

ASSESSING HYPERSPECTRAL AND METABOLOMIC METHODOLOGIES FOR
HERBICIDE RESISTANCE DETECTION IN MONTANAN *AVENA FATUA*

by

Connor Lochlann Nelle

A thesis submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Plants Science

MONTANA STATE UNIVERSITY
Bozeman, Montana

May 2026

©COPYRIGHT

by

Connor Lochlann Nelle

2026

All Rights Reserved

ACKNOWLEDGEMENTS

This thesis was made possible by the collaboration and hard work of many people. Thank you Dr. Jennifer Lachowiec for your unwavering support and mentorship. Thank you Dr. Barbara Keith for the many years of work and the sage advice throughout this research. Thank you Dr. Paul Nugent, Dr. Bill Dyer, and Dr. Tim Seipel for advising and supporting me. To Dr. Donald Smith and Jesse Thomas at the MSU Mass Spectrometry Facility, thank you for the many hours training me and processing samples. Thank you to all the members of the Lachowiec and Nugent labs, the student workers, the MSU Plant Growth Center, and the folks at the Western Triangle Agricultural Research Center/Post Farm, who dedicated their time to this research. I would also like to thank Kaelin Smith and Lucas Wright for the many hours spent helping me.

To my fellow students, thank you for creating a welcoming workplace full of comradery. Thank you to my family for always looking out for me and for feeding me motivation. A very special thank you to Callan Malone for her belief in my ability to complete this degree, her steadfast support, and for a few late nights in the greenhouse.

Finally, this research would not have been possible without several sources of funding from organizations: USDA, Montana INBRE/RAIN Tech Access, and the Montana Wheat and Barley Committee.

TABLE OF CONTENTS

1. INTRODUCTION.....	1
Overview of Herbicide Resistance	1
Metabolomics	3
Hyperspectral Imagery.....	4
Wild Oat (<i>Avena fatua</i>)	5
Research Objectives	7
2. METABOLOMIC ANALYSIS OF HERBICIDE RESISTANT <i>AVENA</i> <i>FATUA</i>	9
Introduction.....	10
Methods	14
Creation and Phenotyping of RIL Populations.....	14
Plant Growth Conditions and Tissue Collection.....	18
Sample Preparation and Mass Spectrometry	19
Post-processing of Mass Spectrometry Data.....	20
Analysis of Metabolomics Data.....	21
Results.....	23
Experiment One	24
Experiment Two.....	26
Experiment Three.....	31
Discussion	37
Future Hypotheses	44
Conclusion.....	45
3. BENCHTOP HYPERSPECTRAL IMAGERY TO DISCRIMINATE AMONG MULTIPLE HERBICIDE RESISTANCE BIOTYPES OF AVENA FATUA.....	46
Summary	46
Introduction.....	47
Methods	51
Plant Growth.....	51
Benchtop Hyperspectral Imaging	52
Spectral Post-processing	52
Modeling And Analysis.....	53
Results.....	56
Two-class Discrimination.....	56
Biotype Discrimination	58
Distance Metrics	60

TABLE OF CONTENTS CONTINUED

Discussion	61
Conclusion	66
 4. HYPERSPECTRAL ANALYSIS IN RECOMBINANT INBRED LINES OF HERBICIDE RESISTANT <i>AVENA FATUA</i>	67
Summary	67
Introduction	68
Methods	71
Benchtop Imagery	71
Airborne Hyperspectral Imagery	71
Post-processing of Airborne Hyperspectral Imagery	73
Analysis	74
Results	76
Discussion	82
Data source causes strong spectral effects	82
Different models may pick up on subtle spectral signal	83
Does the mechanism of resistance in 06R4 and 93R confer a spectral signature?	84
Conclusion	85
 5. ASSESSMENT OF METABOLOMIC AND HYPERSPECTRAL APPROACHES FOR HERBICIDE RESISTANCE PHENOTYPING IN <i>AVENA FATUA</i>	86
Metabolomics	86
Hyperspectral Imagery	87
Implications for Resistance Screening	88
REFERENCES CITED	90
APPENDICES	103
SUPPLEMENTAL FIGURES	104

LIST OF TABLES

Table	Page
1. Table 2.1: Collection and resistance information of wild oat biotypes.	14
2. Table 2.2: Resistant and susceptible RILs selected for 06R x S2 metabolomic analysis, binary resistance classification indicated by letter in parentheses.	17
3. Table 2.3: Resistant and susceptible RILs selected for 93R x S2 metabolomic analysis, binary resistance classification indicated by letter in parentheses.	17
4. Table 2.4: Solvent flow over LC-MS injection.	20
5. Table 2.5: Annotations for selected 06R x S2 experiment metabolites. The mean retention time and mean M/z values were calculated between Experiment 1 and Experiment 2.	31
6. Table 2.6: Spectra, FDR adjusted p-value, and log ₂ (fold change) of metabolites retained after filtering for consistent abundance shifts between resistance groups.	36
7. Table 3.1: Counts of greenhouse grown biotypes imaged between A and B replicates.	51
8. Table 3.2: Model assessment metrics and tuning parameters for resistant vs. susceptible models.	58
9. Table 3.3: Model performance and tuning parameters for biotype classification average across outer folds	59
10. Table 3.4: Resistance phenotypes of herbicide resistant wild oat.	62
11. Table A1: Number of samples and respective metabolite counts after filtering for metabolomic experiments.	105
12. Table A2: Model sensitivity, specificity, precision, and support for biotype prediction models.	107

LIST OF FIGURES

Figure	Page
1. Figure 2.1: Metabolite features detected in Experiment 1 vary in expression between S2 and 06R. Features are colored by $\log_2$ (fold change) with less abundant features in 06R shaded blue and more abundant features shaded red. The size of the circle indicates the FDR corrected p-value.	24
2. Figure 2.2: Clustering of 06R and S2 samples by top metabolomic features. Columns represent biotypes samples as indicated by 06R (red) and S2 (green). Metabolite features are shown across rows following the naming scheme “Retention time “_” M/z. Relative abundance per individual samples (Exp. 1) are shown in each cell with positive values (purple) indicating higher abundance as compared to the “mean” of all plants from Experiment 1 for that metabolite feature.	25
3. Figure 2.3: Differential abundance of metabolic features across 06R versus S2 and herbicide resistant versus sensitive RILs. $\log_2$ (fold change) of metabolic features is indicated in the cells, with more positive values (red) indicating higher abundance in R biotypes and more negative values (blue) indicating lower abundance in R biotypes. Relative abundance for each herbicide tested based on recategorized RILs is shown in the latter four columns (see Table 1.2).	27
4. Figure 2.4: Boxplots of mass normalized abundance per metabolite, across all experiments and RILs for 06R/S2. RILs are colored by flucarbazone percent biomass of control. Batch is the order of the experimental days the sample was collected on.....	29
5. Figure 2.5: Boxplots of putatively annotated tri-acylglycerol metabolites mass-normalized abundance across all experiments and RILs for 06R/S2. RILs are colored by tralkoxydim percent biomass of control. Batch is the order of the experimental days the sample was collected on.	30
6. Figure 2.6: Volcano plot of metabolite features detected in Experiment 3 93R and S2 biotypes. The direction of comparison is 93R/S2. Features are colored by $\log_2$ (fold change) and sized by FDR corrected P-value.	32

LIST OF FIGURES CONTINUED

7. Figure 2.7: Clustering of 93R and S2 samples by top features. Columns represent samples, with 93R samples colored red and S2 samples colored green. Features are shown in rows.....33
8. Figure 2.8: Heatmap of $\log_2(\text{fold change})$ values for the 93R and S2 biotypes and each herbicide response group (triallate and difenzoquat). Metabolites were selected for plotting if they differed significantly between the 93R and S2 biotypes or between the resistant and susceptible RILs for either herbicide response.35
9. Figure 2.9: Boxplots of mass normalized abundance per feature for the 93R/S2 experiments. RILs are colored by the percent biomass of control for triallate, and the points are shaped by the order of experimental day the sample was collected on.37
10. Figure 3.1: A) Hyperspectral analysis flow chart and B) schematic for nested cross validation.....55
11. Figure 3.2: Mean reflectance of biotypes and mean first derivative spectra of biotypes.....56
12. Figure 3.3: Confusion matrices for generalized resistance class models. The top-left to bottom right diagonal indicates correct predictions by each model.57
13. Figure 3.4: Confusion matrices summed across outer test folds and repeats for biotype classification models. Top-left to bottom-right diagonal indicates correct predictions.59
14. Figure 3.5: Canberra and cosine distance matrices for biotypes. Distance is represented by a heatmap, with color intensity of the square representing similarity between the corresponding biotypes on the axis.61
15. Figure 4.1: Georectified reflectance datacubes captured on 07-09-2025. Polygons used to extract RIL spectra are overlain for the 06R (red) and 93R (purple) RIL plants.....73

LIST OF FIGURES CONTINUED

16. Figure 4.2: UMAP of convolutional bottleneck features for model trained on flucarbazone percent biomass of control across all 06R4 x S2 datasets. Data source was visualized by shape, and the percent biomass of control for flucarbazone per RIL observation was visualized with color. Purple indicates a susceptible RIL and yellow indicates a resistant RIL.....	77
17. Figure 4.3: UMAP of convolutional bottleneck features for model trained on mean difenzoquat rating across all 93R x S2 datasets. Data source is visualized by shape and the mean difenzoquat rating for each RIL is visualized by color. Purple indicates a susceptible RIL and yellow indicates a resistant RIL.....	78
18. Figure 4.4: Spearman's correlation between predicted and observed herbicide response metric for airborne and benchtop datasets. The percent biomass of control is the response for flucarbazone, pinoxaden, tralkoxydim, and triallate. The mean visual rating is the response for difenzoquat.....	79
19. Figure 4.5: Scatterplot of triallate resistance predictions for 06R4 benchtop data. Each point represents a prediction of a RIL that occurred outside the training folds within the nested cross-validation.....	80
20. Figure 4.6: Scatterplot of difenzoquat resistance predictions for 93R replicate B benchtop data.....	81
21. Figure 4.7: Spearman's correlation between predicted and observed herbicide response metric for pixel-level airborne and benchtop datasets.....	82
22. Figure A1: MS/MS spectra for conjugate features in the 06R biotype experiments.....	106
23. Figure A2: 1D convolutional neural net architecture used to develop models predicting herbicide resistance in recombinant inbred lines.....	108

ABSTRACT

Producers spend millions of dollars each year combating herbicide resistant wild oat, often requiring high rates of application to successfully control populations. We used recombinant inbred lines to assess the viability of both untargeted metabolomics and hyperspectral imagery as potential diagnostic tools to identify resistant biotypes of wild oat. Untargeted metabolomics identified several features consistently more abundant in susceptible than resistant plants between both field-collected biotypes and their derived recombinant lines. Identification of these metabolites could provide insights into mechanisms and lead to development of resistance screening assays. Hyperspectral imagery was able to discriminate between field-collected resistant and susceptible strains, though it struggled to discriminate among similar resistance phenotypes. Within the recombinant lines, hyperspectral approaches could not predict resistance, suggesting that different computational methods are necessary to predict resistance. These insights into pre-spraying resistance detection methods are invaluable for directing future research in herbicide resistance detection in wild oat and other resistant weeds.

CHAPTER ONE

INTRODUCTION

Overview of Herbicide Resistance

Global population is projected to reach 10 billion by 2050, and food demand is projected to increase 50-60 percent over food production in 2019, with demands for multiple major essential grains increasing by up to 95 percent (Falcon et al. 2022). Essential grains such as wheat, barley, and oat are critical for both direct consumption and feed production; increasing yields for these crops requires synthesis of biotechnology, genetics, and management practices specific to cropping environments (Parry and Hawkesford 2010; Albahri et al. 2023). Crop losses due to weed pressure are predicted to maintain at 28 percent under projected environmental changes in CO₂ levels and drought (Vilà et al. 2021). Herbicide use is a necessity for effective control of weeds and preservation of crop yields. However, overuse of herbicides has led to evolution of herbicide-resistant weeds (Gaines et al. 2020). While prudent application of herbicides alongside integrated management practices can slow the evolution of herbicide resistance (Ofosu et al. 2023), development of new herbicides is extremely slow and costly (He et al. 2022). Thus, the development of technologies that reduce selection pressure for herbicide resistance and increase the efficiency of herbicide application contributes to addressing resistant weed infestations and thus global food security.

Herbicide resistance can be conferred in multiple ways, making control measures and discovery of mechanism-of-resistance challenging. Genetic changes that confer herbicide resistance can be split into the categories target site resistance and non-target site resistance.

Target site resistance stems from gene mutations that prevent herbicides from binding to target sites or result in gene amplification and overexpression of target enzymes (Devine 1997; Powles and Yu 2010; Gaines et al. 2020). Non-target site resistance manifests as reduction in herbicide uptake, metabolism of herbicides in the plant, lowering of the rate of herbicide activation, and sequestration of herbicides (Powles and Yu 2010). Mechanisms of non-target site resistance can be challenging to discover, and the co-occurrence of target site and non-target site resistance further complicates identifying sources of herbicide resistance (Jugulam and Shyam 2019).

Applying herbicides is time consuming and costly for stakeholders and can be rendered ineffective if the targeted plants are resistant. Overreliance and overuse of herbicides apply strong selection pressure for herbicide resistance in weeds. If an herbicide is to remain a potent tool for producers, herbicide applications should have high success rates. Detection of herbicide resistance prior to application would facilitate targeted herbicide applications. This could reduce selection pressure for resistance, mitigate off-target effects, and reduce producer expenditures. Rapid herbicide resistance detection approaches using biochemical, molecular, analytical, and genetic approaches (Beffa et al. 2012; Han et al. 2016; Kutasy et al. 2021) have been developed for a variety of weeds; however the toolkit for pre-spray detection remains relatively underdeveloped. Within the field of precision agriculture, hyperspectral imagery has shown promise for resistance detection in several weeds (Scherrer et al. 2019; Mensah et al. 2025), though the ability of these approaches to differentiate among many herbicide resistant phenotypes is unstudied. We assessed both untargeted metabolomics and hyperspectral imagery as methods for rapid pre-spray detection of herbicide resistance.

Metabolomics

Originally coined by Oliver et al. (1998), a metabolome can be defined as “the complete complement of small molecules present in an organism” (Oliver et al. 1998; Hall 2006). Untargeted metabolomics specifically refers to a methodological approach which seeks to obtain a comprehensive view of a metabolome as opposed to quantification of a selected subset of metabolites. Plant metabolomics is an established approach for genotype-to-phenotype association, species discrimination, population screening, physiological pathway inference, and many other applications (Hall 2006; Ma and Qi 2021). Alongside other -omics approaches, such as transcriptomics and proteomics, metabolomics has become a powerful tool for assessing herbicide response. A review of integration using metabolomic and transcriptomic approaches applied to herbicide resistant weeds by Sen et al. (2023) identified nine studies which used metabolomic approaches to assess herbicide response in resistant weeds (Sen et al. 2023). These studies generally focused on exploring metabolomic differences between biotypes after herbicide application and identifying mechanisms of resistance rather than identifying potential constitutive differences between resistant and susceptible plants. However, metabolomics could also be used to rapidly identify resistance if the abundance of certain metabolites could be linked to a resistance mechanism before application of herbicides. The biomarker use-case for metabolomics has been demonstrated in the medical field for human diseases: metabolomics has been used to assess risk, treatment response, and disease presence for cancers and other human health factors (Günther 2015; Luque de Castro and Priego-Capote 2018).

Of the common platforms for metabolomics, liquid chromatography-mass spectrometry (LC-MS) is among the most widely used and has relatively high sensitivity and simple sample

preparation (Fang and Gonzalez 2014). LC-MS is often used for profiling unknown features, and metabolite identification can be aided by the acquisition of MS/MS spectra (Xiao et al. 2012). Chromatography and spectrometry methods can be fine-tuned to focus on resolution of different areas of the metabolome. After an untargeted metabolomics method has been selected and data collected from LC-MS runs has been processed, multivariate statistics can reveal metabolomic markers which can then be annotated, facilitating the creation of targeted metabolomic protocols which collect qualitative concentrations of target compounds (Allwood et al. 2021).

Hyperspectral Imagery

Though metabolomics offers detailed insight into compound abundance and plant phenotypes, it may be difficult to scale across many herbicide resistant phenotypes. Metabolomics requires time-consuming sample collection and expensive instruments. In contrast to metabolomics, the field of precision agriculture is already established in cropping environments worldwide. The bloom of precision agriculture in the 1990s, enabled by GPS, GIS, sensors, control of system inputs (fertilizers, pesticides, irrigation, tillage, etc.), and computing, introduced a new paradigm for managing crops (Pierce and Nowak 1999). However, it is undeniable that present-day ability to collect and effectively leverage gargantuan amounts of data has once again transformed the landscape of precision agriculture. Sensor networks, cloud computing, and the application of powerful artificial intelligence are among many fields of research enabling management at extremely local scales (Shafi et al. 2019; A. Sharma et al. 2021). Many imagery-based approaches are being used in herbicide resistant systems in order to detect herbicide damage, assess resistant weeds post-herbicide application, and develop new modes of action (Niu et al. 2025). One technology used to quantify subtle plant reflectance

characteristics is hyperspectral imaging. Hyperspectral approaches capture reflectance values across large wavelength ranges (~400-1000 nanometers) with narrow spectral bands (~2 nanometers), resulting in extremely large images known as data cubes (Sarić et al. 2022).

In plants, remote, non-destructive proxies for other plant traits have been developed from hyperspectral imagery. Hyperspectral imagery has been used for metabolite prediction (Vergara-Diaz et al. 2020; Shu et al. 2025), nutrient, sugar, and carbohydrate abundance (Grzybowski et al. 2021), water stress detection (Hurley et al. 2019), and disease detection (Nesar et al. 2025), among many other applications. Critically, hyperspectral imagery is effectively integrated into unoccupied aerial systems (UASs), taking advantage of UASs' ability to fly over large areas, making hyperspectral imagery a relatively scalable technology spatially. Most current efforts using UASs for weed management are focused on discriminating weeds from crops (Wang et al. 2019). Pre-application detection of herbicide resistance using spectral methods are anchored in these weed detection approaches. If weeds are already segmented from crops and soil using computer vision methods, then models predicting resistance can act on the segmented images. Hyperspectral imagery has been used to classify herbicide resistant biotypes pre-application (Scherrer 2019; Mensah et al. 2025); however, it is unclear if current models directly capture herbicide resistance signal in spectra, making these approaches generalizable to new environments.

Wild Oat (*Avena fatua*)

Wild oat (*Avena fatua*) is a winter annual weedy grass which affects many crops, including wheat, barley, legumes, and peas (Beckie et al. 2012a; Boyd et al. 2022). Herbicide resistant wild oat is a problem worldwide: populations in Canada (Beckie et al. 2012b; Mangin et al. 2017), the

United States (Kern et al. 2002; Keith et al. 2015), Mexico (Torres-García et al. 2018), Australia (Owen and Powles 2009; Ahmad-Hamdani et al. 2012, 2013), and China (Han et al. 2023) have screened positive for resistance to herbicides, often exhibiting cases of multiple herbicide resistance. Not including the above countries, fifteen other countries have reported resistant wild oat according to the International Herbicide-Resistant Weed Database (Heap 2026). Wild oat populations can be extremely difficult to eradicate due to seed shatter, extended seed dormancy periods of around nine years, and the ability for seeds to self-bury (Somody et al. 1985; Van Acker 2009; Beckie et al. 2012a). Wild oat is a major problem for producers worldwide and will continue to be as food and feed demand increases.

To assess both untargeted metabolomics' and hyperspectral imagery's abilities to leverage constitutive differences of herbicide resistant weeds to detect resistance pre-application, we utilized several resistant and susceptible lines of wild oat and their derived recombinant inbred lines. The use of recombinant inbred lines (RILs) was critical to infer whether metabolites and spectral profiles were related to resistance of the plant rather than indicative of non-specific phenotypic differences between resistant and susceptible biotypes. Recombinant inbred lines are most commonly used to associate phenotypic traits to causative genetic loci; in other words to map quantitative genetic loci (Bailey 1971; Pollard 2012). By crossing a resistant strain with a susceptible strain of wild oat and selfing the offspring over many generations, a population of lines with stable genomes is created. The resulting population has a mosaic of the parental genomes and is assumed to be nearly homozygous. In genetic mapping populations, including RIL populations, if the abundance of metabolites or spectral signatures are highly correlated with resistance, both are underpinned by the same genetic loci. Prior metabolomic and hyperspectral

studies on herbicide resistance in weeds have compared unrelated biotypes, making it difficult to parse out metabolites and spectra from diverse genetic backgrounds. The use of RILs allows us to assume a degree of causality when assessing relationships between herbicide resistance and metabolites/spectra.

Research Objectives

This thesis has three objectives:

1. Determine if any metabolites have consistent abundance shifts between both parental biotypes and the derived recombinant inbred lines from two resistant biotypes of wild oat using Liquid Chromatography-Mass Spectrometry (LC-MS). These metabolites could be related to resistance and possibly leveraged in rapid detection methods.
2. Determine if several resistant and susceptible wild oat biotypes can be discriminated with benchtop hyperspectral imagery. This would show that resistant biotypes may have unique spectral signatures which could be leveraged for detection.
3. Determine if hyperspectral imagery can be used to predict resistance in two populations of recombinant inbred lines. This can determine if models built using hyperspectral data are detecting fingerprints indicative of resistance. If models can detect a spectral fingerprint, then detection of resistance in field settings is a more attainable goal.

There is a need for precise and efficient weed management approaches. The results of this research can be used to direct future efforts in rapid detection of herbicide resistance. Metabolite patterns and spectral signatures in RILs will have important implications for how pipelines for rapid resistance detection are built. This research lays the foundation for developing a program

which maps herbicide resistance in Montanan wild oat, an invaluable tool for producers facing constant pressure.

CHAPTER TWO

METABOLOMIC ANALYSIS OF HERBICIDE RESISTANT

AVENA FATUA

This study aimed to explore untargeted metabolomics as a method for detection of herbicide resistance between resistant and susceptible lines of wild oat. An untargeted metabolomics protocol was used to assess metabolomic differences in semi-polar and non-polar metabolites between two biotypes of multiple herbicide resistant wild oat and a susceptible biotype. The same protocol was used to assess metabolomic differences within extreme phenotypes selected from populations of recombinant inbred lines derived from both resistant biotypes and the susceptible biotype. Metabolite features that had consistent shifts in abundance in both the resistant-susceptible comparison and the derived RIL comparison were investigated as metabolomic markers possibly linked to resistance. Within the 06R x S2 biotype experiments, one feature was found to have lower abundance in both the 06R biotype and the resistant derived RILs. Two conjugates also differed in a similar pattern. Within the 93R x S2 biotype experiment, three features had consistent non-detections in the 93R biotype and derived RILs. Though these features also shared very similar abundance patterns, no evidence was found that they were conjugates. Features identified as potentially important were eluted very early in the LC-MS runs in all experiments, meaning they are likely semi-polar in nature. This also made annotation difficult as LC-MS/MS spectra were not collected for all features. This research contributes to the development of metabolomic assays for herbicide resistance by highlighting areas of the wild oat metabolome which may be of interest. Future research can improve these methods by

screening a denser array of recombinant inbred lines, tuning the LC-MS protocol for semi-polar metabolites, and by annotating important metabolites.

Introduction

Herbicides are a critical management tool for global agriculture, but their usefulness has been curtailed by the emergence of herbicide resistance. Following initial documentation in the 1950s, the number of herbicide resistant weeds has increased steadily (Heap 2026). Herbicide resistance causes yield loss and increases management costs (Heap 2014). Identification of herbicide resistance is important to control populations effectively and to work towards mechanistic understanding of resistance.

The standard protocols for detection of herbicide resistance in weeds are greenhouse bioassays. A suspected resistant weed is collected from the field, then grown for seed and screened against an array of herbicides to determine resistance (Seefeldt et al. 1995; Squires et al. 2021). While simple and effective, bioassays can be time consuming and costly. Management actions resulting from these methods often cannot be applied until the following growing season, allowing resistant progeny to establish. Alternative genotyping assays are being developed to decrease identification time for specific herbicide resistant genotypes (Grossmann 2005; Beffa et al. 2012). Han et al., (2016) screened for resistance in Australian *Lolium rigidum* populations using single gene sequencing (Han et al. 2016). KASP assays have been developed for resistant hybrid identification in *Myriophyllum spicatum* (Patterson et al. 2017). One downside to these approaches is that they cast a relatively shallow net, only screening for known mutations or genetic differences.

Metabolomic approaches show potential for detection of herbicide resistance and discovery of underlying mechanisms (Hall 2006). Metabolomics measures hundreds to thousands of metabolites to interrogate metabolic pathways. Several studies have used mass-spectrometry to compare herbicide abundance between resistant and susceptible weeds post-application with the goal of demonstrating enhanced metabolism (Beffa et al. 2012; Anagnostopoulos et al. 2020; Li et al. 2024). Previous studies using metabolomics reported the up-regulation of specific amino acid species and transport sugars, along with general metabolome alterations in response to herbicides (Sen et al. 2023; Jin et al. 2024).

