The role of rhizosphere signalling in the plant-cyst nematode interaction

Vlaar, L.E.

Publication date
2022

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 4

A combination of metabolomics and machine learning results in the identification of a new cyst nematode hatching factor

Lieke E. Vlaar¹, Benjamin Thiombiano¹, Davar Abedini¹, Mario Schilder¹, Yuting Yang¹ and Lemeng Dong¹

¹ Plant Hormone Biology Group, Swammerdam Institute for Life Sciences (SILS), University of Amsterdam, Amsterdam, The Netherlands

Manuscript submitted to Metabolites
Abstract
Potato Cyst Nematodes (PCNs) are an economically important pest for potato growers. A crucial event in the life cycle of the nematode is hatching, after which the juvenile will move towards the host root and infect it. Hatching of PCN is induced by known and unknown compounds in the root exudate of host plant species, called hatching factors (HF, induce hatching independently), such as solanoeclepin A (solA), or hatching stimulants (HS, enhance hatching activity of HF). In this study, we used a new approach to identify new HF and HS for PCN in potato. Hereto, root exudates of a series of different potato cultivars were analysed for their PCN hatch inducing activity and their solA content. The exudates were also analysed using untargeted metabolomics, and subsequently the data was integrated using machine learning, specifically Random Forest feature selection, and Pearson’s correlation testing. As expected, solA highly correlates with hatching. Furthermore, this resulted in the discovery of a number of metabolite features present in the root exudate that correlate with hatching and solA content, and one of these is a compound of m/z 526.18 that predicts hatching even better than solA with both data methods. This compound’s involvement in hatch stimulation was confirmed by fractionation of three representative root exudates and hatching assays with the resulting fractions. Moreover, the compound shares mass fragmentation similarity with solA, and we therefore assume it has a similar structure. With this work, we show that potato likely produces a solA analogue and we contribute to unravelling the hatch-inducing cocktail exuded by plant roots.

Keywords
Cyst nematode – hatching – metabolomics – machine learning – Pearson’s correlation – random forest
Introduction

Potato (*Solanum tuberosum*) is an important staple food and a major source of starch. Its production has increased tremendously in the past decades, making it one of the crops that are feeding the growing world population, especially in developing countries [1]. With the adaptation of potato to cultivation in a range of environments, combined with the variation available in the form of genetic resources, there is potential to further optimise potato production, and, for example, more effectively combat pests [1].

Plant-parasitic nematodes are one of the most harmful pests in agriculture, since they result in losses of up to US$80 billion annually [2]. Among them are the cyst nematodes, comprising the genera *Globodera* and *Heterodera*, together the second most important parasitic nematode species economically [3]. Their host range is highly specific, as opposed to other species such as the root-knot nematodes. For example, potato cyst nematode (PCN, *Globodera pallida* and *Globodera rostochiensis*) only parasitises species of the Solanaceae.

The life cycle of cyst nematodes is divided into 4 larval stages and one adult stage, separated by moults. Juvenile stage 2 (J2) larvae use chemotaxis to locate a suitable host, penetrate the host root and subsequently establish a feeding structure, the syncytium. After moulting into J3, J4, and adults, the female stays sedentary, whereas the male is migratory and finds the female to fertilize her eggs. During egg production, the female body swells and protrudes from the root epidermis. Upon death, the female body will dry and harden and thus turns into a cyst that remains in the soil when the crop is harvested. In the cyst, the embryos in the eggs moult from J1 into J2, and then arrest developmentally.

Cysts with J2's can stay dormant in the soil for many years, until they perceive a cue from their host [4]. This is especially true for *Globodera spp.* that rely almost exclusively on host root exudate for hatching, while other cyst nematodes, such as the beet cyst nematode (*Heterodera schachtii*), may also hatch through rehydration only [5]. Root exudates contain hatching factors (HF), which induce hatching independently, and hatching stimulants (HS), which by itself are inactive but enhance the hatching activity of HF [6]. Up to date, the strongest HF for PCN that has been identified is solanoeclepin A (solA) [7], which can induce up to 80% hatching in *G. rostochiensis* in nanomolar concentrations [8]. SolA has been identified in the root exudate of a range of solanaceous species, including potato and tomato [7,9]. In kidney bean, three eclepins have been identified: glycinoeclepin A, B and C [10,11]. These eclepins induce hatching of the Soybean Cyst Nematode (SCN) of the genus *Heterodera*, but not to the same extent: whereas glycinoeclepin A (glyA) is active at 10⁻¹¹ to 10⁻¹² g/mL, glyB and glyC are active only at higher concentrations, around 10⁻⁸ to 10⁻⁹ g/mL [10,11]. The variation in glycinoeclepin structure suggests that more solanoeclepins might exist and that their hatch inducing activity can vary. Besides solA, some weaker HFs have been identified, such as α-chaconine, α-solanine, solasodine and solanidine [12,13]. Apart from hatch inducing compounds, also hatch inhibitors (HI) are known. They are, for example, present in the root exudate of young plants preventing egg hatching before enough root material is available for a successful infection [6]. Once the ratio HF:HI is high enough, hatching will be induced.

