Metabolomics data analysis of herbicide susceptible and resistant populations of black-grass (*Alopecurus myosuroides*)

M. Dickinson
The Food and Environment Research Agency, York, United Kingdom.

nonlinear DYNAMICS
A Waters Company

Overview

- A user-friendly workflow based data analysis approach has been used to process metabolite profile information from different sample sets over a time course study using Progenesis. Analysis took less than 2 hours per experiment for 43 high resolution MS data files (average file size 200MB).
- Multivariate statistics provided by Progenesis have been used to visually inspect the data and identify trends within this.
- 62 compounds with significantly different abundances between sample sets have been established from the analysis using the univariate statistics provided by Progenesis. A proportion of the compounds have been assigned tentative identifications based on searches against an in-house database with MetaScope.
- Data reproducibility of significant compounds found is good with low % CV's in a pooled QC sample (23 <30% and 39 <10% CV).

Introduction

The control of grass weeds in wheat represents one of the greatest challenges to sustainable intensification in arable agriculture in Northern Europe[1]. Since the early 1980s, this problem has been compounded by the rapid spread of herbicide resistance in the problem weed black-grass (*Alopecurus myosuroides*)[2]. Infestation with herbicide resistant black-grass has now been reported at over a 1000 sites in the UK, with resultant crop losses calculated at between 15-35% due to yield and quality losses[2]. Prior to 1982, the major type of resistance determined in black-grass was due to mutations in the proteins targeted by herbicides, notably acetyl CoA carboxylase (ACCase) and photosystem II (PSII). Termed target site resistance (TSR), such mutations render herbicides less effective as inhibitors. TSR in black-grass now extends to the more recently introduced sulphonyurea herbicides which act on acetolactate synthase (ALS). In addition to TSR, the last 3 decades have seen a proliferation in an additional type of resistance termed non-target site, or multiple herbicide resistance (MHR)3. MHR causes an increased resistance to herbicides irrespective of their mode of action, apparently due to an enhanced rate of their detoxification3. MHR is very difficult to control, as unlike TSR it cannot be overcome by rotating graminicide mode of action1. As such the UK is now witnessing a proliferation of herbicide resistance in black-grass and a diminishing access to herbicide chemistries as a consequence of EU environmental directives which now threaten further sustainable intensification in cereal production[2].

In the face of herbicide resistance, farmers now need rapid and robust diagnostic technologies in order to deploy alternative weed control strategies within a growing season. For example recognizing the difference between TSR and MHR would inform the choice of which, if any, herbicides should be applied, or whether alternative control measures are required. Currently, assays for herbicide resistance consist of collecting weed seed, growing on seedlings and spray trials with different graminicides; a process which typically takes 6 weeks and can only inform control options for the following season. This note describes a data analysis approach using Progenesis for the initial metabolomic profiling analysis between TSR, MHR and control (susceptible) black grass.

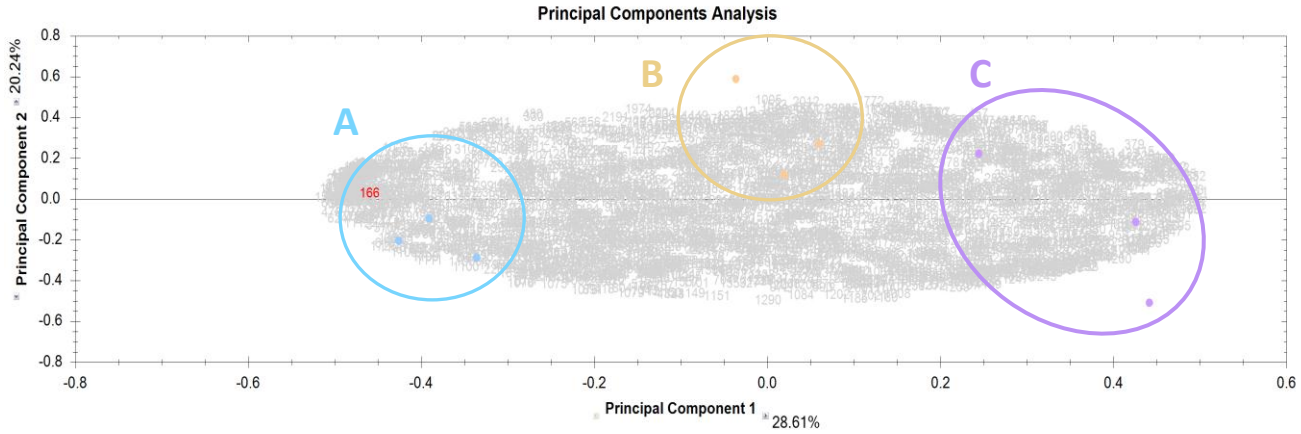

Figure 1: PCA showing the metabolic relationship between black-grass plants of varying herbicide resistance 13 days after the herbicide application based on data-analysis with Progenesis™ software. Key: **(A)** Herbicide Susceptible (SUS); **(B)** Multiple herbicide resistance (MHR) and **(C)** Target site resistant (TSR). Compounds having significant abundance were selected for validation using NMR and / or MS/MS.