In contrast, leveraging metabolomics for early detection of herbicide resistance requires sampling prior to application. While this has not been an explicit objective of studies, several studies have included relevant metabolomics on untreated resistant weeds. Maroli et al., (2015) found that metabolomic differences in *Amaranthus palmeri* were much greater post-glyphosate treatment, though some metabolomic separation between untreated control plants was observed, possibly related to phenylpropanoid abundance in the resistant biotype, which may be complementary to antioxidative resistance mechanisms (Maroli et al. 2015). A similar study by Zulet-Gonzalez et al. (2023) reported that metabolite profiles between resistant and susceptible *A. palmeri* biotypes were very similar across multiple experiments (Zulet-Gonzalez et al. 2023). In *A. palmeri*, Sandhu et al. found that glyphosate resistant biotypes had a higher abundance of primary metabolites such as nonaromatic amino acids, sugars, and organic acids, but lower abundance of secondary metabolites (Sandhu et al. 2023). Boonchaisri et al. (2021) reported that glyphosate resistant buffalo grass (*Stenotaphrum secundatum* L.) was nearly identical to a wild type buffalo grass based on the untreated metabolome but did report that metabolomes diverged

in response to treatment (Boonchaisri et al. 2021). Overall, it appears that metabolomes between untreated resistant and susceptible weeds are generally very similar, though sometimes several important metabolite groups may be detected, as in the case of Sandhu et al., (2023). Several studies corroborated their findings with sequencing or separate assays to investigate mechanisms of herbicide resistance (Maroli et al. 2015; Sandhu et al. 2025). From the studies cited above, it is difficult to claim that metabolites are related to the mechanism of resistance itself when comparing untreated resistant and susceptible plants, as there is no link between the genotype of the plants and the metabolite abundance. Currently, the application of metabolomics to rapid detection of herbicide resistance remains relatively understudied.

Few studies have evaluated metabolomic approaches as diagnostic tools to distinguish untreated resistant and susceptible weeds. Such tools would exploit constitutive differences between biotypes, which have been demonstrated to exist in herbicide resistant wild oat (*Avena fatua* L.) using other -omics approaches. Transcriptomic and proteomic analyses of a resistant wild oat biotype compared to a susceptible biotype identified suites of genes and proteins whose abundances varied prior to any herbicide application (Keith et al. 2017), suggesting that metabolomics may identify informative metabolites indicative of resistance in wild oat. Metabolomic analysis may identify metabolites that are not only useful as a diagnostic tool, but when combined with other -omics techniques, may help inform the mechanisms involved in resistance.

Wild oat is a winter annual weedy grass which infests various systems including wheat, barley, legumes leading to substantial yield losses (Beckie, Francis & Hall, 2012). Wild oat is ranked as one of the world's worst weeds (Holm et al. 1977), and increasing instances of

herbicide resistance in wild oat have been reported across modes of action (Ahmad-Hamdani et al. 2013; Keith et al. 2015; Mangin et al. 2017). Multiple populations of herbicide resistant wild oat have been reported over time in north central Montana, United States with increasing resistance. In 1993, after reports of triallate resistance in wild oat, the 93R biotype was collected and found to be resistant to Very Long Chain Fatty Acid (VLCFA) synthesis inhibitor herbicide triallate and difenzoquat, an herbicide with an unknown mode of action (Kern et al. 2002). In 2006, the first year pinoxaden was commercially available, wild oats in the same region were reported to be resistant to pinoxaden. The collected 06R biotype (Keith et al., 2017) is resistant to four modes of action via non-target site resistance (NTSR), possibly conferred by P450-mediated metabolism for flucarbazone (ALS inhibitor), imazamethabenz (ALS inhibitor), and difenzoquat (unknown), but non-P450 resistance to tralkoxydim (ACCase inhibitor), fenoxaprop p-ethyl (ACCase inhibitor), and triallate (VLCFA synthesis inhibitor) (Lehnhoff et al. 2013; Keith et al. 2015).

Thus far only treated wild oat has been examined using DIESI-MS (a high-throughput mass spectrometry technique used for metabolomic fingerprinting) to trace the origins of resistant wild oat populations in Mexico based on possible detoxification pathways detectable in treated samples (Tafoya-Razo et al. 2022). To explore whether metabolomics could enable accurate prediction of wild oat resistance to one or more herbicides based on specific metabolite signatures in untreated samples, we used mass-spectrometry in conjunction with recombinant inbred line (RIL) populations. To cast a wide net for possible metabolites related to resistance, an untargeted metabolomics LC-MS protocol was used in two RIL populations, derived by crossing the resistant wild oat biotypes 06R and 93R with the same herbicide susceptible biotype S2. We

hypothesized that metabolites detected would differ between herbicide resistance and susceptible biotypes as well as between lines identified as the most susceptible and resistant from the RIL populations. Furthermore, within the RILs showing the strongest ranked herbicide resistance and sensitivity phenotype, metabolite abundances would differ depending on the herbicide considered.

Methods

Table 2.1: Collection and resistance information of wild oat biotypes.

Population	Date	Location	Resistance
93R	1993	Teton, County, Montana	triallate, difenzoquat
06R (alias 06R4)	2006	Teton County, Montana	triallate, difenzoquat, pinoxaden, flucarbazone, imazamethabenz, tralkoxydim, fenoxaprop-p-ethyl
S2	1995	Canada	None detected

Creation and Phenotyping of RIL Populations

To investigate the linkage between expression of metabolites and herbicide resistance phenotypes, two recombinant inbred lines (RILs) populations were developed from two resistant inbred lines (93R and 06R) both crossed to the inbred susceptible line S2 (Table 2.1). The 06R

line referenced in this chapter is also labeled as the 06R4 biotype in later chapters. S2 is an established line (Johnson et al. 1995) susceptible to all herbicides tested. Hybrid progeny from the 93R x S2 cross and the 06R x S2 cross were then self-pollinated to generate F2 offspring. To generate RILs, 120 F2 progeny from 06R x S2 cross and 122 progeny from the 93R x S2 cross were self-pollinated and maintained via single-seed descent until the F7 generation.

All plants were grown unless otherwise stated in the Plant Growth Center at Montana State University. For recombinant inbred line screens plants were grown in the Plant Growth Center at Montana State University in greenhouse soil mix [1:1:1 (by vol) Bozeman silt loam:Sunshine mix #12 (Sun Gro Horticulture, Inc., Bellevue, WA):perlite] in 10.2 square pots under a 16-hr photoperiod of natural sunlight supplemented with mercury vapor lamps ($165 \mu\text{mol m}^{-2} \text{s}^{-1}$ PAR) at 25 ± 4 C, fertilized weekly with 100 ppm Jack's Professional 20-20-20 fertilizer (JR Peters Inc., Allentown, PA), and watered as needed.

The 06R x S2 RILs were screened for resistance to several herbicides. For foliar applied herbicides (flucarbazone, pinoxaden, tralkoxydim, and difenzoquat), plants were treated at Zadoks stage 13 (third leaf 8-10 cm long) (Zadoks et al. 1974) using a moving nozzle sprayer in 94 L ha^{-1} of water at 275.6 kPa pressure. Shoots were visually rated and harvested 21 days after herbicide application, dried at 50°C for 7 days, and dry weights recorded. For triallate, 8 seeds from each RIL or parental line were placed on 6 cm of greenhouse soil mix and covered with 2.5 of either untreated soil (control) or soil treated with triallate. Seedlings were grown for 21 days at which time seedlings were visually rated for damage, then seedlings were harvested, dried at 50°C for 7 days, and weighed. The herbicide rates used for screening are as follows: flucarbazone at 30 and $60 \text{ g a.i. ha}^{-1}$ (Everest 2.0, Arysta LifeScience North America, Cary, NC),

pinoxaden at 30 and 45 g a.i. ha⁻¹ (Axial XL, Syngenta Crop Protection, LLC, Greensboro, NC), tralkoxydim at 100 and 200 g a.i. ha⁻¹ (Achieve, Dow AgroSciences Canada, Calgary, AB), difenzoquat at 170 g a.i. ha⁻¹ (Avenge, AMVAC, Newport Beach, CA) and triallate in both the greenhouse at 1.12 and 2.24 kg a.i. ha⁻¹ and at the Arthur H. Post Agronomy Farm (111.158438°W 45.677669°N, RCB design with three blocks) at 2.24 kg a.i. ha⁻¹ (Far-GO, Gowan Co., Yuma, AZ). The percent biomass of control, the biomass of the RIL when treated with an herbicide divided by the mass when untreated, was used to assess the resistance of a RIL to an herbicide. A high percent biomass of control indicated the RIL's growth was unaffected under treatment, while a low percent biomass of control indicates the herbicide successfully controlled the RIL.

The 93R x S2 RILS were screened in the greenhouse for difenzoquat (foliar applied) and triallate (soil applied) as described above. Reduction in biomass is not a robust metric of control for difenzoquat, thus control was assessed by visual rating. For both populations, RILs were ranked by resistance across all conditions, then a set of RILs which included lines ranked as highly resistant and lines ranked as highly susceptible were selected for metabolomic analysis ((Table 2.2, Table 2.3).

Table 2.2: Resistant and susceptible RILs selected for 06R x S2 metabolomic analysis, binary resistance classification indicated by letter in parentheses.

RIL ID	Mean flucarbazone % biomass	Mean pinoxaden % biomass	Mean tralkoxydim % biomass	Mean triallate % biomass
12	81.29 (r)	69.55 (r)	18.83 (s)	81.34 (r)
37	92.52 (r)	102.56 (r)	33.32 (s)	93.1 (r)
42	90.64 (r)	64.8 (s)	93.12 (r)	89.83 (r)
72	73.82 (r)	156.03 (r)	69.16 (r)	94.83 (r)
77	104.14 (r)	79.66 (r)	86.98 (r)	100 (r)
88	92.22 (r)	116.53 (r)	76.92 (r)	101.49 (r)
27	22 (s)	35.94 (s)	21.8 (s)	0 (s)
34	20.51 (s)	31.77 (s)	28.72 (s)	0 (s)
14	29.01 (s)	52.34 (s)	20.62 (s)	0 (s)
28	19.26 (s)	27.63 (s)	32.48 (s)	0 (s)
63	25.59 (s)	45.37 (s)	23.64 (s)	107.23 (r)
109	25 (s)	25.56 (s)	19.09 (s)	0 (s)

Table 2.3: Resistant and susceptible RILs selected for 93R x S2 metabolomic analysis, binary resistance classification indicated by letter in parentheses.

RIL ID	Mean triallate % biomass	Mean difenzoquat rating
78	101.57 (r)	3.7 (r)
83	124.86 (r)	4 (r)
88	115.38 (r)	4 (r)
59	95.19 (r)	3.5 (r)
75	3.43 (s)	1.6 (s)
76	3.77 (s)	2.9 (s)
89	7.84 (s)	2.4 (s)
122	5.98 (s)	1.9 (s)

Plant Growth Conditions and Tissue Collection

One seed was sown in a 4-in square pot containing standard greenhouse soil mix [1:1:1 (by vol) Bozeman silt loam:Sunshine mix #12 (Sun Gro Horticulture, Inc., Bellevue, WA):perlite in a temperature and light-controlled growth room at the Montana State University Plant Growth Center (16-hour photoperiod, $498.05 \mu\text{mol m}^{-2} \text{s}^{-1}$). Temperature (set to $25 \pm 4 \text{ }^{\circ}\text{C}$) and humidity were recorded using a MOCREO ST6 hygrometer (MOCREO) to ensure constant conditions within an experiment.

Three experiments were performed to identify metabolite features that differed between susceptible and resistant biotypes. In Experiment 1, twelve 06R and ten S2 plants were grown as described above and arranged in an alternating grid to minimize position effects (Table A1). In Experiment 2, twelve 06R plants, nine S2 plants and 54 derived RIL plants (Table 2.2, Table A1) were grown, and in Experiment 3, nine 93R plants, nine S2 plants, and 35 derived RIL plants (Table 2.3, Table A1) were grown. For Experiments 2 and 3, due to instrument availability constraints, an incomplete design was used where each block (herein “trial”) was planted on a different day. S2 and 93R or 06R were present in every block, and four different RILs were present per trial. Each biotype or RIL was grown with five replicates in a 6 x 5 grid with the positioning rotating per column to minimize position effects.

Leaf tissue was harvested at Zadok’s stage 13, at the unfolding of the third leaf (Zadoks et al. 1974; Tottman et al. 1979). To avoid possible sample degradation, the harvest and extraction protocols were tuned to minimize sample exposure to light and heat: leaf clippings (200 mg) were taken from the base of the first leaf, weighed, and placed in black microfuge

tubes and flash-frozen in liquid nitrogen. The flash-frozen leaf tissue was bead-beaten and placed in a -80 °C freezer before extraction.

Sample Preparation and Mass Spectrometry

A protocol was developed based on several plant matter extraction procedures and the constraints of the mass spectrometry system (Simonovska et al. 2013; Petibon and Wiesenberg 2022). A solution of 10% DI water, 16.5% dichloromethane, and 73.5% acetone was added to the tube containing the bead-beaten leaf tissue. The tube was then vortexed for two minutes, then centrifuged for five minutes at 4 °C at 21010 RCF. Ten percent volume of supernatant was pipetted from the solution and diluted at a 5:1 ratio in acetone, then re-centrifuged for five minutes at 4 °C at 21010 RCF. The diluted solution (50 uL) was pipetted into the liquid chromatography vial with a 200 uL insert (Fisherbrand™, Cat No: 03-391-35), then placed on wet ice to transfer to the mass spectrometry facility at Montana State University.

Liquid chromatography-mass spectrometry (LC-MS) was selected as suitable method for separating and detecting wild oat metabolites (De Vos et al. 2007). For every trial of samples scanned, equal volumes of each sample were pooled and scanned using an MS-MS protocol. At the start of every experimental day, a standard of beta-carotene (DHS, Batch No. 135) was also scanned to assess column flow. Samples within a trial were randomly ordered and run on an Agilent 1290 UPLC (Santa Clara, CA) coupled with an Agilent 6538 Q-TOF mass spectrometer (Santa Clara, CA), using an Agilent Eclipse C18 (Santa Clara, CA; 2.1 x 100 mm, 1.8 mm particle size) column. Scans were collected using electrospray ionization in positive-ion mode, with a mass-to-charge ratio (M/z) range of 105-1700 at 3 Hz. Data was collected from 0-8 minutes of the runtime. The liquid chromatography method used two solvents: Solvent A: 50%

acetonitrile, 50% water, 0.1% FA (all Fisher Optima LC-MS grade), Solvent B: 90% isopropyl alcohol, 10% acetonitrile, 0.1% FA, (all Fisher Optima LC-MS grade), with a 10 uL injection volume and a flow of 0.4 mL/min (Table 2.4).

Table 2.4: Solvent flow over LC-MS injection.

Time (min)	%A	%B
Initial	50	50
4	5	95
5	1	99
6.5	1	99
6.51	50	50
8.0	50	50

Post-processing of Mass Spectrometry Data

The software MSConvert (Chambers et al. 2012) was used to covert files to mzML format, with parameters peak picking MS Levels 1-2, and Threshold Peak Filter = 2000 most intense.

The software MS-DIAL (Tsugawa et al. 2015) was used to process the raw LC/MS data. MS-DIAL was used to 1) perform spectral deconvolution and alignment, 2) annotate features using reference databases and peak similarities, and 3) perform initial data visualization and accuracy checks. Default parameters were chosen unless specified. Features were only considered up to six minutes of retention time, as features detected thereafter were likely contaminants from the column. Retention time alignment was performed using a set of manually curated peaks which had low tailing and high abundance across all samples. Each experiment was aligned separately. Public spectral libraries were used to annotate features: the positive-

mode MSMS library from MS-DIAL (Tsugawa et al. 2015), the positive-mode lipid library from MS-DIAL (Tsugawa et al. 2015), and the full GNPS library (Wang et al. 2016). Adducts considered were $[M+H]^+$, $[M+NH_4]^+$, and $[M+Na]^+$. A 0.3-minute retention time tolerance for alignment was specified to reduce over-annotation, along with a 0.02 Dalton mass tolerance. Features were removed using blank information using a 5-fold change (sample mean/blank mean) cutoff. Features with low signal to noise ratios ($\sim < 25$) were excluded from the dataset, as well as features only present in standard samples.

The web platform MetaboAnalyst (Pang et al. 2024b, a) was used to clean and analyze metabolite abundance data. An R script (R Version 4.5.2)(R Core Team 2025) was used to format data for input into the Batch Effect Correction module, where the EigenMS method was used to account for batch effects between experimental days. For Experiments 2 and 3, data for the parental biotypes was split from the RIL data. A separate dataset was created for each herbicide response per cross. This led to three datasets for the 06R population: 06R and S2 samples from Experiment 1, 06R and S2 samples from Experiment 2, and RIL samples from Experiment 2. Experiment 3 had two resulting datasets: 93R and S2 samples from Experiment 3 and RILs from Experiment 3. Analyses were conducted with the one-factor statistical analysis module. Features were filtered with a 25% IQR and 25% mean intensity filter. Features were log-transformed and auto-scaled. For the 06R and 93R biotype datasets, sample sum normalization was performed.

Analysis of Metabolomics Data

To identify important features, several visualization and analysis steps were performed in MetaboAnalyst. Between the 06R/93R biotypes and the S2 biotype, Welch's t-test per metabolite was combined with fold change in volcano plots. P-values were corrected for False Discovery

Rate (FDR) using the Benjamini-Hochberg procedure (Benjamini and Hochberg 1995).

Metabolites were retained as potentially important if they had an absolute fold change greater than one and a corrected p-value equal to or less than 0.1. Heatmaps were created using the top 25 differentially abundant features (Pang et al. 2024a).

For each herbicide, each RIL was labeled as R (indicating resistant) if it had a percent biomass greater than the mean of all RILs, and S (indicating susceptible) if less than that threshold (Table 2.2, Table 2.3). To avoid pseudo-replication of the RILs, the median abundance was taken for each feature per RIL. Then fold-change based on the mean of the medians within the R and S categories were used as a metric of potential feature importance. Fold change per feature was calculated for each feature for every herbicide screened after filtering. The direction of change was always R versus S, meaning a positive fold change indicated a higher value in the resistant line.

Screening for intersections of potential important features was conducted using R version 4.5.2 (R Core Team 2025). Features identified in the 06R, 93R, and S2 biotypes were cross-referenced to see if they shared similar patterns in RIL populations. Experiment 1 and Experiment 2 failed to align with MS-DIAL, so putative annotations and mass values were manually compared by screening Experiment 2 for features found to be different between 06R and S2 in Experiment 1. Data were joined by annotation and metabolite mass value rounded to two decimal places, then filtered to only contain features selected as important by the criteria above (greater than one-fold change and $p\text{-value} \leq 0.1$). The ggplot2 (Wickham 2016) package was used to plot fold change data.

The 93R data were subject to the same pipeline, with the difference that 93R and S2 biotypes used to filter features for exploratory analysis were intermixed with RIL samples, allowing features to be aligned across all datasets. A heatmap was created showing any metabolite across all datasets with an absolute fold change greater than one to explore patterns between trials. The data was then subset to contain only features selected as important in the 93R and S2 biotype dataset with the same criteria as above.

For both experiments, the remaining features were then filtered to features which had the same sign for the fold-change value across all herbicide response groups, meaning the direction of abundance change had to be the same between resistant and susceptible parental biotypes and RILs. Fold change and corrected p-values were reported for the remaining features, and box-plots were used to visualize normalized abundances of the original samples.

Potential metabolite annotations were reported from the MS-DIAL output and querying both the public libraries listed above with MS-DIAL and the MS-MS annotation function in MetaboAnalyst.

Results

Untargeted metabolomics has the potential to identify metabolites related to herbicide resistance in *A. fatua* (Maroli et al. 2015; Sandhu et al. 2025). To corroborate that metabolites differing in abundance between the 06R or 93R biotypes and susceptible S2 biotype were likely to be related to resistance, metabolomic profiling was also performed on RILs derived from both resistant biotypes and the susceptible biotype.

Experiment One

To identify candidate metabolites related to resistance in 06R, we profiled a total of 705 metabolite features using the LC-MS protocol and retained 297 metabolite features after data cleaning in both MS-DIAL and MetaboAnalyst (Table A1). While most features were expressed at similar levels in both biotypes, 20 features (6.7%) showed varied abundance between S2 and 06R ($p < 0.1$, Figure 2.1). Of these features, 5 were less expressed in 06R than S2 and 15 were more expressed in 06R. One feature (M/z: 0.687_147.0771) stood out as less abundant in the resistant biotype with a $\log_2(\text{fold change})$ less than -1.

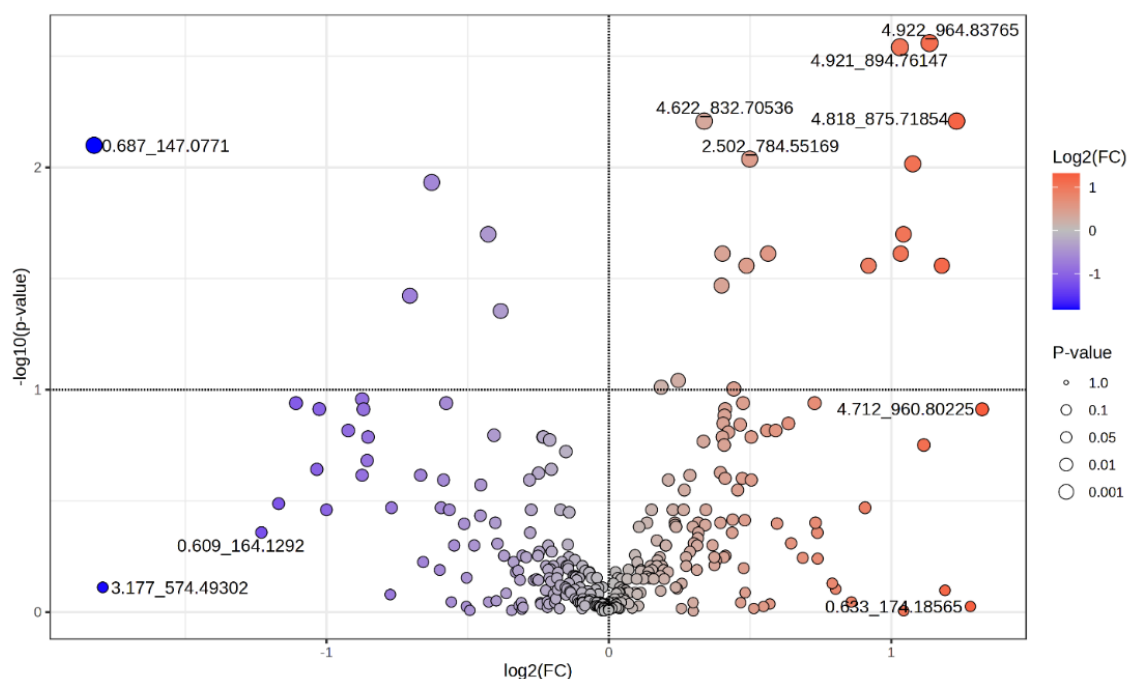


Figure 2.1: Metabolite features detected in Experiment 1 vary in expression between S2 and 06R. Features are colored by $\log_2(\text{fold change})$ with less abundant features in 06R shaded blue and more abundant features shaded red. The size of the circle indicates the FDR corrected p-value.

Patterns of metabolite abundance could be used to group similar samples (Chong and Xia 2020). Therefore, we explored how a suite of metabolite features could distinguish 06R and S2. The top twenty-five most differentially abundant features based on statistical significance were used to cluster samples. The 06R and S2 biotypes clustered into two clades, and features clustered into two clades (Figure 2.2). The upper clade of features tended to be higher in 06R samples. The bottom clade appeared to have generally higher abundance in the S2 samples. Several features that were identified as potentially important in the volcano plot (Figure 2.1) are present: feature 0.687 147.0771 is in the bottom clade alongside several other metabolites

Patterns of metabolite abundance could be used to group similar samples (Chong and Xia 2020). Therefore, we explored how a suite of metabolite features could distinguish 06R and S2. The top twenty-five most differentially abundant features based on statistical significance were used to cluster samples. The 06R and S2 biotypes clustered into two clades, and features clustered into two clades (Figure 2.2). The upper clade of features tended to be higher in 06R samples. The bottom clade appeared to have generally higher abundance in the S2 samples. Several features that were identified as potentially important in the volcano plot (Figure 2.1) are present: feature 0.687 147.0771 is in the bottom clade alongside several other metabolites

eluting around 0.65 minutes. Many compounds eluting around 4.8 minutes make up a subclade within the upper metabolite clade and generally are more abundant in 06R samples. The top clade of metabolites is less consistent than the other clusters but also appears to be more abundant in the 06R samples.

Experiment Two

A second set of 06R and S2 plants were planted alongside the selected RILs for Experiment 2. The 06R/S2 biotype comparison for Experiment 2 yielded no significant differences between the biotypes for any specific metabolite, likely due to the reduced and uneven sample size (Table A1). Thus, the 06R and S2 samples from Experiment 2 are only used as a check and to corroborate the findings from Experiment 1. From the initial 778 features detected using the LC-MS protocol, 354 metabolite features were retained for analysis after data processing and cleaning (Table A1).