Although PCN can infect a range of species of the Solanaceae, potato is economically most affected. Little is known about the natural variation in hatch induction rate of exudates and solA production in potato, which leads to unpredictable levels of PCN infection. In the present study we analysed the solA content in the root exudate of 51 potato cultivars, and, on a selection of these cultivars, also conducted hatching assays and untargeted metabolomics analysis. We found that solA alone does not completely predict hatch inducing activity of potato root exudates, and therefore expected to find more, hitherto unknown, HFs and/or HSs. Through
machine learning with random forest (RF) feature selection and Pearson’s correlation analysis, we identified several metabolic features that predict PCN hatching and correlate with solA. Furthermore, we chromatographically fractionated three exudates with different solA content, and tested the resulting fractions with hatching assays. Fractions that induced hatching were once more analysed on LC-MS. This study identified a new potential HF for PCN, which will advance our understanding towards plant-nematode interaction at the pre-parasitic stage and possibly offer alternative solutions to develop resistant crops to these detrimental pests.

Materials and methods

Chemicals
A standard of solanoeclepin A was kindly provided by prof. Keiji Tanino (Hokkaido University, Japan). KOH and NH₄OH for sample extraction and purification were obtained from Merck KGaA (Darmstadt, Germany). Methanol, acetonitrile, deionized water and formic acid for LC-ESI-QTOF-MS and UPLC-MS/MS analysis were all hypergrade for LC-MS (Biosolve BV, The Netherlands). Milli-Q water was prepared using a water purification system Milli-Q® (Merck Millipore, Burlington, MA, USA).

Plant and nematode materials and growth conditions
Tubers of S. tuberosum were obtained from three different potato breeding companies: Avebe (Veendam, The Netherlands), HZPC Holland B.V. (Joure, The Netherlands) and Meijer Potato (Rilland, The Netherlands). Cysts of G. rostochiensis Mierenbos pathotype Ro1 were obtained from Aska Goverse, Laboratory of Nematology, Wageningen University. Cysts had been reared in the greenhouse, and dormancy was broken by storage in -80°C for several months, after which they were stored at 4°C until use.

Plants were grown in the greenhouse, in pots (8 cm bottom diameter, 12 cm top diameter and 9 cm height) with standard greenhouse compost nr. 3, using one tuber per pot. Root exudates were collected from 5 replicates for each cultivar by pouring distilled water on the soil in the pot and collecting 200 mL of flow-through per pot. The root exudates were stored at 4°C until further processing. Roots were collected from the soil; all soil particles were carefully removed through washing, the roots dried with paper tissue, and subsequently stored at -80 °C.

Sample extraction and purification
Root exudates were filtered over filter paper to remove soil particles. Root exudates for solA measurement were purified by solid phase extraction (SPE) using vacuum through Oasis® MAX cartridges (3 cc/60 mg, Waters, Milford, MA, USA) cartridges according to the protocol from Guerrieri et al. [9] For metabolomics analysis, C18 cartridges (6 cc/500 mg, Waters, Milford, MA, USA) were conditioned and equilibrated as per the manufacturer’s instructions. Of each sample, 25 mL was loaded on the cartridge, after which the cartridge was washed with 3 mL MQ. Elution was achieved with 3 mL methanol. Samples were dried under vacuum, redissolved in 20% methanol, and filtered using 750 μl 0.22 μm Nonsterile Micro-Centrifugal Filters (Fisher Scientific, Landsmeer, Netherlands) before LC-ESI-QTOF-MS analysis.

Roots were ground to a fine powder using liquid N₂ and subsequently 100 mg of this powder was extracted with 80% methanol according to the protocol from De Vos et al. [14]. Extracts were filtered using 750 μl 0.22 μm Nonsterile Micro-Centrifugal Filters (Fisher Scientific, Landsmeer, Netherlands) before LC-ESI-QTOF-MS analysis.

UPLC- multiple reaction monitoring-MS/MS analysis
SolA analysis was executed according to Guerrieri et al. [9]. In short, 5 μl of purified root exudate was injected onto the reversed-phase UPLC column (Acquity UPLC® Ethylene Bridged Hybrid (BEH) C18 column, 2.1 × 100 mm, 1.7 μm particle size, Waters), and the eluent was
introduced into the ESI ion source of the mass spectrometer. SolA was analysed in positive mode as [M+H]+, using transitions 499>83, 499>315, 499>399, and 499>453. The instrument MS data acquisition and processing were carried out by MassLynx™ software, version 4.2 (Waters, Milford, MA, USA).

**LC-ESI-QTOF-MS analysis**

Untargeted metabolomics was executed according to Guerrieri et al. [9]. Briefly, root exudates and extracts and UPLC fractions were analyzed using a QTOF-MS equipped with a dual-stage trapped ion mobility separation cell (timsTOF pro Bruker Daltonics Inc, Billerica, MA, USA). Sample injection (20 μl) and LC separation were performed on an Ultimate RS UPLC system (Thermo Scientific, Germeringen, Germany) with an Acquity UPLC CSH C18 130 Å, 1.7 μm, 2.1 mm × 100 mm protected by a VanGuard 2.1 mm × 5 mm of the same material. A gradient from 1% acetonitrile to 99% acetonitrile in 18 min was applied. Eluting compounds were sprayed in positive and negative ion mode by an Apollo II ion funnel ESI source (Bruker Daltonics Inc). The resulting data were analyzed using DataAnalysis ver. 4.3 (Bruker Daltonics) and MetaboScape® 5.0 (Bruker Daltonics).

