2018

Predicting Herbivore Induced Phytochemical Shifts in Helianthus using Spectral Reflectance

Jessica Jowais
University of Central Florida

PREDICTING HERBIVORE INDUCED PHYTOCHEMICAL SHIFTS IN HELIANTHUS USING SPECTRAL REFLECTANCE

By

JESSICA JEAN JOWAIS

A thesis submitted in partial fulfillment of the requirements

for the Honors in the Major Program

in the Department of Biology

in the College of Sciences

at the University of Central Florida

Orlando, Florida

Fall Term

2018

ABSTRACT

Induced defense responses in plants vary greatly among species, with many species exhibiting strong upregulation of secondary metabolites under attack by herbivores or pathogens. Secondary metabolite responses are most commonly analyzed using nuclear magnetic resonance or mass spectroscopy, though such approaches are costly and time-intensive. This study explores the use of hyperspectral reflectance as a more time- and cost-efficient method of detecting herbivore-induced secondary metabolite responses in plants. A diverse cross-section of wild sunflowers (genus *Helianthus*) were grown under controlled conditions and challenged with insect herbivory. Hyperspectral reflectance data was collected and analyzed using a principal component analysis in conjuncture with a support vector classification model to detect herbivore-induced versus control plants. The best model had a 93% accuracy rate at predicting whether a sample came from an induced or control plants when using data from all species tested. However, the changes in hyperspectral reflectance under herbivore induction varied greatly across species.

ACKNOWLEDGEMENTS

I would like to acknowledge Thesis Chair Dr. Bohlen and Committee Members Dr. Chase Mason and Jordan Dowell for being an integral part of this entire thesis process. From planning the experiment to drafting the thesis their contributions were vital to the success of this project.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

APPENDIX

INTRODUCTION

Agricultural pests, such as pathogens and herbivores, destroy one sixth of global crop production annually, estimated at approximately 650 million tons (Bebber *et al.*, 2014). While research into agricultural pest control is making strides in tackling this issue for major staple crops (i.e.maize, rice, and soy), there still are many less prominent crops for which we lack a formal understanding of biotic interactions and effects on yield. Furthermore, the application of current methods of analysis for quantifying plant defensive chemistry is time and cost intensive (Tankeu S *et al.*,2014). The lack of accessibility of cost-effective analytic tools limit the inclusivity of the scientific community to research diverse crops. Spectral reflectance is an alternative method of chemical analysis that has been used in plants for decades however, predominantly for single metabolite concentrations (Road and Dun, 1989). This method of secondary metabolite analysis is both extremely time efficient and cost effective. Using machine learning techniques to comparatively analyze wavelength data is an extremely novel application of spectral reflectance data that could allow us to data to predict entire phytochemical shifts in plant defensive chemistry or even entire secondary metabolite profiles.

This research addresses patterns of plant defensive chemistry in *Helianthus* and its wild relatives, while also serving as a proof of concept in developing a set of methods to reduce the limitations on phytochemical sampling capacity via this novel approach of spectral reflectance analysis–increasing access to rapid cost-effective methods of plant secondary metabolite analysis.

Secondary Metabolite Diversity

Secondary metabolites were for a long time considered to be waste products (Hartmann, 2007). Since then, many secondary metabolite functions for plants have been discovered. Some secondary metabolite plant functions, other than defense, include interacting with pollinators and mycorrhizal fungi, plant to plant signaling, and response to abiotic stressors (Dixon and Paiva, 1995). This wide range of functions can be altered by variation in individual secondary metabolite presence/absence and concentration among plant individuals (Moore *et al*., 2013). Variation can also be found among tissue types and ontogenetic stages within an individual (Moore *et al., 2013*).

There are many theories as to why secondary metabolite profiles and distribution in plants are so variable. All such theories attribute the effects of genes and genome duplication to producing mutations that are the source of new secondary metabolites. It is also widely accepted that most secondary metabolites originated from a small group of precursor compounds that were once, or are still, members of the primary metabolism (Speed *et al.,* 2015). For example, isoprene sub units make up terpenes which are common secondary metabolites and all 30,000 isoprenoid compounds originated from pyruvate and D-glyceraldehyde 3-phosphate, members of the primary metabolism (Lange *et al.,* 2000). However, any of these new compounds produced, as a result of mutation, have a low probability of being bioactive. According to the "Screening Hypothesis", the many non-biologically active compounds produced are then still retained because they increase the probability of producing new active compounds by acting as precursors (Speed *et al.,* 2015). Thus, plants must have high degrees of chemical diversity to be able to create new molecules capable of deterring herbivores and pathogens (Jones and Firn,

1991). The name is appropriate because plants would have to be capable and constantly "screen" new metabolites for biological activity. This would explain why there are so many secondary metabolites present in plants that have no known biological activity. Still, for this hypothesis to be valid it requires negligible costs to production of new metabolites as to not create a short-term disadvantage (Speed *et al*., 2015). This has yet to be proven.

Another explanation for the diversification for secondary metabolites is the coevolution between plants and their natural enemies (Speed *et al*., 2015). It can be argued that a simple pairwise arms race of coevolution being the only source of diversity does not explain the extreme variability of secondary metabolites, as strong and consistent directional selection should erode quantitative variation (Moore *et al.*, 2013). However, there are some important details that make coevolution still a viable option for the main source of diversity. Mainly, plants are coevolving with not only many enemies at once but also pollinators, mutualists and competitors. This changes the evolutionary process greatly. The concept of diffuse coevolution has been developed to refer to plants coevolving with many influences at once, in contrast to pairwise coevolution (Stamp, 2003). Diffuse coevolution allows for much more diversity than just simple pairwise evolution (Stamp, 2003). It is also thought that pairwise evolution in plants is likely rare as they have so many natural enemies and mutualistic relationships acting on them at any one time (Stamp, 2003). A positive correlation can be seen between diverse selection pressures and more diverse secondary metabolite composition, supporting this theory (Moore et al., 2013). Furthermore, many defensive compounds work synergistically together, and this adds to their capacity for diverse evolution because it selects for retaining more than one compound to complete a single function (Challis and Hopwood, 2003). From this we can conclude diffuse

coevolution is likely a significant source, or at least plays a significant role, in the extreme diversity of secondary metabolites. Overall, the main cause for the evolution of such immense diversity in secondary metabolites is still up for debate.

Defense in Secondary Metabolites and Induced Response

Secondary metabolite defenses can be either constitutive or induced. In a constitutive defense the defensive compounds would always be present in the plant, where as in an induced response the plant only upregulates, or even produces the defensive compound in response to being attacked. These responses can differ for each plant and natural enemy relationship. Inducible responses can have a significant impact on the resistance of a species to their natural enemies (Adler *et al.,* 1995). An induced response can be expressed locally, where the attack has occurred, or systemically across the whole plant. A systemic induction requires a signal from the infected site to cause a similar or even very different induced response to be used on un-eaten leaves (Choudhary *et al.*, 2007). This type of response focuses on preventing the spread of the attack to other parts of the plant (Choudhary *et al.*, 2007).

Helianthus and Resistance

Little is known about the specific secondary metabolite defense responses across the genus *Helianthus,* despite it being an important economic crop, making *Helianthus* a prime example of a significant and yet understudied crop. Improving *Helianthus* crop resistance would have significant economic impact as in 2008/09 sunflower production in the United States had a farm-gate value of $669 million (USDA, 2017). Species in the genus *Helianthus* in particular show abundant variation in secondary metabolite defense traits (Mason *et al*., 2015). This

variation provides an interesting opportunity to evaluate a genus-wide variation of secondary metabolite defense methods. Understanding the variation of secondary metabolite defense methods in *Helianthus* can be useful for plant breeders working to increase the resistance of *Helianthus* crop varieties, whether for the oilseed, confectionary, or cut flower industry (Wink, 1987).

Common Methods of Measuring Secondary Metabolites

Because secondary metabolite defense mechanisms vary greatly with their individual metabolite make up, synergistic combinations, and ratios of concentrations it is necessary to consider both qualitative and quantitative information when measuring diversity of secondary metabolite defense mechanisms. Nuclear magnetic resonance (NMR) and mass spectroscopy (MS) are the most common analytical tools in metabolomics research in general, as well as, in plant secondary metabolite analysis specifically (Emwas, 2015). While these methods are widely accepted, they each have their own limitations. Moreover, it is important to note that there is no single analytical platform that can completely quantify and identify all of the molecules in a sample (Emwas, 2015).