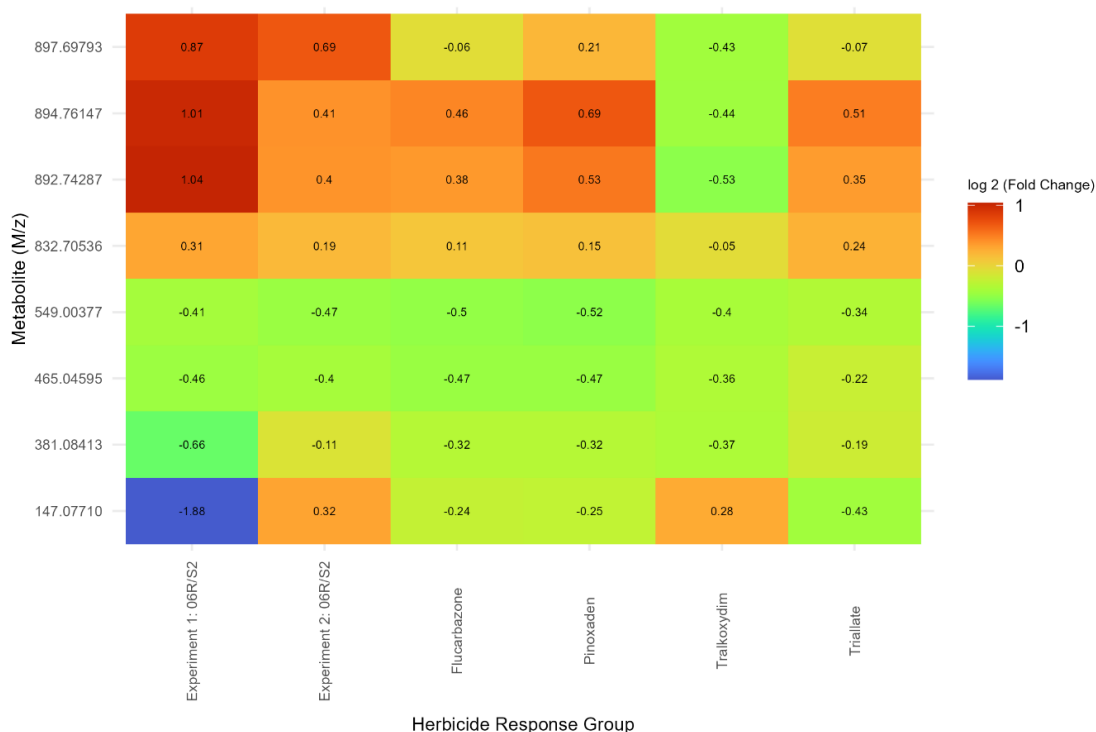


Figure 2.3: Differential abundance of metabolic features across 06R versus S2 and herbicide resistant versus sensitive RILs. Log₂(fold change) of metabolic features is indicated in the cells, with more positive values (red) indicating higher abundance in R biotypes and more negative values (blue) indicating lower abundance in R biotypes. Relative abundance for each herbicide tested based on recategorized RILs is shown in the latter four columns (see Table 1.2).

Resistance to different herbicides is not always conferred by the same mechanism within a mode of action (Gaines et al. 2020). Therefore, RILs were recategorized into resistance and susceptible classes per herbicide tested. To identify features with robust association with herbicide resistance, by-herbicide fold change was calculated from Experiment 2 and were merged with putative annotations and rounded mass values from Experiment 1, leaving eight features with putative annotations (Figure 2.3). For six of the features, the direction of the fold change between 06R and S2 tended to be the same for the RILs categorized for the herbicides flucarbazone, pinoxaden, and triallate, indicating that these features were consistent identifiers of herbicide resistance. Three of the six features were higher in the resistant lines. However, the

direction of the fold change varied for the herbicide tralkoxydim. The features with M/z values 892.74287 and 894.76147 had positive fold changes in the 06R versus S2 comparison but had negative fold change values for tralkoxydim resistance in the RILs. To explore this, we noted that the RILs were intermediate in their tralkoxydim responses (Table 2.3) unlike for the other three herbicides. Features were subsequently filtered with the requirement that fold change occurred in the same direction across Experiment 1 and 2 for all four herbicides, leaving three metabolites.

To understand potential patterns in feature abundance and variation across the 06R, S2, and individual RILs, feature abundance across samples were examined (Figure 2.4). Relative feature abundance differences between the 06R and S2 biotypes were similar between Experiments 1 and 2, though the differences appear to be stronger in magnitude in Experiment 1 (Figure 2.4). Further, variability in the resistant RILs appeared to be greater than variability in the susceptible RILs. Across the RILs, the abundance patterns across the three features were very similar. The M/z for these three features followed a pattern. The three identified metabolites are spaced 83.96 ± 0.01 M/z apart, indicating they are likely conjugates of some base compound, potentially the 381.084 feature, which was annotated as an O-glycosyl compound (Table 2.5). The MS-MS spectrum for the 465.045 feature shows a strong peak at 381 M/z, which is consistent with the above logic (Figure A1). By comparing the MS-MS spectra to the reference spectra, we conclude that 465.045 and 549.005 are the conjugates of a base compound.

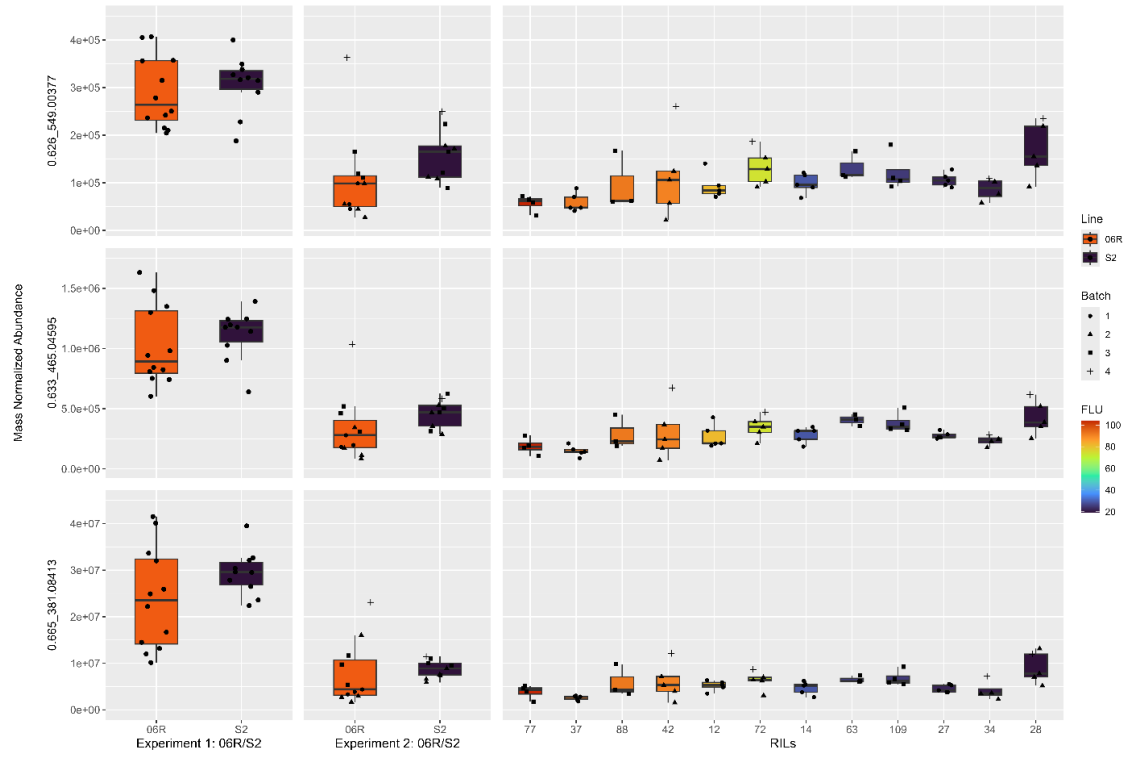


Figure 2.4: Boxplots of mass normalized abundance per metabolite, across all experiments and RILs for 06R/S2. RILs are colored by flucarbazon percent biomass of control. Batch is the order of the experimental days the sample was collected on.

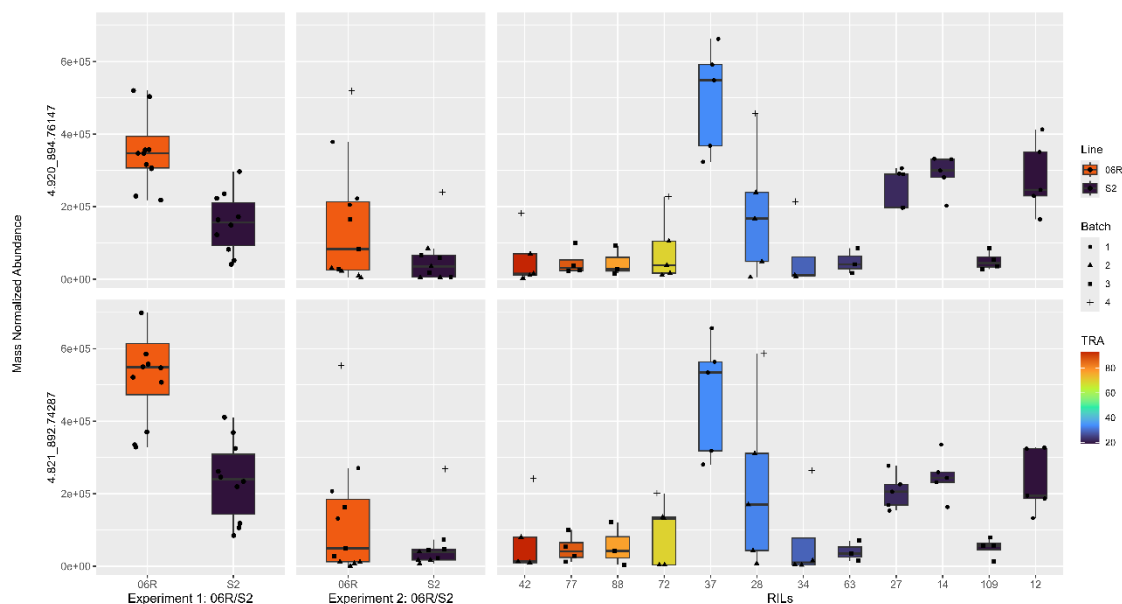


Figure 2.5: Boxplots of putatively annotated tri-acylglycerol metabolites mass-normalized abundance across all experiments and RILs for 06R/S2. RILs are colored by tralkoxydim percent biomass of control. Batch is the order of the experimental days the sample was collected on.

Relative metabolite abundance was visualized using boxplots for the two features (M/z: 892.74 and 892.76) which had consistent fold changes for all herbicides excluding tralkoxydim (Figure 2.3). These metabolites were putatively annotated as tri-acylglycerols (TG) using positive ion-mode lipid library in MS-DIAL, using theoretical MS-MS spectra created in-silico (Table 2.5). Strong positive relative abundance differences were visible between the 06R and S2 biotypes in Experiment 1, though this difference was much more subdued in Experiment 2 (Figure 2.5). The shift in the direction of fold change for tralkoxydim seems to be driven by the intermediate levels of tralkoxydim resistance for RILs 37 and 28. RIL 37 is highly resistant to flucarbazone, pinoxaden and triallate, but relatively susceptible to tralkoxydim, having a percent biomass of control of 33.32 (Table 2.2). RIL 28 is among most susceptible to all other herbicides,

though it has a middling level of tralkoxydim resistance (32.48 percent biomass of control) compared to the other RILs (Table 2.2). Both RILs have high relative abundances of both triacylglycerol metabolites, though the RILs 63 and 109 have relatively middling abundances of these metabolites (Figure 2.5). Given that the relative abundance of both metabolites appears correlated across samples and RILs, and the difference between the two metabolites is 2.02 Da, it is likely that the 894.76 M/z metabolite has one additional double bond within the compound.

Table 2.5: Annotations for selected 06R x S2 experiment metabolites. The mean retention time and mean M/z values were calculated between Experiment 1 and Experiment 2.

Mean Retention Time	Mean M/z	Annotation	Adduct	MSMS Acquired	Ontology
0.689	381.08362	low score: β -Gentiobiose	[M+K] ⁺	TRUE	O-glycosyl compounds
0.659	465.045495	No annotation	[M+H] ⁺	TRUE	NA
0.651	549.00518	No annotation	[M+H] ⁺	FALSE	NA
4.47	892.74	TG	[M+NH ₄] ⁺	TRUE	TG
4.56	894.76	TG	[M+NH ₄] ⁺	FALSE	TG

Experiment Three

The 93R biotype has a different herbicide resistance phenotype than 06R (Kern et al. 2002; Keith et al. 2015), so we hypothesized that different metabolites would be associated with herbicide resistance in Experiment 3. The 93R biotype is resistant to the herbicides triallate and difenzoquat (Table 2.1). From an initial 722 features detected in the LC-MS scan, a total of 402 features were retained after data processing and cleaning (Table A1). Only five features (1.2%) varied between the biotypes ($p < 0.1$, Figure 2.6). Four features had lower abundances in 93R, and one of which had higher abundance in 93R than S2 (Figure 2.6).

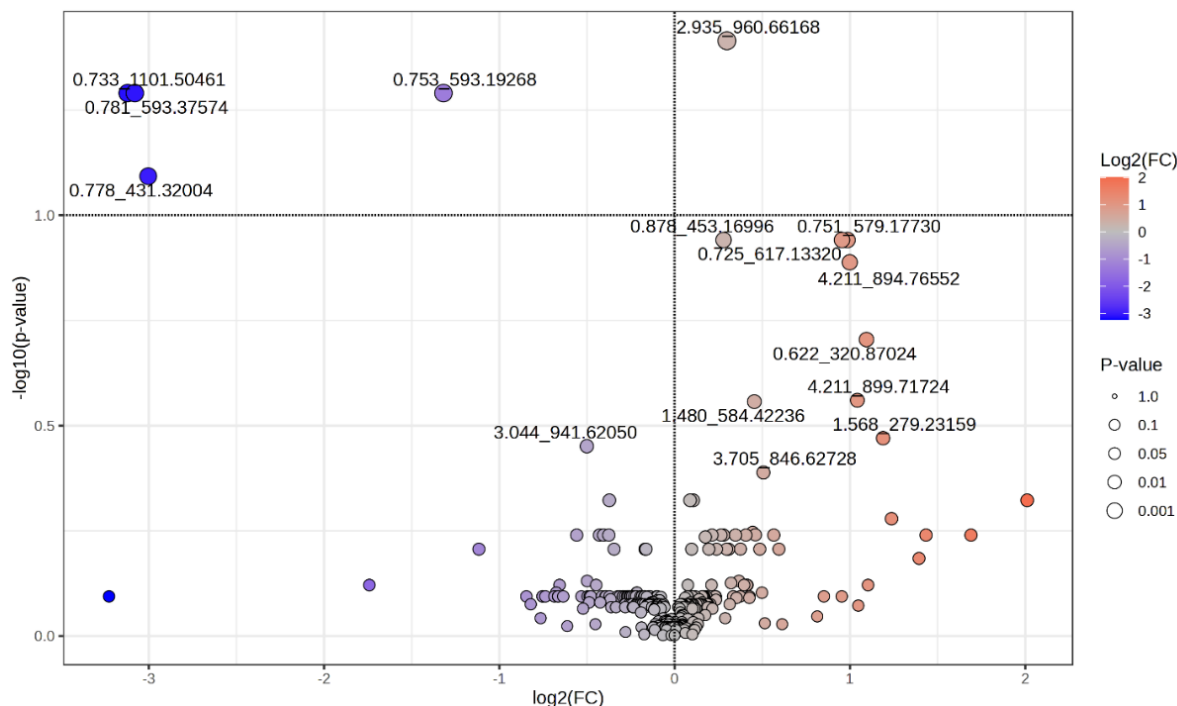


Figure 2.6: Volcano plot of metabolite features detected in Experiment 3 93R and S2 biotypes. The direction of comparison is 93R/S2. Features are colored by $\log_2(\text{fold change})$ and sized by FDR corrected P-value.

A suite of features was also examined for their combined effect to distinguish 93R and S2. These two biotypes 93R and S2 clustered according to resistance when using the top 25 most significant features for clustering except for two 93R samples that grouped with S2 (Figure 2.7). The metabolite features are grouped into two clades. The bottom clade of features included the four features found to be less abundant in 93R (Figure 2.6). The two misclustered 93R samples had lower abundance of features present in the bottom cluster, though they appear as outliers for features in the upper clade, which were more abundant in the 93R biotype (Figure 2.7).

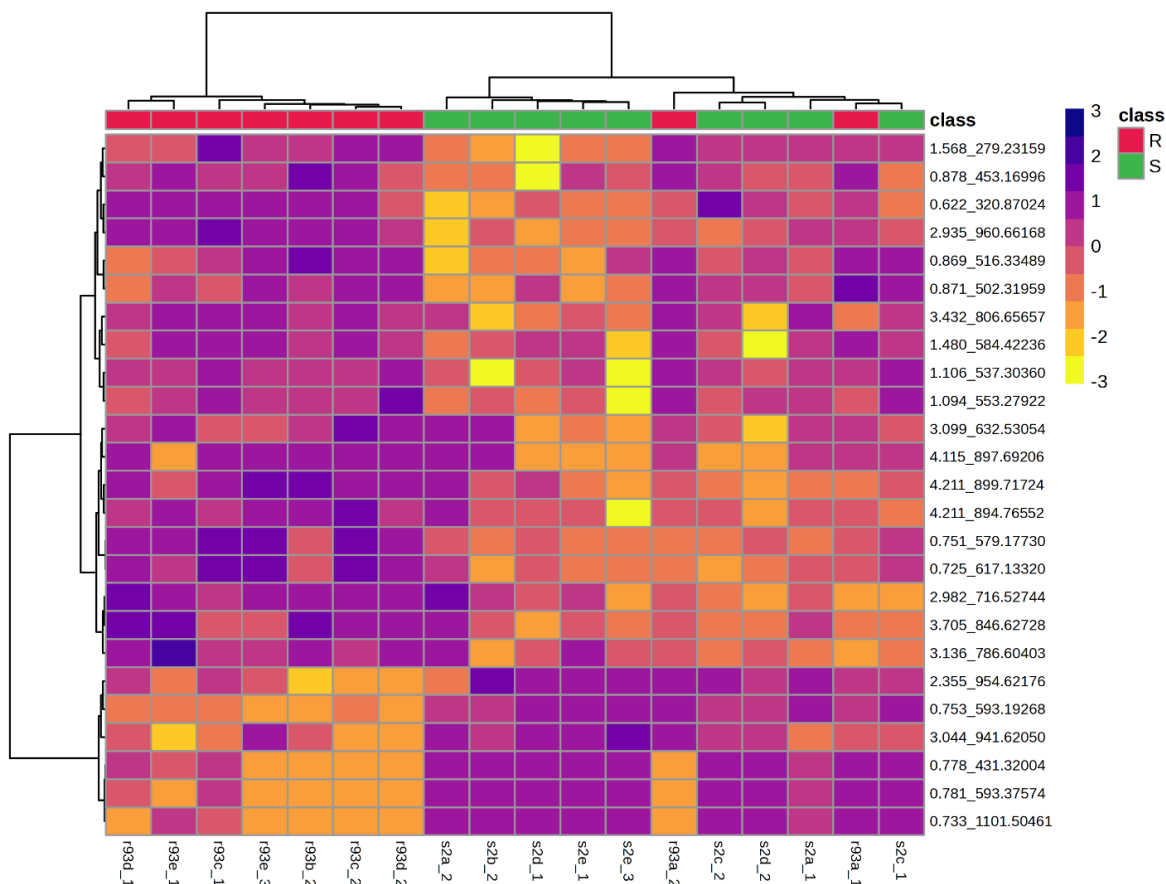


Figure 2.7: Clustering of 93R and S2 samples by top features. Columns represent samples, with 93R samples colored red and S2 samples colored green. Features are shown in rows.

To assess whether fold changes in the RILs were also associated with herbicide resistance in a similar manner to 93R and S2, the RILs were categorized based on their herbicide responses to triallate and difenzoquat. Using the uncorrected p-value from a Welch's t-test, 22 features were identified as differentially abundant for triallate or difenzoquat; the fold change of these features were compared to the features identified in the 93R/S2 comparison (Figure 2.8). Several features (M/z: 899.72, 894.77, 599.42) had positive $\log_2$ (fold change) in abundance between the 93R and S2 samples, but negative fold changes in the RILs. Only one feature (M/z: 593.19) had a negative $\log_2$ (fold change) between the 93R and S2 samples, but there were no strong changes in

the RILs. To screen for features that differed in abundance similarly between resistance classes in both parents and RILs, only features with an FDR corrected p-value < 0.1 and the same sign of the fold change in both 93R versus S2 and in the RILs. Several features did have similar positive fold change values (M/z: 279.23 and 537.30) but were not below the p-value threshold (Figure 2.6). Three features were identified, with M/z values of 1101.50, 593.38, and 431.32 (Table 2.6), and all three had lower abundances in resistant plants compared to susceptible plants.

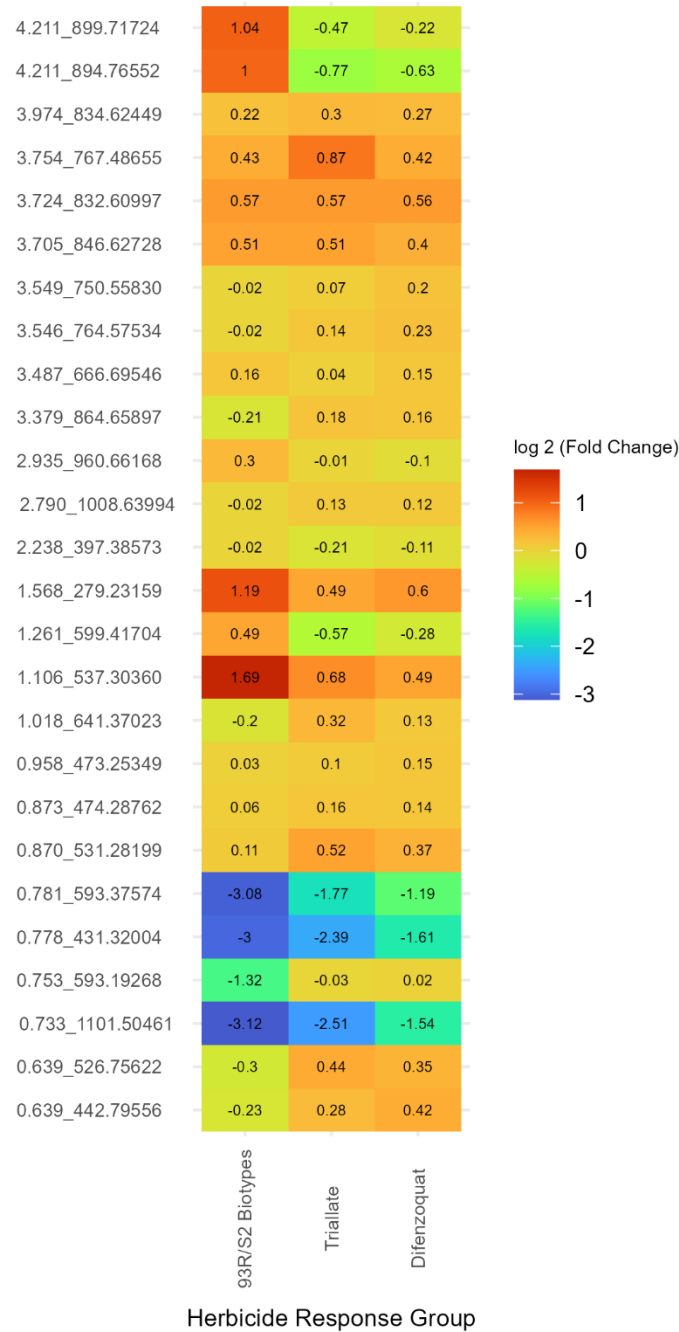


Figure 2.8: Heatmap of $\log_2$ (fold change) values for the 93R and S2 biotypes and each herbicide response group (triallate and difenzoquat). Metabolites were selected for plotting if they differed significantly between the 93R and S2 biotypes or between the resistant and susceptible RILs for either herbicide response.

Table 2.6: Spectra, FDR adjusted p-value, and $\log_2$ (fold change) of metabolites retained after filtering for consistent abundance shifts between resistance groups.

Herbicide response group	Metabolite		
	0.733 1101.50461	0.781 593.37574	0.778 431.32004
Adjusted p-value	0.051255	0.051255	0.080755
93R/S2 Biotypes	-3.1209	-3.0799	-3.0035
Triallate	-2.5127	-1.7744	-2.3856
Difenzoquat	-1.5389	-1.1851	-1.6149

The three features were screened against the public GNPS databases and MetaboAnalyst MS-MS annotation. Only one of the selected features (M/z: 1101.50461) was detected during MS-MS scans; however, no annotation was made. To understand the possible relationships between resistance and feature abundance for the selected features, normalized feature abundance across RILs was examined. The trends in variance and abundance were similar between features, as observed in the S2 and 06R comparisons as well. The features selected had consistent non-detections for both the 93R biotype and the resistant RILs (Figure 2.9). This pattern occurs across experimental days (batches), meaning that features had consistent detection patterns for plants grown separately and processed on different days. There are four points which do not follow the pattern, both the 93R and S2 biotypes contain a sample which follows the pattern of the reciprocal group. RIL 88 and RIL 122 both have samples which do not match the pattern for resistant RILs or susceptible RILs respectively. Interestingly, RIL 59 was classified as resistant to both triallate and difenzoquat, but the abundance of the identified features more similar to the susceptible RILs than the resistance RILs. This is consistent with RIL 59 having an intermediate response to triallate and flucarbazone compared to the other RILs (Table 2.3).