Nonlinear Dynamics Ltd.
Keel House, Garth Heads, Newcastle upon Tyne, NE1 2JE, UK **[T]** +44 (0) 191.230.2121 **[F]** +44 (0) 191.230.2131
Americas Sales Office
2530 Meridian Parkway, 3rd Floor, Durham, NC 27713 U.S.A.
[T] +1 919.806.4401 **[TOLL FREE]** +1 866.435.7872 **[F]** +1 919.806.4301 **[E]** info@nonlinear.com **[W]** www.nonlinear.com

Method

36 black grass plants were grown under controlled conditions in a greenhouse at the Food and Environment Research Agency for 19 days. The samples included 12 x MHR type, 12 x TSR type and 12 x Susceptible (Sus). After 19 days of growth the samples were sprayed with the acetyl-coA carboxylase inhibitor Clodinafop-Propargyl at field spray rates. Samples were collected at time points Day 0 (pre spray), Day 4, Day 8 and Day 12 post spray with each time point having 3 biological replicates.

Samples were immediately frozen after harvest in liquid nitrogen and freeze dried for 48 hours. Dried material was ground into a fine powder and extracted into 1:1 methanol: water. Extracts were analysed in a random order by Liquid Chromatography-High Resolution Mass Spectrometry (LC-HRMS). LC analysis was performed on an Accela 1250 High Speed LC system from Thermo Fisher Scientific (Waltham, Massachusetts, USA). Extracts were analysed in a random order, separated on a ACE AQ (Advanced Chromatography Technologies, UK) 150mm x 3 mm , 300Å and detected on a Orbitrap Velos Pro hybrid mass spectrometer (ThermoFisher, Waltham, Massachusetts, USA) set at 50,000 resolution at 250 m/z in both positive and negative mode (separate runs). For quality control a pooled extract of each sample was prepared and analysed every 6 injections.

Including QC, 43 raw thermo data files (each with an approximate size of 200 MB) were uploaded into Progenesis® software for both ionisation modes. The data from each mode were analysed using two separate Progenesis experiments. For each experiment the data was uploaded, visually inspected for any obvious anomalies, retention time aligned (using a quality control sample) and all files put through the peak picking process. A limit of 3000 absolute ion intensity was applied in the peak picking algorithm. For compound identification all detected features were searched, using MetaScope, against an in-house accurate mass database of approximately 300 compounds relating to plant stress (using a 5 ppm mass tolerance on theoretical mass). Data was analysed using the statistics included in Progenesis (Principal Components Analysis (PCA), ANOVAs and mean fold change between groups) for various experimental designs to assess the difference between grass lines and time. Compounds of interest were discovered by filtering on p ≤0.01, fold change ≥ 2 and %CV between replicates < 30%. .

Results

After adduct deconvolution 2,741 compounds were found in the positive mode data of which 131 were identified from Metascope. 1,187 compounds were found in the negative mode data of which 62 compounds were identified from Metascope. The easy to use automatic workflow of Progenesis took approximately 2 hours per experiment to complete the whole analysis. Figures 2 and 3 below describe two PCA from all profile positive mode data where some initial definitions of sample type can be seen.

Figure 2 shows all data separated by colour according to sample line. Figure 3 is identical data but with sample groups defined according to the time experiment only. The flexibility of altering experimental design in minutes and generating new statistics greatly eases the challenge of exploring complicated data sets such as this. The overall repeatability of the QC response was good which can also be visually assessed on the PCA.

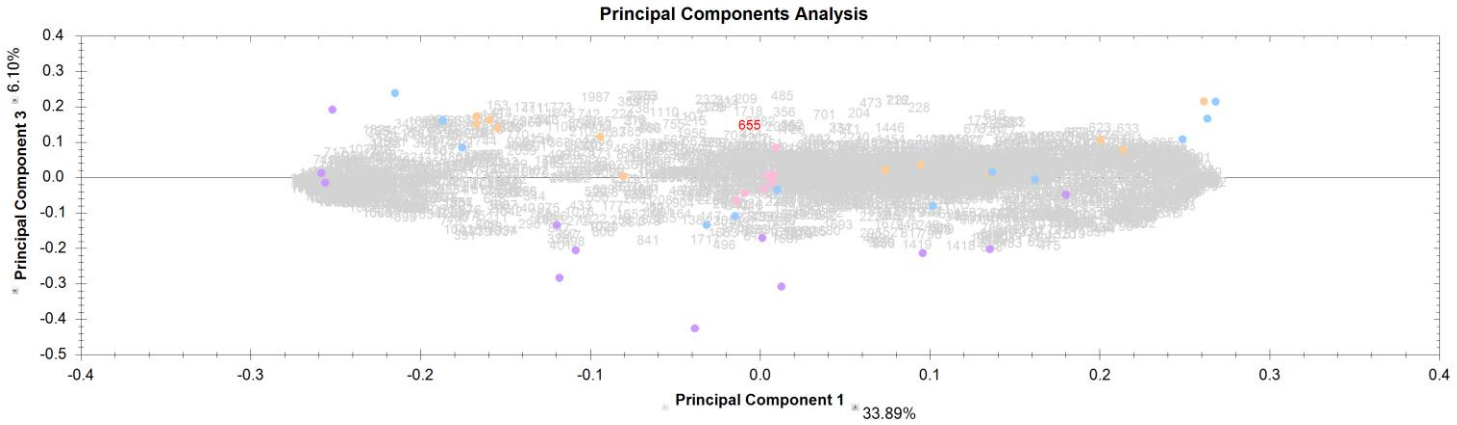

Figure 2. PCA of all positive mode data based on groups selected by plant phenotype.
Key: Purple=TSR, Orange=MHR, Dark Blue = Susceptible plants, and Pink = QC. Data is not filtered and includes all features detected.

