Bombs and cocaine: detecting nefarious nitrogen sources using remote sensing and machine learning

Christopher Adams¹, Soo Mei Chee², David Bell² and Oliver P.F. Windram^{1,3}

1. Grand Challenges in Ecosystems & the Environment, Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot, SL5 7PY, UK

2. SynbiCYTE, Imperial College London, South Kensington, London, SW7 2AZ

3. Corresponding author: o.windram@imperial.ac.uk

Abstract

Plants are treated with synthetic or organic nitrogen sources to increase growth and yield, the most common being calcium ammonium nitrate. However, some nitrogen sources are used in illicit activities. Ammonium nitrate is used in explosive manufacture and ammonium sulphate in the cultivation and extraction of the narcotic cocaine from *Erythroxylum spp.* Here we show that hyperspectral sensing, multispectral imaging and machine learning image analysis can be used to visualise and differentiate plants exposed to different nefarious nitrogen sources. Metabolomic analysis of leaves from plants exposed to different nitrogen sources reveals shifts in colourful metabolites that may contribute to altered reflectance signatures. Overall this suggests that different nitrogen feeding regimes alter plant secondary metabolism leading to changes in the reflectance spectrum detectable via machine learning of multispectral data but not the naked eye. Our results could facilitate the development of technologies to monitor illegal activities involving various nitrogen sources and further inform nitrogen application requirements in agriculture.

Introduction

Synthetic or organic nitrogen sources applied to soils or plant tissue provide essential nutrients to enhance and facilitate plant growth. Variable nitrogen abundance can alter gene expression and metabolic pathways (Stitt, 1999). These pathways alter root growth, influencing nutrient uptake, the structure of the plant and nitrogen distribution in the surrounding environment. Assimilation of nitrogen within plant tissue can vary dependent on the presence of certain enzymes (Foyer, 2018), which in turn are dependent on the nutrient status of the plant. Nitrogen deficiency can shift secondary metabolism from nitrogen rich alkaloids to carbon rich molecules. Overall, the transcriptome, proteome and the metabolome are all influenced by nitrogen availability (Fritz *et al.*, 2006).

64% of the 0.86 million tonnes of nitrogen used annually in the UK is in the form of Ammonium Nitrate (AN) or Calcium Ammonium Nitrate (CAN) (Harty *et al.*, 2016). Recently, there has been a noticeable shift towards using CAN in agriculture as AN is a detonatable oxidiser that can be used to construct simple yet highly destructive explosives (Brust *et al.*, 2015; Djerdjev *et al.*, 2018). Infamously, used by terrorists AN is inexpensive⁶ and small amounts are needed to create explosives (Figuli *et al.*, 2016), which have been used in numerous attacks (Anil Ananthaswamy, 2004; Henry Dodd, 2018). Subsequently, Ireland has banned the sale of AN based fertilisers (The Government Inspector of Explosives, Ireland, 2016), while the European commission prohibits the distribution of fertilisers containing more than 16% AN (ECHA, 2009). The USA is also introducing new regulation to limit its use (CISA, 2019).

Ammonium sulphate (AS) is another nitrogen source associated with illegal activity. AS is used in the production of cocaine from *Erythroxylum coca* and *Erythroxylum novogranatense* (U.S. Department of Justice, 1991; Mallette *et al.*, 2017). *Erythroxylum* species prefer moderately acidic soil conditions (Johnson, Campbell and Foy, 1997), therefore, AS is used lower soil pH

enhancing yields (Johnson, Campbell and Foy, 1997). Additionally, the extraction of cocaine from coca leaves produces AS as a by-product after Sulfuric acid is added to coca leaf paste and then washed with Ammonia (U.S. Department of Justice, 1991; Mallette *et al.*, 2017).

Remote sensing represents an appealing strategy to monitor dangerous objects or illicit activity as it is non-invasive and has the potential to scale up using airborne sensors. Remote sensing measures shifts in electromagnetic reflectance, absorption and/or fluorescence to infer the chemical nature of objects under study (Schott, 2007). Close range remote sensing has been used to detect explosives such as TNT and DNT (Keith J. Albert *et al.*, 2001; Jason M. Conder *et al.*, 2003). Similar approaches have also been used to detect cocaine in acetonitrile solution (Wren *et al.*, 2014). Furthermore, high resolution spectroscopy has also been developed to detect a wide range of explosives (De Lucia *et al.*, 2003; Federici *et al.*, 2005; Liu *et al.*, 2006) and differentiate cocaine concentrations in caffeine, glucose and rum (Ryder, O'Connor and Glynn, 2000; Eliasson, Macleod and Matousek, 2008). However, these methods require expensive fine-tuned instruments and access to physical samples with their adaption as airborne imaging systems appearing limited.

Determination of nitrogen uptake by plants and nitrogen content in soil is an area of intense interest in agricultural research. Several remote sensing and imaging methods have been developed to estimate the nitrogen status of plants (W. C. Bausch and H. R. Duke, 1996; Serrano, Filella and Pen˜uelas, 2000; Smith *et al.*, 2003; Schlemmer *et al.*, 2005; Pagola *et al.*, 2009). However, in general the focus of this research has been on estimating total nitrogen content *in planta* and relating its impact to yield, rather than detection of different nitrogen compound applications. For this satellite and airborne imaging systems have been developed to help estimate nitrogen abundance in crops (Swain, Jayasuriya and Salokhe, 2007; Rama Rao *et al.*, 2008).

Given the role of different nitrogen compounds in illicit explosive and narcotic manufacture and that they are readily bioassimilated by plants, we were curious what impact their presence has on plant physiology, and if this could be used to detect these substances. In this paper, we show that different nitrogen sources alter plant leaf reflectance and plant metabolism. Through nitrogen application experiments using *Arabidopsis thaliana*, *Solanum lycopersicum* (tomato) and mixed *Panicum* species (lawn grass) we reveal with hyperspectral sensing how different nitrogen source applications alter plant leaf reflectance. Then we show how customised multispectral imaging using conventional RGB cameras and machine learning image analysis can distinguish plants exposed to different nitrogen compounds. Finally, we show through metabolomic profiling that different nitrogen compounds alter the plant leaf metabolome and suggest how some of these secondary metabolites might influence plant leaf reflectance.