NMR has low sensitivity, meaning it needs a higher than sometimes optimum concentration of the compound to be indentified (Emwas, 2015). This sensitivity can be improved with a higher field strength magnet (Emwas, 2015). However, these magnets are already extremely costly a standard 600 MHz NMR costs roughly $800,000, but a more sensitive 900 MHz sells for about $5 million (Constans A, 2000). This makes NMR an extremely costly method of analysis, especially if high sensitivity is needed. Most sampling done with NMR is through extraction,

while the sample prepared can be recovered and stored for a long time, it still requires the destruction and altering of the original tissue sample. However, high-resolution magic-angle spinning (HRMAS) NMR, is capable of sampling tissues still, these tissue samples must be small and brought to the NMR machine (Emwas, 2015). This limits the ability to take larger organismal samples, multiple samples of an organism over time without destruction and samples in the field. Furthermore, running an NMR sample can take anywhere from 5 minutes to multiple hours a sample, depending on what the NMR is measuring (hydrogen, carbon, ect), the size of the sample and the components of the sample (Forseth and Schroeder, 2010). The amount of time necessary for one sample directly limits the number of samples that can be run for one study. Especially, considering many universities share NMR machines due to their immense cost, limiting accessibility to the technology (Constans A, 2000).

MS has a much higher sensitivity than NMR, with a detection limit at nanomolar resolution and is more cost effective (Emwas, 2015). However, MS has more debilitating limitations regarding reproducibility and sampling preparation. MS has moderate reproducibility as compared to NMR which is very high, this may in part be because of the particular conditions necessary in the sample preparation for MS (Emwas, 2015). A MS sample cannot be run on tissue and requires extraction, this extraction must be at optimal ionization conditions as well as run through particular columns for different polarities of metabolites (Emwas, 2015). Furthermore, each sample is destroyed after its use, making MS require higher amounts of samples, while also being limited in that it must be extracted. MS also can take significant time to run per sample ranging from an average of 2 minutes to 30 minutes per sample (Grebeand and Singh, 2011). While this is on average less time, the NMR and MS is more accessible, due to

affordability. Both technologies still poses a limitation when upwards of hundreds of samples are necessary for a study.

Spectral Reflectance

Different surfaces will reflect and absorb light in different ways. These differences in reflectance are present on large surfaces down to the atoms of a molecule. Because of this it is possible to identify secondary metabolites by analyzing the spectral reflectance signatures. Many foliar biochemical, physiological, structural and morphological properties have been successfully quantified using reflectance spectroscopy (Couture *et al.,* 2016). The estimation of biochemical concentrations from reflectance spectroscopy relies on variations in absorption as a consequence of vibrational excitation of molecular bonds, primarily C–H, N–H and O–H bonds at specific wavelengths in the visible (400–700 nm), near-infrared (NIR, 700–1100 nm) and shortwave infrared (SWIR, 1100–2400 nm) (Couture *et al.,* 2016). Most notably chlorophyll concentrations in vegetation have been measured using spectral reflectance for decades now (Richardson *et al.,* 2002).

The method of using spectral reflectance for analysis is extremely cost and time efficient as well as noninvasive. Spectral reflectance machines are a small fraction of the cost of NMR. Furthermore, they don't require as much maintenance as MS, which needs new columns and extraction materials. A reflectance sample is taken within seconds, by simply exposing the sample to light and capturing the seemingly instantaneous reflection of that light. This sample can be taken directly on tissue, even live tissue. It requires no extraction or manipulation of the sample. These advantages can have huge implications for the future of chemical analysis.

Spectral reflectance could take samples on living organisms at different time points without altering them and it has the potential to be taken out to the field to take mass samples in the environment. The almost instantaneous sample processing time exponentially increases the amount of samples that a single study can feasibly take on and finally its cost efficiency makes it accessible to universities and institutions of all financial status.

Until recently, spectral reflectance could only be used for metabolite analysis to identify a single specific compound or class of compound present in a sample (Road and Dun, 1989). This limitation is what makes spectral reflectance noncompetitive with NMR and MS. Novel methods of spectral reflectance data analysis have recently been employed that allow all reflectance signatures of all present compounds in a sample to be taken into account by using machine learning techniques to evaluate patterns within the spectral data.

Research Question

The goal of this study is to better understand the diversity of secondary metabolite induced systemic response in *Helianthus*, while also testing the efficacy of using spectral reflectance to analyze entire phytochemical shifts. To address these interests this experiment was designed to answer the following questions: 1. How effective is spectral reflectance at predicting induction in Helianthus across the genera and within species? 2. Is the response of induction different between species?

METHODS

Plant Growth

A broad cross-section of wild sunflower species were selected for inclusion in this study. Eight replicates each of 20 species were grown under identical high-resource conditions in a greenhouse (Figure 1). Seeds were purchased from USDA Germplasm Resources Information Network. Each species was scarified with a razor and placed in a petri dish with soaked filter paper for 24 hours in a dark cabinet. After 24 hours the seed coats were removed using tweezers and the filter paper was replaced and also soaked with water. After the seed coats were removed the seeds were left in the cabinet with no light until the radicle developed root hairs. The seeds were then moved to an LED lit room where they were kept watered until the first two green leaves appeared. After this germination the seedlings were moved to seedling trays with sand and watered from a tray underneath. The seedlings were grown in the trays until the first true leaves appeared. The seedlings were then transferred into tree pots with a 1:1 sand and soil mix in a green house and watered daily. All plants were grown to 8 to 10 leaf pairs.

Species Name	PI Number
H. agrophyllus	673306
H. petiolaris	673325
H. debilis	673213
H. praecox	435847
H. agrestis	673202
H. mollis	673318
H. occidentalis	673323
H. angustifolius	673210
H. atrorubens	649940
H. giganteus	664647
H. grosseserratus	613793
H. divericatus	664645
H. arizonensis	653549
H. exilis	649895
H. nuttallii	531053
H. lacianatus	653562
H. gracilentus.	649987
H. silphioides	664795
H. salicifolius	664768
H. maxamiliani	613794

Table 1: Shows the different species of *Helianthus* grown

Herbivory Treatment and Data Collection

Vanessa cardui (an Asteraceae generalist caterpillar species) eggs were ordered from Carolina Biological Supply. The eggs were grown from egg on a constant diet in a controlled incubator for 1 to 2 weeks until the majority of caterpillars reached approximately 1.5 inches. Once 8 to 10 leaf pairs were reached, groups of 4 replicates from each species were induced by *Vanessa carudi.* Five *Vanessa cardui* caterpillars were placed on the stem of each plant and were left to eat the leaves of each species for 24-48 hours. The plant was considered to be induced when at least 2 leaves had approximately 30% total leaf area eaten. Leaf samples were taken from the most recently fully expanded leaves (MRFELs) of the induced and remaining control plants immediately after that species' treatment group was considered induced. The MRFLELs selected were un-eaten to ensure that the induced response measured was systemic for all of the species. Samples were taking by cutting the base of the petiole with scissors, placed the leaves in plastic bags and taken directly to the laboratory. All samples were then immediately analyzed with an Ocean Optics Spectral Reflectance from UV to Infrared light (200-2500nm, by 0.5nm) to get reflectance data across wavelengths. Three measurements were taken on each leaf from the tip, center and base.

Data Analysis

Spectral reflectance data was analyzed using R version 3.5.1. A PCA was run to develop a model that can accurately classify a control versus induced state. PCA is an unsupervised learning technique, commonly used in exploratory forms of analysis, especially with large datasets that are often difficult to interpret. This technique is used to identify groupings and

variance within a dataset, highlighting the relationships between samples and variables (Jolliffeand Cadima, 2016). The utility of a PCA stems from the reduction of dimensionality and collinearity of the dataset, increasing the interpretability but at the same time minimizing information loss (Jolliffeand Cadima, 2016). SVC was run on the resulting principal components with a radial kernel. A Support Vector Classification (SVC) is a discriminative classifier defined by a separating hyperplane (Shihong *et al.*, 2003). This means, given the PCs the algorithm outputs an optimal hyperplane which categorizes the data. A hyperplane is a subspace whose dimension is one less than that of its ambient space. SVC uses the kernel trick where data is projected N+1 dimensions higher to find the hyperplane because this allows for the SVC to deal with overlaps in the data. Hypertuning was used to determine the best parameters for the SVC by iteratively running cost values from .01-100 and gamma values from .001-2. From SVC results Out-Of-Bag classification errors (OOB) for PCs 1:2, 1:3 and 1:4 for all species were determined.

RESULTS

Plots of the PCs from every individual species as well as all *Helianthus* considered together show clear differences in patterns of separation of induced versus control amongst the PCs. The majority of species, such as, *H. maximiliani*, *H. longifolius* and *H. mollis* have multiple examples of clear visual separation amongst the PCs (Figures 1, 2 and 3). Only two species, *H. silphioidies* and *H. salicifolius*, were poorly visually separable on the majority of their PC plots (Figures 4 and 5).