Nonlinear Dynamics Ltd.
Keel House, Garth Heads, Newcastle upon Tyne, NE1 2JE, UK **[T]** +44 (0) 191.230.2121 **[F]** +44 (0) 191.230.2131
Americas Sales Office
2530 Meridian Parkway, 3rd Floor, Durham, NC 27713 U.S.A.
[T] +1 919.806.4401 **[TOLL FREE]** +1 866.435.7872 **[F]** +1 919.806.4301 **[E]** info@nonlinear.com **[W]** www.nonlinear.com

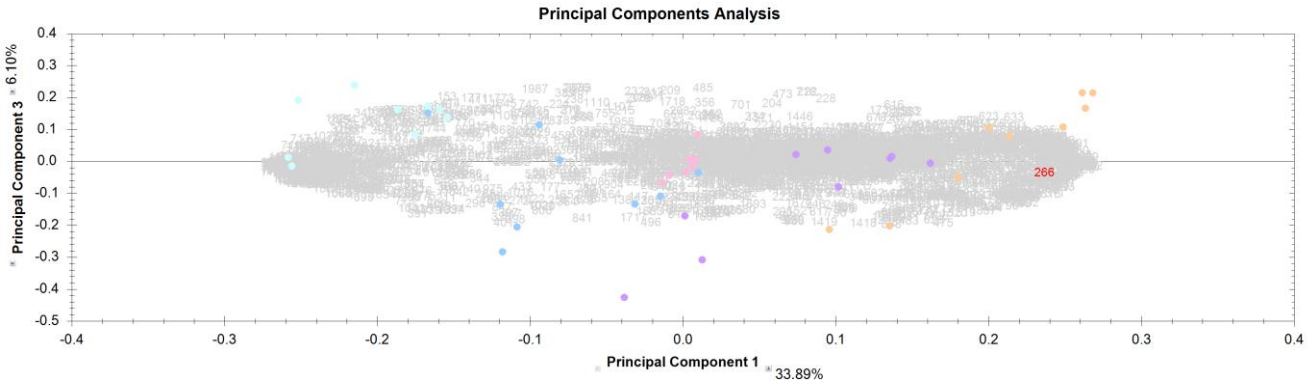

Figure 3. PCA of all positive mode data based on groups selected by day of sample collection.
Key: Light Blue = Day 0, Dark Blue = Day 4, Purple=Day 8, Orange=Day 13 and Pink = QC. Data is not filtered and includes all features detected.

Once all the data was processed changing the experimental design was quick and easy. A number of different designs were therefore produced and the statistics updated within seconds to reflect potential significant compounds between different groups. For example, Figure 4 describes a PCA of only day 13 positive mode data looking at the metabolic relationship between TSR, MHR and Susceptible plant lines 13 days after the herbicide application. Clear distinctions can be seen between the grass types.

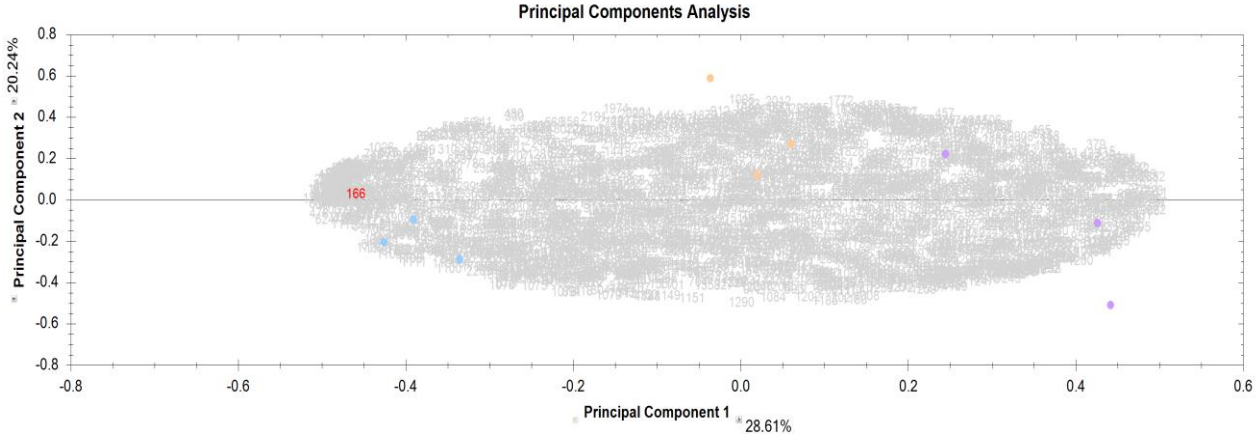

Figure 4. PCA of all positive mode data based on samples collected on day 13 of study.
Key: Light Blue = Day 0, Dark Blue = Day 4, Purple=Day 8, Orange=Day 13 and Pink = QC. Data is not filtered and includes all features detected.