Results and Discussion

Hyperspectral reflectance reveals significant differences in plants exposed to different nefarious nitrogen compounds

To determine if plant leaf reflectance can be used to distinguish between plants exposed to different nitrogen compounds, we first looked at total plant leaf reflectance using hyperspectrometry. *Arabidopsis*, tomato and lawn grass were treated with equimolar (4mM) amounts of either AN (explosives), AS (cocaine manufacture), CAN (safe agricultural fertiliser) or a water control. One-week later leaf hyperspectral reflectance data was collected for all species (Fig. 1a and Supplementary Fig. 1). As part of the aim of our study was to adapt conventional RGB cameras to facilitate distinction of nitrogen treatments we chose to focus on the visual part of the electromagnetic spectrum (VIS-EM) 400 – 700nm. Reflectance intensity in this region

follows a trend across species with relative intensity differences between treatments being hard to distinguish.

Figure 1. Hyperspectral variance reveals differences in plants treated with different nitrogen sources. Mean reflectance intensity from 400-700nm for **a** tomato one week after application of either 4mM of AN (green), CAN (red), AS (yellow) or H₂O (blue), 1000 readings taken from 24 plants per treatment. ANOVA and Tukey's HSD following PCA of reflectance data (400-700 nm) performed on PCs 1-10 **b** tomato **c** Arabidopsis **d** lawn grass. Lines between groups are labelled with PCs which contain variation that can significantly distinguish treatment group pairs. * indicates p -value for Tukey's HSD < 0.05, ** < 0.01 and *** < 0.001.

We performed principle component analysis (PCA) on VIS-EM hyperspectral reflectance data (Fig 1b, c & d). Overall, in Arabidopsis 93.15%, in tomato-95.74% and in grass-79.23% of the variation can be explained by the first and second principle components (PCs), while in Arabidopsis-94.58%, in tomato-97.53% and in grass-83.79% of total variation is captured in the first 10 PCs (Supplementary fig 2). It is likely that the lower total level of variance explained for grass is influenced by the small size of grass leaves and the need to sample over swards versus single leaves captured for Arabidopsis and tomato.

Next, we analysed the first 10 PCs for each plant by analysis of variance followed by a Tukey's Honest Significance Difference test (Fig. 1b-d and Supplementary Fig. 3). The four nitrogen treatments can be significantly distinguished by variation captured within the first 10 PCs. It is common practice to perform analysis of variance on the first or second PC often discarding PCs with lower eigenvalues (Climaco Pinto *et al.*, 2008; Zwanenburg *et al.*, 2011). However, it has been rightly noted (Jolliffe, 1982) that under certain conditions PCs that capture only small amounts of variance can be important. Here we argue that consideration of PCs with small eigenvalues can be important, when distinguishing plant leaf reflectance groups. This is shown in our data where we see PCs with small eigenvalues capable of distinguishing our treatment classes (Fig 1 and Supplementary Fig 2 & 3).

Ensemble machine learning regression highlights VIS-EM regions with high variability

Having established that sufficient reflectance variation exists within the VIS-EM to distinguish plant leaves treated with different nitrogen compounds, we now wanted to identify key regions to target with multispectral imaging. For this we used an ensemble machine learning approach (Feilhauer, Asner and Martin, 2015), combining partial least square, random forest and support vector machine regression to identify VIS-EM regions exhibiting high spectral variance across nitrogen treatments (Fig. 2). The approach was adapted to identify variance within and across species that differed significantly between nitrogen treatments. This identified multiple VIS-EM regions. In particular, we note that 400-410, 445-455, 540-560, 610 and 675-700 nm are consistently identified across all species. Overall, this highlights a complexity of VIS-EM regions exhibiting altered reflectance patterns across nitrogen treatments, both unique and conserved across species. Supporting this we observe eigenvalues of PC highlighted in our ANOVA derive high variance contribution from these regions (Supplemental Fig. 4).

Figure 2. Ensemble regression identifies key spectral signatures that distinguish nitrogen treatments. **a** Ensemble regression of reflectance data from 400-700nm for tomato, Arabidopsis and lawn grass leaves 1-week after application of either 4mM of AN, CAN, AS or H₂O, 1000 readings from 24 plants or grass swards per nitrogen treatment. Important wavelengths shown in black, selected for separate species if the absolute sum of regression coefficients is greater than two-standard deviations from the mean. The consensus all species bands were selected if the absolute sum of regression coefficients of all experimental replicates, for all species, was greater than two-standard deviations from the mean. **b** Vegetation ratio indices generated from hyperspectral data used to monitor nitrogen content (RVI, RV12), water content (PWI, fWBI) and chlorophyll content (PSSRa, PSSRb). A nested ANOVA and Tukey HSD test was performed for each species with the experiment as the block to compare nitrogen treatments using specific indices. Significant differences are colour coded by p-value.

We expected total nitrogen abundance to vary across comparisons with the water control and were curious to see if variable nitrogen uptake could influence treatment classification. Thus, we chose to assess nitrogen abundance using a range of reflectance ratios that have previously been used to estimate nitrogen abundance in leaves such as RV1 and RV2 (D. M. El-Shikha *et*

al., 2008; Zhu *et al.*, 2008). Overall, we found that these simple indices could not be used to reliably distinguish any treatments groups (Fig 2b, Supplementary Fig 5).

However, reflectance ratios developed to estimate chlorophyll abundance via absorption of chlorophyll b (PSSRb-800nm/635nm) (Blackburn, 1998) could in some cases distinguish treatments, but not measures of chlorophyll a abundance (PSSRa-800nm/680nm) (Blackburn, 1998). Unusually, a water index measurement (fWBI- 900nm/min(930:980nm) (Strachan, Pattey and Boisvert, 2002) could partially resolve some comparisons across our species, however this was not recapitulated with PWI (970nm/900nm) another water content estimator (Penuelas *et al.*, 1997). This shows that our nitrogen treatments cannot be fully distinguished using these simplified ratio comparisons, while our regression approach and PCA analysis suggest multiple EM regions are needed to distinguish our nefarious nitrogen sources and controls.