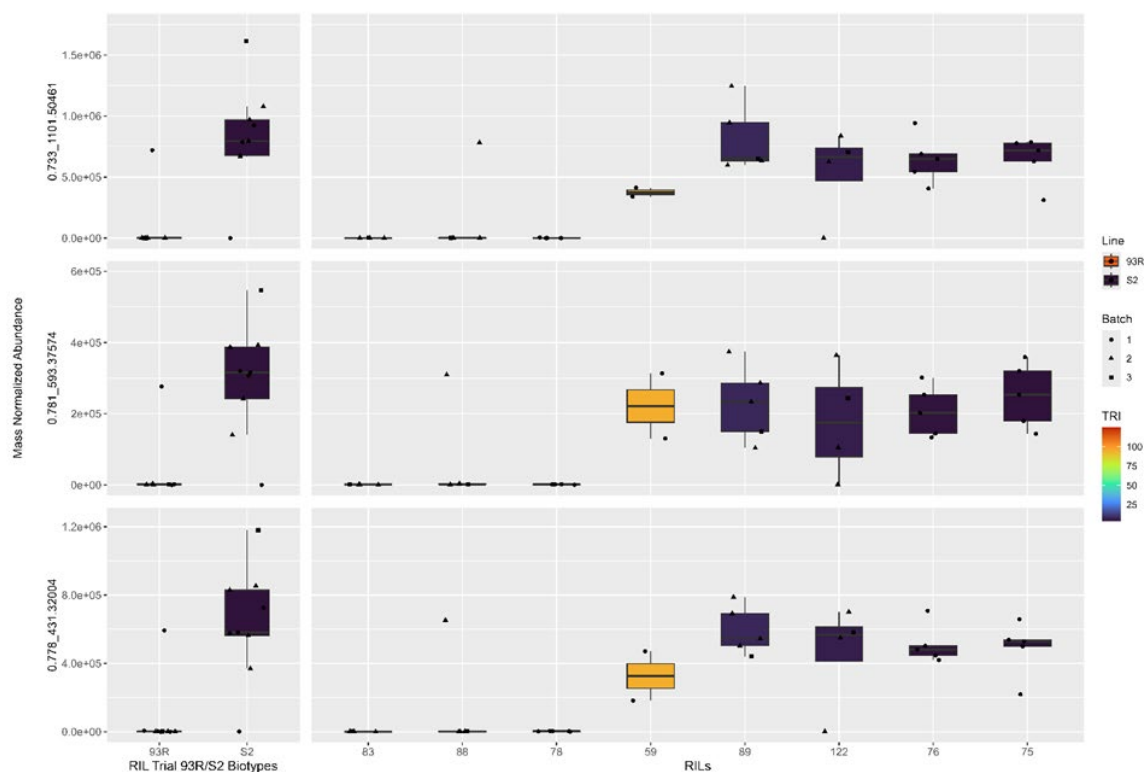


Figure 2.9: Boxplots of mass normalized abundance per feature for the 93R/S2 experiments. RILs are colored by the percent biomass of control for triallate, and the points are shaped by the order of experimental day the sample was collected on.

Discussion

Annotation of metabolites and identification metabolite features related to resistance in wild oat could lead to development of metabolite assays used to screen for herbicide resistance before chemical application. Historically, resistance phenotypes have been assessed by dose-response assays, and while effective at characterizing resistance, these can take time and resources (Beckie et al. 2000; Squires et al. 2021). Importantly, bioassays and other diagnostic tests necessitate that herbicide be applied to characterize resistance. Contrasting herbicide resistance and susceptible lines of wild oat with no herbicide exposure enabled the discovery of

metabolite features associated with susceptibility in both the 06R versus S2 and 93R versus S2 biotype comparisons and the derived RIL populations.

Comparing the metabolomes of any two random genotypes of wild oat will reveal many differences among them. Therefore, contrasting a small number of resistant and susceptible biotypes does not prove that metabolites are reflective of resistance itself. This problem can be addressed by obtaining high-quality annotations and subsequent pathway analysis of metabolomes under induced herbicide response to identify metabolites directly causal to resistance (Ma and Qi 2021; Sen et al. 2023). Instead, we utilized recombinant inbred lines (RILs) to associate metabolites with a resistance phenotype, allowing for stronger inference when concluding that a metabolite is related to resistance even when annotation is limited. To address sample size limitations in performing metabolomics, we focused on RILs exhibiting the most extreme phenotypes, examining whether features exhibited consistent shifts in abundance between resistant and susceptible biotypes.

Unlike previous studies, annotating metabolite features was limited by the procedures used in this study. The extraction was designed to isolate non-polar plant metabolites such as pigments; however, many of these compounds are poorly annotated. The LC-MS was tuned for these non-polar compounds, and thus semi-polar compounds eluted early within runs, making MS-MS collection and therefore annotation difficult. The 06R and S2 biotype experiments yielded three features potentially indicative of resistance across all herbicides screened, though only two had MS-MS scans (Table 2.5). In both Experiment 1 and Experiment 2, a M/z spacing of 83.96 ± 0.01 was observed between the three compounds, which is strong evidence that they are conjugates. A 381 M/z peak was visible in the MS-MS spectrum of the second conjugate,

suggestive of the base metabolite (Figure A1). However, the identity of the potential group being added to the 381 M/z feature to create the conjugates was not discovered. The strongest annotation was for M/z 381.08, which was annotated as β -gentiobiose (Table 2.5). However, many other sugars also match well with the MS-MS spectrum, so this annotation is not exact. For the 93R experiments, no annotations were made for the identified metabolite features.

The wild oat metabolite features varying due to herbicide resistance all had lower abundance in resistant samples relative to the susceptible for both the 06R and 93R biotypes. This observation is similar to other metabolic studies in other herbicide resistant weeds. A higher abundance of *N*-acetylhexosamine and dihydrouracil was reported in susceptible Palmer amaranth compared to a resistant biotype, though the $\log_2$ (fold change) value was less than one (Sandhu et al. 2023). For both the 06R/S2 and 93R/S2 comparisons, it is consistent with the literature that global metabolome differences between resistant and susceptible biotypes were minimal. For the 06R biotype experiments, the abundance shifts for the retained features were all low magnitude negative fold changes (Figure 2.3). The degree of difference in abundance varied between the parents of the RIL populations and the RIL lines. For example, feature 381.08 M/z had the strongest fold change between 06R and S2 biotypes (-0.66, p-value 0.05), but the weakest values of $\log_2$ (fold change) across the RILs compared to the other conjugates, ranging from -0.37 to -0.19 (Figure 2.3). The other two conjugate features had fold changes that were more consistent between the 06R/S2 biotypes and the RILs. Given that these are very likely conjugates of the same metabolite, a pattern emerges. If feature 381.08 M/z is the parent compound, it has the highest normalized abundance (centered around 2.5×10^7 in Experiment 1), followed by the first conjugate (465.05 M/z, centered around 1.0×10^6 in Experiment 1), and the

last conjugate centered around 3.0×10^5 (Figure 2.4). There is a greater abundance of the base metabolite, and this feature has the highest magnitude $\log_2(\text{fold change})$ values across the 06R and S2 samples, excluding Experiment 2 (Figure 2.3). This is likely noise from the compressed scale of low-abundance differences artificially inflating $\log_2(\text{fold change})$ values. Strong abundance differences are present between Experiment 1 and Experiment 2, and there is visual grouping of abundance values inside the batches of Experiment 2. Batches 1 and 2 tend to have lower relative abundance for the metabolites than Batches 3 and 4, likely due to column effects (Figure 2.4, Figure 2.5). This provides strong rationale for the EigenMS batch correction procedure applied to the data, though plotting the batch corrected data is not visually meaningful due to artifacts caused by the correction method.

The 93R and S2 biotype comparison showed higher magnitudes of $\log_2(\text{fold change})$ values across the selected features, each feature having a $\log_2(\text{fold change})$ value with a magnitude greater than three (Table 2.6). The RILs showed weaker magnitude $\log_2(\text{fold change})$ values, though all exceeded the value of one. While there is not an obvious conjugate relationship between these features, it is apparent they are highly correlated. As opposed to a difference in abundance, these features appear to be important because they were frequently unobserved in the resistant samples (Figure 2.9). So, while the batch correction method does make fold change value uninterpretable in for Experiment 3, it does reflect a real difference in metabolite abundance.

Plotting these features' abundances shows potential reasoning for the lack of strong differences. In 06R experiments, the features selected all share very similar abundance patterns across samples, indicating that they could be highly correlated with each other (Figure 2.4,

Figure 2.5). Within the 06R x S2 RILs, resistant RILs tend to have higher range of abundances than susceptible RILs for all three selected features that had negative fold changes, and this pattern is also evident between the 06R and S2 biotypes (Figure 2.4). RILs 12 and 72 both have similar abundances to that of the susceptible RILs, while RIL 34 has lower abundances than the other susceptible RILs for the selected features (Figure 2.4). This tracks with the reported percent-biomass of control values: RILs 12 and 72 have relatively lower percent biomass than other resistant RILs. RIL 12 had the lowest tralkoxydim percent biomass of control (18.83 percent), and RIL 72 had the lowest flucarbazone percent biomass of control (73.82 percent) (Table 2.2). However, this relationship did not track with susceptible RILs: though RIL 34 had a lower relative abundance, it was not notably different in mean percent biomass scores. RIL 63, which is resistant to triallate, did not associate with resistant RILs. Differences in Experiment 1 were more accentuated than those in the 06R and S2 samples collected alongside the RILs.

In the 93R experiment, a different pattern was observed: the resistant 93R parent and some of the RILs had much lower variation in feature abundance given that the features were not detected (Figure 2.9). The selected features were not detected in RILs 83, 88, and 78, and all had mean percent biomass scores greater than 100 and mean difenzoquat rating scores of nearly four (Table 2.3). RIL 59 had abundances similar to or slightly lower than that of the susceptible RILs (Figure 2.9) and had lower resistance to triallate and difenzoquat than the other RILs, though it was still classified as resistant (Table 2.3). This could indicate these features may have abundances which correlate with resistance across varying levels, however no true intermediate RILs were used for this experiment from the 93R population, so this needs to be validated with experiments including more RILs.

Previous work has shown that metabolomic profiles of the 06R wild oat biotype when unstressed had higher measures of variance than when under stress (Steward et al. 2023). However, no broad variance changes were observed between the resistant and susceptible biotypes or RILs. Canalization of specific pathways has not been studied in wild oat. The patterns of variance may be technical artifacts, though it seems unlikely that such a pattern would be present across multiple experiments and multiple lines.

The patterns of per-herbicide resistance to metabolite abundance for all experiments are moderate at best, indicating that there is possibly still segregation in the RIL populations. Without greater density of RILs, individual herbicide association is difficult. For example, there is a large group of features with M/z values ~900 identified as generally more abundant in the 06R resistant parent, constituting features in the upper right quadrant in Figure 2.1 and the first clade of features in Figure 2.2. Based on putative annotations, these are likely tri-acylglycerols. Two of these features (894.76 and 892.75) had positive $\log_2(\text{fold change})$ values for both the 06R/S2 comparison and each herbicide, excluding tralkoxydim. Fold change indicated that these features were less abundant in RILs resistant to tralkoxydim, though the relationship is inconsistent when abundances per RIL are visualized (Figure 2.5). RILs 28 and 37 were classified as susceptible to tralkoxydim and shifted the fold change negative when reclassified (Table 2.2). These RILs had a high range of relative abundance for these features. This could be evidence that the mechanism of resistance for tralkoxydim is possibly different than for the other screened herbicides. This is consistent with Keith et al.'s (2015) findings that resistance to tralkoxydim in 06R was non-P450 related (Keith et al. 2015).

In the 93R experiment, there were many features with positive fold change values between the 93R and S2 biotypes and between RILs, however none met the p-value cutoff (Figure 2.6). Many of these features were eluted early relative to the bulk of features. No features were found that had strong fold change values between the resistant and susceptible RILs yet had strong opposite direction values in the 93R/S2 biotype comparison (Figure 2.8). This is at least an indication that the metabolites with strong fold change values conserved between the 93R and S2 biotypes and the RILs are not merely present by chance after profiling close to 1000 features.

To validate that these features are possibly useful for screening for resistance, metabolomics could be performed on biotypes that are suspected to have similar mechanisms of resistance to the 06R and 93R biotypes. The 06R biotype has a more complex NTSR resistance phenotype, and this was possibly reflected in the less drastic shifts in abundance for the metabolome (Keith et al. 2015). An LCMS-MS protocol tuned to more semi-polar compounds could reveal the identity of the un-annotated metabolites and reveal new metabolites. Future work should include higher density RIL sampling to capture continuous resistance responses per herbicide, which would allow for the use of machine learning methods to predict resistance. For example, the *party* package can be used to fit conditional inference trees, which suit the highly correlated predictor space of metabolomics data (Hothorn et al. 2006). The *permimp* package can be used to assess feature importance robustly (Debeer et al. 2025). In addition, it is evident that concentrated sampling efforts are paramount for success in large metabolomics datasets, as the procedures used to merge datasets and correct for batch effects likely hindered statistical power in this analysis.

The features identified could be considered as potential candidates for targeting extraction and spectrometry approaches in new resistant biotypes of wild oat. Metabolites need to be validated as indicative of resistance in new lines as well as in re-designed LC-MS pipelines for both the 06R and 93R biotypes and derivative RIL populations.

Future Hypotheses

In the case of 06R, the lower abundance, but not absence of large building block metabolites across both parents and RILs could indicate that these metabolites are already being used in stress-response pathways in the resistant plants. Due to the general absence of the identified metabolites in the 93R biotype, it may be the case that the 93R biotype is responding with a completely different pathway than the S2 biotype. To test both these hypotheses, future work could expand metabolomic screening to plants under various stressors.

Mitigating pseudo-replication in the RILs reduced the selection available analysis tools. With few observations, we were limited to very simple fold-change tests to maintain statistical rigor. However, there was variance structure in metabolite abundance between resistant and susceptible plants. This indicates that replication may be necessary to capture metabolites related to resistance in susceptible wild oat plants.

Pathway analysis via the *mummichog* algorithm (Li et al. 2013) and metabolite annotation via MetDNA (Shen et al. 2019) may be powerful tools for understanding the metabolome differences in wild oat, a plant with low representation in metabolomic databases. As research resistance detection in wild oat using metabolomics progresses, utilization of pipeline tools such as Nextflow and Nextflow4MS-DIAL could make data processing,

metabolite annotation, and decision-making easily reproducible and locally hosted (Du et al. 2025).

Conclusion

Herbicide resistance in wild oat is poorly understood compared to many other weeds, even though wild oat infests millions of acres of crucial crops (Beckie et al. 2012b). Controlling herbicide resistant wild oat is difficult because resistance may evolve quicker than we can document it. However, this research indicates that metabolites related to resistance in parental biotypes may be conserved in derived RIL populations. Future work can improve detection and annotation of these metabolites in higher density RIL populations, facilitating the creation of predictive resistance modeling and insight into mechanisms of resistance.

CHAPTER THREE

BENCHTOP HYPERSPECTRAL IMAGERY TO
DISCRIMINATE AMONG MULTIPLE HERBICIDE
RESISTANCE BIOTYPES OF AVENA FATUA

Summary

Preapplication detection of herbicide resistance in weeds could be leveraged to target, and thus limit herbicide application to where it will be effective, saving producers money and reducing selection pressure for resistant biotypes. A combination of hyperspectral imaging and machine learning has been successful in discriminating resistant and susceptible strains of kochia, johnsongrass, and marestail. We evaluated this approach for multiple biotypes of herbicide resistant wild oat (*Avena fatua*), a pernicious winter annual weedy grass which infests Montanan wheat, barley, and legume crops. Four resistant and two susceptible biotypes of wild oat were scanned with a benchtop hyperspectral camera for high-resolution imagery. Among the resistant biotypes, 06R3 and 06R4 shared similar resistance phenotypes and were collected from nearby fields in Montana, 2006. The 20R2 biotype was collected from the same location in 2020 and is resistant to an additional herbicide on top of shared resistance with the 06R3 and 06R4 biotypes. The 93R biotype exhibited the most distinct resistance profile and has been shown to have a different resistance phenotype than the other resistant biotypes. The two susceptible biotypes, S1 and S2, were verified against herbicides used for resistant biotype selection. With limited observations ($n = 62$), we tested k-nearest neighbors, support vector machine, partial least squares discriminant analysis, and random forest models using mean leaf spectra. Random forest

performed best with 64% accuracy on the test set. Across all models, classification was most difficult between the closely related 06R3 and 06R4 biotypes and between the S1 and S2 biotypes. When biotypes were pooled into resistant versus susceptible classes, model performance improved to 90–100% accuracy. Models were more successful discriminating the resistant 20R2 and 93R biotypes from the other biotypes. These results suggest that more advanced models may be required to resolve subtle differences in spectra among resistant biotypes in the field, as simple resistant or susceptible class labels could mask plant response to herbicides. In addition, mean leaf spectra appear to contain necessary information to classify resistance in wild oat, which is more feasible for field data collection than by-pixel scans. This research creates baseline knowledge for herbicide resistance detection in wild oat in the field, information which could be vital for discriminating other resistant weeds.

Introduction

The occurrence of herbicide resistance in weeds is a global threat for food production and sustainable farming. Selective pressure on weeds due to herbicide application has led to evolution of biotypes resistant to multiple herbicide modes of action in 273 weed species (Heap 2026). Managing resistant weeds with herbicides represents time and money ineffectively spent. Further excess herbicide use exacerbate unintended side effects, such as crop injury, human health costs, herbicide drift and runoff, and further selection for resistance in untargeted weed populations (Peterson et al. 2018; Tangamornsuksan et al. 2019).

Detection of herbicide resistance prior to application would allow tailored herbicide management, mitigating the negative effects of ineffective herbicide application (Ghatrehsamani et al. 2023). Cutting-edge approaches to discriminate between resistant and susceptible weeds

leverage hyperspectral imagery, a technology which allows for scalable imaging at high spectral resolution across wavelengths (Sarić et al. 2022). Hyperspectral datasets include hundreds of bands that can capture subtle differences among plant biotypes. Supervised machine learning approaches are then used to navigate a high-dimensional feature space and classify a plant into a resistance class based on training sets built from reference resistant and susceptible plants. However, these approaches differ greatly in feature engineering techniques and model choice due to differences in goals and data types. Mensah et al. (2025) achieved ~85% accuracy classifying resistant biotypes of *Bassia scoparia* (L.) A.J. Scott (kochia) in crops with support vector machine (SVM) and random forest (RF) using stepwise feature selection paired with linear discriminant analysis (LDA) (Mensah et al. 2025). Huang et al. (2022) utilized k-nearest neighbors (K-NN), RF, and SVM models wrapped with a successive projections algorithm (SPA) to discriminate glyphosate-resistant *Sorghum halepense* (L. Pers.) (johnsongrass), achieving up to 77% accuracy (Huang et al. 2022). Extending beyond traditional machine learning approaches, Scherrer et al. (2019) applied a feedforward neural network to full-spectrum hyperspectral scans of resistant and susceptible kochia, *Chenopodium album* L. (lambsquarters), and *Conyza canadensis* (L.) Cronquist (marestail) within crop plants and found that stepwise classification yielded accuracies between 83% and 99% between resistant biotypes (Scherrer et al. 2019). Building pipelines which explore these modeling approaches in other resistant weeds is an important step towards applying resistance detection to the field.

Classifying resistance to different modes of herbicide action or mechanisms of resistance within a weed species would further support herbicide management decisions. Scherrer et al. (2019) showed that a neural network paired with hyperspectral imaging was able to discriminate

between two resistant biotypes in kochia. The presence of multiple herbicide resistance necessitates models which can predict specific resistance phenotypes. It is unknown if modeling approaches struggle to discriminate among biotypes differing slightly in resistance phenotypes. Mutations that confer herbicide resistance can be split into target site resistance (TSR) and non-target site resistance (NTSR). TSR stems from a mutation which alters the target site so that the herbicide no longer binds or the targeted enzyme is overexpressed (Powles and Yu 2010). NTSR mechanisms of resistance include reduced uptake, altered translocation, sequestration, and enhanced metabolism (Jugulam and Shyam 2019). Machine learning pipelines for herbicide resistance detection in the context of TSR and NTSR have not been explored. To investigate the application of herbicide resistance detection in NTSR weeds, we identified wild oat as a suitable candidate weed.

Avena fatua (L.) (wild oat) is a winter annual allohexaploid monocot weed with widespread herbicide resistant biotypes (Beckie et al. 2012b, a). Wild oat infests varied agricultural systems including wheat, barley, and legumes and has intense yield reduction effects in these crop systems, accounting for yearly losses greater than \$1 billion (Evans et al. 1991; Beckie et al. 2012a). Wild oat seeds disperse through seed shatter and harvest contamination and have a seed bank longevity up to nine years, making resistant varieties difficult to eradicate (Van Acker 2009; Beckie et al. 2012a). Vigorous infestation of cropland by wild oat is made more dire due to wide variety of resistant biotypes present in the field, which include multiple herbicide resistant (MHR) populations (Beckie et al. 2012b; Lehnhoff et al. 2013). The pervasiveness and variety of herbicide resistance in wild oat makes it a prime candidate to explore pre-application detection of resistance.

Extensive multiple herbicide resistance has been characterized in wild oat from a small geographic region in Montana, USA. The 93R biotype was described in 1993, and was found to be resistant to triallate and difenzoquat with two recessive alleles conferring resistance (Kern et al. 2002). The resistant biotypes 06R3 and 06R4 were derived from seeds collected in 2006 in neighboring fields from plants that failed to be controlled by pinoxaden (Lehnhoff et al. 2013); at the same time, the susceptible biotype S1 was identified from an untreated adjacent field (Keith et al. 2015). NTSR metabolism may be involved in multiple herbicide resistance in 06R4 (ACCase, ALS inhibitors, difenzoquat, triallate) and 06R3 (ACCase, ALS inhibitors, difenzoquat, triallate, paraquat), but that cytochrome P450-mediated NTSR was not involved in resistance to tralkoxydim, fenoxaprop-P-ethyl, or triallate (Keith et al. 2015). The 20R2 biotype was collected in 2020 near the origin of the 06R4 and 93R lines, and found to be resistant to glyphosate, ACCase inhibitors, and ALS inhibitors, while being slightly less resistant to triallate and difenzoquat than the 06 biotypes [Keith et al., unpublished]. Thus, the stable, inbred resistant biotypes used in this research have varying degrees of similarity and overlap in herbicide resistance. The resistance phenotypes of these biotypes may be represented in their respective hyperspectral signatures.

The goal of this study is to determine if machine learning models can discriminate resistant biotypes of wild oat. This research seeks to expand on previous research by increasing the number of resistant biotypes to classify and introducing context about the resistance phenotype of the biotypes. We used a dataset with a small number of high-quality hyperspectral scans and evaluated the robustness of machine learning models trained on mean spectra from resistant and susceptible wild oat plants. This research shows the potential for machine learning

hyperspectral approaches to discriminate between several resistant biotypes in wild oat to be transferred to field environments. Furthermore, it highlights the need to create models which predict resistance on spectral signatures related to the specific resistance of weeds, allowing for the correct herbicide to be selected for application.

Methods

Plant Growth

Wild oat biotypes were grown in a greenhouse in the Plant Growth Center at Montana State University. The six biotypes included two lines susceptible to all herbicides tested: S1 (Keith et al. 2015), and S2 (Johnson et al. 1995), and four lines demonstrating varied levels of multiple herbicide resistance: 93R (Kern et al. 2002), 06R3 (Keith et al. 2015), 06R4 (Lehnhoff et al. 2013; Keith et al. 2015), and 20R2 [Keith, unpublished]. Two replications of six individuals for each of the six parental biotypes were planted in 3.8 x 21 cm cone containers in standard greenhouse soil mix [1:1:1 (by vol) Bozeman silt loam:Sunshine mix #12 (Sun Gro Horticulture, Inc., Bellevue, WA):perlite]. Plants were grown under a 16-hour photoperiod of natural light supplemented with mercury vapor lamps ($165 \mu\text{mol m}^{-2} \text{s}^{-1}$ PAR), at $25 \pm 4^\circ\text{C}$, fertilized weekly with 100 ppm Jack's 20-20-20 (JR Peters Inc., Allentown PA). Plants were grown until the third leaf measured 4-6 cm in length, roughly Zadok's stage 13 (Zadoks et al. 1974; Tottman et al. 1979), then were moved to a dark room to be imaged.

Table 3.1: Counts of greenhouse grown biotypes imaged between A and B replicates.

	20R2	93R	06R3	06R4	S1	S2	Total
A Replicate	5	5	5	5	5	5	30
B Replicate	5	5	6	6	6	4	32
Total	10	10	11	11	11	9	62

Benchtop Hyperspectral Imaging

A Resonon Pika-L (Resonon Inc, Bozeman, MT) imager paired with a moving stage was used to obtain hyperspectral images of the wild oat biotypes. Resonon's Spectronon software was used to interface with the imager. Scans were obtained from the third leaf in a dark room with a halide line light. One leaf was imaged per plant, obtaining a high-resolution scan containing roughly 50,000-90,000 pixels. Dark correction, using dark reference cubes collected during imaging, and response correction were performed at the time of imaging using built-in Spectronon functions. The scans ranged from 400 nm to 1000 nm and were spectrally binned with a factor of two for a total of 300 bands with a spectral resolution of 2 nm. The resulting data cubes had dimensions 1190 x 900 x 300. A color card with a known spectral response was imaged after each batch to ensure the instrument consistency. Across six biotypes, 62 total plants were imaged (Table 3.1).

Spectral Post-processing

Hyperspectral leaf scans were imported to the Spectronon software for preprocessing. Datacubes were interpolated to a reference set of spectra collected from a Pika-L Aerial platform, reducing the spectral resolution by a factor of two. Dark correction during imaging was confirmed. As the ends of the spectrum were noisy, the cubes were cropped spectrally to a range of 450 to 950 nm. The spectra were min-max normalized, as the dark room conditions coupled with a simple background profile indicated that it was an acceptable method (Mazdeyasna et al. 2025). A Savitzky-Golay filter with a window length of 13 points, third-degree polynomial, and a derivative of 0 was passed over each scan to remove noise (Savitzky and Golay 1964). Leaf

pixels were extracted from the background using the built in Foreground Object Detection method, passing in a sample cube built from the background of the scan, with structuring size 5, edge erasure, and an object size cutoff of 5000 pixels. Resulting datacubes were passed into a Python script which extracted individual pixels values using the *SPy* module (Boggs et al. 2022), after being passed through a loose NDVI filter of 0.3 to remove any outlier pixels present within the leaves. Means were computed per-leaf across all pixels using R 4.4.2 on the MSU Tempest Research cluster (R Core Team 2025). The derivative of the mean spectra was computed with a Savitzky-Golay filter, using a polynomial order of 2, window length of 3, and the first derivative (signal developers 2023). This step was critical as the derivative transformation reorients the spectra to a common baseline, potentially increasing the predictive power of the models. The processed data had 234 spectra used as predictors in the machine learning models.