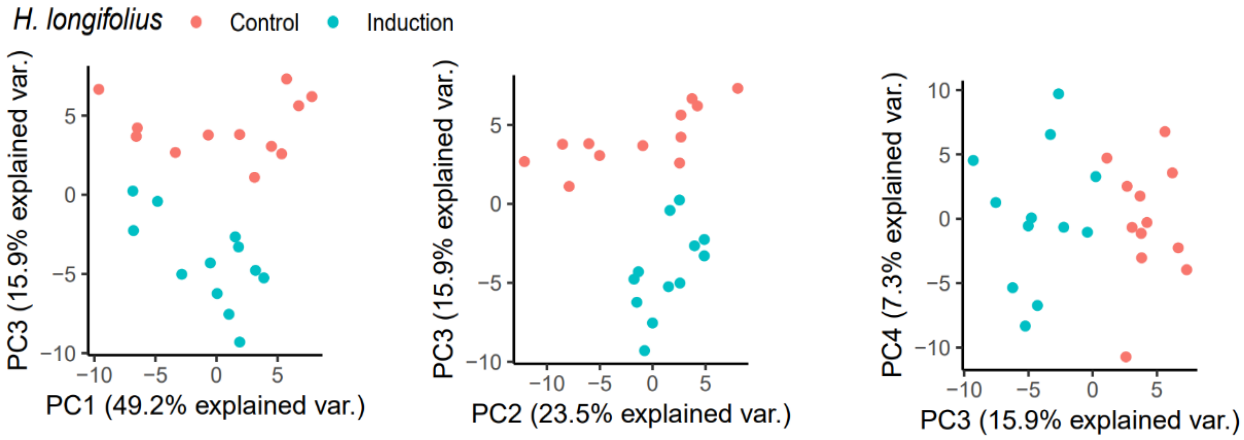

Figure 1: PC plots of *H. longifolius* using PCs 1 and 3 expressing 65.1% of total variance, PCs 2 and 3 expressing 39.4% of total variance and PCs 3 and 4 expressing 23.2% of total variance. These 3 of the 6 plots of *H. longifolius* best illustrate the clear separation of induced versus control with little to no overlap, as well as a clear liner separation.

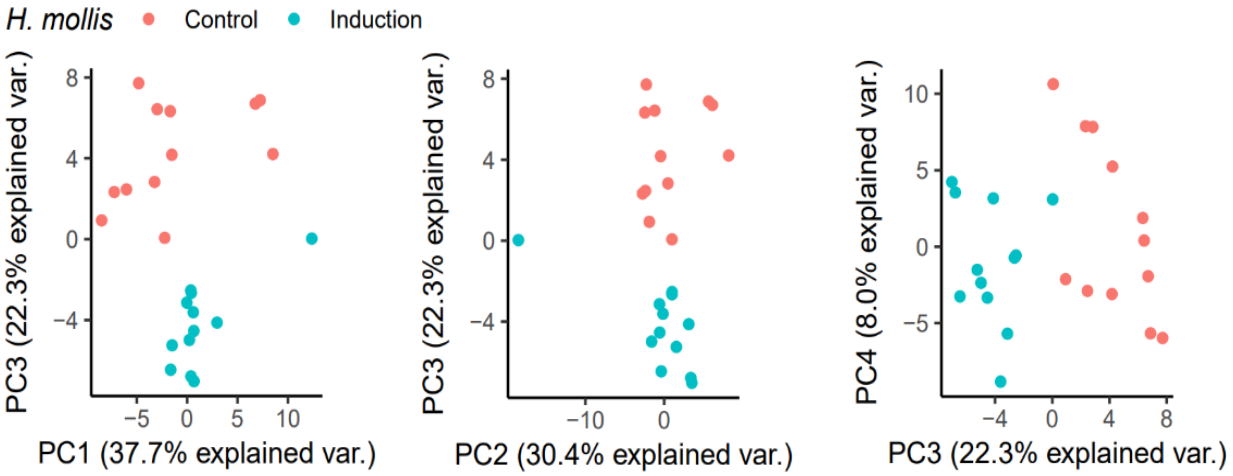

Figure 2: PC plots of *H. mollis* using PCs 1 and 3 expressing 60% of total variance, PCs 2 and 3 expressing 52.7% of total variance and PCs 3 and 4 expressing 30.3% of total variance. These 3 of the 6 plots of *H. mollis* best illustrate the clear separation of induced versus control with little to no overlap, as well as a clear liner separation.

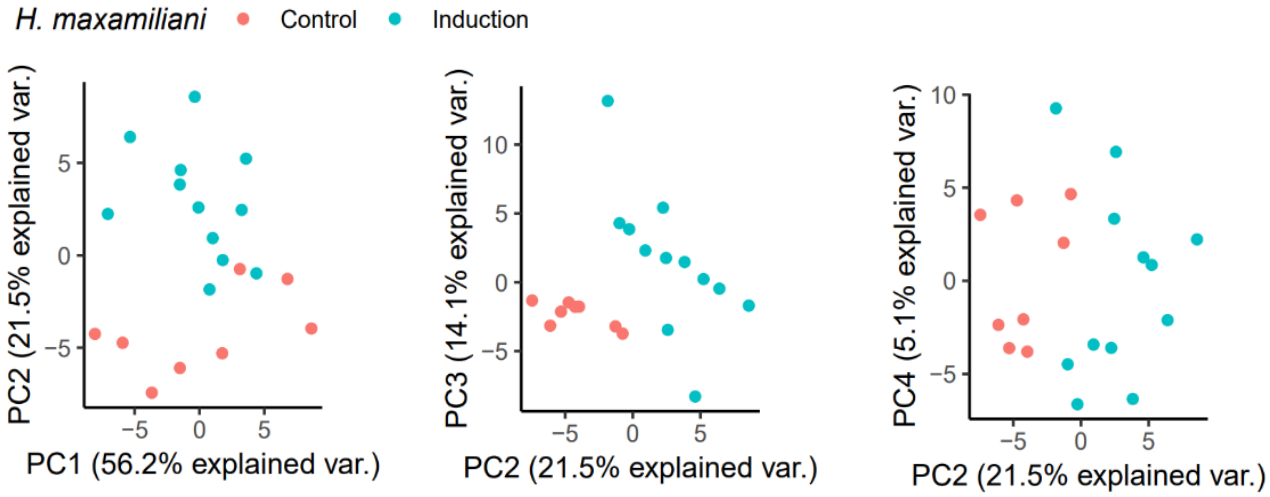

Figure 3: PC plots of *H. maximiliani* using PCs 1 and 2 expressing 77.7% of total variance, PCs 2 and 3 expressing 35.6% of total variance and PCs 2 and 4 expressing 26.6% of total variance. These 3 of the 6 plots of *H. maximiliani* best illustrate the clear separation of induced versus control with little to no overlap.

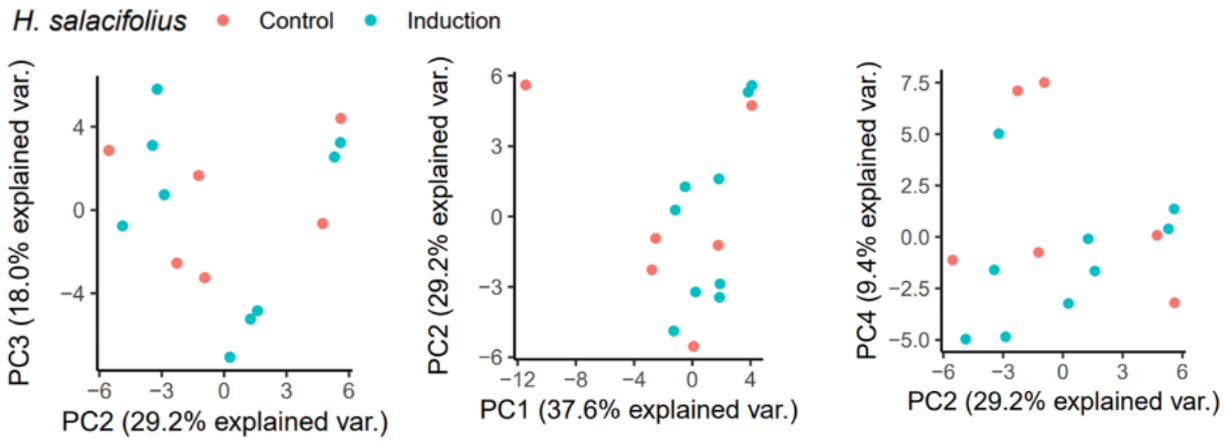

Figure 4: PC plots of *H. salicifolius* using PCs 2 and 3 expressing 47.2% of total variance, PCs 1 and 2 expressing 66.8% of total variance and PCs 2 and 4 expressing 38.6% of total variance. These 3 of the 6 plots of *H. salicifolius* best illustrate the lack of visual separation of control versus induced in the plots.