**Fractionation of root exudates**

For UPLC fractionation of root exudates, 50 mL of the exudate was lyophilized, and salts were precipitated by dissolving the dried exudate in 3 mL of methanol. The samples were centrifuged at 4000 rpm for 3 min and the supernatant was transferred to another vial after which it was dried by vacuum evaporation. Subsequently, the residue was dissolved in 150 μl 25% acetonitrile and injected onto the same Waters UPLC equipped with the same column, as described above. Separation of the samples was achieved using the same gradient as used for solA quantification. The eluent was fractionated using an analytical fraction collector (WFM-A) for UPLC systems and the operating software Empower v.3.6.0 (Waters, Milford, MA, USA). Each sample was injected 8 times, and 19 pooled 30 sec fractions were collected between 0.5 and 10 min. Of each fraction 12μl was lyophilized and then diluted 33-fold to a volume of 0.4 ml in 2% ethanol and then used to test hatching activity. This resulted in the fractions being 33-fold more concentrated than the crude root exudate, which is similar to the concentration used for UPLC-MS/MS analysis, obtained through SPE sample processing. The remaining fraction sample was freeze dried once again and dissolved in 150 μl 25% ACN for untargeted metabolomics using LC-ESI-QTOF-MS.

**Nematode hatching assay**

Hatching assays were carried out as described in Guerrieri et al. [9]. In short, cysts were hydrated for 7 days and, subsequently, eggs were released from the cyst by manually opening each cyst separately. 100 μL of egg suspension, containing around 100 eggs, was distributed to the wells of a glass-coated 96-well plate. Subsequently, 100 μL of test solution (dried root exudate that was filtered and partially purified by SPE dissolved in 2% EtOH in tap water, dried fraction of root exudate dissolved in 2% EtOH in tap water, solA in 2% EtOH in tap water, or 2% EtOH in tap water) was added to each well. Eggs and juveniles were counted at t = 0 and t = 14 days. Hatching percentage was calculated according to the formula:

\[
\frac{(J_{14\text{-t}} - J_{t})}{E_0} \times 100
\]

where \( J_{t} \) is the number of hatched J2 after 14 days of treatment, \( J_0 \) is the number of hatched J2 at the start of the assay, and \( E_0 \) is the number of eggs at the start of the assay.

**Data analysis**

The resulting feature table was filtered by removing features with intensities lower than or equal to the empty pot control (both average of five replicates), occurring in less than three
replicates for any cultivar. Furthermore, peaks areas below 200 were considered to be absent. For Principal Component Analysis (PCA), feature tables were transformed using variance stabilising normalisation (vsn) method, and scaled using autoscaling method. These data preprocessing were performed with the packages vsn (v3.58.0) [15] and mdatools (v0.12.0) [16] using R (v4.0.2) in Rstudio (v1.3.1093). Furthermore, the package ggplot2 (v3.3.5) to plot the trendline of Fig. 1 was used [17]. Pearson’s correlation plots, which show the correlation between solA content and metabolite features and the hatching percentages and metabolite features, were assembled using the package corrrplot (v0.90) [18]. Only the top 10 correlating features are displayed.

The log2 transformed metabolomic data was used to build a multivariate model using a random forest-based MUVR (v0.0.975) in R [19]. Variable selection and validation in multivariate model were conducted using an algorithm that simultaneously identified minimal, optimal and all relevant variables for regression analysis. The model was built using the following parameters: nRep=14, nOuter=8, varRatio=0.85. The model was evaluated with the fitness estimator, Q2, which was used for regression analysis, whereas the number of misclassifications was used for classification and multilevel analysis. Permutations were obtained by 100 and 150 (exudates and root extract respectively) times repeated random sampling of the original metabolomic matrix. To assess modelling performance, permutation p-values were thus calculated with the population of fitness metrics.

Molecular networks were calculated using the online platform Global Natural Products Social Molecular Networking (GNPS) using the Molecular Networking option [20]. Raw mzXML files of the LC-ESI-QTOF-MS data of the fractions 8, 9, 10 and 11 of cultivars Avatar, Desiree and Seresta were uploaded to their server and networks, separated by cultivar, were constructed using Small Data Preset and further default settings. The resulting network was further modified using MolNetEnhancer (default settings) [21], downloaded and opened in Cytoscape (v3.9.0) [22] for rendering.

Detailed scripts can be found in R markdown files in the Supplemental materials.

Results

Natural variation in the solA content of root exudate obtained from potato cultivars is associated with differences in hatching activity

Solanoeclepin A content was analysed in the root exudate of 51 potato cultivars (Fig. 1A) and differed greatly among the cultivars, ranging from 1.1 (Merenco) to 112.3 pmol/g FW (Avatar). SolA was not detected in bulk soil (Fig. 1, empty pot). Twenty cultivars with representative solA content were selected for further experiments (Fig. 1A). Hatching activities of root exudates of these cultivars were evaluated through a hatching assay with G. rostochiensis eggs. All root exudates induced more hatching than the flowthrough of pots with only soil. Hatching correlated positively with SolA content according to a logarithmic relationship, similar as observed previously for (wild) tomato (S. lycopersicum, S. sisymbriifolium, S. habrochaites, S. pennellii and S. pimpinellifolium) species [9]. Although the p-value shows the relationship is significant (>0.001), the R^2 is low (0.14) (Fig. 1B). This low correlation coefficient suggests that there are more compounds present in the root exudate than only solA that influence the hatching activity, such as HS, HI and other HFs [6].
Fig. 1, A: solA concentrations in pmol/g FW of 51 commercial potato cultivars measured by UHPLC-MRM-MS/MS (solA_MRM). Colours indicate whether the species was selected for further experiments: red yes, green no. B: solA concentrations of selected cultivars from A plotted against average hatching percentage. A logarithmic relationship is detected with R2 value of 0.14 and a significant p-value.