Customised multispectral imaging reliably distinguishes nitrogen treatments

PC and ensemble regression analysis suggest that numerous regions within the VIS-EM are needed to separate nitrogen treatments. So, we devised a highly customisable multispectral image analysis pipeline to acquire and analyse image data for plant leaves exposed to different nitrogen sources. We used simple colour pass filters that selectively transmit light within regions of interest identified by the ensemble regression analysis. Colour pass filters are placed in front of a smart phone camera during image acquisition (Supplemental Fig. 6). Images for multiple filters are registered and combined into a multichannel image stack. We then train random forest pixel and object classifiers (Sommer *et al.*, 2011) on these stacks to build a machine learning system capable of distinguishing plants exposed to different nitrogen sources (Fig. 3 a, b & c). An initial pixel level classifier is first trained to distinguish plant leaves or leaf regions from background. This defines leaves as objects. An object classifier is then used to classify leaves by nitrogen treatment. This is trained using 2-6 regions of grass swards or tomato or Arabidopsis leaves from each treatment. This training data is then used to predict the treatment of the remaining leaves in the image.

The object classifier was capable of predicting nitrogen treatment with impressive accuracy (Fig 3d). Furthermore, multispectral image data created a classifier with significantly higher accuracy than models built using standard 3-channel RGB images, determined with a t-test (AN $p = 0.018$; CAN $p = 0.049$; AS $p = 0.00029$ or H₂O $p = 0.014$). This shows that the regression analysis was useful in helping to target VIS-EM regions for subsequent multispectral analysis, with multispectral data providing superior classification performance compared to simple RGB images.

Figure 3. Machine learning classifiers built from multispectral data demonstrate superior performance. Multispectral image classification of **a** tomato, **b** Arabidopsis and **c** grass. Images acquired 1-week after application of either 4mM of AN, CAN, AS or H₂O. Images were stacked and registered producing a multispectral image data cube. A random forest pixel classifier was used to remove background pixels and define leaf objects. A random forest object classifier was then used to predict nitrogen treatment in leaves. **a** 2-3 leaves, **b** 3-4 leaves and **c** 4-6 grass sward regions per treatment were used as training data. **a** & **b** rows 1&2 – H₂O control; 3&4 – CAN; 5&6 – AN; and 7&8 – AS. **c**: top-left – H₂O; top-right – CAN; bottom-right – AN; and bottom-left – AS. Predictions are blue- H₂O control, red-CAN, green-AN, yellow-AS with training labels shown in solid colour. Confusion matrices showing mean accuracy for **d** multispectral and **e** RGB data for 3 experimental replicates from each species (9 experiments in total). Overall accuracy for multispectral classifier AN = 89%, AS = 86.51%, CAN = 81.18%, H₂O = 94.79%. Accuracy of a machine learning classifier built using RGB images AN = 44.76%, AS = 54.28%, CAN = 58.36%, H₂O = 70.12%. Accuracy was significantly different between RGB and multispectral images determined with a t-test (AN - $P = 0.018$; CAN - $P = 0.049$; AS - $P = 0.00029$ or H₂O - $P = 0.014$).

Untargeted metabolomics identifies shifts in metabolites between nitrogen treatments

Our hyperspectrometry and multispectral image analysis show clear shifts in reflectance across the VIS-EM. We reasoned this is likely due, in part, to differences in levels of chlorophyll and other nitrogen rich metabolites. However, we also reasoned that our treatments may alter plant secondary metabolism not directly influenced by nitrogen abundance. Irrespective of this we expect that shifts in chromophores across treatments influenced VIS-EM reflectance. To investigate this, we performed untargeted metabolomics on plant leaves one week after treatment with either 4mM AN, CAN, AS or H₂O. Leaf tissue extracts were run on a HPLC-Q-TOF-MS according to De Vos *et al.* (De Vos *et al.*, 2007) with the acetonitrile gradient modified to 100%. Monoisotopic masses of detected metabolites were filtered for count values >5000, fold change ≥ 2 and significant after a one-way ANOVA ($P < 0.05$) in all replicates per nitrogen treatment when compared to either Water or CAN as a control. PCA performed on differentially

expressed monoisotopic masses (DEMMs) across biological replicates for each species show DEMMs group based on nitrogen treatment (Supplementary Fig 7). Hierarchical clustering was used to compare DEMMs. This compared expression for 163 DEMMs in Arabidopsis, 21 DEMMs in tomato and 31 DEMMs in grass with CAN as a control against AN and AS (Fig 4 a-c). Secondly, we compared expression for 72 DEMMs in Arabidopsis, 18 DEMMs in tomato and 62 DEMMs in grass for Water control against CAN, AN and AS (Fig 4 d-f). AN and AS appear more similar to each other compared to CAN and water control.

Figure 4. Hierarchical clustering of DEMMs reveals similarities between AS and AN treatment. tomato - **a** & **d** -tomato; **b** & **e** -Arabidopsis; and **c** & **f** - grass 1-week after application of either 4mM of AN, CAN, AS or H₂O. 3 biological replicates per experiment and 2 experimental replicates per species. Extracted metabolites were run on HPLC-TOF-MS, Monoisotopic masses were filtered (count value >5000, fold change ≥ 2 , significant after a one-way ANOVA p -value < 0.05) in all replicates per nitrogen source. **a**, **b** & **c** CAN or **d**, **e** & **f** H₂O used as a control.

We searched for DEMMS in species specific PlantCyc databases (Schlpfer *et al.*, 2017) using a neutral monoisotopic mass tolerance of ± 50 ppm. The majority of the DEMMs had no match (Fig. 5 a). However, we identified 21 differentially expressed known metabolites (DEKMs). Certain DEKMs classes contain chromophores known to influence plant colour, such as flavonoids (Panche, Diwan and Chandra, 2016) and light absorption, such as porphyrins (Lee *et al.*, 2018) (Fig 5b). Porphyrins (a key component of chlorophyll (Mauzerall, 1977; Lee *et al.*, 2018)) were upregulated in tomato leaves applied with AS compared to CAN. This supports our regression approach which identified the green (500-575 nm) region of the VIS-EM as key in helping distinguish variation between nitrogen treatments (Fig 2a).