Modeling And Analysis

With 62 observations, we selected models appropriate for small datasets and resistant to overfitting. SVM models have shown strong classification results, and pair well with dimensionality reduction techniques (Nugent et al. 2018; Mensah et al. 2025). We fit the SVM models with a Radial Basis Function (RBF) kernel to capture non-linear relationships, tuned with the *c* and *sigma* hyperparameters. Random forest models have shown promise due to their robustness to overfitting and low complexity (Breiman 2001). The RF models were tuned with hyperparameters *mtry* and *min node size*. PLS-DA models have built-in dimensionality reduction and were tuned using the *ncomp* parameter. The K-NN models were fit with the *knn* package (Schliep and Hechenbichler 2016), with the tuning parameters *kmax*, *distance*, and *kernel*. Cohen's Kappa was used as the metric for model parameter tuning. Accuracy and F1 score were

recorded for performance analysis, as well as per-class sensitivity and specificity. Modeling was performed in R 4.4.2 on the MSU Tempest Research cluster primarily using the *tidymodels*, *parsnip*, *recipes*, and *rsample* packages (Kuhn and Wickham 2020; Frick et al. 2024; Kuhn and Vaughan 2024; Kuhn et al. 2025).

The six biotypes were aggregated into two classes, susceptible (S1 and S2) and resistant (93R, 06R3, 06R4, 20R2) to test models' ability to discriminate general resistance. Test and training sets of 21 and 41 plants respectively were partitioned for the aggregated biotype model, stratified by resistance class. Five-fold cross validation was repeated five times for hyperparameter tuning, and the best hyperparameter combination was refit to the training data and used to predict the test set (Figure 3.1).

For classifying across the six biotypes, nested cross-validation was chosen as an appropriate technique for hyperparameter tuning and model assessment given the low number of observations per biotype (Figure 3.1B). Nested cross-validation performs hyperparameter tuning inside the training set of a fold within cross validation, meaning that test metrics are not positively biased to a set of hyperparameters (Wainer and Cawley 2018). Data was split into five folds for the outer test loop, and each outer loops training data were split into five folds for the inner hyperparameter tuning loop, both stratified by biotype. This process was repeated five times to attain robust performance estimates for the models. Average outer fold metrics from the best tuned respective model were calculated to assess how well each model was able to classify biotypes.

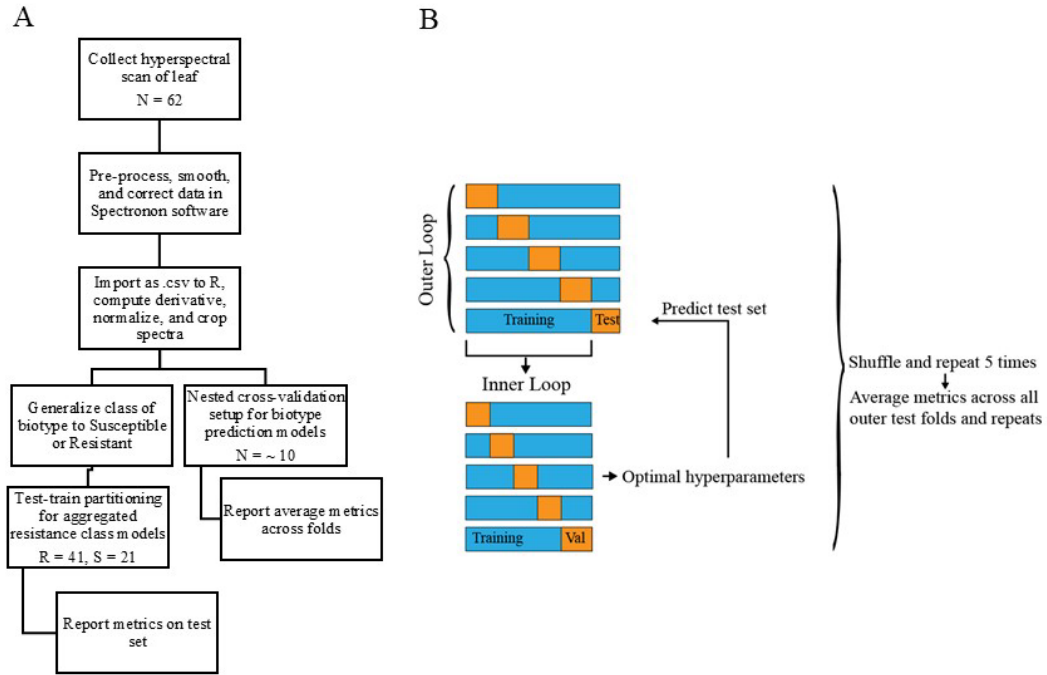


Figure 3.1: A) Hyperspectral analysis flow chart and B) schematic for nested cross validation

In addition to using the spectra to predict resistance, distance metrics were also computed between biotypes mean spectra to build inference about biotype separability. Canberra and cosine distance were chosen (Deborah et al. 2015). Canberra distance was chosen as it had a high saturation point and satisfied the definition of a true metric (Deza and Deza 2009; Deborah et al. 2015). Cosine distance does not satisfy the identity of indiscernibles property, as two spectra which differ only by a scalar value will have a distance of zero, meaning cosine distance picks up on differences in spectral shape (Deza and Deza 2009; Deborah et al. 2015). Non-normalized spectra were used to compute Canberra distance, with the goal of retaining possibly subtle differences in the bands. Normalized spectra were used to calculate cosine distance, minimizing potential magnitude differences in spectra to isolate spectral shape.

Results

The biotypes exhibited visual differences in spectra in both the pre-processed spectra and the derivative spectra (Figure 3.2). At ~550 nm, the resistant biotypes (93R, 06R3, 06R4, 20R2) appear to group at higher mean reflectance than the susceptible lines, which is mirrored in the higher gradient in the derivative spectra. Changes in the slope of the spectra are more obvious visually in the derivative spectra at ~730 nm. At ~750 nm in the derivative spectra, there are notable apparent differences among the biotypes, coinciding with the flattening of the spectral profile.

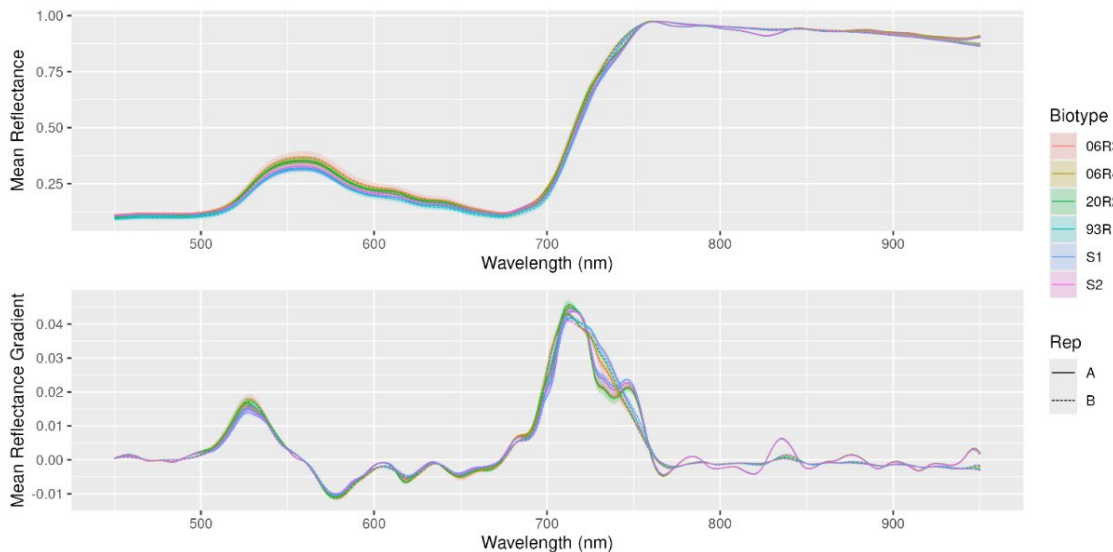


Figure 3.2: Mean reflectance of biotypes and mean first derivative spectra of biotypes.

Two-class Discrimination

Given the visually detectable differences among spectra, machine learning models may successfully discriminate biotypes. To build a foundation of biotype separability into two classes (susceptible versus resistant), four models (K-NN, RF, PLS-DA, SVM) were evaluated. Resistant

biotypes (06R3, 06R4, 93R, 20R2) were generalized to the “Resistant” class. The susceptible biotypes were generalized to the “Susceptible” class. The generalized models only misclassified a few observations in the test set (Figure 3.3). Model performance was similar across all models. PLS-DA was the top performing model with perfect accuracy. The SVM and K-NN models performed the worst, with accuracies of 90% (Table 3.2). Using mean spectra derived from a small training set, the models were successful at broadly classifying resistant and susceptible wild oat.

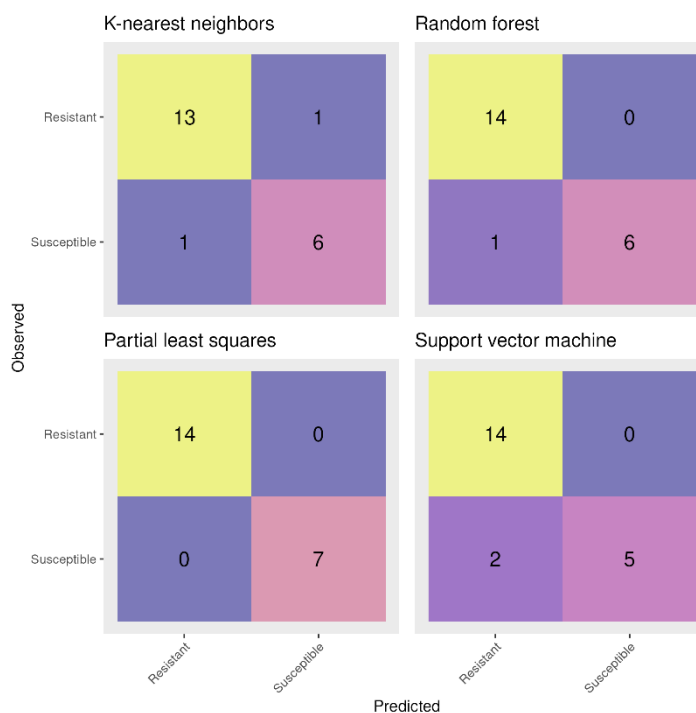


Figure 3.3: Confusion matrices for generalized resistance class models. The top-left to bottom right diagonal indicates correct predictions by each model.

Table 3.2: Model assessment metrics and tuning parameters for resistant vs. susceptible models.

Model	Accuracy	Kappa	F1 Score	Tuning Parameters
K-NN	0.90	0.79	0.89	$k = 1$, <i>biweight</i> , <i>power</i> = 1
RF	0.95	0.89	0.94	<i>mtry</i> = 2, <i>min n</i> = 1
PLS-DA	1.0	1.0	1.0	<i>num components</i> = 5
SVM	.90	0.77	0.88	<i>cost</i> = 21.5, <i>sigma</i> = 0.46

Biotype Discrimination

The wild oat biotypes differ in both their range and degree of herbicide resistance (Kern et al. 2002; Keith et al. 2015). Therefore, models were evaluated for their performance discriminating among the six biotypes (Table 3.3). Overall, models tended to accurately classify biotypes with notable patterns in how they failed, tending to confuse specific pairs of biotypes (Figure 3.4). The K-NN model failed to discriminate between 06R3 and 06R4 and between S1 and S2. It had low accuracy and F1 scores, 52% and 0.52 respectively. The PLS-DA model had similar scores, 52% accuracy and 0.53 F1. The SVM model misclassified S1 and S2 plants as 20R2 and 93R, and had an average accuracy of 63% and an F1 score of 0.64. The RF model performed the best, classifying almost all 93R plants correctly, and had better results for the the S1 and S2 biotypes than the other models. The RF model had an accuracy of 64% and F1 score of 0.64.

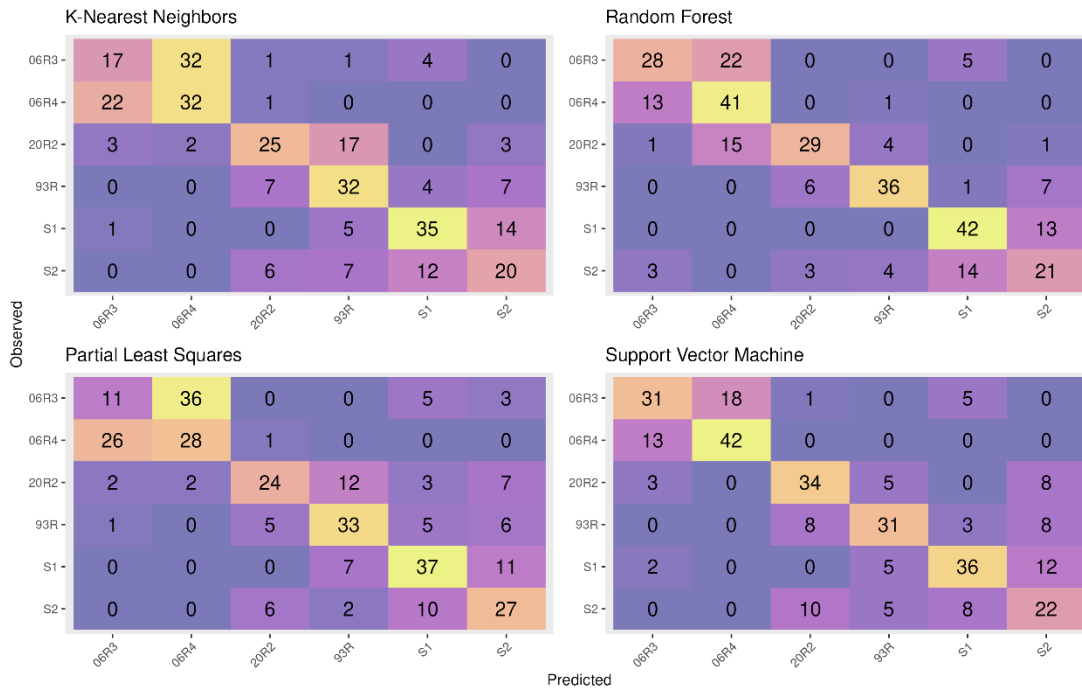


Figure 3.4: Confusion matrices summed across outer test folds and repeats for biotype classification models. Top-left to bottom-right diagonal indicates correct predictions.

Table 3.3: Model performance and tuning parameters for biotype classification average across outer folds

Model	Accuracy	Kappa	F1 Score
K-NN	0.52	0.42	0.52
RF	0.64	0.56	0.64
PLS-DA	0.52	0.42	0.53
SVM	0.63	0.56	0.64

Given there are six classes, the performance metrics require close consideration. Though the accuracy for each model is relatively low (Table 3.3), after examining the confusion matrices (Figure 3.4), it is apparent that most misclassification occurs between the 06R3-06R4 and S1-S2 groups across all models. However, the RF and SVM models were relatively more successful

when discriminating among the 06R3/06R4 biotypes than the PLS-DA and KNN models. For all models, the bulk of predictions fall along the diagonal.

Distance Metrics

To provide a basis for why spectral models may struggle to discriminate between certain biotypes, distance matrices were computed among the mean spectra for the biotypes. Both Canberra and normalized cosine distance were computed to assess spectral differences and curve differences respectively. Both distance metrics showed emerging patterns between the resistant and susceptible biotypes (Figure 3.5). 20R2 appears to isolate from the susceptible biotypes more than the resistant biotypes. In the cosine distance plot, it has a lower distance to 06R3 and 06R4 than 93R. 93R has lower distance to susceptible biotypes than the resistant biotypes, and this difference is accentuated in cosine distance. In both plots, 06R3 and 06R4 have low distances between them, and high distances to the susceptible biotypes. Canberra distance shows a middle level for S1 to S2, but they appear more closely related in the cosine distance plot.

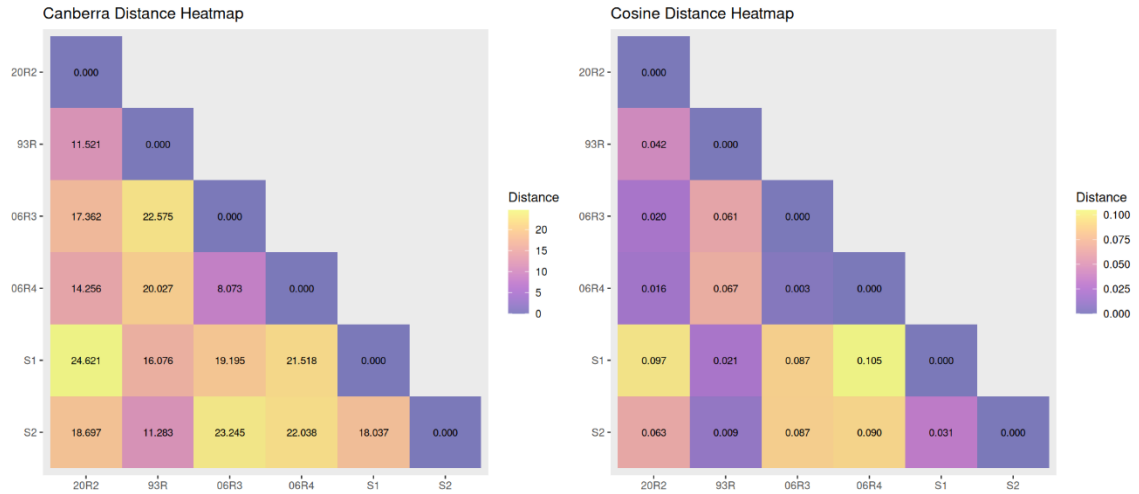


Figure 3.5: Canberra and cosine distance matrices for biotypes. Distance is represented by a heatmap, with color intensity of the square representing similarity between the corresponding biotypes on the axis.

Discussion

Establishing the cases for which pre-application detection of herbicide resistance is successful is critical to translate these spectral approaches for use in the field. Previously, resistance in weeds such as lambsquarters, kochia, marestail, and johnsongrass has been detected spectrally using advanced machine learning models paired with pixel-level data (Huang et al. 2018; Nugent et al. 2018; Mensah et al. 2025). Previous research has shown that detection could be performed in the presence of crops and that hyperspectral imagery can separate different plant species with very high accuracy (Scherrer 2019). The next step is to classify among different resistance phenotypes in a single weed species since herbicide resistance evolves across populations. In this research, classification success across six biotypes of wild oat varied in degree based on phenotype similarity.

Traditional machine learning models (K-NN, RF, PLS-DA, SVM) discriminated between resistant and susceptible wild oat biotypes when binary resistant or susceptible labels were applied to observations (Figure 3.3). All models had classification accuracies between 90% and 100%, with PLS-DA performing the best with an accuracy of 100% (Table 3.2). However, these binary categorizations are overly simplistic. The biotypes vary not only in the range of herbicides to which they are resistant but also vary in the degree of response (Table 3.4). For example, a 20R2 plant classified as resistant exhibits very different herbicide resistances than 93R (Kern et al. 2002). However, this analysis does indicate that spectral features may be similar among herbicide resistant biotypes. We note that relative to the global, or even United States, population of wild oat, these biotypes are only a small sample, and future data collection that captures the diversity of wild oat across environmental conditions is necessary to increase the generalizability of this approach.

Table 3.4: Resistance phenotypes of herbicide resistant wild oat.

Biotype	Location	Resistance
06R3	Teton County	ACCase, ALS-inhibitor, difenzoquat, triallate, paraquat
06R4	Teton County	ACCase, ALS-inhibitor, difenzoquat, triallate
20R2	Teton County	ACCase, ALS-inhibitor, difenzoquat, triallate, glyphosate
93R	Teton County	triallate, difenzoquat
S1	Teton County	None detected
S2	Canada	None detected

To evaluate the ability of these models to distinguish more subtle differences in herbicide resistance, we evaluated models for classifying among biotypes. These models performed worse

than the binary classification in general, which is expected given the increase in the number of classes—and points to a need for more subtle tools for model evaluation when relationships among classes are present. For all models, most misclassification occurred between 06R4 and 06R3 (Figure 3.4). RF and SVM performed the best classifying between 06R4 and 06R3, while the K-NN and PLS-DA models had more misclassifications between the two biotypes (Figure 3.4). The 06R3 and 06R4 biotypes are very similar spectrally based upon visual inspection of the spectra (Figure 3.2) and the low Canberra and cosine distances (Figure 3.5). Furthermore, these biotypes were collected from adjacent fields in Teton County, MT in the same year, and only differ in resistance to one tested herbicide, paraquat (Keith et al. 2015). It is likely that these biotypes are similar genetically; however, this is the case where specific detection of herbicide resistance is needed. Paraquat is a common herbicide used to control wild oat, and thus failing to resolve the 06R3 and 06R4 biotypes could lead to failure in herbicide application. Importantly, 06R3 and 06R4 were rarely confused with S1. S1 was also collected the same year in the same location as 06R3 and 06R4. S1 was confused with S2, which was identified in the 1970s in Canada (Johnson et al. 1995). These patterns suggest that patterns of herbicide resistance, and not only geographic location, were driving correct classifications.

An important performance metric for the biotype models is the propensity for the model to predict that a resistant biotype is one of the two susceptible biotypes. If these models were to be applied to the field, a false negative prediction could lead to herbicide being applied to a resistant plant. Across all models, the random forest model only predicted that a resistant observation was susceptible 13 times (Figure 3.4). The other models had varying levels of false negatives for resistant biotypes: the K-NN model had 18, the PLS-DA model had 29, and the

SVM model had 24. To further investigate the models' ability to classify specific biotypes, per-biotype metrics for precision, sensitivity, specificity, and support were calculated for each outer fold (Table A2). 06R3 was the hardest class to predict, with consistently low precision and sensitivity scores, with SVM having the highest precision of 0.56 and the highest sensitivity of 0.57. Spectral similarity to 06R4 drives most of the misclassification, however 06R3 was also more commonly predicted as susceptible biotypes than 06R4. 20R2 and 93R were both distinguishable from susceptible biotypes by RF and SVM, though PLS-DA and K-NN struggled more. K-NN and PLS-DA models had sensitivity and precision scores 0.2 points less than the RF and SVM models for 06R3 and 06R4. The PLS-DA model did have marginally better metrics when predicting the S1 and S2 biotypes and had the highest number of correct S2 predictions (27). Between the RF and SVM models, the SVM model tended to have slightly higher metrics for the 06R3, 06R3, and 20R2 biotypes, though slightly lower metrics than RF for the 93R, S1, and S2 biotypes (Table A2).

Shared patterns of herbicide resistance were also potentially driving misclassifications of 93R. The SVM and K-NN models both frequently predicted 93R as susceptible (S1 or S2)(Figure 3.4). Given the low cosine distance between 93R to either the S1 or S2 biotypes, misclassification is expected. In addition, 93R is the oldest herbicide resistant biotype in this study and shows resistance to the least number of herbicides among the resistant biotypes, meaning it has the most similar resistance phenotype to the susceptible biotypes in this study.

In comparison to previous work (Nugent et al. 2018; Scherrer 2019), we trained traditional machine learning models on coarser data, using mean spectra from a leaf rather than multiple pixels from the same plant. Using the mean spectra was sufficient for distinguishing

wild oat biotypes, and modeling with this approach provided several benefits. First, training models on mean spectra rather than pixels ensured no data leaking between test and training sets, avoiding overfitting that can happen if pixels are not subset appropriately. Second, the potential for identifying herbicide resistant weeds using hyperspectral imagery in producer fields may include lower resolution data, such as those collected by unmanned aerial systems, a technology with substantial interest in both commercial and research settings (Lachowiec et al. 2024). Therefore, weeds may only be represented by a few pixels, which is similar to the mean spectra approach herein.

An important caveat to these results is that the models may be picking up on genetic background differences between the lines that are unrelated to resistance. Disentangling real spectral signature related requires that a model would be able to predict resistance between biotypes which differ in resistance mechanisms or resistance phenotypes but share a genetic background. This could be accomplished with breeding populations.

Overall, these results lay the groundwork for early detection of herbicide resistance in wild oat. Similar biotypes were misclassified as one another, especially when resistance phenotypes are only slightly different. If resistance phenotypes are to be distinguished, more complex models paired with more data may be necessary to resolve biotypes. More complex models such as multilayer perceptrons (MLP), convolutional neural networks (CNN), and the growing suite of deep learning techniques can derive deeper features and complex non-linear relationships with the potential to better resolve phenotypes (Polder et al. 2024). These results are a crucial step toward application of pre-application resistance phenotyping in wild oat.

Conclusion

Resistant wild oat biotypes exhibit spectral differences which can be leveraged for successful herbicide resistance classification. Classification accuracy was high when biotypes were generalized to binary resistant or susceptible classes. Model shortcomings were largely due to within resistance-class misclassification among biotypes with similar patterns of herbicide resistance. This technology has the potential to save producers time and money while reducing selection pressure for highly resistant weeds. Subsequent studies focusing on generalizability to the field could leverage more complex techniques alongside broader biotype selection to apply these pipelines.

CHAPTER FOUR

HYPERSENSPECTRAL ANALYSIS IN RECOMBINANT INBRED
LINES OF HERBICIDE RESISTANT *AVENA FATUA*Summary

Prediction of herbicide resistance traits in weeds prior to herbicide application would be a boon to producers, reducing costs and saving time. Hyperspectral imagery of plant reflectance may capture signatures that are associated with the degree of herbicide resistance present in a weed. However, field collected imagery is subject to the noise of such an uncontrolled environment and may swamp signal. To explore the utility of hyperspectral data to describe the magnitude of resistance in a plant, both aerial and benchtop hyperspectral data were collected on two recombinant inbred line (RIL) populations, each derived from a different multiple herbicide resistant biotype of wild oat. These RILs displayed variation in herbicide resistance along a spectrum. Several modeling techniques were used to test if spectra could be used to predict resistance in RILs. UMAP performed on both raw spectra and convolutional bottleneck features did not show any clustering of resistant or susceptible RILs within imagery collected in benchtop and aerial trials or when trials were aggregated together. Partial Least Squares regression (PLSR) models were unable to model resistance across trials for any of the herbicides screened. It cannot be concluded that any spectral signature related to resistance exists for these populations of wild oat. This may be a technical or data related problem, though these possibilities have important implications for research in this area. It cannot be assumed that models are picking up on any signal related to resistance. Future research should attempt to validate these findings in other

herbicide resistant weeds, particularly those with resistance mechanisms which may confer spectral changes. To build generalizable models, programs attempting to build pipelines for herbicide resistance detection may want to focus on capturing a wide range of resistance phenotypes.