Root exudates contain solA and solA-like compounds that predict hatching
In order to say more about the possible role of such other compounds in hatching, the root exudates of the twenty selected potato cultivars were also analysed by LC-ESI-QTOF-MS. PCA was used to visualize the differences in the root exudate metabolome of the potato cultivars (Fig. S1). Although variability between replicates was high for some of the cultivars, PC1 and PC2 together explained 55.29% and 73.11% of the total variance of features for positive and negative mode, respectively. Exudate from Fontane was clearly chemically different from exudate collected from Axion, for example.

Next, the relationship between metabolomics features and solA content determined by UPLC-MS/MS analysis was studied using Pearson’s correlation analysis (Fig. 2A, B) and RF feature selection (Fig. 2C, D), which was statistically validated through 100 permutation tests (p value <
RF results in a feature importance score between 0 to 1, which is calculated from the inverse of the rank. In the Pearson's correlation test for positive mode, 6 features, and in negative mode, 2 features, highly correlated with solA, with the highest coefficient reaching 0.89 (Table 1). In RF analysis, 60 and 75 features, in positive and negative mode, respectively, were selected to contribute significantly (p-value < 0.05) to solA content. Among the features thus selected was also solA that we were able to find back in the metabolomics data (Table 1). This feature correlates, as expected, to a high degree with the solA concentration determined by UPLC-MS/MS analysis (solA_MRM), with a Pearson's correlation coefficient of 0.81 and RF feature importance values of 0.65 and 0.51 for positive and negative mode, respectively (Fig. 2C).

<table>
<thead>
<tr>
<th>Molecular mass</th>
<th>Rt</th>
<th>Measured m/z</th>
<th>Calculated mass</th>
<th>Δm/z (ppm)</th>
<th>Ion</th>
<th>Predicted formula</th>
<th>Detected in</th>
<th>Putative name</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>248.0508</td>
<td>9.86</td>
<td>247.0435</td>
<td>247.0448</td>
<td>5.4</td>
<td>[M-H]-</td>
<td>C₂₅H₂₉O₁₀</td>
<td>Exudate</td>
<td>hatch</td>
<td></td>
</tr>
<tr>
<td>248.11</td>
<td>6.38</td>
<td>249.1108</td>
<td>249.1121</td>
<td>5.4</td>
<td>[M-H]-</td>
<td>C₂₅H₂₉O₁₀</td>
<td>Exudate</td>
<td>prenyl caffeate</td>
<td>hatch</td>
</tr>
<tr>
<td>263.0641</td>
<td>0.91</td>
<td>262.0568</td>
<td>262.0577</td>
<td>-4.0</td>
<td>[M-H]-</td>
<td>C₁₆H₁₇NO₅</td>
<td>Extract</td>
<td>ascorbic acid</td>
<td>hatch</td>
</tr>
<tr>
<td>265.9186</td>
<td>0.56</td>
<td>266.9259</td>
<td></td>
<td></td>
<td>[M-H]+</td>
<td></td>
<td>Extract</td>
<td>hatch</td>
<td></td>
</tr>
<tr>
<td>498.1887</td>
<td>7.66</td>
<td>497.1814</td>
<td>497.1806</td>
<td>-1.6</td>
<td>[M-H]-</td>
<td>C₂₇H₁₉O₉</td>
<td>Exudate, fractions</td>
<td>solA</td>
<td>solA, hatch</td>
</tr>
<tr>
<td>526.1843</td>
<td>7.95</td>
<td>525.1771</td>
<td>525.1755</td>
<td>-3.0</td>
<td>[M-H]-</td>
<td>C₂₉H₂₀O₁₀</td>
<td>Exudate, fractions</td>
<td>solA-like</td>
<td>solA, hatch</td>
</tr>
<tr>
<td>572.2254</td>
<td>6.26</td>
<td>571.2182</td>
<td>571.2174</td>
<td>-1.4</td>
<td>[M-H]-</td>
<td>C₃₀H₂₀O₁₁</td>
<td>Extract</td>
<td>solA precursor?</td>
<td>solA</td>
</tr>
<tr>
<td>580.38</td>
<td>13.95</td>
<td>581.3835</td>
<td>581.3837</td>
<td>0.3</td>
<td>[M-H]+</td>
<td>C₂₉H₂₁O₁₁</td>
<td>Exudate</td>
<td>hugonone A</td>
<td>hatch</td>
</tr>
<tr>
<td>589.2527</td>
<td>6.26</td>
<td>590.2600</td>
<td>590.2596</td>
<td>-0.7</td>
<td>[M-H]-</td>
<td>C₂₉H₂₁O₁₁</td>
<td>Extract</td>
<td>plateocin A13</td>
<td>solA</td>
</tr>
</tbody>
</table>

Moreover, a feature [M-H]- 525.1771 (Rt 7.95, Table 1) correlated to an even higher extent with solA (r²= 0.89, Fig. 2A2, Table S2) and had an RF feature importance score of 0.67 and 1 in positive and negative mode, respectively (Fig. 2C). Intriguingly, the MS2 spectrum of this feature showed strong similarities with the solA MS2 spectrum with several of the same fragments: 481.184, 453.187 and 83.0487 (Fig. S2) with the molecular formulae C₂₇H₂₉O₁₀, C₂₅H₂₉O₁₀⁺ and C₅H₇O⁺, respectively. The feature has a predicted molecular formula of C₂₉H₂₀O₁₀ and a mass difference of plus 27.9955 Da compared with SolA, which theoretically could correspond to methylation of one of the alcohols (+14) and double oxidation to a carbonyl (+14) of solA. Since both share the m/z 83.0487 fragment, the carbonyl may be present on ring A, B, C, D or E (the allylic position on ring D is a good candidate position), while the methylation may occur on the alcohols of the C or E-ring. In conclusion, we expect that this compound will be structurally similar to solA and we will call this compound solaneolepin B (solB).