Figure 5. Differentially expressed metabolites include known chromophore containing molecules. DEMMs matched to Plantcyc (Schläpfer *et al.*, 2017) using a mass tolerance of ± 50 ppm. **a** Most of the DEMMs identified across all species across all nitrogen treatments were unknown. **b** Pie chart of classes of the known metabolites. A total of 21 known metabolites were identified, each with a weighting of 1 in the pie chart. If a known mass could be more than a single metabolite group the value was divided accordingly among the possible groups. General structures shown for chemical classes containing known chromophores capable of influencing VIS-EM reflectance.

One specific DEKM is kaempferol-3-rhamnoside-7-rhamnoside a yellow flavonoid previously isolated from delphiniums (Kim *et al.*, 2019). This DEKM is down regulated in Arabidopsis treated with AN and AS and up regulated with CAN compared to H₂O. Yellow bands VIS-EM (560-590 nm) regions are highlighted by our ensemble regression analysis in Arabidopsis and lawn grass. Another DEKM, 3,8-divinyl protochlorophyllide is involved in chlorophyll and porphyrin biosynthesis (Nagata, Tanaka and Tanaka, 2007) and was upregulated in grass treated with AS vs H₂O. Additionally, anthrone noranthrone was also upregulated in AS vs H₂O, this DEKM is likely to be yellow as metabolites in the same structural family are yellow (Morse, 1947). Another yellow DEKM was the carboxylic acid 1-O-sinapoyl- β -D-glucose, this DEKM was upregulated in CAN against H₂O and downregulated in AS and AN compared to CAN. Interestingly, 1-O-sinapoyl- β -D-glucose is known to be involved in phenylpropanoid synthesis (Strack, 1982). Phenylpropanoids are the starting point for the biosynthesis of flavonols, isoflavonoids and anthocyanins (Nanda *et al.*, 2017), all of which are known to regulate plant pigments across the spectra from purple and blue through to red and yellow (Forkmann and Martens, 2001; Katsumoto *et al.*, 2007; Nakatsuka *et al.*, 2007; Chiu *et al.*, 2010; Falcone Ferreyra, Rius and Casati, 2012).

In conclusion, we have developed a sensitive multispectral image analysis machine learning classifier capable of distinguishing plants exposed to various nitrogen sources that are known to be involved in the production of illicit substances. We found that multiple regions within the VIS-EM are critical for facilitating nitrogen class distinction in a way that more simple reflectance ratios were incapable of doing. Exposure to different nitrogen sources induces differential expression of an array of metabolites including chromophore containing molecules. This helps to explain why the ensemble regression analysis identified numerous spectral regions across the VIS-EM as key for distinguishing nitrogen treatments.

We developed the system to facilitate detection and monitoring of illicit activities that utilise nefarious nitrogen sources, such as explosives and narcotic manufacture. Our results suggest significant potential for developing lab and remote sensing-based solutions to help extend this more fully. Also, this study highlights how plants interact differently with different nitrogen sources altering leaf biochemistry and suggests variable bioassimilation of nitrogen compounds. This could have important consequences in future studies of ecological and agricultural systems and biogeochemical cycling.

Materials and Methods

Plant Growth:

Three different plant species were used in this study; *Arabidopsis thaliana* Col-0 (Arabidopsis), *Solanum lycopersicum* moneymaker, Suttons Seeds - tomato and mix of *Panicum* present UK lawn grass seed mix (Multipurpose lawn seed, Evergreen). Arabidopsis seeds were placed in 0.01% agar (Sigma Aldrich, UK) then at 4°C for 3 days before sowing, grass and tomato were sown from the packet. Seeds were planted in John Innes No.1, calcerite and perlite (6:1:1). Plants were grown under a 16-hour light dark cycle, with illumination of $150\mu E\ m^{-2}s^{-1}$ at 23°C and watered when needed. 1-week-old grass, 1.5-week-old Arabidopsis and 2-week-old tomato were watered with a single 4mM application of either calcium ammonium nitrate ((NO₃)₃CaNH₄), ammonium sulphate ((NH₄)₂SO₄), ammonium nitrate (NH₄NO₃) or water. One-week later leaves chosen at random or whole plants in each sample group were used for reflectance analysis or frozen in liquid nitrogen for metabolomic analysis.

Reflectance spectrometry:

Hyperspectral reflectance data was collected from 200 - 1025nm at a 0.4nm resolution, with a Flame-S-XR1, Ocean Optics, using a constant light source (DH-2000-BAL, Ocean Optics) and Ocean view software (Ocean Optics). The software was set to average 10 acquisitions for each spectrum recorded correcting for electric dark and non-linearity with a boxcar width of 2. 100 readings were taken per treatment group per experiment.

Hyperspectral data analysis was performed in R (Team, 2017), including and the adapted ensemble regression approach (Feilhauer, Asner and Martin, 2015). The ensemble regression was adapted to identify wavelength variance within and across species that significantly differed between nitrogen treatments. The reflectance intensity of the sample is first converted to a weighted coefficient before regression so a reflectance shift from 100 to 110 is weighted greater than a shift of 1000 to 1010, as the former is a 10% increase and the latter a 1% increase. Wavelengths were then selected for separate species if the absolute sum of regression coefficients is greater than two-standard deviations from the mean. The consensus all species bands were selected if the absolute sum of regression coefficients of all experimental replicates in the condition was greater than two-standard deviations from the mean.

Multispectral image acquisition:

Multispectral Images were taken in direct or simulated sunlight (iPhone 5s, Apple) using colour pass filters placed in front of the phone camera lens with no further modification. The colour filters used were #12, #23, #39, #65, #3407, #4990, #2005, #3441, #3202, #4290, #4690, #4460, #4790, #318, #376, #41, #378, #77, #93, #2002 (roscolux). Filters were chosen that selectively filtered across EM regions of interest identified in the regression analysis. Several filters cover redundant regions but also selectively filter within different EM regions. This results in an information rich data cube where single channels highlight intensity variance across some but not all treatment groups.