Compounds discovered to have significant abundances across the lines using univariate statistics were investigated further using different experimental designs across the different time points. For example, Figure 5 shows a compound abundance box plot for a compound discovered to be significantly differently expressed between sample lines on day 8 only. Visually the standard deviations are apparent between the replicates and it is clear that there is significantly less of this compound present in the TSR line. Figure 6 shows 3D peak profiles generated by the software to accentuate this conclusion and are shown overleaf.

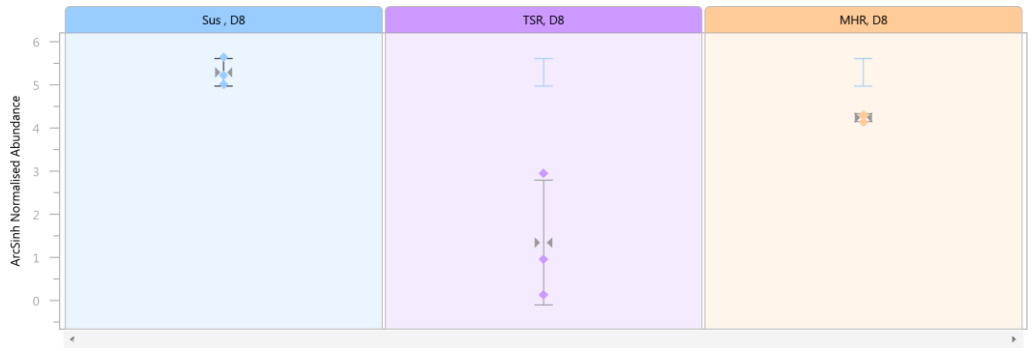

Figure 5. Box plot detailing compound abundances and standard deviations (samples from day 8 only, positive mode) between replicates for a compound already identified as significant from Progenesis.

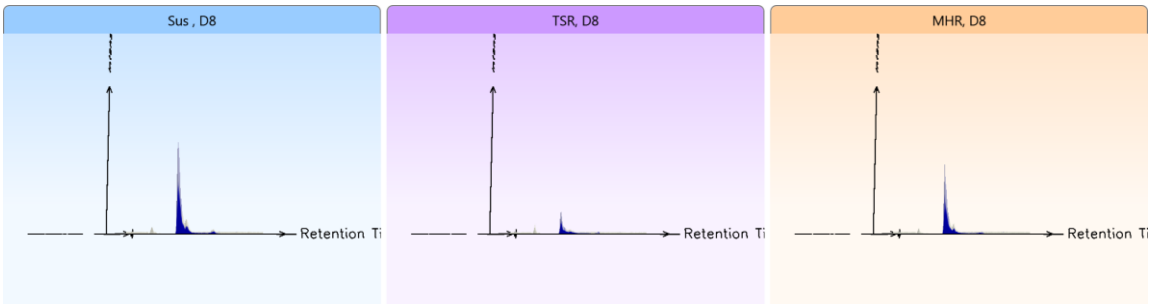

Figure 6. 3D peak montage for a compound discovered to be significantly differently expressed between sample lines on day 8 only.

Compounds discovered as having significant abundances between lines by Progenesis and tentatively identified by MetaScope were collated for further confirmation using NMR and / or MS/MS (15 compounds currently confirmed). The table below lists some of the details of the 62 compounds found, and in particular for quality purposes the reproducibility of the QC extracts (compound identities have had to be removed due to confidentiality purposes as data is not yet published).

