.45 | C48H78O17 | [M]-  [M-H]- | 0.05 | 1 | Saikosaponin BK1 |  | Leaf | Negative | Non-reported for soybean  (Reported in *Bupleurum kunmingense* Y.) - [Ref. 8](https://www.tandfonline.com/doi/abs/10.1080/00021369.1987.10868230) |
| 9 | 85.09 | 2.04, 2.08 | C5H11N | [M]+  [M+K+HCOOH]+  [M+Na+HCOOH]+  [M+HCOONa]+  [M+H+HCOOH]+ | 0.70 | 1 | Piperidine |  | Leaf, seed | Positive | Reported in soybean - [Ref. 9](https://www.tandfonline.com/doi/abs/10.1080/00021369.1966.10858696) |
| 10 | 146.0368 | 6.92, 6.94 | C9H6O2 | [M]+  [M+H]+ | 0.03 | 1 | Coumarin |  | Leaf, seed | Positive | Reported in soybean - [Ref. 10](https://onlinelibrary.wiley.com/doi/full/10.1111/tpj.14426) |
| 11 | 194.0942 | 13.22 | C11H14O3 | [M]+ | - | 1 | Zingerone |  | Leaf, seed | Positive, Negative | Non-reported for soybean  (Reported in Ginger) - [Ref. 11](https://www.tandfonline.com/doi/abs/10.3109/00498257609151654) |
| 12 | 235.1205 | 2.59, 2.63 | C13H17NO3 | [M+K]+  [M+Na]+ | 0.91 | 1 | Lophophorine |  | Leaf, seed | Positive | Non-reported for soybean  (Reported in Lophophora) - [Ref. 12](https://science.sciencemag.org/content/199/4336/1437.abstract?casa_token=PiLNnSE2sKkAAAAA:E1BLGXX_qs3YFdBeH2hY-_oqQfY5LFdtTaiYCspxTsx9pyzzb61pua6aQKG60ifnb4b2msiozagmOw) |
| 13 | 238.0738 | 7.10, 7.11 | C14H10N2O2 | [M]- | - | 1 | Halfordinol |  | Leaf, seed | Negative | Non-reported for soybean  (Reported in *Aeglopsis Chevalieri* Swing.) - [Ref. 13](https://pubs.acs.org/doi/pdf/10.1021/jo01273a076?casa_token=wnNx9SZrT1wAAAAA%3AnxIGPNQbshtz_heAzqPmxaGSmfACJvHQ7b0acfYmEAJGmdD-bL2iFf0hswPOMtz10F-6fNYZo6whYSI&) |
| 14 | 268.0361 | 3.15 | C15H8O5 | [M+2Na]2+  [M+NH3]+ | 0.11 | 1 | Coumestrol |  | Leaf, seed | Negative | Reported in soybean- [Ref. 14](https://www.sciencedirect.com/science/article/abs/pii/S0021967300004441) |
| 15 | 273.2663 | 12.73 | C16H35NO2 | [M]+  [M+H]+ | 0.73 | 1 | Hexadecasphinganine |  | Leaf, seed | Positive | Non-reported for soybean  (Reported in *Manduca sexta*) - PMID: |
| 16 | 287.0548 | 6.51, 6.52 | C15H10O6 | [M+H]+ | 0.76 | - | Aurantinidin |  | Leaf, seed | Positive | Non-reported for soybean  (Reported in ) - [Ref. 15](http://192.248.16.117:8080/research/handle/70130/103) |
| 17 | 301.2977 | 13.53 | C18H39NO2 | [M]+  [M+H]+ | 1.70 | 1 | Sphinganine |  | Leaf, seed | Positive | Reported in soybean - [Ref. 16](https://aocs.onlinelibrary.wiley.com/doi/abs/10.1007/BF02535357) |
| 18 | 303.05 | 7.93, 7.95 | C00389 | [M+H]+ | 0.79 | - | Quercetin |  | Leaf | Positive | Reported in soybean - [Ref. 17](https://link.springer.com/article/10.1007%2Fs00394-019-01992-9) |
| 19 | 330.1211 | 8.14, 8.16 | C17H18N2O5 | [M]+  [2M+Na]+  [M+Na]+ | 0.17 | 1 | Miraxanthin-III |  | Leaf, seed | Positive, Negative | Non-reported for soybean  (Reported in *Beta vulgaris* L.) - [Ref. 18](https://pubs.acs.org/doi/abs/10.1021/jf035491w) |
| 20 | 162.0316 | 5.42 | C9H6O3 | [M]+  [M+H]+ | 0.83 | 1 | Umbelliferone |  | Leaf, seed | Positive | Reported in soybean - [Ref. 19](https://link.springer.com/article/10.1007/s11104-009-0127-6) |