Multispectral image data cubes were constructed by registering images. Image registration was performed in Matlab. The images were separated into red, green or blue single-channel tif images. Tif images were then filtered by eye selecting only tif images with meaningful contrast to the background and/or between leaves across treatments. Selected tifs were stacked into a multispectral image stacks and processed in Ilastik (Sommer *et al.*, 2011). A pixel level classifier was built to distinguish plant tissue from background. The background channel data was discarded, and the leaf channel data segmented into separate objects ($\sigma = 1.2$, making each leaf or grass sward a separate object). An object classifier was then used to classify leaves by nitrogen treatment which was trained using 2-6 regions of grass swards; tomato leaves; or Arabidopsis leaves from each treatment. This training data was then used to predict the treatment of the remaining leaves in the image using a random forest classifier using single and neighbourhood pixel ($x = 40$, $y = 40$) intensities, defects and sharpness. This was used to generate predictions for each nitrogen treatment for the unlabelled image leaf objects.

Metabolomics:

1-week after the application of nitrogen treatments leaves were frozen in liquid nitrogen. Metabolites were extracted from either single frozen tomato leaves; 1g of pooled frozen grass leaves; or 1g of pooled frozen Arabidopsis. Metabolite extraction was performed according to De Vos *et al.* (De Vos *et al.*, 2007) modified by extending the acetonitrile gradient to 100%. Extracted metabolites were run on HPLC-(129-, Agilent, Santa Clara, CA), Q-TOF-MS (6650, Agilent). Each nitrogen treatment had 3 biological replicates per experiment with 2 experimental replicates overall.

Metabolomics data was analysed using mass profiler professional (Agilent). Identified metabolites were filtered to have a count value greater than 5000 and a fold-change greater than 2 compared to the control; the control used was CAN against AN and AS or water as a control against CAN, AN and AS. Masses were then filtered to be significant after a one-way ANOVA ($P < 0.05$). Filtering was performed to ensure the differentially expressed metabolite was present in the above conditions in all 6 biological replicates from the 2 experimental replicates for the nitrogen treatment for the plant species. This was done for all three species separately.

Hierarchical clustering was performed on the significant differentially expressed monoisotopic masses in each plant species with Euclidian similarity distance and Ward's linkage rule using CAN as a control against AN and AS or water as a control against CAN, AN and AS. Principal component analysis was performed on differentially expressed monoisotopic masses for each species. Differently expressed monoisotopic masses were used to query species specific databases (PlantCyc, mass tolerance = ± 50 ppm) (Schlöpfer *et al.*, 2017) to attempt to identify the metabolite.

References

- Anil Ananthaswamy (2004) *Defusing fertiliser may make bomb-building harder* | *New Scientist*, *NewScientist*. Available at: <https://www.newscientist.com/article/dn4782-defusing-fertiliser-may-make-bomb-building-harder/> (Accessed: 29 August 2019).
- Blackburn, G. A. (1998) 'Spectral indices for estimating photosynthetic pigment concentrations: A test using senescent tree leaves', *International Journal of Remote Sensing*. Taylor & Francis Group, 19(4), pp. 657–675. doi: 10.1080/014311698215919.
- Brust, H. *et al.* (2015) 'Isotopic and elemental profiling of ammonium nitrate in forensic explosives investigations', *Forensic Science International*. Elsevier, 248, pp. 101–112. doi: 10.1016/J.FORSCIINT.2014.11.024.
- Chiu, L.-W. *et al.* (2010) 'The purple cauliflower arises from activation of a MYB transcription factor.', *Plant physiology*. American Society of Plant Biologists, 154(3), pp. 1470–80. doi: 10.1104/pp.110.164160.
- CISA (2019) *Ammonium Nitrate Security Program* | *Homeland Security*. Available at: <https://www.dhs.gov/cisa/ammonium-nitrate-security-program> (Accessed: 29 August 2019).
- Climaco Pinto, R. *et al.* (2008) 'Using ANOVA-PCA for discriminant analysis: Application to the study of mid-infrared spectra of carraghenan gels as a function of concentration and temperature', *Analytica Chimica Acta*. Elsevier, 629(1–2), pp. 47–55. doi: 10.1016/J.ACA.2008.09.024.
- D. M. El-Shikha, D. M. *et al.* (2008) 'Remote Sensing of Cotton Nitrogen Status Using the Canopy Chlorophyll Content Index (CCCI)', *Transactions of the ASABE*. American Society of Agricultural and Biological Engineers, 51(1), pp. 73–82. doi: 10.13031/2013.24228.
- Djerdjev, A. M. *et al.* (2018) 'The mechanism of the spontaneous detonation of ammonium nitrate in reactive grounds', *Journal of Environmental Chemical Engineering*. Elsevier, 6(1), pp. 281–288. doi: 10.1016/J.JECE.2017.12.003.
- ECHA (2009) *Ammonium nitrate (AN) - Substances restricted under REACH - ECHA*. Available at: <https://echa.europa.eu/substances-restricted-under-reach/-/dislist/details/0b0236e1807e2f04> (Accessed: 8 October 2019).
- Eliasson, C., Macleod, N. A. and Matousek, P. (2008) 'Non-invasive detection of cocaine dissolved in beverages using displaced Raman spectroscopy', *Analytica Chimica Acta*. Elsevier, 607(1), pp. 50–53. doi: 10.1016/J.ACA.2007.11.023.
- Falcone Ferreyra, M. L., Rius, S. P. and Casati, P. (2012) 'Flavonoids: biosynthesis, biological functions, and biotechnological applications', *Frontiers in Plant Science*. Frontiers, 3, p. 222. doi: 10.3389/fpls.2012.00222.
- Federici, J. F. *et al.* (2005) 'THz imaging and sensing for security applications—explosives, weapons and drugs', *Semiconductor Science and Technology*. IOP Publishing, 20(7), pp. S266–S280. doi: 10.1088/0268-1242/20/7/018.
- Feilhauer, H., Asner, G. P. and Martin, R. E. (2015) 'Multi-method ensemble selection of spectral bands related to leaf biochemistry', *Remote Sensing of Environment*, 164, pp. 57–65. doi: 10.1016/j.rse.2015.03.033.
- Figuli, L. *et al.* (2016) 'Influence of homemade ammonium nitrate and fuel oil explosives charge shapes on blast wave propagation', *International Journal of Computational Methods and*

Experimental Measurements. WIT Press, 4(3), pp. 213–220. doi: 10.2495/CMEM-V4-N3-213-220.