Introduction

The internal state of a plant is manifest in its spectral reflectance. The field of precision agriculture has contributed research to leverage the relationship between spectral reflectance and plant characteristics of interest. Spectra can be indicative of the molecular profile of a plant (Amigo and Grassi 2019). Across the regions of the plant spectrum different inferences concerning the molecular source of spectral variability may be made. The visible portion of the plant spectrum (400-700 nm) is dominated by absorption of plant pigments such as chlorophylls, carotenoids, and anthocyanins (Blackburn 2007). The near-infrared portion of the plant spectrum (700-1300) is associated with cell structure and leaf morphology, while the mid-infrared spectra (1300-2500 nm) is affected by water absorption (Sahoo et al. 2015; Kothari and Schweiger 2022). Representation of additional plants traits in spectra can be extremely difficult due to subtle and overlapping adsorption features (Blackburn 2007; Kothari and Schweiger 2022). Passive sensing techniques can collect plant reflectance spectra at scale.

Multispectral and hyperspectral imaging are two technologies used to collect spectral reflectance data. Multispectral imagery collects spectral signals in several discrete bands and has been an effective tool for agricultural purposes when integrated on imagery platforms ranging from close-range to satellite-based (Tsoulias et al. 2023). Multispectral imagery has been used to assess many plant attributes such as chlorophyll content (Singhal et al. 2019; Qi et al. 2021; Li et

al. 2025), biomass (Wang et al. 2021), and nitrogen content (Shu et al. 2024). In contrast to multispectral imagery, hyperspectral imagery collects many narrow contiguous spectral bands across a wide range of spectra (typically from 350-2500 nm), capturing fine-scale spectral features (Sahoo et al. 2015). Hyperspectral imagery has been used to collect information about the biomass (Mariotto et al. 2013), nitrogen status (Vigneau et al. 2011; Sun et al. 2017), metabolite presence (Vergara-Diaz et al. 2020; Raghav et al. 2026), and many other traits indicative of a plant's state, often with superior performance to that of multispectral approaches (Lee et al. 2004; Lu et al. 2020; Sarić et al. 2022). Plants that harbor herbicide resistance can exhibit constitutive differences in their cellular milieu relative to susceptible plants (Dyer 2018; Giacomini et al. 2018), which may be captured by measuring a plants reflectance using hyperspectral imagery (Scherrer et al. 2019; Mensah et al. 2025; Singh et al. 2026). A fundamental goal of machine learning in understanding the hyperspectral reflectance of weeds is to build models which can generalize to new instances of weeds in field settings (Domingos 2012). Many factors influence plant reflectance in the field, meaning that successful models must parse out signal from noise across many imaging environments. Two obvious solutions to this problem include the following: 1) Train models to discriminate between many resistant biotypes across multiple environments and 2) Train models to identify spectral features indicative of a mechanism of resistance rather than indicative of a biotype-environment interaction.

Chapter 3 of this research addresses Solution 1 in part: we showed that models can classify among some resistant biotypes. However, these models would likely perform very poorly on any new observations of these biotypes in field environments as the models lack context of the spectral profile under different imaging and growth conditions. Three publications

have utilized similar hyperspectral approaches for biotype prediction across several resistant biotypes. Singh et al. (2026) discriminated both glyphosate and fluroxypyr resistant kochia from susceptible observations, and demonstrated that models trained on greenhouse grown plants had moderate generalization to field conditions (Singh et al. 2026). Mensah et al. (2025) discriminated between dicamba-resistant and glyphosate-resistant kochia in greenhouse conditions and noted the limited generalizability of the model to field application (Mensah et al. 2025). Scherrer et al. (2019) also discriminated between dicamba-resistant and glyphosate resistant kochia, though this research was solely performed in field conditions using both ground-level and drone imagery. Models trained on data collected from the drone-platform performed worse than ground-level data when discriminating among resistant biotypes (Scherrer et al. 2019). Scherrer et al. (2019) also noted that to assess the efficacy of the herbicide resistance detection methods developed, future research needs to apply the trained model to new data which varied in lighting, imaging methods, and environmental conditions.

To date, no published research has been reported which seeks to assess Solution 2. If models could learn spectral signatures related to the mechanism of resistance, they may be able to generalize well with new observations which share a mechanism of resistance. It is also important that both solutions are intertwined; models trained across many biotypes may pick up on spectral signatures, given that they share a mechanism of resistance and that the mechanism confers a spectral change. This research tests the hypothesis that the mechanisms of resistance in the 06R4 and 93R biotypes confer a spectral change.

To investigate the hypothesis that machine learning models trained on hyperspectral data may be able to detect mechanisms of herbicide resistance, two populations of recombinant inbred

lines (RILs) derived from 06R4 x S2 and 93R x S2 were planted in greenhouse and field environments. If models are unable to predict resistance in RIL populations, that suggests that the models developed in Chapter 3 may be classifying biotypes based spectral features unrelated to resistance.

Methods

Benchtop Imagery

Under the same conditions and methods as described in the Chapter 3 methods section, RILs (as described in Chapter 2) were planted in the MSU Plant Growth Center and imaged using a benchtop Resonon Pika-L hyperspectral camera (see benchtop description in Chapter 3). Plant pixels were extracted using the *SPy* module (Boggs et al. 2022), *SciPy* module (Virtanen et al. 2020), and the *scikit-image* module (Walt et al. 2014). Plant pixels were isolated with an NDVI filter of 0.55, based upon manual examination of masks generated in the Spectronon software. Both pixel dilation and erosion filters were applied to the cubes. Spectra were smoothed using a Savitzky-Golay filter with window length 11, second order polynomial order, and a derivative of one (Savitzky and Golay 1964). Mean spectra were extracted for two reps of the 93R RIL population and one rep of the 06R4 population.

Airborne Hyperspectral Imagery

Two years of airborne hyperspectral data were collected on both RIL populations. For the first field season, wild oats were planted on June 28, 2024, at the Western Triangle Agricultural Research Center in Conrad, Montana. Prior to planting, the field was rototilled. For both RIL populations, every RIL were planted and one individual of each parent in a randomized complete

block design, with three blocks. Plants were spaced two feet apart along the line, and blocks were one meter apart. Five seeds were sown per RIL using a hill plot seeder. Every other RIL was marked with a colored stake, staggered in adjacent lines to create a grid which could be used to identify RILs in airborne imagery. The field was watered in strong dry spells, and non-target weeds were hand-pulled over the course of the experiment.

Airborne hyperspectral data was collected with two systems. In 2024, an Aurelia X8 Max (Aurelia Aerospace, Las Vegas, NV) was paired with a Resonon Pika-L hyperspectral camera (Resonon Inc, Bozeman, MT). The open-source software MissionPlanner (version 1.3.82) (Osborne 2025) was used to plan flights for the Aurelia X8 Max drone. Plots were flown at the minimum feasible air speed of 0.5 meters per second at an altitude of 15 meters. This required an imager framerate of 107 frames per second and achieved a ground pixel resolution of one centimeter. Flights were performed at 10 am to minimize wind exposure while still flying under a high sun angle. A two-panel reflectance tarp provided by Resonon (Type 822, Group 8 Technology, Inc., Provo, UT) was used to calibrate raw data to reflectance.

Four flights were performed over the course of the experiment, though only one was successful in clean data collection. The first flight, performed on 07-24-2024, was successful, though plants were too small to reliably be identified in georectified data. Two other flights, performed on 08-01-2024 and 08-15-2024 captured the target plots but were unable to be georectified. However, the flight performed on 08-09-2024 had high-quality georectified data but the calibration tarp was not captured due to incorrect MissionPlanner settings. To convert the data to reflectance, a ground-control point captured in all flights was used as a reflection standard.

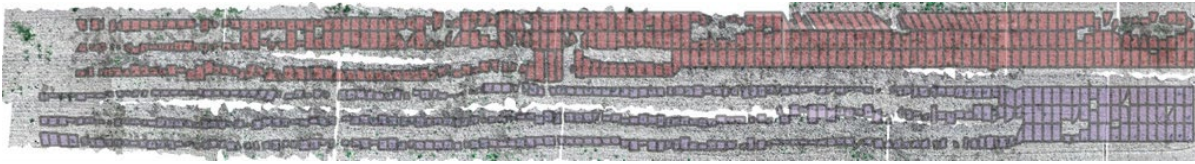


Figure 4.1: Georectified reflectance datacubes captured on 07-09-2025. Polygons used to extract RIL spectra are overlain for the 06R (red) and 93R (purple) RIL plants.

For the 2025 field season, wild oat RILs were planted on June 10th, 2025, at the Post Research Farm in Bozeman, MT. A similar planting scheme was used as at Conrad, though RIL location within a replicate line was re-randomized and a 2.5-foot distance between plants was used. In the 2025 season, a Vision Aerial Vector (Vision Aerial, Four Corners, MT) was paired with a Resonon Pika-L GigE hyperspectral camera with a downwelling sensor provided by Ocean Optics (Ocean Optics). The Vision Aerial Vector drone was flown with the proprietary software on the transmitter. Three flights were performed, and though all were successful, the last flight performed on 07-09-2025 was used as the plants were largest (Figure 4.1). The use of a downwelling sensor and tarp allowed calculation of correlation coefficients to better account for downwelling sensor noise when converting to reflectance.

Post-processing of Airborne Hyperspectral Imagery

Using the Spectronon software (v17.8, Resonon Inc., Bozeman, MT), datacubes were georectified and converted to reflectance. Plant masks were then calculated using red edge normalized difference vegetation index (NDRE), and vegetation pixels were isolated using a threshold of 0.4. Wavelengths were cropped the 400-950 nanometer range, and spectra were smoothed with a Savitzky-Golay filter (derivative of 1, window length of 13, third degree polynomial) (Savitzky and Golay 1964). Reflectance gradient (first derivative) datacubes were created use downstream plant pixel extraction and subsequent model training.

To extract a plant pixel data, the numerical order of RILs was joined to a shapefile constructed over a mosaic of the datacubes, each represented by an individual polygon. Polygons were constructed in QGIS using the *Research Grid* feature alongside manual curation of polygon boundaries to remove non-target plant matter from polygons (Figure 4.1) (QGIS Development Team 2026). The colored stake grid was used to assign RIL IDs to polygon IDs based on the field planting order. The resulting shapefile and masked datacubes were imported into R version 4.5.2 (R Core Team 2025) on the MSU Tempest research cluster, where the *terra* package (Hijmans 2024) was used to extract both individual pixel data and mean spectra per plant. Mean spectra were only retained if the source polygon had more than 50 masked pixels. Prior to modeling, spectra between all airborne and benchtop datasets were cropped to a range of 430 to 950 nm, interpolated to match wavelengths between datasets, and standard normal variate (SNV) normalized in Python (Barnes et al. 1989; Mazdeyasna et al. 2025).

Analysis

The python packages *scikit-learn* (Pedregosa and others 2011) and *torch* (Paszke et al. 2019) were used to assess whether spectra contained information which could be used to predict herbicide resistance in the field. First, bottleneck features were extracted from a convolutional neural network (CNN) trained on the herbicide response metric for each herbicide screened for each RIL population. The input data dimensions were: 1 (the number of channels, in this case the single spectral dimension) and 242 (the number of spectral bands after interpolation and cropping). A standard batch size of 32 was used. The network architecture was designed to learn spectral patterns across varying levels of spectral detail. The first convolutional layer used a kernel size of 21, the second used a kernel size of 7, and the third used a kernel size of 5.

Between the convolutional layers, inputs were processed with a batch normalization, rectified linear unit (ReLU) activation function, max pooling, and dropout of 0.3 (Figure A2). After the third convolutional layer, inputs were batch normalized, followed by a ReLU activation function, then adaptive average pooling to obtain dimensions of (32, 128, 1). Inputs were then flattened to a bottleneck dimension and passed through two fully connected dense layers with dropout and ReLU activation, with the final layer outputting one prediction of resistance for a given herbicide response. Bottleneck features ($n = 128$) were used to map RILs observations using UMAP, a manifold learning algorithm that is known to preserve local and global data structure (McInnes et al. 2020). UMAP has been used for phenotype prediction using UAS imagery (Farag et al. 2024). The goal of this analysis is to determine if the spectral features the convolutional network learned are related to resistance.

To test whether any experiment alone could be used to model herbicide resistance in the RILs, a partial least squares regression (PLSR) model was fit within every dataset for each herbicide response. PLSR is a common baseline machine learning model that is common for plant phenotyping approaches (Grzybowski et al. 2021). A nested-cross validation approach was used as described in Chapter 3 was used. The cross-validation was grouped by RIL, so that no RIL was present within both the training and validation sets. The inner hyperparameter tuning loop used 3-fold grouped cross validation, and an outer 5-fold grouped cross-validation was used for model evaluation. A grid from 2 to 12 components was used to tune the models. Spearman's correlation was used to assess model performance between the true herbicide resistance value and the predicted. Spearman's correlation between the predicted herbicide response metric and the observed herbicide response metric was used to assess model performance.

Due to the nature of the plant pixel extraction method for the airborne datasets, pixel-level data per plant was readily available. In a similar manner as for the mean-spectra analysis, PLSR models were trained on pixel-level data for each herbicide response for the airborne datasets. The number of pixels per plant was capped at 200. The final dataset count for each year was: 51643 pixels for the 06R4 2024 airborne data, 32233 pixels for the 06R4 2025 airborne data, and 27487 pixels for the 93R 2025 airborne data. The cross-validation scheme was grouped by RIL, meaning that no RIL and therefore no pixels from the same RIL could leak from the training set to the validation set. The model was trained using a grouped K-fold cross-validation scheme with 10 PLSR components. To predict a given RIL's herbicide response (percent biomass of control or mean visual rating) in the validation folds, the mean response across the respective observation's pixels was reported. Spearman's correlation between the observed response and the predicted responses was used to assess model performance.

Results

For the 06R4 datasets, two years of airborne data and one year of benchtop data were collected. After processing the datacubes, extracting plant pixels, and taking the mean spectra of plants with a pixel count greater than 50, 657 mean spectra were collected for 06R4 RIL population. The 2024 airborne dataset had 334 mean spectra, the 2025 dataset had 211 mean spectra, and the benchtop dataset had 112 mean spectra. For the 93R RIL population, 380 mean spectra were collected across one airborne dataset ($n=182$) and two benchtop datasets (replicate A: $n=100$ and replicate B: $n=98$). For the training of convolutional neural networks, the datasets from a respective RIL population were aggregated.

If plant spectra contained subtle differences which were associated with resistance, bottleneck features from a convolutional network may encode that information. Thus, convolutional bottleneck features from models trained on each herbicide response were used as features for UMAP projections of samples.

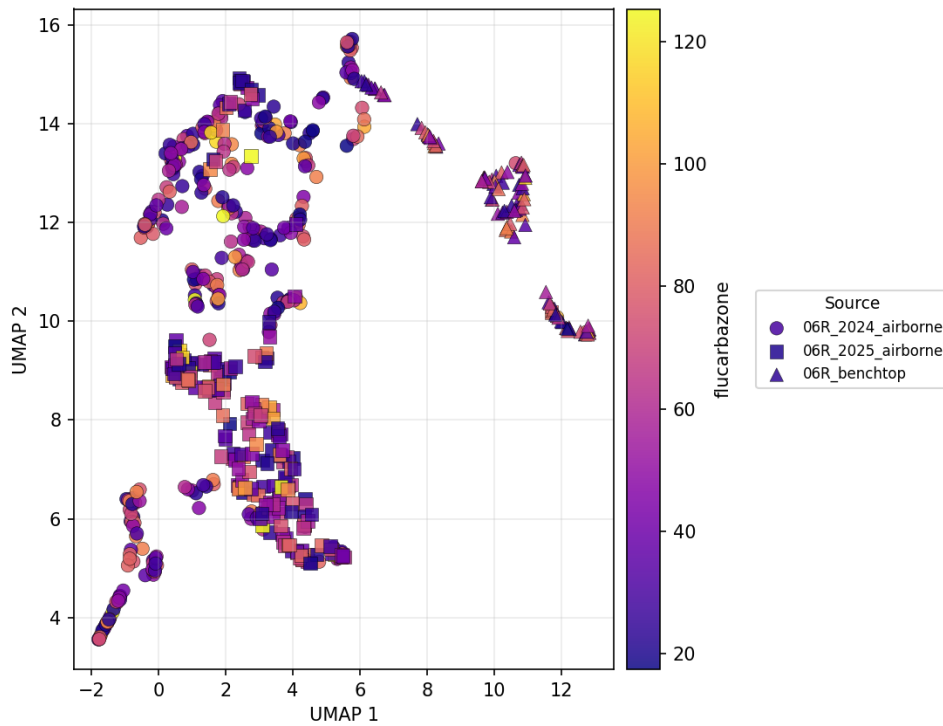


Figure 4.2: UMAP of convolutional bottleneck features for model trained on flucarbazone percent biomass of control across all 06R4 x S2 datasets. Data source was visualized by shape, and the percent biomass of control for flucarbazone per RIL observation was visualized with color. Purple indicates a susceptible RIL and yellow indicates a resistant RIL.

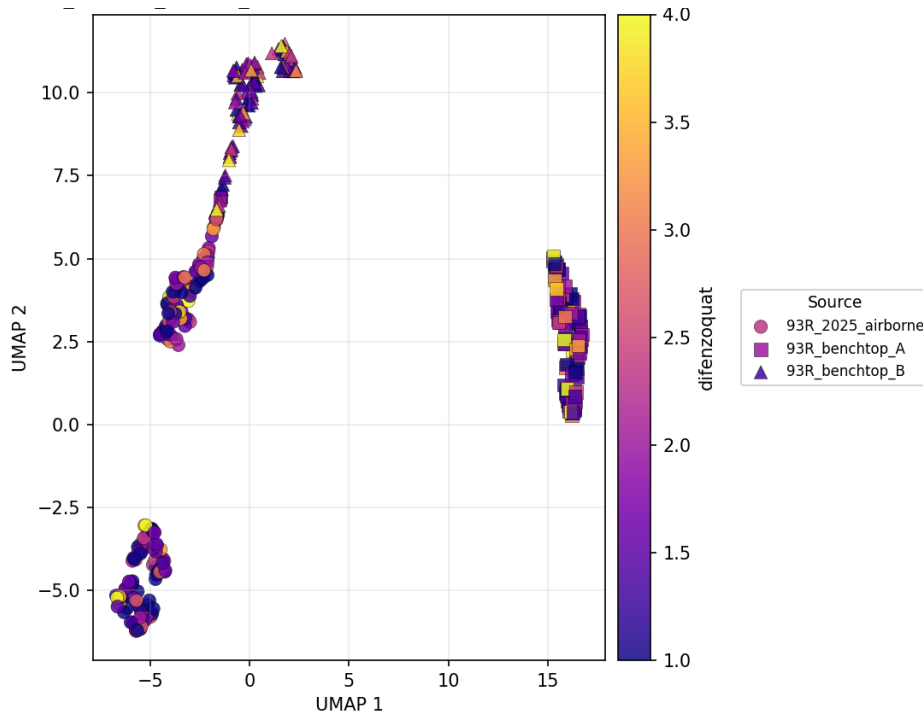


Figure 4.3: UMAP of convolutional bottleneck features for model trained on mean difenzoquat rating across all 93R x S2 datasets. Data source is visualized by shape and the mean difenzoquat rating for each RIL is visualized by color. Purple indicates a susceptible RIL and yellow indicates a resistant RIL.

Visualization of convolutional bottleneck features did not indicate that herbicide resistance was captured in plant spectra for any herbicide response for either the 06R4 or 93R RIL populations. Instead, the samples largely separated based on data source. For the 06R4 UMAP, both aerial datasets clustered separately from the benchtop dataset (Figure 4.2), indicating that collection method and/or growth environment had strong effects on the spectra. The two airborne flights showed some overlap in UMAP space. In the 93R UMAP, the differences between datasets were much more pronounced. The A benchtop replicate was largely separate of the B replicate despite being collected by the same instrument (Figure 4.3). The 2025 airborne data also clustered away from either replicate of the benchtop data, and there is some

clustering within the airborne observations (Figure 4.3). Based on the visualization of herbicide response across samples, no clustering of resistance was observed for any herbicide.

To test if any of the datasets encoded spectral signals for any of the herbicides screened, a partial least-squares regression (PLSR) model was trained for every combination of dataset and herbicide. The Spearman's correlation between the model's predicted herbicide response metric and the observed herbicide response metric across all RILs in that dataset was used to assess model performance.

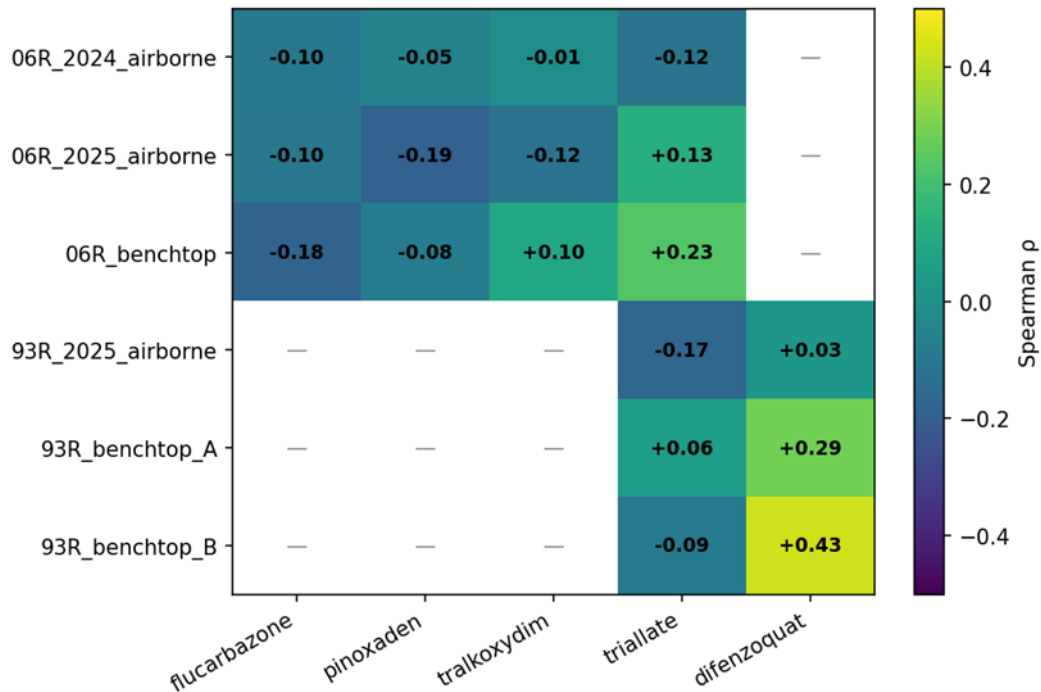


Figure 4.4: Spearman's correlation between predicted and observed herbicide response metric for airborne and benchtop datasets. The percent biomass of control is the response for flucarbazone, pinoxaden, tralkoxydim, and triallate. The mean visual rating is the response for difenzoquat.

The 06R4 RIL population did not have any strong positive correlations of model performance for any of the datasets or herbicides. The strongest correlation was for triallate in the benchtop data with Spearman's ρ of 0.23 (Figure 4.4). However, a scatterplot of predictions

reveals that the correlation is artificially inflated due to the large presence of triallate susceptible individuals in the RIL population with a percent biomass of control value of zero (Figure 4.5).

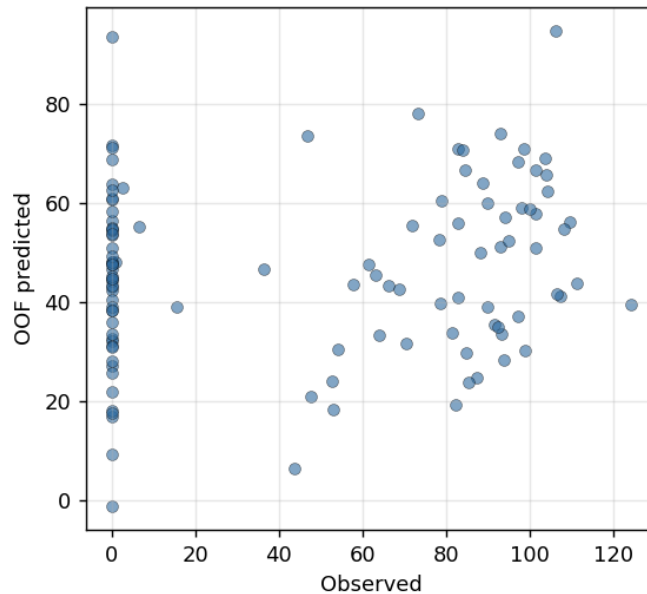


Figure 4.5: Scatterplot of triallate resistance predictions for 06R4 benchtop data. Each point represents a prediction of a RIL that occurred outside the training folds within the nested cross-validation.

For the 93R data, relatively strong correlations of 0.29 and 0.43 were observed for benchtop reps A and B respectively; however, when plotting the predictions it is clear this is an artifact of the rating system for difenzoquat (Figure 4.6). Difenzoquat resistance was visually rated for herbicide injury on a scale of 1-4 (Ex: 2.33). The predictions across difenzoquat ratings are still mostly centered around 2, indicating the lower number of high-resistance RILs is artificially inflating the correlation (Figure 4.6).