The correlation and predictive value of features for hatching were generally lower, with only 1 and 2 moderately correlating features in positive and negative mode, and no highly correlating features in Pearson's correlation tests (Fig. 2B). RF feature selection for correlation with PCN hatching rendered less compounds: 24 for both positive and negative mode, of which 3 and 7 were also detected with the Pearson's correlation test (Fig. 2D). The highest correlators for Pearson and predictors found in RF are solA and solB (Fig. 2B, D). Interestingly, in negative mode, RF feature importance is almost three times higher for solB than for solA (Fig. 2D2), suggesting that solB predicts hatching better than solA. Further analysis of the results yielded three more interesting features: [M-H]+ 581.3835 and [M-H]- 249.1108 had a feature importance > 0.1, and [M-H]- 247.0435 had a feature importance of 0.2 (Fig. 2 B, D, Table 1). These features were not significant in the Pearson's correlation analysis, so may have a non-

Table 1: metabolites detected in root extract, exudates and fractions of exudates that are of particular interest because of high correlation (RF and/or Pearson’s) with solA content of root exudates or PCN hatching.
Features with a feature importance lower than 0.2 were not further considered in this manuscript.

**Fractionation of root exudates shows specific hatching activity**
To confirm the hatching activity of some of the features we identified, three cultivars were selected that induced high (Avatar and Desiree, ~21%) and low (Seresta, 10.5%) hatching in PCN, and contained high (Avatar, 95.42 pmol / g FW and Desiree, 81.58 pmol / g FW) and low (Seresta, 9.37 pmol / g FW) levels of solA in their root exudates. These three root exudates were fractionated into 19 fractions using UPLC that were tested for hatching activity (Fig. 3A). In Avatar and Desiree, fraction 8 and 9 induced hatching comparable to the unfractionated root exudate (~60%). Furthermore, fractions 10 (both genotypes) and fractions 11 (only Desiree) also induced significant hatching. For Seresta, fraction 9 induced similar hatching as the unfractionated root exudate, while fraction 10 also induced some hatching (Fig. 3A).

**Metabolomics of fractions with high hatching activity confirm discovery of a new hatching factor**
To assess if we could detect the presence of some of the putative hatching factor candidates identified above, in the hatch-inducing fractions, fractions 8, 9, 10 and 11 were analysed using LC-ESI-QTOF-MS. PCA analysis shows that, in both positive and negative mode, the same fractions of the different genotypes cluster together (Fig. 3B). Using the GNPS tool, classical molecular networks were constructed of the LC-ESI-QTOF-MS data of these three cultivars (Fig. 4). This showed that the three genotypes are chemically different, especially in the amount of

---

**Figure continues on the next page**
Fig. 2 – a fully readable version of this figure is available in the supplementary materials - A, B: m/z values of features in root exudate samples detected by LC-ESI-QTOF-MS analysis linearly correlating (Pearson) with (A) solA and (B) hatching in positive mode (A1, B1) and negative mode (A2, B2). The features in the (Pearson) correlograms are ordered according to the correlation coefficient using “hclust” method (corrplot package in R). C, D: random forest feature selection from LC-ESI-QTOF-MS root exudate data for (A) positive and (B) negative mode with solA measured by UPLC-MRM-MS/MS (solA_MRM), and positive (C) and negative mode (D) with hatching. The features that were also detected with Pearson’s correlation analysis, in the same mode (positive or negative), are marked in red. Features that were detected with Pearson’s correlation analysis
 CHAPTER 4

in the opposite mode are marked in blue. Features that were detected with Pearson’s correlation analysis in the same as well as the opposite mode are marked in green. Their Pearson’s correlation coefficient is displayed in or next to the bar. The cut-off for the Pearson’s correlation coefficient was 0.1. Only statistically significant (p < 0.05) Pearson’s correlations are marked. The degree of significance is marked by stars: p < 0.001, ***, p < 0.01, **, p < 0.05, *. Retention times of features in root extracts and root exudates are not comparable, since they were run on a different column, although of the same type.

organic acids and derivatives. Avatar exudate contains the most phenylpropanoids and polyketides, followed by Desiree. Desiree exudate contains much more benzenoids and alkaloids than the other two cultivars.