Forkmann, G. and Martens, S. (2001) 'Metabolic engineering and applications of flavonoids', *Current Opinion in Biotechnology*. Elsevier Current Trends, 12(2), pp. 155–160. doi: 10.1016/S0958-1669(00)00192-0.

Foyer, C. H. (2018) 'Mitochondrial Redox State, Nitrogen Metabolism and Signalling', *Annual Plant Reviews online*, 42. doi: <https://doi.org/10.1002/9781119312994.apr0459>.

Fritz, C. *et al.* (2006) 'Regulation of secondary metabolism by the carbon-nitrogen status in tobacco: nitrate inhibits large sectors of phenylpropanoid metabolism', *The Plant Journal*. John Wiley & Sons, Ltd (10.1111), 46(4), pp. 533–548. doi: 10.1111/j.1365-313X.2006.02715.x.

Harty, M. A. *et al.* (2016) 'Reducing nitrous oxide emissions by changing N fertiliser use from calcium ammonium nitrate (CAN) to urea based formulations', *Science of The Total Environment*. Elsevier, 563–564, pp. 576–586. doi: 10.1016/J.SCITOTENV.2016.04.120.

Henry Dodd (2018) *Why is it so easy to buy bomb ingredients?* - BBC News, BBC News. Available at: <https://www.bbc.co.uk/news/uk-42690177> (Accessed: 29 August 2019).

Jason M. Conder, *,† *et al.* (2003) 'Nondestructive, Minimal-Disturbance, Direct-Burial Solid-Phase Microextraction Fiber Technique for Measuring TNT in Sediment'. American Chemical Society . doi: 10.1021/ES0260770.

Johnson, E. L., Campbell, T. A. and Foy, C. D. (1997) 'Effect of soil pH on mineral element concentrations of two *Erythroxylum* species', *Journal of Plant Nutrition*. Taylor and Francis Inc., 20(11), pp. 1503–1515. doi: 10.1080/01904169709365352.

Jolliffe, I. T. (1982) 'A Note on the Use of Principal Components in Regression', *Applied Statistics*, 31(3), p. 300. doi: 10.2307/2348005.

Katsumoto, Y. *et al.* (2007) 'Engineering of the Rose Flavonoid Biosynthetic Pathway Successfully Generated Blue-Hued Flowers Accumulating Delphinidin', *Plant and Cell Physiology*. Narnia, 48(11), pp. 1589–1600. doi: 10.1093/pcp/pcm131.

Keith J. Albert, † *et al.* (2001) 'Field-Deployable Sniffer for 2,4-Dinitrotoluene Detection'. American Chemical Society . doi: 10.1021/ES010829T.

Kim, S. *et al.* (2019) 'PubChem 2019 update: improved access to chemical data', *Nucleic Acids Research*. Narnia, 47(D1), pp. D1102–D1109. doi: 10.1093/nar/gky1033.

Lee, T.-C. *et al.* (2018) 'Eliminating interference by anthocyanins when determining the porphyrin ratio of red plant leaves', *Journal of Photochemistry and Photobiology B: Biology*. Elsevier, 187, pp. 106–112. doi: 10.1016/J.JPHOTOBIO.2018.08.007.

Liu, H.-B. *et al.* (2006) 'Detection and identification of explosive RDX by THz diffuse reflection spectroscopy', *Optics Express*. Optical Society of America, 14(1), p. 415. doi: 10.1364/OPEX.14.000415.

De Lucia, F. C. *et al.* (2003) 'Laser-induced breakdown spectroscopy analysis of energetic materials', *Applied Optics*. Optical Society of America, 42(30), p. 6148. doi: 10.1364/AO.42.006148.

Mallette, J. R. *et al.* (2017) 'The isotopic fractionation of carbon, nitrogen, hydrogen, and oxygen during illicit production of cocaine base in South America', *Forensic Science International*. Elsevier, 270, pp. 255–260. doi: 10.1016/J.FORSCIINT.2016.10.016.