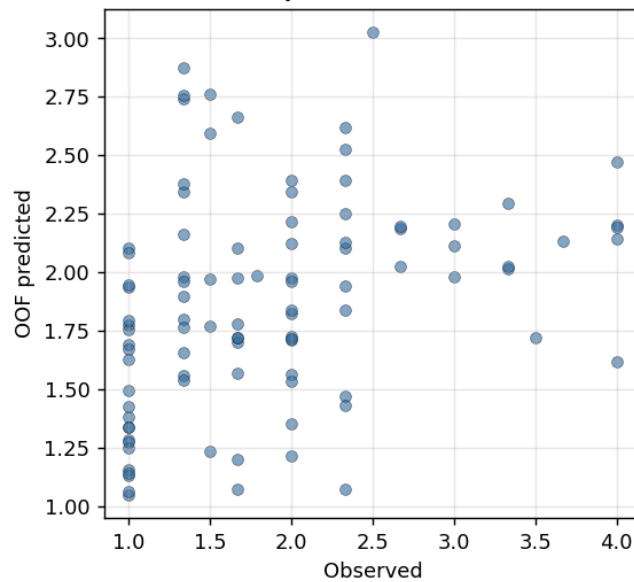


Figure 4.6: Scatterplot of difenzoquat resistance predictions for 93R replicate B benchtop data.

PLSR models were trained using pixel-level spectra from the airborne data for both 06R4 and 93R RIL populations, however no improvement in model performance was observed (Figure 4.7). Several of the instances which had slight positive correlations for the mean spectra PLSR models (Ex: prediction of triallate resistance for 06R4 2025 airborne data) had negative correlations for pixel-level models. Several other candidate models, such as random forest and support vector machine were tested, and were unable to pick up any signal related to resistance in the data.

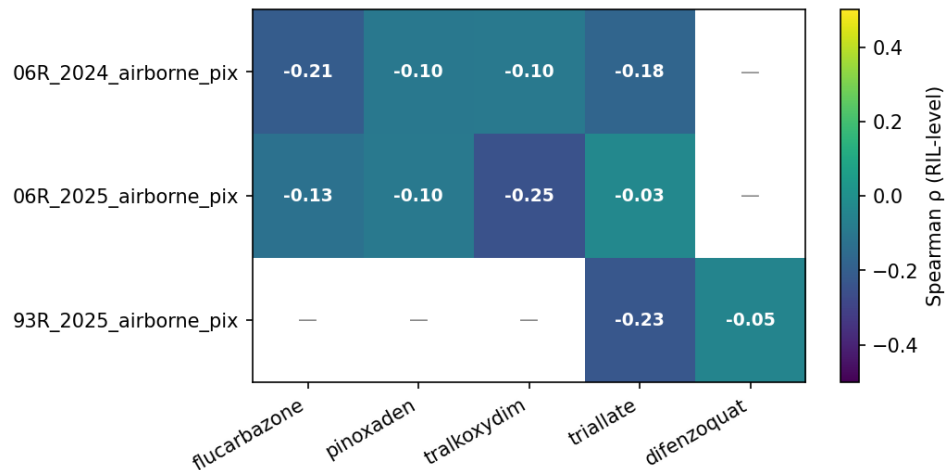


Figure 4.7: Spearman's correlation between predicted and observed herbicide response metric for pixel-level airborne and benchtop datasets.

Discussion

Several machine learning approaches were used to test whether models trained on hyperspectral imagery could detect spectral signatures of resistance in recombinant inbred lines. The approaches used represent a spectrum of complexity: a convolutional neural network may pick up on very fine spectral features at multiple levels of granularity in the spectra if they exist (A. Sharma et al. 2021; Pourdarbani et al. 2021). PLSR models are the standard for many hyperspectral imagery tasks and have demonstrated high performance in many applications (Grzybowski et al. 2021). Neither the CNN nor PLSR models detected any signal related to herbicide resistance in RIL populations leads to several biological and methodological hypotheses.

Data source causes strong spectral effects

Samples clustered strongly by experiment in the UMAP visualizations (Figure 4.2, Figure 4.3). The models may have been unable to parse spectral signal associated with resistance from

the inter-experiment spectral differences. While SNV normalization was applied, other normalization techniques may be necessary to account for changes in spectra between experiments. The inability for by-pixel models to predict resistance indicates spectral differences within an experiment could be masked by noise, at least for airborne data (Figure 4.7). In the field experiments, lighting shifts over the course of a flight may have been too strong to account for with downwelling correction. Across benchtop scans, sensor drift may have masked spectral signals.

Many other methods of spectral data collection exist. Determining the correct approach for resistance detection in wild oat will require balancing scalability and resolution. Handheld hyperspectral cameras, such as the Specim IQ, would be able to collect scans in greenhouse settings with relative ease (Singh et al. 2026). New pipelines would need to be developed to integrate this data.

Different models may pick up on subtle spectral signal

Alternatively, not enough data was collected to fit models complex enough to model a spectra-resistance relationship. There may be ways to leverage by-pixel reflectance information for better prediction of herbicide resistance. A simple approach would be to use a majority vote method across all pixels in a scan. Machine learning methods such as multi-dimensional convolutional networks can also use spatial information for prediction (Ang and Seng 2021). Though the benchtop scans contain tens of thousands of pixels per leaf which could be leveraged for model training, the spectral similarity between pixels is much higher. Models trained on homogenous data are prone to overfitting (Domingos 2012), and the mean leaf reflectance is likely a good representation of the datacubes.

A missed opportunity of this research was to assess whether models could discriminate 93R RILs from 06R4 RILs. This could provide valuable information as to how these models generalize between biotypes. However, because the RILs were planted in separate blocks in the field and flown sequentially, the spectral change over time confounds this analysis completely. In the benchtop imagery, the RILs were imaged in sequential blocks, also confounding any possible between-RIL models with experimental effects. In the future, simply setting a flight path perpendicular to the block design could facilitate these experiments.

Another aspect to consider when building these models is the response variable. The high proportion of RILs wholly susceptible to triallate in the 06R4 population and the discrete scale used to assess difenzoquat resistance in the 93R population both artificially inflated model performance metrics (Figure 4.5, Figure 4.6). Furthermore, because the response variables were collected on RILs not imaged in this study, replicates of each RIL had the same numeric response, possibly inducing overfitting in the models due to pseudo-replication. This could be mitigated in future research by collecting spectra of plants being grown for dose-response assays.

Does the mechanism of resistance in 06R4 and 93R confer
a spectral signature?

This is a critical result for anyone developing approaches for pre-application herbicide resistance detection. This indicates that models will likely generalize very poorly to new biotypes collected, especially if those biotypes are not closely related to biotypes which the model was trained on. Neither of the UMAP projections showed any clustering of resistance for any of the herbicides tested (Figure 4.2, Figure 4.3). The range of kernel sizes implemented (from 21 to 5) means that the CNN models had the ability to pick up on very fine spectral features related to resistance. The CNN architecture is widely used and has demonstrated high performance across

many computer vision and hyperspectral tasks with the ability to resolve fine spectral features (Ang and Seng 2021; A. Sharma et al. 2021). Because aggregating RIL spectra across datasets may introduce too much noise to isolate signal, convolutional bottlenecks were also mapped using UMAP for each individual dataset. No clustering of resistance was observed, though the small number of observations per dataset means the models were likely overfitting.

The research that has been published has generally used machine learning and hyperspectral imagery to classify resistant and susceptible weeds that have discrete resistance labels (Scherrer et al. 2019; Mensah et al. 2025; Singh et al. 2026). This work facilitates detection of these specific resistant biotypes; however, the model's performance is arbitrary if it is unknowingly used to predict different biotypes in a field setting. This is reasonable concern, as the true landscape of herbicide resistance is likely much more complex than a binary system. Thus, future research should not assume that spectral differences in herbicide resistant biotypes are spectral “signatures” related to resistance itself.

Conclusion

This study has important implications for future research on hyperspectral detection of herbicide resistance. We have shown that for two NTSR biotypes of wild oat, models trained on hyperspectral imagery collected in both field and greenhouse environments were unable to parse out any spectral signal related to resistance in recombinant inbred lines. Future studies should apply hyperspectral imaging of RIL populations in other herbicide resistant weeds to validate this finding.

CHAPTER FIVE

ASSESSMENT OF METABOLOMIC AND HYPERSPECTRAL
APPROACHES FOR HERBICIDE RESISTANCE
PHENOTYPING IN *AVENA FATUA*

Metabolomics

While it is a documented practice in medical fields, the use of metabolomics for biomarker discovery in herbicide resistant weeds is understudied. Pre-application detection of herbicide resistance could be a transformative tool in management and monitoring of herbicide resistant weeds and has been explored using other methods. Lowe et al. (2024) identified RNA and protein biomarkers related to metabolic NTSR in resistant black-grass (*Alopecurus myosuroides*) (Lowe et al. 2024). Metabolomic biomarkers have also been researched for crops. Kumaraswamy et al. (2011) used an LC-MS approach to identify several metabolites related to fusarium head blight (*Fusarium graminearum*) in barley (Kumaraswamy et al. 2011). In a similar vein, Tafoya-Razo et al. (2022) showed that several herbicide resistant wild oat biotypes with different mutations conferring resistance had distinct metabolic profiles (Tafoya-Razo et al. 2022). In the future, these profiles could be leveraged to detect differences in populations and make predictions about the resistance phenotypes of collected biotypes.

While this research did not confirm that metabolomics could be used for biomarker discovery in herbicide resistant wild oat, it did set a foundation for future research in a new approach for herbicide resistance detection. The use of RILs demonstrated that certain

metabolomic patterns were conserved between biotypes and derived RILs, indicating that biomarkers could be indicative of mechanisms of resistance rather than specific biotypes. Future research on different areas of the wild oat metabolome could validate and annotate the features described in this research and point to new metabolic biomarkers of resistance. Subsequent workflows could greatly decrease LC-MS runtime by targeting specific areas of the wild oat metabolome.

Hyperspectral Imagery

The results of this research imply that future work leveraging imagery approaches must exercise a degree of caution when assessing the generalizability of models. This research demonstrates that resistance in RILs does not confer a spectral signal related to resistance for two resistant biotypes of wild oat. An alternative explanation is that current hyperspectral imaging platforms are unable to reliably pick up on such a subtle spectral change if one exists. This implies that models classifying resistant biotypes are strictly valid for only the biotypes imaged and may perform poorly if spectra change due to evolving resistance.

An effective framework to assess the hyperspectral models developed in Chapters 3 and 4 is in the context of field implementation. Starting from the generalized resistant/susceptible models presented in Chapter 3: these models assume that all resistant and susceptible biotypes have similar spectral signatures. These models may be powerful within a small geographic region where the assumption of spectral similarity is met. However, as the model is applied to new biotypes, they may not meet that assumption. Thus, the models trained to predict the individual biotype may be more effective to map resistance across larger regions.

This conclusion is only applicable to these biotypes of wild oat. An assumption of these hypotheses is that resistance mechanisms confer a spectral change, possibly due to altered chemical composition in leaves (Mensah et al. 2025). Several important plant compounds conferring spectral changes could be leveraged for detection of herbicide resistance post-application (Dayan and Zaccaro 2012; Jones et al. 2023). It follows that for many other weeds with mechanisms of action potentially altering leaf spectra, these approaches could be valid.

Implications for Resistance Screening

Both approaches assessed attempt to mitigate the need to screen suspected resistant weeds against an assay of herbicides. For future work, both metabolomic and hyperspectral approaches necessitate labeled response data, meaning that screening for herbicide resistance with dose-response assays is still required. This bottleneck is an advantage, as collecting metabolomic and spectral data on individual plants affords stronger inference than biotype-level data, also capturing potential environmental interactions altering herbicide resistance (Dyer 2018).

It has been proposed that using sensors to differentiate weeds among crops imposes selection pressure for crop mimicry in weeds (McElroy 2014). Coleman et al. (2023) proposed that one of the requirements for this type of mimicry is that the weed has morphological similarity to the crop during the weeding period (Coleman et al. 2023). This is certainly the case for wild oat in barley and wheat fields. This inference could be extended to discriminating herbicide resistant weeds from herbicide susceptible weeds, under the condition that a model predicts an herbicide resistant weed as susceptible and that weed has heritable morphological traits.

Weeds are going to be a persistent problem in cropping systems. Pre-application detection of herbicide resistance could help producers outpace evolved resistance. This research is a critical first step in developing metabolomic and hyperspectral methodologies for herbicide resistance detection that can be applied to field environments. Future research should consider the limitations of the approaches described and the suggested alterations to facilitate effective research.

REFERENCES CITED

- A. Sharma, A. Jain, P. Gupta, V. Chowdary (2021) Machine Learning Applications for Precision Agriculture: A Comprehensive Review. *IEEE Access* 9:4843–4873.
<https://doi.org/10.1109/ACCESS.2020.3048415>
- Ahmad-Hamdani MS, Owen MJ, Yu Q, Powles SB (2012) ACCase-Inhibiting Herbicide-Resistant *Avena* spp. Populations from the Western Australian Grain Belt. *Weed Technol* 26:130–136
- Ahmad-Hamdani MS, Yu Q, Han H, et al (2013) Herbicide Resistance Endowed by Enhanced Rates of Herbicide Metabolism in Wild Oat (*Avena* spp.). *Weed Sci* 61:55–62.
<https://doi.org/10.1614/WS-D-12-00078.1>
- Albahri G, Alyamani AA, Badran A, et al (2023) Enhancing Essential Grains Yield for Sustainable Food Security and Bio-Safe Agriculture through Latest Innovative Approaches. *Agronomy* 13:. <https://doi.org/10.3390/agronomy13071709>
- Allwood JW, Williams A, Uthe H, et al (2021) Unravelling Plant Responses to Stress—The Importance of Targeted and Untargeted Metabolomics. *Metabolites* 11:558.
<https://doi.org/10.3390/metabo11080558>
- Amigo JM, Grassi S (2019) Configuration of hyperspectral and multispectral imaging systems. In: *Data Handling in Science and Technology*. Elsevier, pp 17–34
- Anagnostopoulos C, Stasinopoulou P, Kanatas P, et al (2020) Differences in metabolism of three *Conyza* species to herbicides glyphosate and triclopyr revealed by LC-MS/MS. *Chil J Agric Res* 80:100–107. <https://doi.org/10.4067/S0718-58392020000100100>
- Ang KL-M, Seng JKP (2021) Big Data and Machine Learning With Hyperspectral Information in Agriculture. *IEEE Access* 9:36699–36718.
<https://doi.org/10.1109/ACCESS.2021.3051196>
- Bailey DW (1971) RECOMBINANT-INBRED STRAINS AN AID TO FINDING IDENTITY, LINKAGE, AND FUNCTION OF HISTOCOMPATIBILITY AND OTHER GENES. *Transplantation* 11:325
- Barnes RJ, Dhanoa MS, Lister SJ (1989) Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Appl Spectrosc* 43:772–777.
<https://doi.org/10.1366/0003702894202201>
- Beckie HJ, Francis A, Hall LM (2012a) The Biology of Canadian Weeds. 27. *Avena fatua* L. (updated). *Can J Plant Sci* 92:1329–1357. <https://doi.org/10.4141/cjps2012-005>
- Beckie HJ, Heap IM, Smeda RJ, Hall LM (2000) Screening for Herbicide Resistance in Weeds. *Weed Technol* 14:428–445. [https://doi.org/10.1614/0890-037X\(2000\)014%255B0428:SFHRIW%255D2.0.CO;2](https://doi.org/10.1614/0890-037X(2000)014%255B0428:SFHRIW%255D2.0.CO;2)

- Beckie HJ, Warwick SI, Sauder CA (2012b) Basis for Herbicide Resistance in Canadian Populations of WildOat (*Avena fatua*). *Weed Sci* 60:10–18. <https://doi.org/10.1614/WS-D-11-00110.1>
- Beffa R, Figge A, Lorentz L, et al (2012) Weed resistance diagnostic technologies to detect herbicide resistance in cerealgrowing areas. A review. *Julius-Kühn-Arch* 434 1 75-80 75–80. <https://doi.org/10.5073/JKA.2012.434.008>
- Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Methodol* 57:289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Blackburn GA (2007) Hyperspectral remote sensing of plant pigments. *J Exp Bot* 58:855–867. <https://doi.org/10.1093/jxb/erl123>
- Boggs T, March D, kormang, et al (2022) spectralpython/spectral: Spectral Python (SPy) 0.23.1
- Boonchaisri S, Rochfort S, Stevenson T, Dias DA (2021) LC–MS untargeted metabolomics assesses the delayed response of glufosinate treatment of transgenic glufosinate resistant (GR) buffalo grasses (*Stenotaphrum secundatum* L.). *Metabolomics* 17:28. <https://doi.org/10.1007/s11306-021-01776-5>
- Boyd NS, Moretti ML, Sosnoskie LM, et al (2022) Occurrence and management of herbicide resistance in annual vegetable production systems in North America. *Weed Sci* 70:515–528. <https://doi.org/10.1017/wsc.2022.43>
- Breiman L (2001) Random Forests. *Mach Learn* 45:5–32. <https://doi.org/10.1023/A:1010933404324>
- Chambers MC, Maclean B, Burke R, et al (2012) A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* 30:918–920. <https://doi.org/10.1038/nbt.2377>
- Chong J, Xia J (2020) Using MetaboAnalyst 4.0 for Metabolomics Data Analysis, Interpretation, and Integration with Other Omics Data. In: Li S (ed) *Computational Methods and Data Analysis for Metabolomics*. Springer US, New York, NY, pp 337–360
- Coleman GRY, Bender A, Walsh MJ, Neve P (2023) Image-based weed recognition and control: Can it select for crop mimicry? *Weed Res* 63:77–82. <https://doi.org/10.1111/wre.12566>
- Dayan FE, Zaccaro ML de M (2012) Chlorophyll fluorescence as a marker for herbicide mechanisms of action. *Pestic Biochem Physiol* 102:189–197. <https://doi.org/10.1016/j.pestbp.2012.01.005>
- De Vos RC, Moco S, Lommen A, et al (2007) Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat Protoc* 2:778–791. <https://doi.org/10.1038/nprot.2007.95>

- Debeer D, Hothorn T, Strobl C (2025) permimp: Conditional Permutation Importance
- Deborah H, Richard N, Hardeberg JY (2015) A Comprehensive Evaluation of Spectral Distance Functions and Metrics for Hyperspectral Image Processing. *IEEE J Sel Top Appl Earth Obs Remote Sens* 8:3224–3234. <https://doi.org/10.1109/JSTARS.2015.2403257>
- Devine MD (1997) Physiological, biochemical and molecular aspects of herbicide resistance based on altered target sites. *Herbic Act Toxicol Biochem Mol Biol*
- Deza E, Deza MM (2009) *Encyclopedia of Distances*. Springer Berlin Heidelberg, Berlin, Heidelberg
- Domingos P (2012) A few useful things to know about machine learning. *Commun ACM* 55:78–87. <https://doi.org/10.1145/2347736.2347755>
- Du X, Dobrowolski A, Brochhausen M, et al (2025) Nextflow4MS-DIAL: A Reproducible Nextflow-Based Workflow for Liquid Chromatography–Mass Spectrometry Metabolomics Data Processing. *J Am Soc Mass Spectrom* 36:433–438. <https://doi.org/10.1021/jasms.4c00364>
- Dyer WE (2018) Stress-induced evolution of herbicide resistance and related pleiotropic effects. *Pest Manag Sci* 74:1759–1768. <https://doi.org/10.1002/ps.5043>
- Evans RM, Thill DC, Tapia L, et al (1991) Wild Oat (*Avena fatua*) and Spring Barley (*Hordeum vulgare*) Density Affect Spring Barley Grain Yield. *Weed Technol* 5:33–39. <https://doi.org/10.1017/S0890037X00033212>
- Falcon WP, Naylor RL, Shankar ND (2022) Rethinking Global Food Demand for 2050. *Popul Dev Rev* 48:921–957. <https://doi.org/10.1111/padr.12508>
- Fang Z-Z, Gonzalez FJ (2014) LC–MS-based metabolomics: an update. *Arch Toxicol* 88:1491–1502. <https://doi.org/10.1007/s00204-014-1234-6>
- Farag F, Huggins TD, Edwards JD, et al (2024) Manifold and spatiotemporal learning on multispectral unoccupied aerial system imagery for phenotype prediction. *Plant Phenome J* 7:e70006. <https://doi.org/10.1002/ppj2.70006>
- Frick H, Chow F, Kuhn M, et al (2024) rsample: General Resampling Infrastructure
- Gaines TA, Duke SO, Morran S, et al (2020) Mechanisms of evolved herbicide resistance. *J Biol Chem* 295:10307–10330. <https://doi.org/10.1074/jbc.REV120.013572>
- Ghatrehsamani S, Jha G, Dutta W, et al (2023) Artificial Intelligence Tools and Techniques to Combat Herbicide Resistant Weeds—A Review. *Sustainability* 15:1843. <https://doi.org/10.3390/su15031843>

- Giacomini DA, Gaines T, Beffa R, Tranel PJ (2018) Optimizing RNA-seq studies to investigate herbicide resistance. *Pest Manag Sci* 74:2260–2264. <https://doi.org/10.1002/ps.4822>
- Grossmann K (2005) What it takes to get a herbicide's mode of action. *Physionomics, a classical approach in a new complexion*. *Pest Manag Sci* 61:423–431. <https://doi.org/10.1002/ps.1016>
- Grzybowski M, Wijewardane NK, Atefi A, et al (2021) Hyperspectral reflectance-based phenotyping for quantitative genetics in crops: Progress and challenges. *Plant Commun* 2:. <https://doi.org/10.1016/j.xplc.2021.100209>
- Günther UL (2015) Metabolomics Biomarkers for Breast Cancer. *Pathobiology* 82:153–165. <https://doi.org/10.1159/000430844>
- Hall RD (2006) Plant metabolomics: from holistic hope, to hype, to hot topic. *New Phytol* 169:453–468. <https://doi.org/10.1111/j.1469-8137.2005.01632.x>
- Han H, Yu Q, Owen MJ, et al (2016) Widespread occurrence of both metabolic and target-site herbicide resistance mechanisms in *Lolium rigidum* populations. *Pest Manag Sci* 72:255–263. <https://doi.org/10.1002/ps.3995>
- Han Y, Sun Y, Ma H, et al (2023) Target-site and non-target-site based resistance to clodinafop-propargyl in wild oats (*Avena fatua* L.). *Pestic Biochem Physiol* 197:105650. <https://doi.org/10.1016/j.pestbp.2023.105650>
- He B, Hu Y, Wang W, et al (2022) The Progress towards Novel Herbicide Modes of Action and Targeted Herbicide Development. *Agronomy* 12:. <https://doi.org/10.3390/agronomy12112792>
- Heap I (2026) weedsdatabase.org/Home.aspx. In: *Int. Herbic.-Resist. Weed Database*. <http://www.weedsdatabase.org/>. Accessed 31 Jan 2026
- Heap I (2014) Global perspective of herbicide-resistant weeds. *Pest Manag Sci* 70:1306–1315. <https://doi.org/10.1002/ps.3696>
- Hijmans RJ (2024) terra: Spatial Data Analysis
- Holm LG, Plucknett DL, Pancho JV, Herberger JP (1977) The world's worst weeds. Distribution and biology. University Press of Hawaii., Honolulu, Hawaii
- Hothorn T, Hornik K, Zeileis A (2006) Unbiased Recursive Partitioning: A Conditional Inference Framework. *J Comput Graph Stat* 15:651–674. <https://doi.org/10.1198/106186006X133933>

- Huang Y, Lee MA, Nandula VK, Reddy KN (2018) Hyperspectral Imaging for Differentiating Glyphosate-Resistant and Glyphosate-Susceptible Italian Ryegrass. *Am J Plant Sci* 09:1467–1477. <https://doi.org/10.4236/ajps.2018.97107>
- Huang Y, Zhao X, Pan Z, et al (2022) Hyperspectral plant sensing for differentiating glyphosate-resistant and glyphosate-susceptible johnsongrass through machine learning algorithms. *Pest Manag Sci* 78:2370–2377. <https://doi.org/10.1002/ps.6864>
- Hurley S, Horney M, Drake A (2019) Using hyperspectral imagery to detect water stress in vineyards. In: Thomasson JA, McKee M, Moorhead RJ (eds) *Autonomous Air and Ground Sensing Systems for Agricultural Optimization and Phenotyping IV*. SPIE, Baltimore, United States, p 6
- Jin W, Xie K, Tang W, et al (2024) Comparative metabolomics and transcriptomics provide new insights into florpyrauxifen-benzyl resistance in *Echinochloa glabrescens*. *Front Plant Sci* 15:. <https://doi.org/10.3389/fpls.2024.1392460>
- Johnson RR, Cranston HJ, Chaverra ME, Dyer WE (1995) Characterization of cDNA clones for differentially expressed genes in embryos of dormant and nondormant *Avena fatua* L. caryopses. *Plant Mol Biol* 28:113–122. <https://doi.org/10.1007/BF00042043>
- Jones EAL, Austin R, Dunne JC, et al (2023) Utilization of image-based spectral reflectance to detect herbicide resistance in glufosinate-resistant and glufosinate-susceptible plants: a proof of concept. *Weed Sci* 71:11–21. <https://doi.org/10.1017/wsc.2022.68>
- Jugulam M, Shyam C (2019) Non-Target-Site Resistance to Herbicides: Recent Developments. *Plants* 8:417. <https://doi.org/10.3390/plants8100417>
- Keith BK, Burns EE, Bothner B, et al (2017) Intensive herbicide use has selected for constitutively elevated levels of stress-responsive mRNAs and proteins in multiple herbicide-resistant *Avena fatua* L. *Pest Manag Sci* 73:2267–2281. <https://doi.org/10.1002/ps.4605>
- Keith BK, Lehnhoff EA, Burns EE, et al (2015) Characterisation of *Avena fatua* populations with resistance to multiple herbicides. *Weed Res* 55:621–630. <https://doi.org/10.1111/wre.12172>
- Kern AJ, Myers TM, Jasieniuk M, et al (2002) Two Recessive Gene Inheritance for Triallate Resistance in *Avena fatua* L. *J Hered* 93:48–50. <https://doi.org/10.1093/jhered/93.1.48>
- Kothari S, Schweiger AK (2022) Plant spectra as integrative measures of plant phenotypes. *J Ecol* 110:2536–2554. <https://doi.org/10.1111/1365-2745.13972>
- Kuhn M, Vaughan D (2024) *parsnip: A Common API to Modeling and Analysis Functions*