**Supplementary** **Table 3**: Common scaffolds identified in soybean small molecules, drug molecules, and scaffold merged network (Fig. 6) with their therapeutic category information (n=10)

| **Sr. No.** | **Chemical Class**  **(a)** | **Scaffold**  **(b)** | **Drug**  **(c)** |
| --- | --- | --- | --- |
| 1. | 3-hydroxybenzoate    Sc ID: 38 |  | Phenylephrine: alpha-1 adrenergic receptor agonist    Sc ID: 05  DB ID: DB00179 |
| 2. | Sinapaldehyde    Sc ID: 15 |  | Chloroxylenol: antiseptic and disinfectant agent    Sc ID: 31  DB ID: DB11121 |
| 3. | Myo-inositol    Sc ID: 01 |  | Lindane: Antiscabies agent    Sc ID: 23  DB ID: DB00431 |
| 4. | Quinolinate    Sc ID: 45 |  | Nicorandil: An orally efficacious vasodilatory drug and antianginal    Sc ID: 121  DB ID: DB00198 |
| 5. | Isomethyleugenol    Sc ID: 23 |  | Midodrine: Vasoconstrictor agent    Sc ID: 02  DB ID: DB00211 |
| 6. | Chavicol    Sc ID: 25 |  | Metoprolol: Antihypertensive agent, anti-arrhythmia agent    Sc ID: 26  DB ID: DB00264 |
| 7. | (Indol-3-yl)acetate    Sc ID: 48 |  | L-tryptophan: Anti-depressive agent, dietary supplement    Sc ID: 07  DB ID: DB00150 |
| 8. | Trans-cinnamate    Sc ID: 42 |  | Phentermine: Sympathomimetic amine anorectic agent    Sc ID: 6  DB ID: DB00191 |
| 9. | Berberine    Sc ID: 36 |  | Berberine: antidiarrheal, antifungal    Sc ID: 165  DB ID: DB04115 |
| 10. | Papaverine    Sc ID: 34 |  | Papaverine: muscle relaxant    Sc ID: 129  DB ID: DB01113 |

(Sc: Scaffold)

**Supplementary** **Table 4**: Virtual library novel molecules with their molecular weight, TPC and PDL, PLL scores (n= 10)

| **Sr. No.** | **VL molecule structure** | **Molecular weight** | **PDL** | **PLL** | **T** | **P** | **C** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 |  | 326.559 | 1.278 | 2.03 | 1 | 21 | 6 |
| 2 |  | 326.559 | 1.278 | 2.03 | 1 | 20 | 6 |
| 3 |  | 326.559 | 1.278 | 2.03 | 1 | 20 | 6 |
| 4 |  | 434.739 | 0.845 | 2.775 | 1 | 26 | 8 |
| 5 |  | 434.739 | 0.512 | 2.593 | 1 | 26 | 8 |
| 6 |  | 434.739 | 0.345 | 1.799 | 1 | 26 | 8 |
| 7 |  | 392.66 | 0.086 | 1.115 | 1 | 26 | 8 |
| 8 |  | 392.66 | 0.086 | 1.115 | 1 | 26 | 8 |
| 9 |  | 350.58 | -0.007 | 0.622 | 1 | 22 | 8 |
| 10 |  | 392.66 | 0.086 | 1.536 | 1 | 26 | 8 |

**Notes**: PDL: Progressive Drug-Like, PDL: Progressive Lead-Like, T: Toxicophore, P: Pharmacophore, C: Chemophore