- Mauzerall, D. (1977) *Porphyryns, Chlorophyll, and Photosynthesis*. In: Trebst A., Avron M. (eds) *Photosynthesis I. Encyclopedia of Plant Physiology (New Series)*. 5th edn. Berlin: Springer, Berlin, Heidelberg. doi: https://doi.org/10.1007/978-3-642-66505-9_5.
- Morse, E. E. (1947) 'Anthrone in Estimating Low Concentrations of Sucrose', *Analytical Chemistry*. American Chemical Society, 19(12), pp. 1012–1013. doi: 10.1021/ac60012a021.
- Nagata, N., Tanaka, R. and Tanaka, A. (2007) 'The Major Route for Chlorophyll Synthesis Includes [3,8-divinyl]-chlorophyllide a Reduction in *Arabidopsis thaliana*', *Plant and Cell Physiology*, 48(12), pp. 1803–1808. doi: 10.1093/pcp/pcm153.
- Nakatsuka, T. et al. (2007) 'Production of red-flowered plants by genetic engineering of multiple flavonoid biosynthetic genes', *Plant Cell Reports*. Springer-Verlag, 26(11), pp. 1951–1959. doi: 10.1007/s00299-007-0401-0.
- Nanda, S. et al. (2017) 'Metabolic Engineering of Phenylpropanoids in Plants', in. Springer, Cham, pp. 485–510. doi: 10.1007/978-3-319-28669-3_30.
- Pagola, M. et al. (2009) 'New method to assess barley nitrogen nutrition status based on image colour analysis: Comparison with SPAD-502', *Computers and Electronics in Agriculture*. Elsevier, 65(2), pp. 213–218. doi: 10.1016/J.COMPAG.2008.10.003.
- Panche, A. N., Diwan, A. D. and Chandra, S. R. (2016) 'Flavonoids: an overview', *Journal of Nutritional Science*. Cambridge University Press, 5, p. e47. doi: 10.1017/jns.2016.41.
- Penuelas, J. et al. (1997) 'Estimation of plant water concentration by the reflectance Water Index WI (R900/R970)', *International Journal of Remote Sensing*. Taylor & Francis Group, 18(13), pp. 2869–2875. doi: 10.1080/014311697217396.
- Rama Rao, N. et al. (2008) 'Estimation of leaf total chlorophyll and nitrogen concentrations using hyperspectral satellite imagery', *The Journal of Agricultural Science*. Cambridge University Press, 146(1), pp. 65–75. doi: 10.1017/S0021859607007514.
- Ryder, A. G., O'Connor, G. M. and Glynn, T. J. (2000) 'Quantitative analysis of cocaine in solid mixtures using Raman spectroscopy and chemometric methods', *Journal of Raman Spectroscopy*. John Wiley & Sons, Ltd, 31(3), pp. 221–227. doi: 10.1002/(SICI)1097-4555(200003)31:3<221::AID-JRS518>3.0.CO;2-5.
- Schläpfer, P. et al. (2017) 'Genome-Wide Prediction of Metabolic Enzymes, Pathways, and Gene Clusters in Plants', *Plant Physiology*, 173(4), pp. 2041 LP – 2059. doi: 10.1104/pp.16.01942.
- Schlemmer, M. R. et al. (2005) 'Remotely Measuring Chlorophyll Content in Corn Leaves with Differing Nitrogen Levels and Relative Water Content', *Agronomy Journal*. American Society of Agronomy, 97(1), p. 106. doi: 10.2134/agronj2005.0106.
- Schott, J. R. (John R. (2007) *Remote sensing : the image chain approach*. Oxford University Press.
- Serrano, L., Filella, I. and Pen˜uelas, J. (2000) 'Remote Sensing of Biomass and Yield of Winter Wheat under Different Nitrogen Supplies', *Crop Science*. Crop Science Society of America, 40(3), p. 723. doi: 10.2135/cropsci2000.403723x.
- Smith, M.-L. et al. (2003) 'Analysis of hyperspectral data for estimation of temperate forest canopy nitrogen concentration: comparison between an airborne (aviris) and a spaceborne (hyperion) sensor', *IEEE Transactions on Geoscience and Remote Sensing*, 41(6), pp. 1332–1337. doi: 10.1109/TGRS.2003.813128.

- Sommer, C. *et al.* (2011) 'Ilastik: Interactive learning and segmentation toolkit', in *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, pp. 230–233. doi: 10.1109/ISBI.2011.5872394.
- Stitt, M. (1999) 'Nitrate regulation of metabolism and growth', *Current Opinion in Plant Biology*. Elsevier Current Trends, 2(3), pp. 178–186. doi: 10.1016/S1369-5266(99)80033-8.
- Strachan, I. B., Pattey, E. and Boisvert, J. B. (2002) 'Impact of nitrogen and environmental conditions on corn as detected by hyperspectral reflectance', *Remote Sensing of Environment*. Elsevier, 80(2), pp. 213–224. doi: 10.1016/S0034-4257(01)00299-1.
- Strack, D. (1982) 'Development of 1-O-sinapoyl- β -D-glucose: L-malate sinapoyltransferase activity in cotyledons of red radish (*Raphanus sativus* L. var. *sativus*)', *Planta*. Springer-Verlag, 155(1), pp. 31–36. doi: 10.1007/BF00402928.
- Swain, K. C., Jayasuriya, H. P. W. and Salokhe, V. M. (2007) 'Suitability of low-altitude remote sensing images for estimating nitrogen treatment variations in rice cropping for precision agriculture adoption', *Journal of Applied Remote Sensing*. International Society for Optics and Photonics, 1(1), p. 013547. doi: 10.1117/1.2824287.
- Team, R. C. (2017) 'R: A language and environment for statistical computing', *R Foundation for Statistical Computing*, Vienna, Au.
- The Government Inspector of Explosives, Ireland, I. (2016) *Guide to Explosives Legislation In Ireland*. Ireland.
- U.S. Department of Justice (1991) *Coca Cultivation and Cocaine Processing: An Overview*.
- De Vos, R. C. *et al.* (2007) 'Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry', *Nature Protocols*. Nature Publishing Group, 2(4), pp. 778–791. doi: 10.1038/nprot.2007.95.
- W. C. Bausch, W. C. and H. R. Duke, H. R. (1996) 'Remote Sensing of Plant Nitrogen Status in Corn', *Transactions of the ASAE*. American Society of Agricultural and Biological Engineers, 39(5), pp. 1869–1875. doi: 10.13031/2013.27665.
- Wren, S. P. *et al.* (2014) 'Preparation of novel optical fibre-based Cocaine sensors using a molecular imprinted polymer approach', *Sensors and Actuators B: Chemical*. Elsevier, 193, pp. 35–41. doi: 10.1016/J.SNB.2013.11.071.
- Zhu, Y. *et al.* (2008) 'Analysis of common canopy vegetation indices for indicating leaf nitrogen accumulations in wheat and rice', *International Journal of Applied Earth Observation and Geoinformation*. Elsevier, 10(1), pp. 1–10. doi: 10.1016/J.JAG.2007.02.006.
- Zwanenburg, G. *et al.* (2011) 'ANOVA-principal component analysis and ANOVA-simultaneous component analysis: a comparison', *Journal of Chemometrics*. John Wiley & Sons, Ltd, 25(10), pp. 561–567. doi: 10.1002/cem.1400.

Supplementary information

Supplementary figure 1. There is low visible hyperspectral reflectance pattern variance between across nitrogen sources Mean reflectance intensity of the electromagnetic spectrum from 400-700nm for **a, b & c** tomato **d, e & f** Arabidopsis and **g, h & i** grass one week after application of either 4mM of AN (green), CAN (red), AS (yellow) or H₂O (blue) (1000 readings, 24 plants or grass swards per nitrogen treatment, 0.4 nm resolution, boxcar width 2).