Table 1. List of 62 differentially observed compounds

Cpd no (Pos).	Main adduct	RT	p-value	MF change	% CV QC	Highest mean	Lowest mean	ID Confirmed
1	M+H	1.8	<1.16 E-16	27	3	Sus	Day 0	
2	M+H	1.8	1.16E-16	3	1	Sus	Day 0	
3	M+H	3.1	9.24E-14	172	21	TSR	Day 0	
4	M+K	1.8	2.45E-12	5	5	Sus	Day 0	
5	M+NH4	1.8	2.51E-11	8	1	Sus	Day 0	Yes
6	M+H	2.1	3.63E-11	19	16	Sus	Day 0	
7	M+H	4.9	1.03E-10	13	3	TSR	Sus	Yes
8	M+H	1.7	1.39E-10	66	8	TSR	Sus	
9	M+H	14.3	2.56E-10	90	6	Day 0	TSR	
10	M+Na	2.1	6.58E-10	69	11	TSR	Sus	
11	M+H	2.1	1.32E-09	3	2	Sus	Day 0	
12	M+H	1.8	1.85E-09	4	19	Sus	Day 0	
13	M+H	5.0	2.97E-09	10	4	Sus	Day 0	
14	M+H	9.7	1.43E-08	7	3	Sus	Day 0	
15	M+Na	2.6	1.94E-08	34	13	TSR	Day 0	
16	M+H	11.0	2.28E-08	9	5	MHR	Day 0	
17	M+H	13.6	1.37E-07	8	10	Day 0	Sus	
18	M+NH4	2.1	4.49E-07	284	11	Sus	Day 0	
19	M+H	1.9	4.74E-07	17	26	MHR	TSR	
20	M+H	10.6	7.19E-07	54	5	Sus	Day 0	Yes
21	M+H	10.6	1.46E-06	249	9	Sus	Day 0	Yes
22	M+H	14.7	1.96E-06	6	7	Day 0	Sus	
23	M+H	2.4	2.16E-06	9	4	Sus	MHR	
24	M+H	10.2	2.36E-06	2040	8	Sus	Day 0	Yes
25	M+H	3.5	2.82E-06	5	3	Sus	Day 0	
26	M+H	11.4	3.96E-06	4	6	Day 0	Sus	
27	M+H	10.6	4.03E-06	29	4	Sus	Day 0	Yes
28	M+H	2.1	6.76E-06	2	3	Sus	Day 0	
29	M+H	2.3	7.75E-06	4	5	Sus	Day 0	
30	M+Na	10.6	9.02E-06	3	7	Sus	MHR	Yes
31	M+H	10.0	1.04E-05	264	3	Sus	Day 0	Yes
32	M+Na	21.5	1.27E-05	20	28	Day 0	Sus	
33	M+NH4	3.5	2.67E-05	3	5	Day 0	Sus	
34	M+H	1.6	2.72E-05	2	10	MHR	TSR	
35	M+H	12.1	3.32E-05	2	5	Day 0	Sus	
36	M+NH4	2.0	6.86E-05	2	12	Day 0	Sus	Yes
37	M+H	2.1	1.17E-04	5	14	Day 0	MHR	
38	M+H	1.7	4.29E-04	3	15	MHR	TSR	
39	M+2H	2.1	1.28E-03	4	10	Sus	TSR	Yes
40	M+H	1.9	2.02E-03	2	16	TSR	MHR	
NEG								
1	M-H	14.6	1.10E-16	596	22	Day 0	TSR	
2	M-H	2.1	4.87E-13	187	22	Sus	Day 0	
3	M-H20-H	2.1	9.65E-11	24	8	Sus	Day 0	
4	M-H	11.0	2.47E-10	188	4	MHR	Day 0	
5	M-H20-H	1.8	5.55E-10	10	9	Sus	Day 0	Yes
6	M-H	1.7	1.11E-07	2	5	Day 0	Sus	
7	M-H	2.8	2.04E-07	82	11	Sus	Day 0	Yes
8	M-H	19.9	4.01E-07	3	14	Day 0	Sus	
9	M-H	2.0	8.76E-07	5	11	Day 0	MHR	
10	M-H20-H	2.3	5.40E-06	12	10	Sus	Day 0	
11	M-H20-H	12.2	1.03E-05	4	18	Sus	MHR	
12	M-H	10.0	1.19E-05	1700	7	Sus	Day 0	
13	M-H	3.5	1.86E-05	2	3	Sus	Day 0	
14	M-H	1.8	5.53E-05	5	12	Day 0	Sus	
15	M-H20-H	10.6	7.94E-05	2810	18	Sus	Day 0	Yes
16	M-H	10.8	1.38E-04	5	21	MHR	Day 0	
17	M-H	2.1	1.39E-04	5	10	Day 0	Sus	
18	M-H	4.7	4.94E-04	3	4	Day 0	MHR	
19	M-H	2.1	8.75E-04	3	9	MHR	Day 0	
20	M-H	11.3	9.14E-04	5	7	MHR	Day 0	
21	M-H	10.6	1.27E-03	Infinite	18	Sus	Day 0	Yes
22	M-H	2.1	7.65E-03	2	4	MHR	TSR	Yes

Nonlinear Dynamics Ltd.
Keel House, Garth Heads, Newcastle upon Tyne, NE1 2JE, UK **[T]** +44 (0) 191.230.2121 **[F]** +44 (0) 191.230.2131
Americas Sales Office
2530 Meridian Parkway, 3rd Floor, Durham, NC 27713 U.S.A.
[T] +1 919.806.4401 **[TOLL FREE]** +1 866.435.7872 **[F]** +1 919.806.4301 **[E]** info@nonlinear.com **[W]** www.nonlinear.com

Conclusions

- A user friendly workflow based data analysis approach has been used to process metabolite profile information from different sample sets over a time course study. Analysis took less than 2 hours per experiment for 43 high resolution MS data files (average file size 200MB).

- Multivariate statistics provided by Progenesis software have been used to visually inspect the data and identify trends within this.

- Compounds with significantly different abundances have been established from the analysis using the univariate statistics provided. 62 of the compounds have been assigned tentative identifications based on an in-house database used with MetaScope. Of these 15 have had their identities confirmed using analytical standards.

- Data reproducibility of significant compounds found is good with low % CV's in a pooled QC sample (23 <30% and 39 <10% CV).