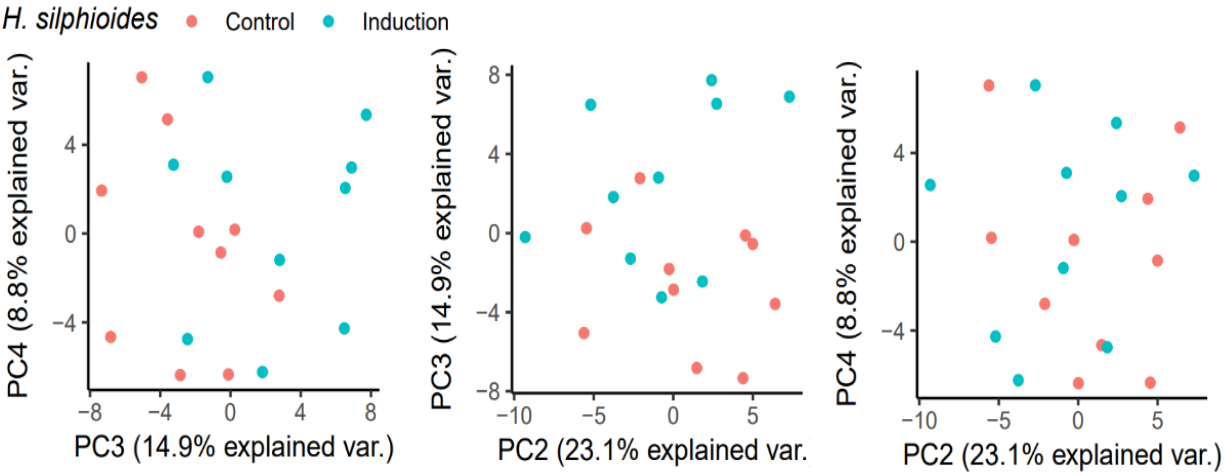

Figure 5: PC plots of *H. silphioides* using PCs 3 and 4 expressing 23.7% of total variance, PCs 2 and 3 expressing 38% of total variance and PCs 2 and 4 expressing 31.9% of total variance. These 3 of the 6 plots of *H. silphioides* best illustrate the lack of visual separation of control versus induced in the plots.

Variation in PC plots separating the control and induced plants in clear linear separation versus a radial separation was seen less often in species. *H. atrorubens* and *H. arizonensis* predominately show separation in a radial pattern amongst the majority of their PC plots (Figures 6 and 7). This was the pattern seen in the majority of species due to overlaps of induced and control samples on the PC plots being very common. However, some species exhibited a mostly linear separation of their control versus induced samples, such as, *H. longifolius* and *H. mollis* as was seen in Figures 3 and 4.

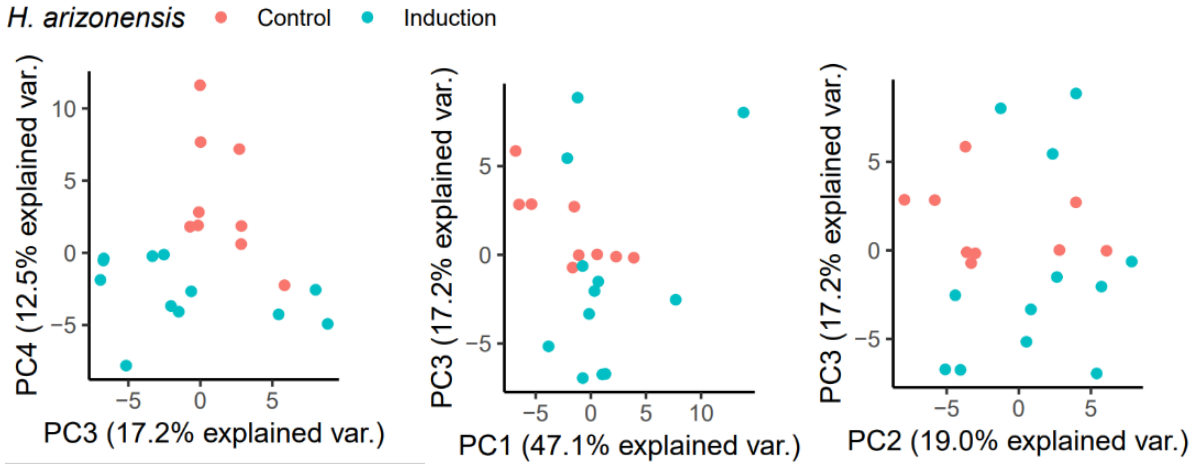

Figure 6: PC plots of *H. arizonensis* using PCs 3 and 4 expressing 29.7% of total variance, PCs 1 and 3 expressing 64.3% of total variance and PCs 2 and 3 expressing 36.2% of total variance. These 3 of the 6 plots of *H. arizonensis* best illustrate the radial separation of control and induced species seen in the plots.

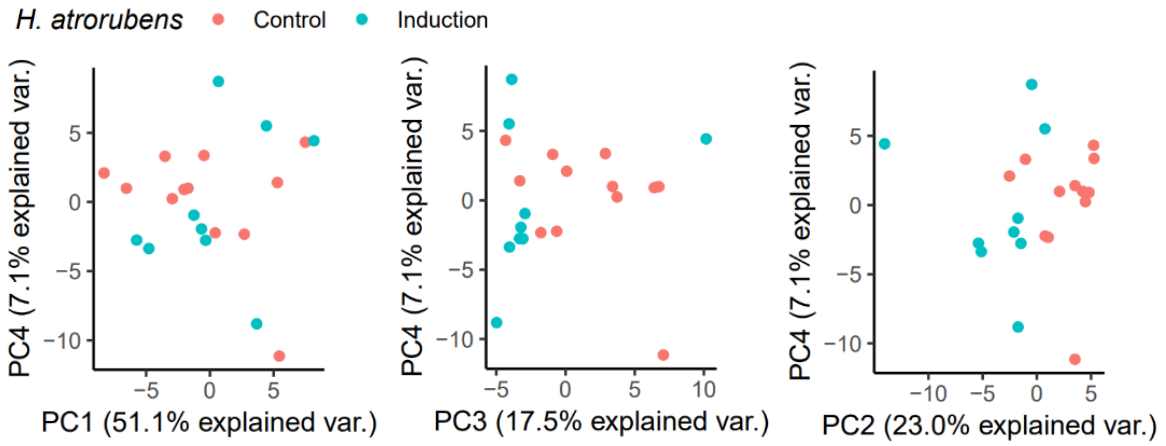

Figure 7: PC plots of *H. atrorubens* using PCs 1 and 4 expressing 58.2% of total variance, PCs 3 and 4 expressing 24.6% of total variance and PCs 2 and 4 expressing 30.1% of total variance. These 3 of the 6 plots of *H. atrorubens* best illustrate the radial separation of control and induced species seen in the plots.

The PC plots of the reflectance data from all of the species is not as easily visually discernable (Figure 8). However, small clusters of control versus induced species can be seen amongst the data. The difficulty there is in just visually discerning between induced and control as compared to the success the SVC exhibited, according to the OOB, exemplifies the capabilities of using this machine learning technique. There was a sampling error for three of the species, *H. agrestis, H. praecox* and *H. exilis*. Their wavelength readings were recorded as only one value for each reading. This can be seen in the PC plots using the reflectance data of all the species on the plot PCs 1 and 2 as a straight line of samples and as far outliers on the plot of PCs 3 and 4 (Figure 8).