Supplementary figure 2. The majority of variance across data is captured in the first principal component. Principal component analysis was performed on the hyperspectral reflectance data from 400-700nm one week after application of either 4mM of AN (green), CAN (red), AS (yellow) or H₂O (blue), 1000 readings taken from 24 plants per treatment. The percentage explained variance across all data captured by each principle component is shown for **a, b & c** Tomato; **d, e & f** Arabidopsis; and **g, h & i** grass. Three separate replicates shown for each species.

Supplementary figure 3. Hyperspectral variance reveals differences in plants treated with different nitrogen sources. Data for three replicates per species shown. ANOVA and Tukey's HSD following PCA of reflectance data (400-700 nm) performed on PCs 1-10 for **a, b & c** tomato; **d, e & f** Arabidopsis; and **g, h & i** lawn grass one week after application of either 4mM of AN (green), CAN (red), AS (yellow) or H₂O (blue), 1000 readings taken from 24 plants per treatment. ANOVA and Tukey's HSD following PCA of reflectance data (400-700 nm) performed on PCs 1-10. Lines between groups are labelled with PCs which contain variation that can significantly distinguish treatment group pairs. * indicates p -value for Tukey's HSD < 0.05 , ** < 0.01 and *** < 0.001 . With the exception of **e, g & h** ANOVA followed by Tukey's HSD was able to resolve all nitrogen treatment comparisons across all replicates and species.

Supplementary figure 4. Loadings of principle components that could significantly separate nitrogen sources capture variation from regions identified by machine learning regression. Principal component analysis was performed on the hyperspectral reflectance data from 400-700nm one week after application of either 4mM of AN (green), CAN (red), AS (yellow) or H₂O (blue), 1000 readings taken from 24 plants per treatment. **a, b & c** tomato; **d, e & f** Arabidopsis; and **g, h & i** lawn grass eigenvalue plots of PCs which could significantly explain variation between nitrogen treatments in hyperspectral data (400-700nm) after ANOVA and Tukey HSD test on single components ($p < 0.05$).

	Equation (nm)	Tomato						<i>Arabidopsis thaliana</i>						Grass					
		H ₂ O vs CAN	H ₂ O vs AN	H ₂ O vs AS	CAN vs AN	CAN vs AS	AN vs AS	H ₂ O vs CAN	H ₂ O vs AN	H ₂ O vs AS	CAN vs AN	CAN vs AS	AN vs AS	H ₂ O vs CAN	H ₂ O vs AN	H ₂ O vs AS	CAN vs AN	CAN vs AS	AN vs AS
RVI	800																		
	670	1	0.17	1	0.11	1	0.14	1	0.49	1	0.49	1	0.49	1	0.49	1	0.49	1	0.49
RVI2	870																		
	660	0.05	0.98	0.24	0.02	0.9	0.11	0.08	0.84	0.64	0.01	0.63	0.18	1	0.49	1	0.49	1	0.49
PSSRa	800																		
	680	0.8	0.94	0.79	0.99	1	0.99	0.96	0.96	0.61	0.74	0.89	0.3	1	0.49	1	0.49	1	0.49
PSSRb	800																		
	635	0.92	0.92	0.94	0.57	1	0.61	0	0.4	0.01	0	0.21	0.42	0.94	0.58	1	0.9	0.99	0.72
fWBI	900																		
	Λ(930 : 980)	0.99	0.02	0	0.01	0.01	0	0.44	0	0.08	0	0.8	0.05	0	1	1	0	0	1
PWI	970																		
	900	0.04	0.06	0.01	1	0.98	0.94	0.75	0.24	0.94	0.81	0.39	0.66	0.78	0.75	0.96	0.21	0.46	0.96

Key P ≤ 0.05 P ≤ 0.01 P ≤ 0.001

Supplementary figure 5. Plant reflectance indices are unable to consistently differentiate nitrogen treatments. Hyperspectral reflectance intensity was taken from either tomato, *Arabidopsis* and lawn grass leaves 1-week after application of either 4mM of AN, CAN, AS or H₂O, 100 readings from 24 plants or grass swards per nitrogen treatment, for 3 experimental replicates. Vegetation ratio indexes that monitor nitrogen content (RVI, RVI2), water content (PWI, fWBI) and chlorophyll content (PSSRa, PSSRb) were then calculated for each experiment. This produced 100 index values per treatment per experiment. Then a nested ANOVA and Tukey HSD test (*P* values shown) was performed on the index values for each species with the experiment used as the block.

Supplementary figure 6. Cropped images of leaves captured using different filters used to form a multispectral cube for analysis. Cropped images of randomly selected tomato leaves from 1 experiment taken 1 week after the application 4mM of NH_4NO_3 with colour pass filters that selectively transmit light in regions of interest. This illustrates the intensity diversity across filters channels in the multispectral cube. Images were registered to align and then split into individual colour pixel channels. The pixel channels without contrast were discarded. The grid of images was then be stacked into a multispectral cube for analysis.

Supplementary figure 7. Significant monoisotopic mass expression identified with untargeted metabolomics cluster after PCA dependant on nitrogen source applied to the plant. Untargeted metabolomics was performed on plant leaves one week after treatment with either 4mM AN, CAN, AS or H₂O according to De Vos et al., (De Vos et al., 2007) with the acetonitrile gradient modified to 100%. Leaf tissue extracts were run on a HPLC-TOF-MS. Monoisotopic masses of detected metabolites were filtered for count values >5000, fold change ≥ 2 and significant after a one-way ANOVA ($P < 0.05$) in all replicates per nitrogen treatment when compared to the control. The control was CAN for AN and AS or water as a control against CAN, AN and AS. PCA was performed on differentially expressed monoisotopic masses in each experiment. The PC scores for the first component were plotted against the second for each experiment, for **a & b** tomato; **c & d** Arabidopsis; and **e & f** grass. The replicates visibly clustered based on the nitrogen treatment applied. Two experimental replicates performed per species.

Author contributions

C.A. performed hyperspectral and multispectral data collection and data analysis, performed metabolite extractions and analyzed metabolomic data, helped design experiments, conceive the project and write the manuscript. S.C and D.B. performed the HPLC helped and advised with metabolomic data analysis. O.W. conceived the project, helped design experiments and wrote the paper.

Acknowledgements

This work is was supported by the following grants: BB/M011178/1 and NE/M018768/1.