- Potential metabolic biomarkers have been initially identified between different lines of herbicide resistant and susceptible *Alopecurus myosuroides,* before and after herbicide application. Further validation and experimental work will be undertaken to confirm the potential markers discovered in this study.

References

1. Cummins, I. and Edwards, R. (2010) The biochemistry of herbicide resistance in weeds. Outlooks on Pest Management. 21, 73-77.
2. Moss S (2012) Developing and promoting more sustainable grass-weed control strategies to combat herbicide resistance. Defra report PS2714.
3. Cummins, I. , Bryant, D.N. and Edwards, R. (2009) Safener responsiveness and multiple herbicide resistance in the weed black-grass (*Alopecurus myosuroides*). Plant Biotechnology Journal. 7, 807-820.

Supplement to the Application Note

Using Progenesis QI to validate putative compound identities

Since the publication of the original application note above, Progenesis QI has been updated to include the ability to confirm the identity of compounds using their MS2 product spectra.

The following extension to this application note describes the procedures used to confirm the identity of significantly changing metabolites tentatively identified from the initial profiling experiment as previously described.

Method

Black-grass extracts were analysed using Liquid Chromatography-High Resolution Mass Spectrometry (LC-HRMS). LC analysis was performed on an Accela 1250 High Speed LC system from Thermo Fisher Scientific (Waltham, Massachusetts, USA). Selected extracts for confirmation purposes were separated on a ACE AQ (Advanced Chromatography Technologies, UK) 150mm x 3 mm , 300Å and detected on an Orbitrap Velos Pro hybrid mass spectrometer (ThermoFisher, Waltham, Massachusetts, USA). A precursor list of potentially significant accurate masses was generated from the profiling analysis and used as precursor ions for MS/MS experiments in both positive and negative ionisation modes. Collision Induced Dissociation (CID) fragmentation was employed using 35 ev. A fragmentation event was only triggered if the precursor mass had a signal of 1000 (au) or greater. MS2 FTMS detection was used in order to obtain accurate mass product ions. A corresponding analytical standard was analysed when available to help confirm compound identity.

Sample and analytical standard data files containing MS2 accurate mass product ion spectra were uploaded into two separate new QI experiments, to cover both negative and positive ionisation modes. Sample alignment and peak picking with the highest sensitivity setting (5 out of 5) were performed in minutes, as not all 43 samples needed to be re-analysed for compound confirmation purposes.

As with the initial profiling experiment, tentative identifications were again acquired using an in-house accurate mass database of plant metabolites. The known analytical standards were then viewed separately (using the experiment design setup feature) with any identifications from these files accepted by ticking the gold star next to the potential identification cell. Once the identification had been established in the analytical standard each compound was separately tagged and included in a list of "identified standard compounds" only. This list was then exported as a fragment database .msp file.

Once the .msp file had been established this was then recursively applied (again using MetaScope) to all sample data files to confirm or reject any tentative identifications from

preliminary accurate mass only searching. As analytical standards had now been analysed retention time information could also be included in this secondary identification step.

For compounds tentatively identified in the first accurate mass search, where an analytical standard could not be sourced or the particular significant *m/z* had a very large list of potential identifications, QI gave a second alternative. A list of structures from the Human Metabolome Database (HMDB) is freely available to download (http://www.hmdb.ca/downloads) and was added to the MetaScope tool for a third search through the MS2 data. When a file (.sdf) containing a list of compounds is included in the search (such as the HMDB structures) theoretical fragmentation patterns are generated which allowed the comparison of these with MS2 information acquired from the samples. Identifications from theoretical fragmentation offer a likely identification but are not unequivocal.

Results

As stated in the original application note 15 compounds had their identities confirmed using MS/MS or nuclear magnetic resonance spectroscopy (NMR) (LC-HRMS/MS data analysed using Xcalibur from ThermoFisher, Waltham, Massachusetts, USA and Mass Frontier from HighChem Ltd, Bratislava, Slovakia). Using the approaches described above, using Progenesis QI, 5 compounds had their identities unequivocally identified with others affirmatively identified using theoretical fragmentation. Figures 1a-f show 6 compounds with analytical standards MS2 product spectra against the sample metabolite MS2 spectra. All but one produced a fragmentation match score of > 70 indicating an unequivocal identification for at least 5 out of the 6 compounds.

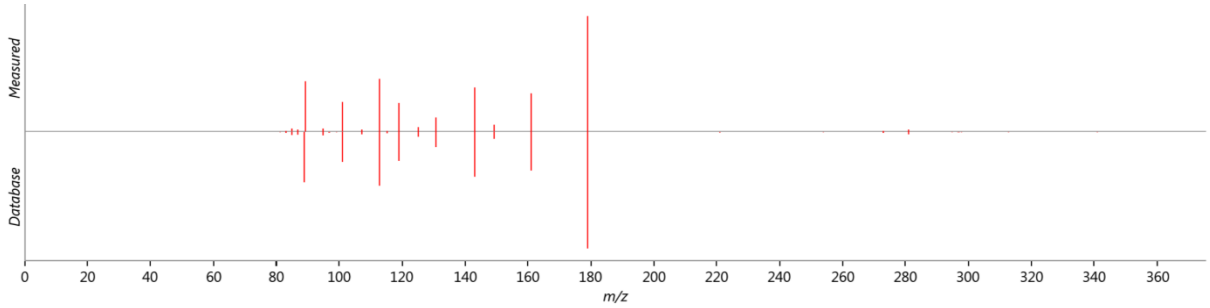

Figure 1 a) Confirmation trace for *m/z* 341.1105, cpd 2 on Table 1 (neg list). Fragmentation score: 100.