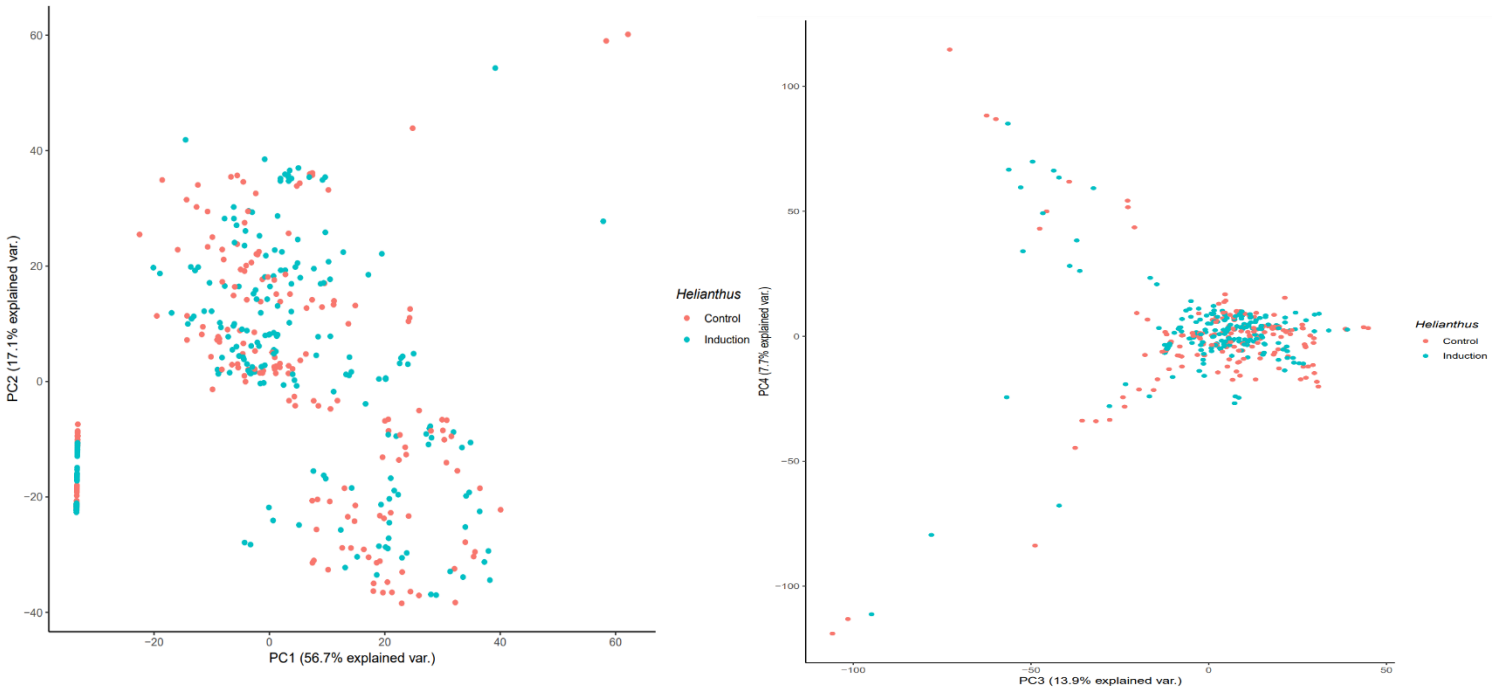

Figure 8: PC plots of all *Helianthus* species using PCs 1 and 2 expressing 73.8% of total variance and PCs 3 and 4 expressing 21.6% of total variance.

The confusion matrix for all of the species shows a low 7% OOB classification error rate when using PCs 1-4 (Table 2). It also shows that consistently for all of the PCs the OOB error rate for classifying a sample as induced when it was actually a control was higher than classifying a sample as being a control when it was actually induced (Table 2). On PCs 1-4 the error rate for classifying a sample as a false induction was 3% higher than classifying a sample as being falsely controlled (Table 2). *H. agrestis*, *H. praecox* and *H. exilis* were removed from the matrices analyzes as a result of sampling error.

Helianthus

Models:					
PC 1-2			Predicted		
			Control	Induction	Classification Error
	Observed	Control	136	46	0.33
		Induction	61	169	0.27
					OOB: 0.26
PC 1-3			Predicted		
			Control	Induction	Classification Error
	Observed	Control	178	25	0.12
		Induction	19	190	0.10
					OOB: 0.11
PC 1-4			Predicted		
			Control	Induction	Classification Error
	Observed	Control	185	16	0.09
		Induction	12	199	0.06
					OOB: 0.07

Table 2: Confusion matrix of all species tested in *Helianthus*. Using PCs 1-2 the SVC exhibited a 26% error rate, an 11% error rate for PCs 1-3 and a 7% error rate for PCs 1-4.

H. salicifolius and *H. silphioides* were the only two species that didn't reach a zero percent error rate in classifying induced versus control when run through the SVC with their data alone (Tables 3 and 4). All other species reached a zero percent error rate on either PCs 1-2, 1-3 or 1-4 (See Appendix). 10 of the species reached zero percent error when using PCs 1-3 (See Appendix). Only 2 species *H. grossesserratus*and *H. giganteus* had a zero percent error rate when using only PCs 1-2 (Tables 5 and 6). The remaining species reached a zero percent error rate by PCs 1-4 (See Appendix).

H. salicifolius

Models:					
PC 1-2			Predicted		
			Control	Induction	Classification Error
	Observed	Control	1	0	0.00
		Induction	5	9	0.55
					OOB: 0.33
PC 1-3			Predicted		
			Control	Induction	Classification Error
	Observed	Control	1	0	0.00
		Induction	5	9	0.55
					OOB: 0.33
PC 1-4			Predicted		
			Control	Induction	Classification Error
	Observed	Control	3	0	0.00
		Induction	3	9	0.33
					OOB: 0.20

Table 3: Confusion matrix of *H. salicifolius*, using PCs 1-2 the SVC exhibited a 33% error rate, a 33% error rate for PCs 1-3 and a 20% error rate for PCs 1-4.

H. silphioides

Models:					
PC 1-2			Predicted		
			Control	Induction	Classification Error
	Observed	Control	4	2	0.50
		Induction	6	8	0.75
					OOB: 0.40
PC 1-3			Predicted		
			Control	Induction	Classification Error
	Observed	Control	9	3	0.33
		Induction	1	7	0.14
					OOB: 0.20
PC 1-4			Predicted		
			Control	Induction	Classification Error
	Observed	Control	8	1	0.13
		Induction	2	9	0.22
					OOB: 0.15

Table 4: Confusion matrix of *H. silphioides* ,using PCs 1-2 the SVC exhibited a 40% error rate, a 20% error rate for PCs 1-3 and a 15% error rate for PCs 1-4.

H. grosseserratus

Models:					
PC 1-2			Predicted		
			Control	Induction	Classification Error
	Observed	Control	12	0	0.00
		Induction	0	12	0.00
					OOB: 0.00
PC 1-3			Predicted		
			Control	Induction	Classification Error
	Observed	Control	12	0	0.00
		Induction	0	12	0.00
					OOB: 0.00
PC 1-4			Predicted		
			Control	Induction	Classification Error
	Observed	Control	12	0	0.00
		Induction	0	12	0.00
					OOB: 0.00

Table 5: Confusion matrix of *H. grosseserratus*, the SVC exhibited a 0% error rate for all of the PCs.

H. giganteus

Models:					
PC 1-2			Predicted		
			Control	Induction	Classification Error
	Observed	Control	12	0	0.00
		Induction	0	12	0.00
					OOB: 0.00
PC 1-3			Predicted		
			Control	Induction	Classification Error
	Observed	Control	12	0	0.00
		Induction	0	12	0.00
					OOB: 0.00
PC 1-4			Predicted		
			Control	Induction	Classification Error
	Observed	Control	12	0	0.00
		Induction	0	12	0.00
					OOB: 0.00

Table 6: Confusion matrix of *H. giganteus*, the SVC exhibited a 0% error rate for all of the PCs.

DISCUSSION

The variant responses to induction seen across the PC plots is likely due to a variance in induced response across the genus. However, exactly how the variance may predict increase or decrease in resistance is beyond the scope of this study. The strong linear separations seen in species may indicate a stronger more simplistic defense response, where the radial separations may be due to a more complex response. It would make sense that if the response was the strong up regulation of few compounds this could be more easily discernable from the control than a response that may be small upregulations of multiple compounds. The two species that were poorly separable in the majority of their PC plots, *H. siphioides* and *H. salicifolius*, may have been due to a weak defense response to the herbivory. This is supported by the fact that these same two species were the only two species that the SVC was unable to classify with zero percent error as induced versus control. Another explanation for the difficulty in separation of these species amongst the PCs and in classification via the SVC could be because these species exhibit a strong constitutive defense. If these species constantly have defensive compounds being produced that may make the need for strong inducible responses decrease. If the response is less strong and congruent with other defensive compounds already present, this could easily make classification between the induced and control more difficult. While the variance of the response among the species is clear, determination of how these responses are differing and what that means for resistance cannot be determined. However, this variance in induced responses does provide opportunities for plant breeders to engineer crops that only upregulate defensive compounds when necessary. If more research is to be conducted concerning the specifics of the secondary metabolites being induced amongst the species, it is almost certain the responses

would vary greatly and correspond to variation in resistance. Conserving metabolic resources in *Helianthus* crops that would normally be used for constitutive responses by breeding for stronger induced responses would improve yields – having significant economic impact.

Three species, *H. agresits, H. praecox* and *H. exilis,* had the same value for all of their wavelength data as a result of sampling error. They were all taken on the same day and when the data was analyzed the error became clear. The error was likely due to failure to open the light on the Ocean Optics spectrophotometer. Another explanation could be failure to remove the cap on the reflectance probe. Either explanation is likely, as they both would explain why only one wavelength value was reflected.