Several m/z values that were selected for their hatching activity by Pearson’s correlation and RF, were detected in these fractions as well (Table 1). Among these are solA ([M-H]- 497.18199) and solB ([M-H]- 525.17611), which highly correlated with both hatching and solA presence in complete root exudates (Fig. 2). Comparing the relative abundance of these compounds in the fractions with the PCN hatching, shows that solA, which is detected in both positive and negative mode, is mostly present in fraction 8 and 9 (Fig. 5). In Avatar and Desiree these fractions induced most hatching. In Seresta, only a trace amount of solA was detected (the metabolomics analysis is less sensitive for solA than the MRM-LC-MS/MS analysis). Since solA concentrations are highest in fraction 9, but the highest hatching in Avatar and Desiree is induced by fraction 8, there likely is an additional HF present in the latter fraction. Indeed, [M-H]- 525.17611 is present in fraction 8 in Avatar and Desiree (Fig. 5). Moreover, it seems the hatching activity in fraction 10 of Avatar and Desiree is caused by solB, as no solA was found. We thus consider solB has hatch inducing activity. The presence of a trace amount of solB in both fraction 8 and fraction 10 suggests these are isomers which are separated by UPLC but not separated by the LC-ESI-QTOF-MS. The relative abundance of this compound in the twenty cultivars that were metabolically analysed, is highly variable (Fig. S3): the highest producer, cultivar Avatar, contains about ten times more of this compound than the root exudate of the lowest producer, Ivory Russet.

Apart from solA and solB, there were other compounds present in fractions 8, 9, 10 and 11, that highly correlate with hatching within the scope of these fractions (Fig. S4). Interestingly, the compound that had the highest correlation with hatching as well as solA content in whole root exudates, solB, is not within the top 10 in the fractions correlation. The highest correlating compound with hatching was solA, followed by compounds [M-H]- 297.13419 and [M-H]- 293.10245. These compounds were not found to correlate with hatching in the whole root exudate dataset. However, compounds that score high in this analysis might do so because they have similar chemical properties as solA and therefore end up in these fractions, and might have nothing to do with hatching. This analysis would benefit from including all fractions and their LC-ESI-QTOF-MS analysis, but this was not attempted in this study.

Root extracts do not contain solA, but some metabolites that correlate with hatching may be its precursors

Because the hatching factors are produced in and, subsequently, exuded from the roots we also performed metabolomics on root extracts, hoping to identify additional features associated with solA and hatching. Roots were ground, extracted and subsequently analysed by LC-ESI-QTOF-MS. Principal Component Analysis (PCA) showed reasonable separation of the cultivars for negative mode and, to a lesser extent, for positive mode (Fig. S5). PC1 and PC2 together explained 16.84% and 25.23% of the variation in the metabolic features measured in negative and positive mode, respectively.
SolA was not detected in root extracts, also not when using MRM-LC-MS/MS, but Pearson correlation and feature selection by Random Forest (RF) machine learning yielded several features that significantly correlate with solA content in the root exudate, and PCN hatching. The top 10 Pearson correlating compounds are displayed in correlograms (Fig. 6A, B). Pearson correlation with solA concentration resulted in 7 and 12 highly correlating compounds (coefficient > 0.5) in positive and negative mode, respectively (Table S1). The two features with the highest coefficients (0.70 and 0.68) are [M-H]⁻ 571.2182 and [M-H]⁺ 590.2600, respectively (Fig. 6A, Table 1, Table S2). The correlation coefficient of root extract features with hatching was lower, less than 0.5. A moderate degree of correlation (coefficients between 0.3 and 0.5) was found for 14 and 8 compounds for positive and negative mode, respectively. The two features with the highest correlation with hatching were [M-H]⁻ 262.0568 (coefficient 0.36) and [M-H]⁺ 266.9259 (coefficient 0.37) (Table 1, Table S2). These two compounds had very short retention times, which indicates they are polar. The top correlating compounds with solA and hatching did not overlap.
Fig. 3: A) hatching of G. rostochiensis eggs under influence of fractions of three root exudates of genotypes Avatar (A), Desiree (D) and Seresta (S). These genotypes had high and low concentrations of solA, respectively (Fig. 1A). Fractions 8, 9, 10 and 11 were selected for LC-ESI-QTOF-MS analysis. The first column, 10xdil, shows the hatching percentage of the 10x diluted unfraccionated root exudates. The concentrations of fractions are comparable to the 10x diluted unfraccionated root exudates. B) PCA plots of LC-ESI-QTOF-MS data positive (top) and negative (bottom) mode, of four selected fractions associated with high hatching inducing activity of root exudates of the genotypes Avatar, Desiree and Seresta.

Whereas Pearson’s correlation only detects linear correlations, feature selection by RF machine learning can also detect other types of relationships. RF results on predictive features for hatching could not be validated by permutation tests (Table S1). Therefore, we only used RF feature selection for solA content, which could be validated by permutation tests (Table S1, Fig. 6C, D). In positive mode, 63 features, and in negative mode, 20 features were selected. Out of these, 27 and 14, respectively, were also found in the Pearson’s correlation test (Fig. 6). The two compounds with highest Pearson’s correlation coefficient ([M-H]: 571.2182 and [M-H]+ 590.2600) were confirmed with RF (Table 1, Fig. 6C, D). Upon further inspection, the compound [M-H]: 571.2182 has an exact mass of 572.2254 and is likely C_{30}H_{35}O_{11} (Table 1). It could possibly be a solA precursor (Fig. S6), such as a cycloartenol-derived abietospiran-type
precursor with a C30 backbone as suggested in Sun et al. (2019) [23]. However, due to the low amount that was detected, we cannot show a reliable MS2 spectrum.

**Fig. 4:** classical GNPS molecular networks of Avatar (A, D), Desiree (B, E) and Seresta (C, F) root exudate LC-ESI-QTOF-MS data, analysed in positive (A, B, C) and negative mode (D, E, F). Networks were constructed using default settings. Non-connected nodes, and grey networks consisting of 4 or less nodes were deleted to improve visibility.