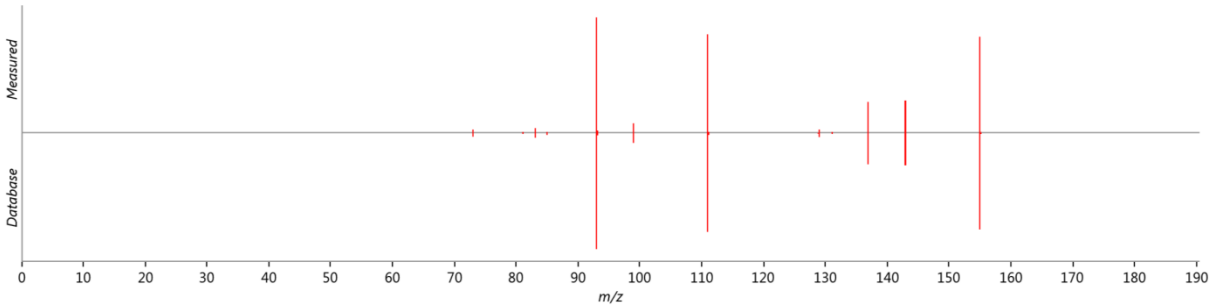

Figure 1 b) Confirmation trace for *m/z* 173.0457, cpd 19 on Table 1 (neg list). Fragmentation score: 100.

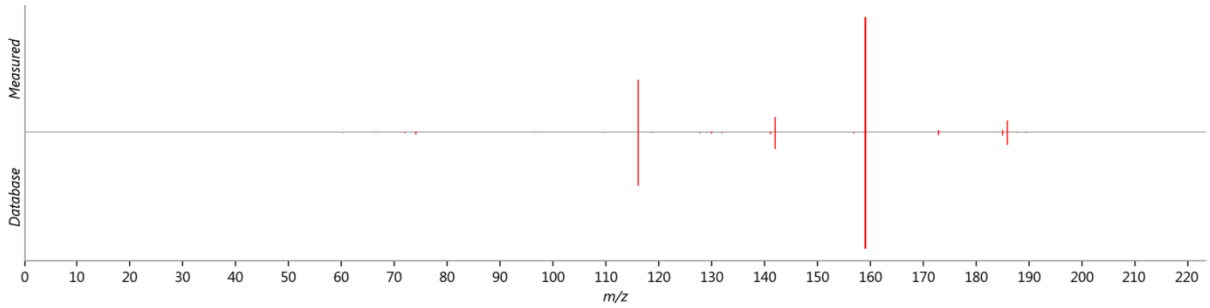

Figure 1 c) Confirmation trace for *m/z* 203.0831, cpd 21 on Table 1 (neg list). Fragmentation score: 100.

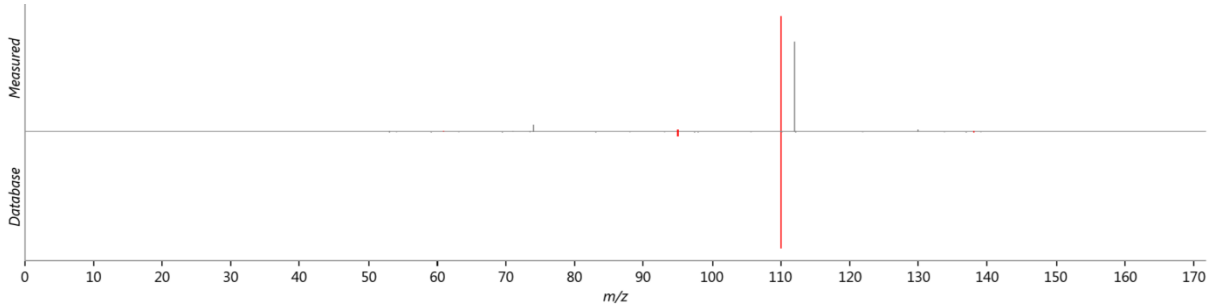

Figure 1 d) Confirmation trace for *m/z* 156.0766, cpd 8 on Table 1 (pos list). Fragmentation score: 71.