The confusion matrix for all of the species consistently showed for all of the PCs the OOB error rate for classifying a sample as induced when it was actually a control was higher than classifying a sample as being a control when it was actually induced. One reason the classification may be erroneous in predicting a sample as induced when its actually control may be due to differences in the environment of samples such as, variation in ontogenetic stages of samples, ratio of leaf area eaten to biomass, other pathogens or herbivores may have adulterated the sample or another variation in the sample's conditions. An explanation for a false control prediction when the sample is actually induced could be due to high constitutive defenses. Similarly, there is a possibility that constitutive defenses may be making the PCs less useful for classification, thus causing the SVC to be less effective in predicting induction. If the defenses are already high in the plant, the attack of the herbivore may not trigger a strong defense. This would cause the SVC to predict the plant as control as it would not differ much from the control since the defensive compounds are constitutively present.

The confusion matrix for all of the species also showed a very low OOB error for PCs 1-3 and PCs 1-4, while PCs 1-4 could be an over fit model, PCs 1:3 still show extremely low error in predicting which samples are controlled versus induced across all the species. Only a 7% error for PCs 1-4 and 11% error for PCs 1-3 without any attempt to expand parameters are exciting results. On a scale within the individual species the error rates were far lower. Only 2 species didn't reach 0% error. The model experienced near perfect results on these tests, making the application of this novel method to species wide studies extremely promising. While more research is needed to improve this technique, this novel approach of chemical analysis could vastly increase researchers' chemical sampling capacity, helping to push forward discovery in all fields of science.

FUTURE STEPS

Moving forward with my data I intend to answer the following questions: 1. How effective is spectral reflectance at analyzing the specific secondary metabolites upregulated for induced defense response? 2. How have secondary metabolite defense responses evolved across the genus *Helianthus*? To answer the first question GC/MS will be run on the sampled leaves that are currently being stored in -80° Celsius. This same reflectance data will be used to predict the GC/MS results in a partial least-squares framework. The validated model will give insight into which combination of reflected wavelengths are predictive of GC/MS data. This in conjuncture with the SVM classification data will be used to test whether or not spectral reflectance can be used to predict concentrations of individual secondary metabolites employed in induced responses. To address the second question, I intend to analyze the induced defense response profiles obtained by GC/MS through a phylogenetic comparison to examine how relative induction response has evolved during sunflower differentiation – has high inducibility evolved many or few times, and how the degree of inducibility evolves with native habitat environmental conditions.

APPENDIX

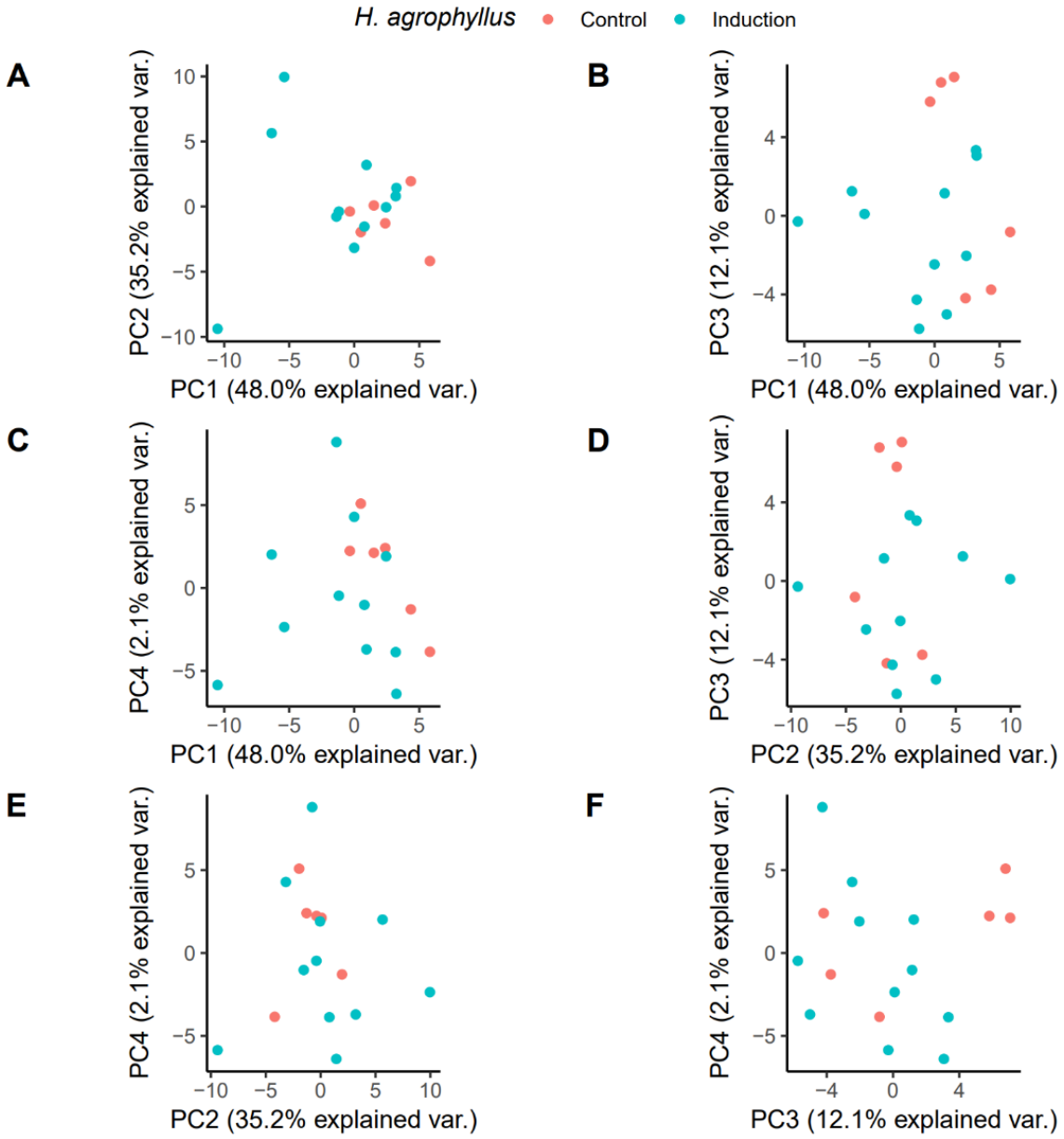

Figure A1: Shows all six PC plots for classifying control versus induced plants in species *H. agrophyllus*