- Kuhn M, Wickham H (2020) Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles.
- Kuhn M, Wickham H, Hvitfeldt E (2025) recipes: Preprocessing and Feature Engineering Steps for Modeling
- Kumaraswamy KG, Kushalappa AC, Choo TM, et al (2011) Mass Spectrometry Based Metabolomics to Identify Potential Biomarkers for Resistance in Barley against Fusarium Head Blight (*Fusarium graminearum*). *J Chem Ecol* 37:846–856. <https://doi.org/10.1007/s10886-011-9989-1>
- Kutasy B, Farkas Z, Kolics B, et al (2021) Detection of Target-Site Herbicide Resistance in the Common Ragweed: Nucleotide Polymorphism Genotyping by Targeted Amplicon Sequencing. *Diversity* 13:118. <https://doi.org/10.3390/d13030118>
- Lachowiec J, Feldman MJ, Matias FI, et al (2024) Adoption of unoccupied aerial systems in agricultural research. *Plant Phenome J* 7:e20098. <https://doi.org/10.1002/ppj2.20098>
- Lee K-S, Cohen WB, Kennedy RE, et al (2004) Hyperspectral versus multispectral data for estimating leaf area index in four different biomes. *Remote Sens Environ* 91:508–520. <https://doi.org/10.1016/j.rse.2004.04.010>
- Lehnhoff EA, Keith BK, Dyer WE, et al (2013) Multiple Herbicide Resistance in Wild Oat and Impacts on Physiology, Germinability, and Seed Production. *Agron J* 105:854–862. <https://doi.org/10.2134/agronj2012.0428>
- Li Q, Wang H, Yu J, et al (2024) Metabolism-Based Herbicide Resistance to Mesosulfuron-methyl and Identification of Candidate Genes in *Bromus japonicus*. *Plants* 13:1751. <https://doi.org/10.3390/plants13131751>
- Li S, Park Y, Duraisingham S, et al (2013) Predicting Network Activity from High Throughput Metabolomics. *PLoS Comput Biol* 9:e1003123. <https://doi.org/10.1371/journal.pcbi.1003123>
- Li X, Zhu B, Li S, et al (2025) A Comprehensive Review of Crop Chlorophyll Mapping Using Remote Sensing Approaches: Achievements, Limitations, and Future Perspectives. *Sensors* 25:.. <https://doi.org/10.3390/s25082345>
- Lowe C, Onkokesung N, Goldberg A, et al (2024) RNA and protein biomarkers for detecting enhanced metabolic resistance to herbicides mesosulfuron-methyl and fenoxaprop-ethyl in black-grass (*Alopecurus myosuroides*). *Pest Manag Sci* 80:2539–2551. <https://doi.org/10.1002/ps.7960>
- Lu B, Dao PD, Liu J, et al (2020) Recent Advances of Hyperspectral Imaging Technology and Applications in Agriculture. *Remote Sens* 12:2659

- Luque de Castro MD, Priego-Capote F (2018) The analytical process to search for metabolomics biomarkers. *J Pharm Biomed Anal* 147:341–349.
<https://doi.org/10.1016/j.jpba.2017.06.073>
- Ma A, Qi X (2021) Mining plant metabolomes: Methods, applications, and perspectives. *Plant Commun* 2:. <https://doi.org/10.1016/j.xplc.2021.100238>
- Mangin AR, Hall LM, Beckie HJ (2017) Triallate-resistant wild oat (*Avena fatua* L.): unexpected resistance to pyroxasulfone and sulfentrazone. *Can J Plant Sci* 97:20–25.
<https://doi.org/10.1139/cjps-2016-0029>
- Mariotto I, Thenkabail PS, Huete A, et al (2013) Hyperspectral *versus* multispectral crop-productivity modeling and type discrimination for the HypSIIRI mission. *Remote Sens Environ* 139:291–305. <https://doi.org/10.1016/j.rse.2013.08.002>
- Maroli AS, Nandula VK, Dayan FE, et al (2015) Metabolic Profiling and Enzyme Analyses Indicate a Potential Role of Antioxidant Systems in Complementing Glyphosate Resistance in an *Amaranthus palmeri* Biotype. *J Agric Food Chem* 63:9199–9209.
<https://doi.org/10.1021/acs.jafc.5b04223>
- Mazdeyasna S, Arefin MS, Fales A, et al (2025) Evaluating Normalization Methods for Robust Spectral Performance Assessments of Hyperspectral Imaging Cameras. *Biosensors* 15:20.
<https://doi.org/10.3390/bios15010020>
- McElroy JS (2014) Vavilovian Mimicry: Nikolai Vavilov and His Little-Known Impact on Weed Science. *Weed Sci* 62:207–216. <https://doi.org/10.1614/WS-D-13-00122.1>
- McInnes L, Healy J, Melville J (2020) UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction
- Mensah B, Betitame K, Zhang Y, et al (2025) Hyperspectral imaging and machine learning for herbicide-resistant kochia identification in sugarbeet. *J Agric Food Res* 23:102284.
<https://doi.org/10.1016/j.jafr.2025.102284>
- Nesar SB, Nugent PW, Zidack NK, Whitaker BM (2025) Unmanned Aerial Vehicle-Based Hyperspectral Imaging for Potato Virus Y Detection: Machine Learning Insights. *Remote Sens* 17:1735. <https://doi.org/10.3390/rs17101735>
- Niu Z, Li X, Zhao T, et al (2025) Advanced Plant Phenotyping Technologies for Enhanced Detection and Mode of Action Analysis of Herbicide Damage Management. *Remote Sens* 17:1166. <https://doi.org/10.3390/rs17071166>
- Nugent PW, Shaw JA, Jha P, et al (2018) Discrimination of herbicide-resistant kochia with hyperspectral imaging. *J Appl Remote Sens* 12:016037–016037
- Oborne M (2025) Mission Planner

- Ofosu R, Agyemang ED, Márton A, et al (2023) Herbicide Resistance: Managing Weeds in a Changing World. *Agronomy* 13:1595. <https://doi.org/10.3390/agronomy13061595>
- Oliver SG, Winson MK, Kell DB, Baganz F (1998) Systematic functional analysis of the yeast genome. *Trends Biotechnol* 16:373–378. [https://doi.org/10.1016/S0167-7799\(98\)01214-1](https://doi.org/10.1016/S0167-7799(98)01214-1)
- Owen MJ, Powles SB (2009) Distribution and frequency of herbicide-resistant wild oat (*Avena* spp.) across the Western Australian grain belt. *Crop Pasture Sci* 60:25–31. <https://doi.org/10.1071/CP08178>
- Pang Z, Lu Y, Zhou G, et al (2024a) MetaboAnalyst 6.0: towards a unified platform for metabolomics data processing, analysis and interpretation. *Nucleic Acids Res* 52:W398–W406. <https://doi.org/10.1093/nar/gkae253>
- Pang Z, Xu L, Viau C, et al (2024b) MetaboAnalystR 4.0: a unified LC-MS workflow for global metabolomics. *Nat Commun* 15:3675. <https://doi.org/10.1038/s41467-024-48009-6>
- Parry MAJ, Hawkesford MJ (2010) Food security: increasing yield and improving resource use efficiency. *Proc Nutr Soc* 69:592–600. <https://doi.org/10.1017/S0029665110003836>
- Paszke A, Gross S, Massa F, et al (2019) PyTorch: an imperative style, high-performance deep learning library. In: *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Curran Associates Inc., Red Hook, NY, USA, pp 8026–8037
- Patterson EL, Fleming MB, Kessler KC, et al (2017) A KASP Genotyping Method to Identify Northern Watermilfoil, Eurasian Watermilfoil, and Their Interspecific Hybrids. *Front Plant Sci* 8:. <https://doi.org/10.3389/fpls.2017.00752>
- Pedregosa F, others (2011) Scikit-learn: Machine Learning in Python. *J Mach Learn Res* 12:2825–2830
- Peterson MA, Collavo A, Ovejero R, et al (2018) The challenge of herbicide resistance around the world: a current summary. *Pest Manag Sci* 74:2246–2259
- Petibon F, Wiesenberger GLB (2022) Characterization of complex photosynthetic pigment profiles in European deciduous tree leaves by sequential extraction and reversed-phase high-performance liquid chromatography. *Front Plant Sci* 13:. <https://doi.org/10.3389/fpls.2022.957606>
- Pierce FJ, Nowak P (1999) Aspects of Precision Agriculture. In: *Advances in Agronomy*. Elsevier, pp 1–85
- Polder G, Blasco J, Cen H (2024) Editorial: Deep learning approaches applied to spectral images for plant phenotyping. *Front Plant Sci* 15:1425310. <https://doi.org/10.3389/fpls.2024.1425310>

- Pollard DA (2012) Design and Construction of Recombinant Inbred Lines. In: Rifkin SA (ed) Quantitative Trait Loci (QTL): Methods and Protocols. Humana Press, Totowa, NJ, pp 31–39
- Pourdarbani R, Sabzi S, Rohban MH, et al (2021) One-Dimensional Convolutional Neural Networks for Hyperspectral Analysis of Nitrogen in Plant Leaves. *Appl Sci* 11:. <https://doi.org/10.3390/app112411853>
- Powles SB, Yu Q (2010) Evolution in Action: Plants Resistant to Herbicides. *Annu Rev Plant Biol* 61:317–347. <https://doi.org/10.1146/annurev-arplant-042809-112119>
- QGIS Development Team (2026) QGIS Geographic Information System. QGIS Association
- Qi H, Wu Z, Zhang L, et al (2021) Monitoring of peanut leaves chlorophyll content based on drone-based multispectral image feature extraction. *Comput Electron Agric* 187:106292. <https://doi.org/10.1016/j.compag.2021.106292>
- R Core Team (2025) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria
- Raghav M, Dubey A, Singh J (2026) Hyperspectral Imaging for Detection and Classification of Plant Primary and Secondary Metabolites: A Review. *Phytochem Anal* 37:4–27. <https://doi.org/10.1002/pca.70029>
- Sahoo RN, Ray SS, Manjunath KR (2015) Hyperspectral remote sensing of agriculture. *Curr Sci* 108:848–859
- Sandhu PK, Kumar R, Nandula V, Tharayil N (2025) Integrated multi-omics analysis reveals divergent molecular responses in Palmer amaranth (*Amaranthus palmeri*) biotypes susceptible and resistant to glyphosate. 2025.07.20.665775
- Sandhu PK, Leonard E, Nandula V, Tharayil N (2023) Global Metabolome of Palmer Amaranth (*Amaranthus palmeri*) Populations Highlights the Specificity and Inducibility of Phytochemical Responses to Abiotic Stress. *J Agric Food Chem* 71:3518–3530. <https://doi.org/10.1021/acs.jafc.2c07162>
- Sarić R, Nguyen VD, Burge T, et al (2022) Applications of hyperspectral imaging in plant phenotyping. *Trends Plant Sci* 27:301–315. <https://doi.org/10.1016/j.tplants.2021.12.003>
- Savitzky Abraham, Golay MJE (1964) Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal Chem* 36:1627–1639. <https://doi.org/10.1021/ac60214a047>
- Scherrer B, Sheppard J, Jha P, Shaw JA (2019) Hyperspectral imaging and neural networks to classify herbicide-resistant weeds. *J Appl Remote Sens* 13:044516–044516

- Scherrer BJ (2019) Weed and crop discrimination with hyperspectral imaging and machine learning. Montana State University - Bozeman, College of Engineering
- Schliep K, Hechenbichler K (2016) kkn: Weighted k-Nearest Neighbors
- Seefeldt SS, Jensen JE, Fuerst EP (1995) Log-Logistic Analysis of Herbicide Dose-Response Relationships. *Weed Technol* 9:218–227. <https://doi.org/10.1017/S0890037X00023253>
- Sen MK, Bhattacharya S, Bharati R, et al (2023) Comprehensive insights into herbicide resistance mechanisms in weeds: a synergistic integration of transcriptomic and metabolomic analyses. *Front Plant Sci* 14:. <https://doi.org/10.3389/fpls.2023.1280118>
- Shafi U, Mumtaz R, García-Nieto J, et al (2019) Precision Agriculture Techniques and Practices: From Considerations to Applications. *Sensors* 19:. <https://doi.org/10.3390/s19173796>
- Shen X, Wang R, Xiong X, et al (2019) Metabolic reaction network-based recursive metabolite annotation for untargeted metabolomics. *Nat Commun* 10:1516. <https://doi.org/10.1038/s41467-019-09550-x>
- Shu M, Harfouche AL, Trtílek M, et al (2025) Leveraging hyperspectral phenotyping for accurate, non-destructive prediction of metabolite profiles in poplar under drought stress. *Environ Exp Bot* 237:106218. <https://doi.org/10.1016/j.envexpbot.2025.106218>
- Shu M, Wang Z, Guo W, et al (2024) Effects of Variety and Growth Stage on UAV Multispectral Estimation of Plant Nitrogen Content of Winter Wheat. *Agriculture* 14:. <https://doi.org/10.3390/agriculture14101775>
- signal developers (2023) signal: Signal processing
- Simonovska B, Vovk I, Glavnik V, Černelič K (2013) Effects of extraction and high-performance liquid chromatographic conditions on the determination of lutein in spinach. *J Chromatogr A* 1276:95–101. <https://doi.org/10.1016/j.chroma.2012.12.032>
- Singh K, Natarajan M, Ravichandran P, et al (2026) Hyperspectral Imaging and Machine Learning to Classify Herbicide-Resistant and -Susceptible Weeds in Field Crops. *Can J Plant Sci*. <https://doi.org/10.1139/cjps-2025-0244>
- Singhal G, Bansod B, Mathew L, et al (2019) Chlorophyll estimation using multi-spectral unmanned aerial system based on machine learning techniques. *Remote Sens Appl Soc Environ* 15:100235. <https://doi.org/10.1016/j.rsase.2019.100235>
- Somody CN, Nalewaja JD, Miller SD (1985) Self-Burial of Wild Oat Florets. *Agron J* 77:359–362. <https://doi.org/10.2134/agronj1985.00021962007700030002x>

- Squires CC, Coleman GR, Broster JC, et al (2021) Increasing the value and efficiency of herbicide resistance surveys. *Pest Manag Sci* 77:3881–3889.
<https://doi.org/10.1002/ps.6333>
- Steward KF, Refai M, Dyer WE, et al (2023) Acute stress reduces population-level metabolic and proteomic variation. *BMC Bioinformatics* 24:87. <https://doi.org/10.1186/s12859-023-05185-4>
- Sun J, Yang J, Shi S, et al (2017) Estimating Rice Leaf Nitrogen Concentration: Influence of Regression Algorithms Based on Passive and Active Leaf Reflectance. *Remote Sens* 9:.
<https://doi.org/10.3390/rs9090951>
- Tafoya-Razo JA, Mora-Munguía SA, Torres-García JR (2022) Diversity of Herbicide-Resistance Mechanisms of *Avena fatua* L. to Acetyl-CoA Carboxylase-Inhibiting Herbicides in the Bajío, Mexico. *Plants* 11:. <https://doi.org/10.3390/plants11131644>
- Tangamornsuksan W, Lohitnavy O, Sruamsiri R, et al (2019) Paraquat exposure and Parkinson's disease: A systematic review and meta-analysis. *Arch Environ Occup Health* 74:225–238.
<https://doi.org/10.1080/19338244.2018.1492894>
- Torres-García JR, Tafoya-Razo JA, Velázquez-Márquez S, Tiessen A (2018) Double herbicide-resistant biotypes of wild oat (*Avena fatua*) display characteristic metabolic fingerprints before and after applying ACCase- and ALS-inhibitors. *Acta Physiol Plant* 40:119.
<https://doi.org/10.1007/s11738-018-2691-y>
- Tottman DR, Makepeace RJ, Broad H (1979) An explanation of the decimal code for the growth stages of cereals, with illustrations. *Ann Appl Biol* 93:221–234.
<https://doi.org/10.1111/j.1744-7348.1979.tb06534.x>
- Tsoulias N, Zhao M, Paraforos DS, Argyropoulos D (2023) Hyper-and multi-spectral imaging technologies. In: *Encyclopedia of Digital Agricultural Technologies*. Springer, pp 629–640
- Tsugawa H, Cajka T, Kind T, et al (2015) MS-DIAL: Data Independent MS/MS Deconvolution for Comprehensive Metabolome Analysis. *Nat Methods* 12:523–526.
<https://doi.org/10.1038/nmeth.3393>
- Van Acker RC (2009) Weed biology serves practical weed management. *Weed Res* 49:1–5.
<https://doi.org/10.1111/j.1365-3180.2008.00656.x>
- Vergara-Díaz O, Vatter T, Kefauver SC, et al (2020) Assessing durum wheat ear and leaf metabolomes in the field through hyperspectral data. *Plant J* 102:615–630.
<https://doi.org/10.1111/tpj.14636>

- Vigneau N, Ecarnot M, Rabatel G, Roumet P (2011) Potential of field hyperspectral imaging as a non destructive method to assess leaf nitrogen content in Wheat. *Field Crops Res* 122:25–31. <https://doi.org/10.1016/j.fcr.2011.02.003>
- Vilà M, Beaury EM, Blumenthal DM, et al (2021) Understanding the combined impacts of weeds and climate change on crops. *Environ Res Lett* 16:034043. <https://doi.org/10.1088/1748-9326/abe14b>
- Virtanen P, Gommers R, Oliphant TE, et al (2020) SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat Methods* 17:261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- Wainer J, Cawley G (2018) Nested cross-validation when selecting classifiers is overzealous for most practical applications
- Walt S van der, Schönberger JL, Nunez-Iglesias J, et al (2014) scikit-image: Image processing in Python. *PeerJ* 2:e453. <https://doi.org/10.7717/peerj.453>
- Wang A, Zhang W, Wei X (2019) A review on weed detection using ground-based machine vision and image processing techniques. *Comput Electron Agric* 158:226–240. <https://doi.org/10.1016/j.compag.2019.02.005>
- Wang M, Carver JJ, Phelan VV, et al (2016) Sharing and community curation of mass spectrometry data with GNPS. *Nat Biotechnol* 34:828–837. <https://doi.org/10.1038/nbt.3597>
- Wang T, Liu Y, Wang M, et al (2021) Applications of UAS in Crop Biomass Monitoring: A Review. *Front Plant Sci* 12:. <https://doi.org/10.3389/fpls.2021.616689>
- Wickham H (2016) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York
- Xiao JF, Zhou B, Ransom HW (2012) Metabolite identification and quantitation in LC-MS/MS-based metabolomics. *TrAC Trends Anal Chem* 32:1–14. <https://doi.org/10.1016/j.trac.2011.08.009>
- Zadoks JC, Chang TT, Konzak CF (1974) A decimal code for the growth stages of cereals. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-3180.1974.tb01084.x>. Accessed 10 Sept 2025
- Zulet-Gonzalez A, Gorzolka K, Döll S, et al (2023) Unravelling the Phytotoxic Effects of Glyphosate on Sensitive and Resistant *Amaranthus palmeri* Populations by GC–MS and LC–MS Metabolic Profiling. *Plants* 12:1345. <https://doi.org/10.3390/plants12061345>

APPENDICES

APPENDIX

SUPPLEMENTAL FIGURES

Table A1: Number of samples and respective metabolite counts after filtering for metabolomic experiments.

	Number of samples				Number of features (raw)	Number of features (cleaned)
Experiment	Biotype: 06R or 93R	Biotype: S2	Resistant RILs	Susceptible RILs		
Experiment 1: 06R and S2	12	10	NA	NA	705	297
Experiment 2: 06R, S2, and RILs	12	9	27	27	778	354
Experiment 3: 93R, S2, and RILs	9	9	16	19	722	402

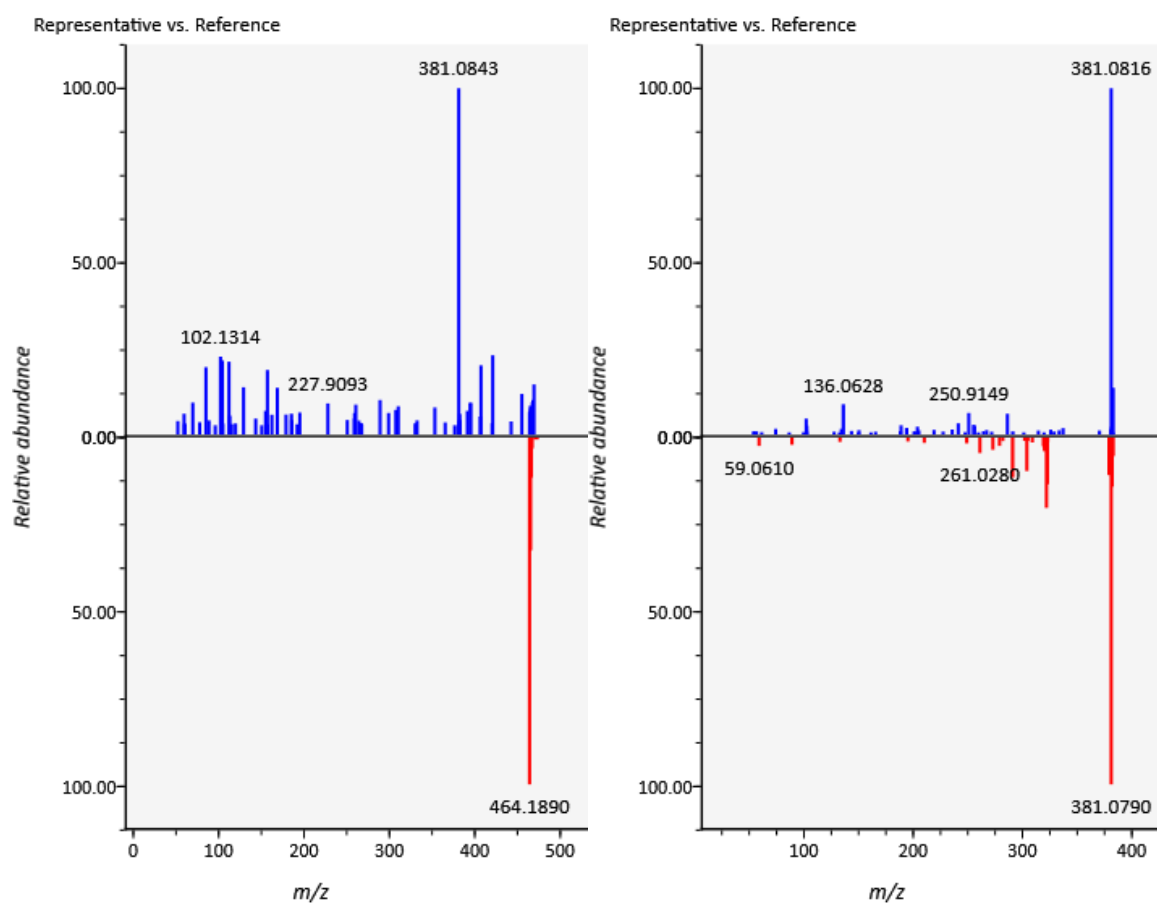


Figure A1: MS/MS spectra for conjugate features in the 06R biotype experiments.

Table A2: Model sensitivity, specificity, precision, and support for biotype prediction models.

Line	Model	Sensitivity	Specificity	Precision	Support
06R3	KNN	0.3133	0.8627	0.3133	55
06R3	RF	0.5067	0.9013	0.5133	55
06R3	PLS	0.2133	0.8427	0.2133	55
06R3	SVM	0.5667	0.9133	0.5600	55
06R4	KNN	0.5533	0.9107	0.5400	55
06R4	RF	0.7267	0.9453	0.7200	55
06R4	PLS	0.5000	0.9000	0.5200	55
06R4	SVM	0.7467	0.9493	0.7400	55
20R2	KNN	0.5000	0.9000	0.5000	50
20R2	RF	0.5800	0.9160	0.5800	50
20R2	PLS	0.4800	0.8960	0.4800	50
20R2	SVM	0.6800	0.9360	0.6800	50
93R	KNN	0.6400	0.9280	0.6400	50
93R	RF	0.7200	0.9440	0.7200	50
93R	PLS	0.6600	0.9320	0.6600	50
93R	SVM	0.6200	0.9240	0.6200	50
S1	KNN	0.6267	0.9253	0.6200	55
S1	RF	0.7600	0.9520	0.7400	55
S1	PLS	0.6733	0.9347	0.6600	55
S1	SVM	0.6400	0.9280	0.6400	55
S2	KNN	0.4800	0.8960	0.4800	45
S2	RF	0.5000	0.9000	0.5000	45
S2	PLS	0.6400	0.9280	0.6400	45
S2	SVM	0.5200	0.9040	0.5200	45

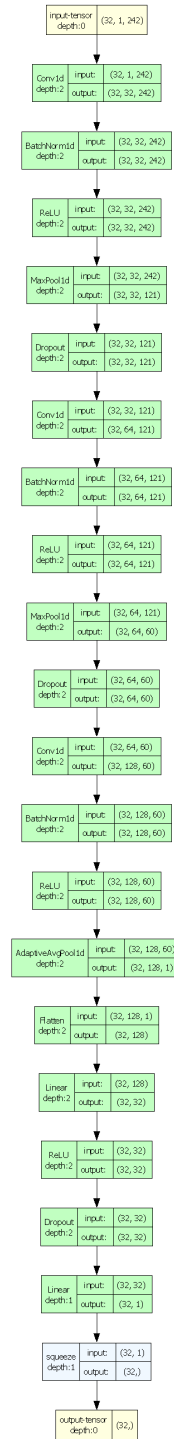


Figure A2: 1D convolutional neural net architecture used to develop models predicting herbicide resistance in recombinant inbred lines.