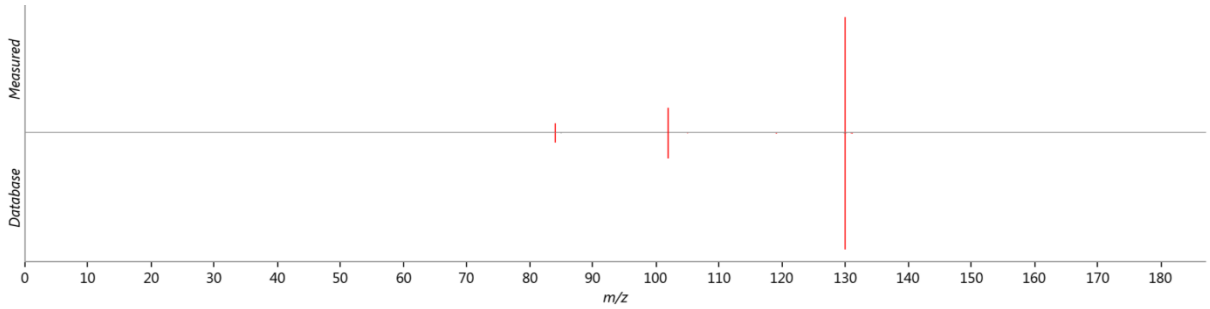

Figure 1 e) Confirmation trace for *m/z* 148.0603, cpd 4 on Table 1 (pos list). Fragmentation score: 98.

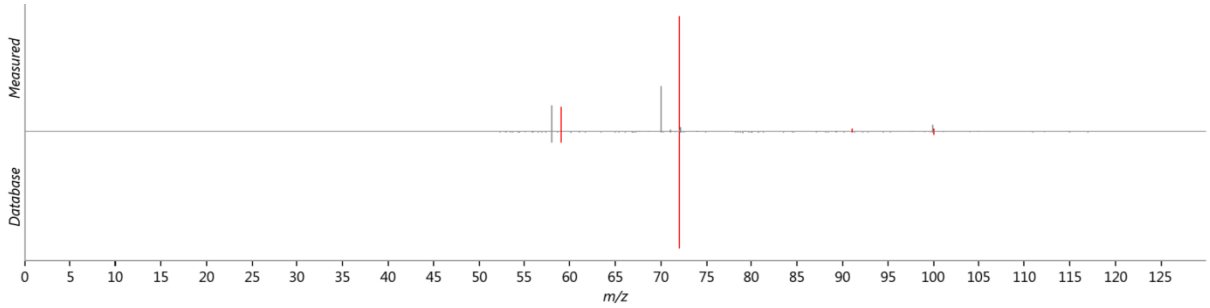

Figure 1 f) Confirmation trace for *m/z* 118.0860, cpd 11 on Table 1 (pos list). Fragmentation score: 60.

Nonlinear Dynamics Ltd.
Keel House, Garth Heads, Newcastle upon Tyne, NE1 2JE, UK **[T]** +44 (0) 191.230.2121 **[F]** +44 (0) 191.230.2131
Americas Sales Office
2530 Meridian Parkway, 3rd Floor, Durham, NC 27713 U.S.A.
[T] +1 919.806.4401 **[TOLL FREE]** +1 866.435.7872 **[F]** +1 919.806.4301 **[E]** info@nonlinear.com **[W]** www.nonlinear.com

For identification purposes an overall identification "score" is produced. This is generated from accurate mass information within the MS1 data, isotopic distribution fit, retention time data (which can be included into the MetaScope search feature described), fragmentation score and collision cross section (CSS) information (not used in this application).

Using the theoretical fragmentation approach, compound 9 from Table 1 in the previous application note was assigned an identification from the HMDB. As shown in figure 2 the fragmentation score is low (40.9) and therefore the identification could only be classed as tentative. The identity score, 45.8 in this case, is awarded for each possible compound which can help to prioritise future identification validation. Although this identification would still be classed as very tentative, this approach gives much greater detail to the data in a matter of minutes. As seen in figure 2, theoretical fragments are also identified with theoretical mass error of the fragment also provided.

Figure 2. Tentative identifications produced from matching theoretical fragmentation patterns generated from Progenesis QI with accurate mass MS2 sample spectra. Compounds for identification / theoretical fragmentation sourced from HMDB.

Conclusions

- **New identification feature tools in Progenesis QI allow metabolites to be easily and quickly identified from a pre-determined list of potentially significant compounds.**

- **The use of Progenesis MetaScope as a search tool employing theoretical fragmentation using the HMDB, for example, can filter out or include possibilities in order to pre-select for future confirmation analysis, thus saving time and potential expense on standards.**

- **The clear fragmentation scoring and visual comparison tools from a generated fragmentation file quickly confirm or discount identifications.**

- **The fragmentation patterns of compounds were consistent whether observed in the standard or matrix greatly improving the ease of identification.**

- **In this application 5 out of 6 compound identities were confirmed using Progenesis QI, agreeing with identifications made from either other vendor software by a MS/MS approach or NMR.**

Nonlinear Dynamics Ltd.
Keel House, Garth Heads, Newcastle upon Tyne, NE1 2JE, UK **[T]** +44 (0) 191.230.2121 **[F]** +44 (0) 191.230.2131
Americas Sales Office
2530 Meridian Parkway, 3rd Floor, Durham, NC 27713 U.S.A.
[T] +1 919.806.4401 **[TOLL FREE]** +1 866.435.7872 **[F]** +1 919.806.4301 **[E]** info@nonlinear.com **[W]** www.nonlinear.com