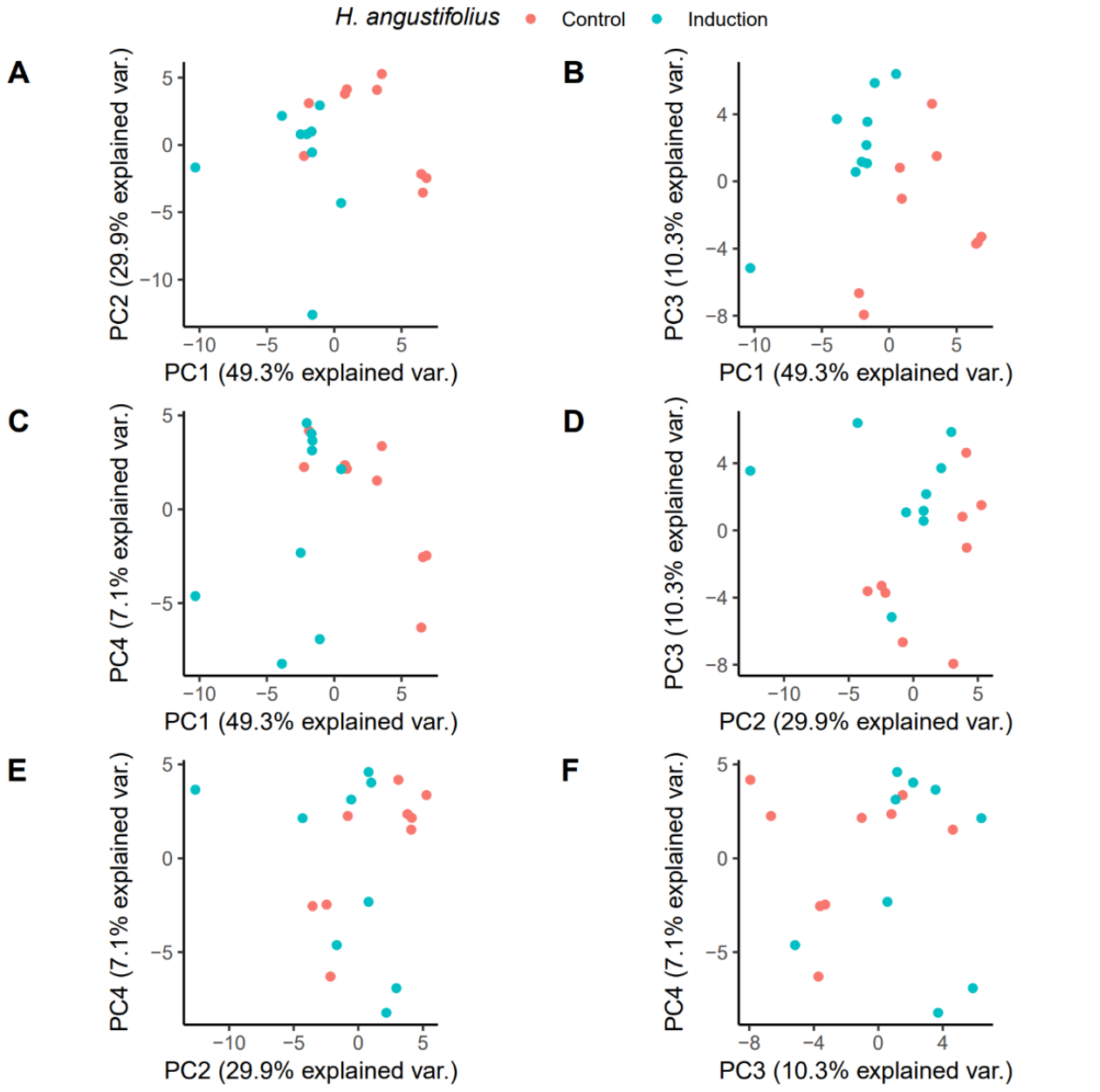

Figure A2: Shows all six PC plots for classifying control versus induced plants in species *H. angustifolius*

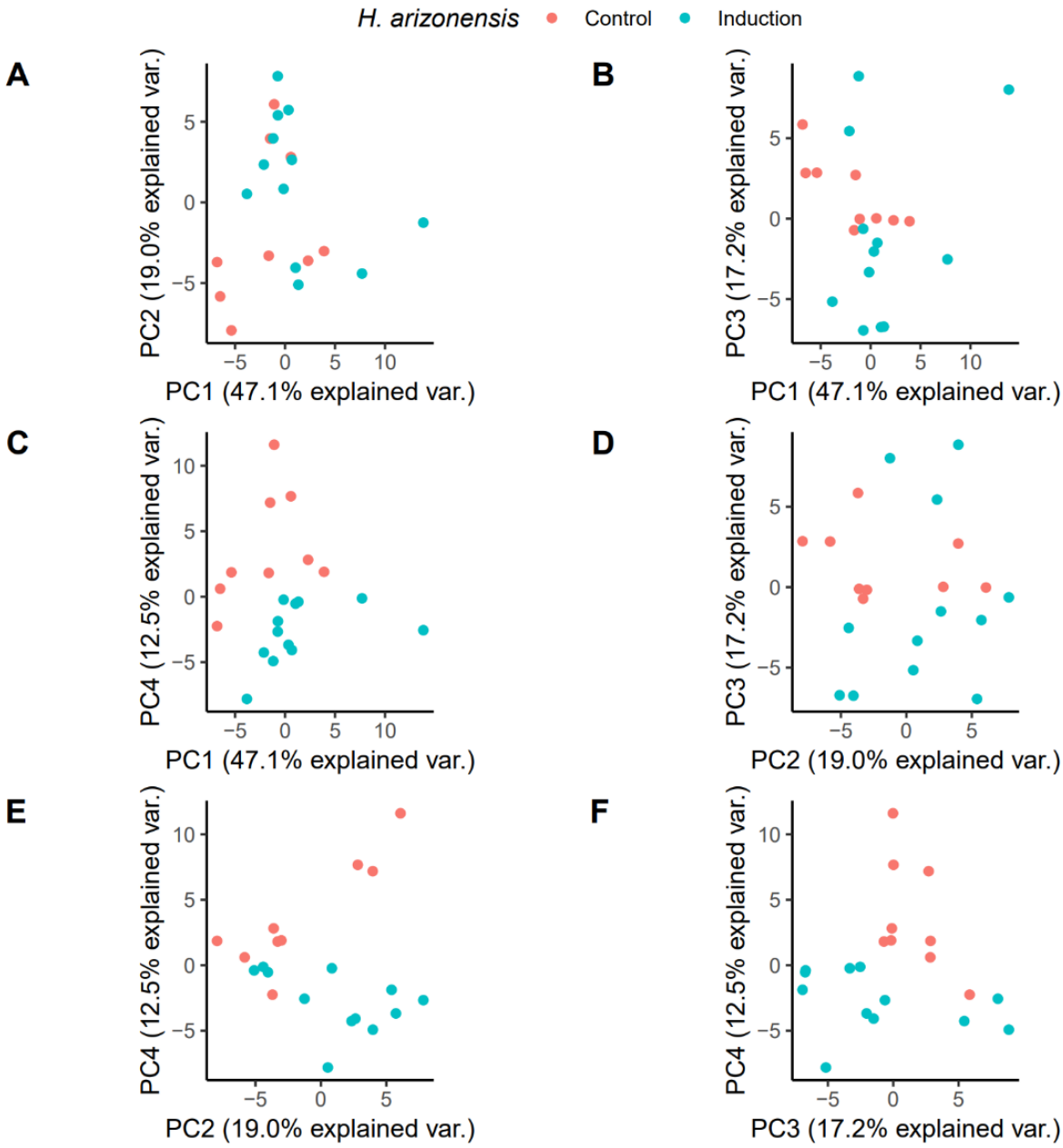

Figure A3: Shows all six PC plots for classifying control versus induced plants in species *H. arizonensis*

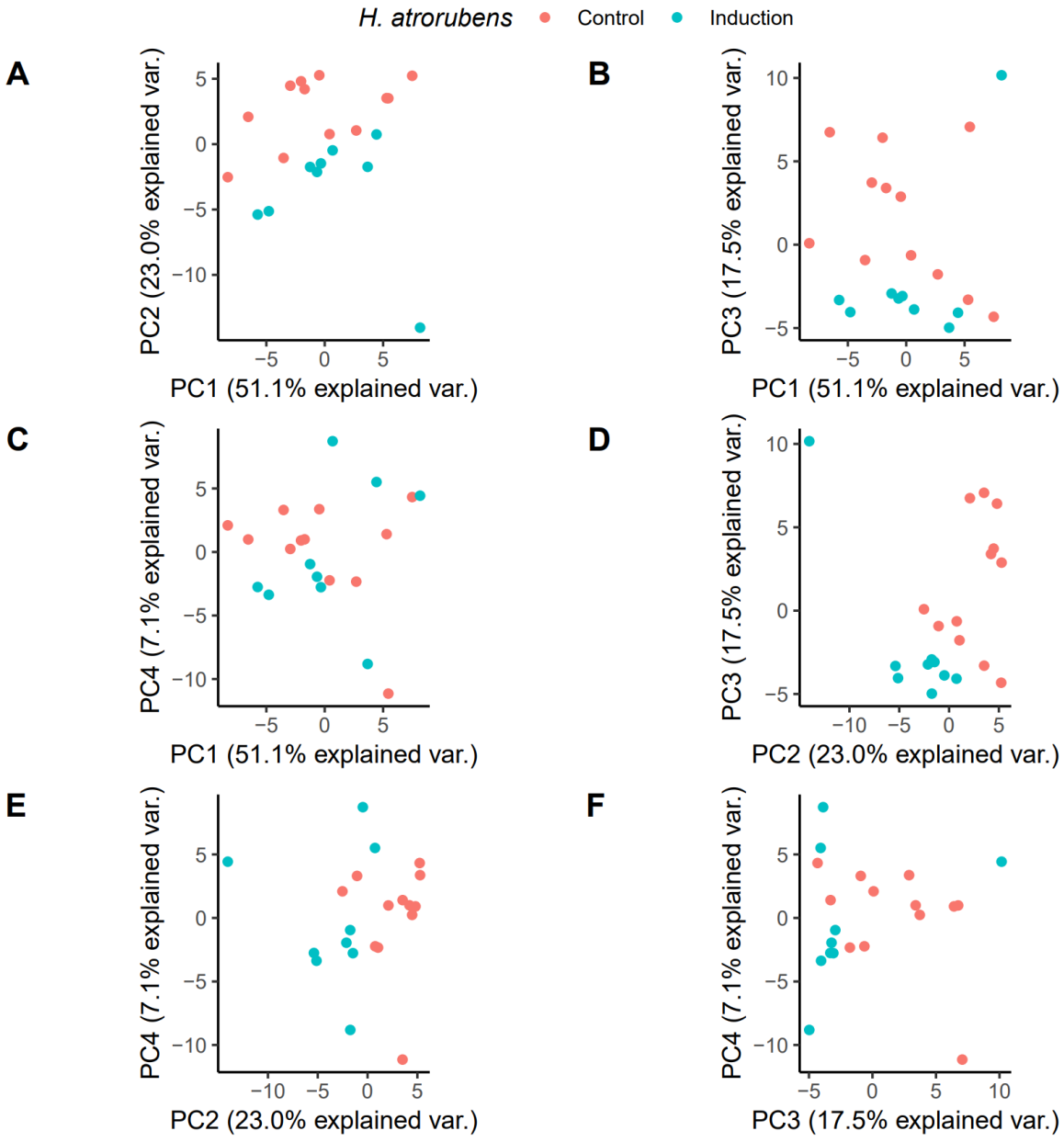

Figure A4: Shows all six PC plots for classifying control versus induced plants in species *H. atrorubens*