**Discussion**

In this study, we identified new putative hatching stimulants and/or factors for potato cyst nematode (PCN), in particular *Globodera rostochiensis*. First, the content of solA, which is the most potently currently known HF for PCN, was determined in the root exudates of 51 commercial potato cultivars by UPLC-MS/MS (Fig. 1) using a previously published method [9]. Hatching assays on *G. rostochiensis* with twenty partly purified root exudates showed that solA content and hatching percentage are logarithmically correlated, but this correlation is weak, which suggests that more HF and HS, and possibly HI, are influencing hatching. In the next experiment, root exudates were analysed through metabolomics on LC-ESI-QTOF-MS and features were computationally linked to solA content and hatching through Pearson’s linear correlation test and RF (Fig. 2). Although solA itself was detected and, as expected, correlated highly with solA content (measured with MRM) and PCN hatching, it did not display the highest correlation with hatching. A second compound of molecular weight 526.17 correlated with solA and hatching to an even higher extent. This exceptionally high correlation (correlation coefficient of 0.89 and RF feature importance of 1, which is the highest possible) suggests that this compound is structurally related to solA. Comparison of the MS2 spectra of this compound, coined solB, with solA revealed three identical fragments lending support to the postulated structural relatedness (Fig. S2). Lastly, we fractionated three selected root exudates that varied
in solA concentration and hatch inducing potency. We tested all fractions in a hatching assay (Fig. 3A) and analysed the fractions by LC-ESI-QTOF-MS. Even though the hatch inducing activity between cultivars varied greatly, in PCA the same fractions of different cultivars clustered together (Fig. 3B). Hence, only a small number of compounds in root exudates influences hatching, and their hatch inducing potency is not a major discriminating factor. SolA and solB were both found in fractions inducing high PCN hatching rates (Fig. 5). Next, the roots from twenty cultivars were harvested and extracted using methanol in order to find solA/solB precursors and/or other HFs. The root extracts were analysed using LC-ESI-QTOF-MS metabolomics, and through Pearson’s correlation test and RF, we could correlate root metabolite features with solA content of the root exudate, some of which showed a high correlation (Fig. 6). The root extracts did not contain solA, but possibly they contain precursors of solA, of which the concentration could correlate with solA content in the exudate. Indeed, especially for a feature of 572.2 Da, a strong correlation with solA content was detected in both data analysis methods. The predicted structural formula of this feature has A, B and C rings that are identical to those of solA (Fig. S6). The correlation of root extract features with hatching was low, and RF models were not significant.

On a sidenote, in this study it was observed that hatching with unpurified but diluted root exudates results in almost three times more (60% versus 21% for Avatar and Desiree, 35% versus 10.5% for Seresta) hatching than the partially purified root exudates (Fig. 1B, Fig. 3A). This was shown before for Sephadex G-10 fractionated root exudates [24]. This is probably because HF work according to an optimum, and concentrations above this optimum do not stimulate, or even inhibit hatching [25]. Furthermore, by dilution of root exudates, HI are diluted as well, which reduces their inhibitory effect. This effect might play a role as well for high hatch inducing fractions (Fig. 3A), which contain mainly HF and HS, whereas the HI might end up in a different fraction based on their chemical properties, and hence do not mitigate the hatch induction of the fraction.

![Graph showing counts levels of individual features per fraction per genotype](image-url)
**Fig. 6:** A fully readable version of this figure is available in the supplementary materials. m/z values and retention times (Rt) of compounds in root extract samples detected by LC-ESI-QTOF-MS analysis that correlate with solA measured by UHPLC-multiple reaction monitoring-MS/MS (solA_MRM) (A, C, D) content and/or hatching (B), detected with Pearson’s correlation test (A, B) and Random Forest feature selection methods (C, D). (A) correlograms showing linear Pearson’s correlations of metabolites with solA_MRM in positive (A1) and negative (A2) mode. (B) correlograms showing linear Pearson's correlations of metabolites with hatching in positive (B1) and negative (B2) mode. Order of the features in all correlograms is reordered according to the correlation coefficient using “hclust” method (corrplot package in R). (C) random forest feature selection for positive mode with solA_MRM. (D) random forest feature selection for negative mode with solA_MRM. The features that were as well detected with Pearson’s correlation analysis in the same mode (positive or negative) are marked in red. Features that were detected with Pearson’s correlation analysis in the opposite mode are marked in blue. Features that were detected with Pearson’s correlation analysis in the same as well as the opposite mode are marked in green. Their Pearson’s correlation coefficient is displayed in or next to the bar. The threshold for the Pearson’s correlation coefficient was 0.1. Only statistically significant (p < 0.05) Pearson’s correlations are marked. The degree of significance is marked by stars: p < 0.001, ***, p < 0.01, **, p < 0.05, *. Other metabolic features that showed to be interesting in our data analyses (Table 1) are hard to identify, partially because of low mass intensities that resulted in poor quality MS2 spectra. So far we can conclude that they do not overlap with previously purified but unidentified hatching factors, such as a 530.5 Da compound identified by Devine and Jones [26], or the 437 Da compound identified by Atkinson et al. [27]. The molecular weights we found in the current study also do not match with glycoalkaloids identified in potato (for example, α-chanonine or α-solanine), which all have molecular weights (MW) of over 800 Da, or their aglycones (solanidine, MW 397.6, solasodine, MW 413.6).
Since the other eclepin group, the glycinoeclepins produced by kidneybean, displays more structural variation [11], we anticipated that also more solanoeclepins exist. Since it is unlikely that a plant produces metabolites that are solely detrimental to itself, it is to be expected, from an evolutionary standpoint, that eclepins also have a beneficial effect on the plant. The diversification of eclepins, therefore, could be caused by an arms race between plant and parasitic nematode. Since the eclepins have a detrimental effect on the plant (the cyst nematode hatch induction), but are indispensable at the same time because of a hitherto unknown beneficial effect, the plant’s eclepin biosynthesis diversifies, leading to the biosynthesis of new eclepins that still have the beneficial effect but induce no or lower cyst nematode hatching. Therefore, the detection of new eclepins using hatch activity guided bioassays might be ineffective.

In the present study, the influence of microorganisms on the production of hatching factors and stimulants was not monitored. However, microorganisms have been reported to have a considerable effect on the hatch-inducing cocktail. For example, mycorrhization increases the production and variety of hatching factors in potato roots [28] and aseptically grown potato plants lack hatching factors that are present in the root exudate of conventionally grown potato [29]. Indeed, it is noteworthy that solA could not be detected in root extract, only in exudate, which suggests that it is exuded as a precursor and metabolized into solA by microorganisms. Therefore, follow-up experiments should take this factor into account by, for example, analysing root exudates from aseptically grown plants. For the current study, it can be assumed that all compounds produced by microorganisms that are present in bulk soil, so independent of the plant, were filtered from the dataset (through the use of an empty pot control). Hence, all features discussed in this manuscript depend on the presence of potato, either directly (biosynthesis in the plant) or indirectly (biosynthesis by potato root-associated microorganisms).

PCN is a pest that causes severe damage and yield losses in potato. By identifying the compounds that induce their hatching, it becomes possible to select for potato cultivars that produce low concentrations of hatching factors, thereby preventing hatch and thus infestation of the roots by PCN. For example, from the cultivars that were analysed in our study, Ivory Russet and Seresta produce the lowest concentrations of solA and solB (Fig. 1, Fig. S3), which would make them good candidates for breeding low HF producers. Furthermore, other plant species that produce these hatching factors, but are not a suitable host for PCN, can then be identified. These species can be used as trap crop, inducing suicide hatch in PCN. This was observed before in the wild tomato relative, Solanum sisymbriifolium, one of the best-known PCN trap crops, that we recently showed to produce solanoeclepin A, but is not a host for PCN [9,30].

In the present study, we used a new approach to detect putative HF and HS for PCN produced by commercial potato cultivars. This approach involved the use of a large number of potato cultivars and using metabolomics in combination with machine learning, RF feature selection, and Pearson’s correlation analysis to uncover correlations between metabolic features and hatching. The involvement of several highly correlating features in hatch stimulation was subsequently confirmed by fractionation of representative root exudates. With this, we contribute to the elucidation of the activity of specific compounds in the hatch-inducing cocktail exuded by plant roots and thereby to possible solutions for this important agricultural problem.
References


Supplemental materials

Full size Fig. 2 and Fig. 6, and Table S2 are available through https://doi.org/10.5281/zenodo.6319660.

Table S1: p values of Random Forest permutations

<table>
<thead>
<tr>
<th>dataset</th>
<th>p value</th>
<th># permutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>exudate (+) hatching</td>
<td>0.042422482</td>
<td>100</td>
</tr>
<tr>
<td>exudate (+) solA</td>
<td>0.042422482</td>
<td>100</td>
</tr>
<tr>
<td>exudate (-) hatching</td>
<td>0.045679047</td>
<td>100</td>
</tr>
<tr>
<td>exudate (-) solA</td>
<td>0.042422482</td>
<td>100</td>
</tr>
<tr>
<td>extract (+) hatching</td>
<td>0.056267724</td>
<td>150</td>
</tr>
<tr>
<td>extract (+) solA</td>
<td>0.042422482</td>
<td>150</td>
</tr>
<tr>
<td>extract (-) hatching</td>
<td>0.150932136</td>
<td>150</td>
</tr>
<tr>
<td>extract (-) solA</td>
<td>0.042422482</td>
<td>150</td>
</tr>
</tbody>
</table>

**Fig. S1:** PCA plots of positive (A) and negative (B) mode of LC-ESI-QTOF-MS data from root exudate samples of selected genotypes (Fig. 1A). As an example, it is shown that the root exudates of cultivars Axion and Fontane vary metabolically.
Fig. S2: MS2 spectra of positive mode solA (top) and positive and negative mode of tentative solB (bottom) including the putative molecular structure. These were extracted from whole root exudate LC-ESI-QTOF-MS data. Overlapping fragments between these two metabolites are marked with a red box.
Fig. S3: mass intensity (corrected for root weight) of compound with m/z 525.18 (-) in all LC-ESI-QTOF-MS analysed samples

Fig. S4: the top 10 Pearson’s correlations between metabolites in positive (A) and negative (B) mode detected by LC-ESI-QTOF-MS analysis of fractions 8, 9, 10 and 11 of Avatar, Desiree and Seresta cultivars and hatching of PCN.
Fig. S5: PCA plots of positive (A) and negative (B) mode of LC-ESI-QTOF-MS data from root extract samples of selected genotypes (Fig. 1A)

Fig. S6: putative structure of compound detected in root extract that highly correlates with solA as well as hatching with [M-H] 571.22