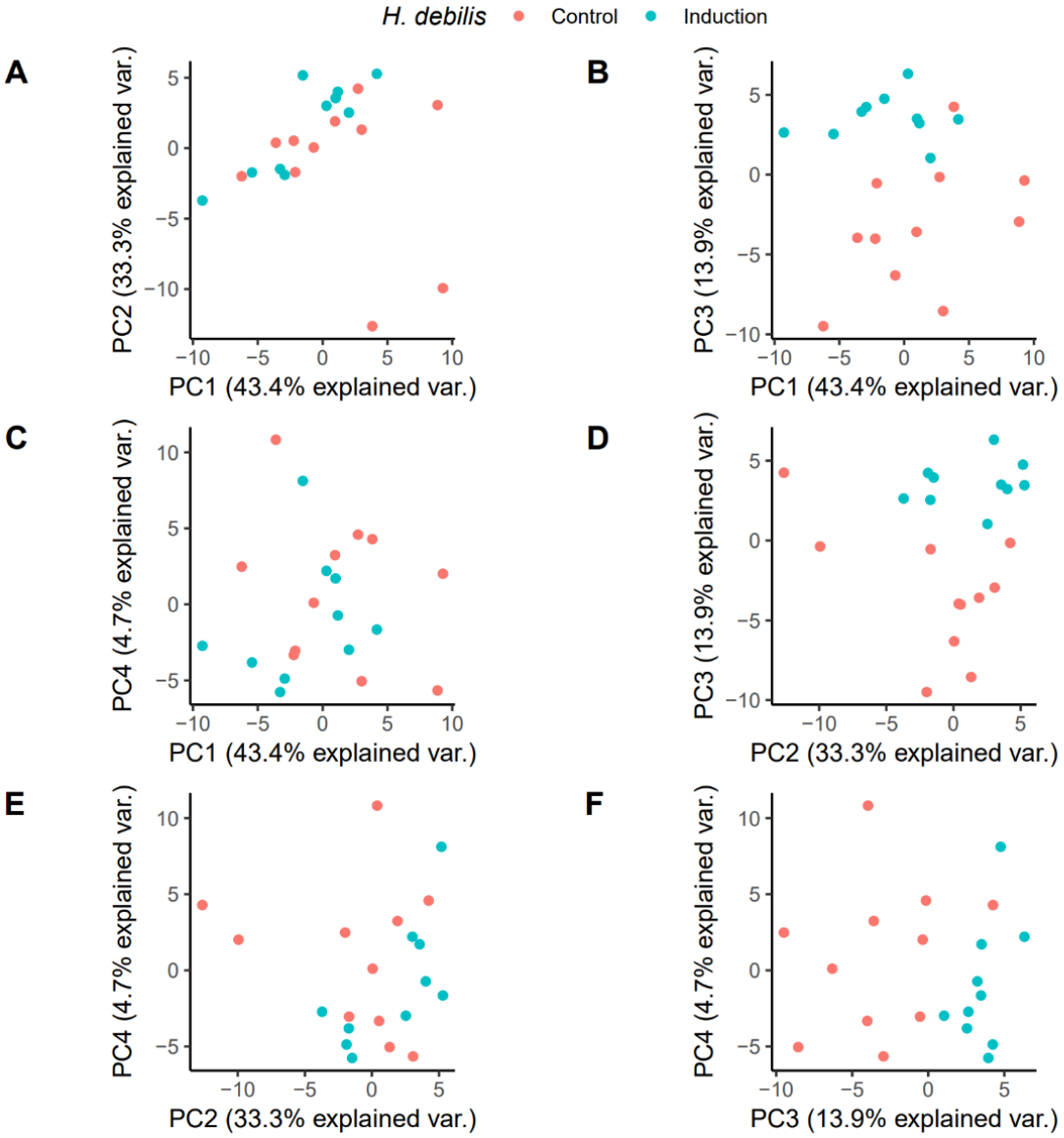

Figure A5: Shows all six PC plots for classifying control versus induced plants in species *H. debilis*

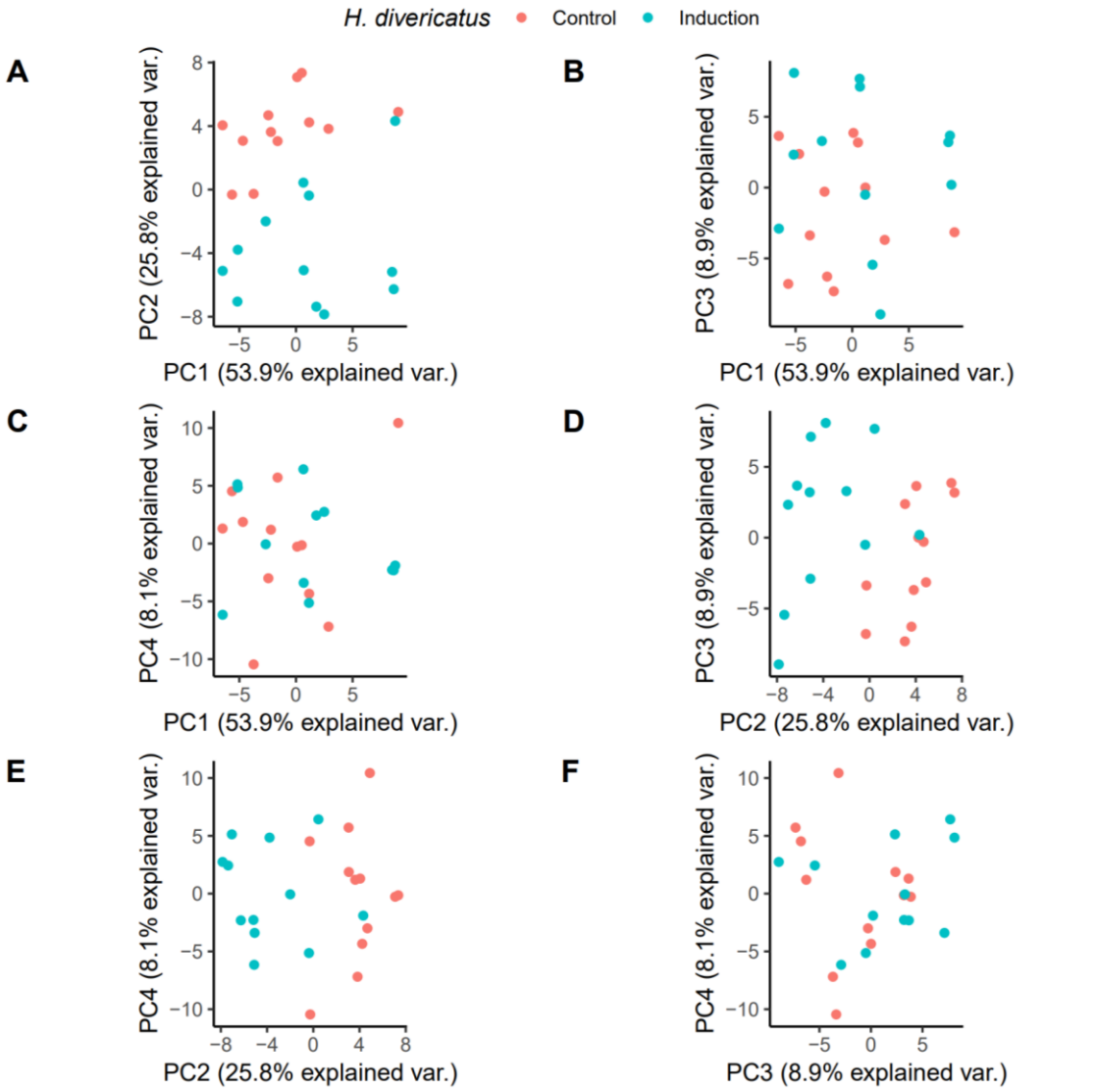

Figure A6: Shows all six PC plots for classifying control versus induced plants in species *H. divericatus*

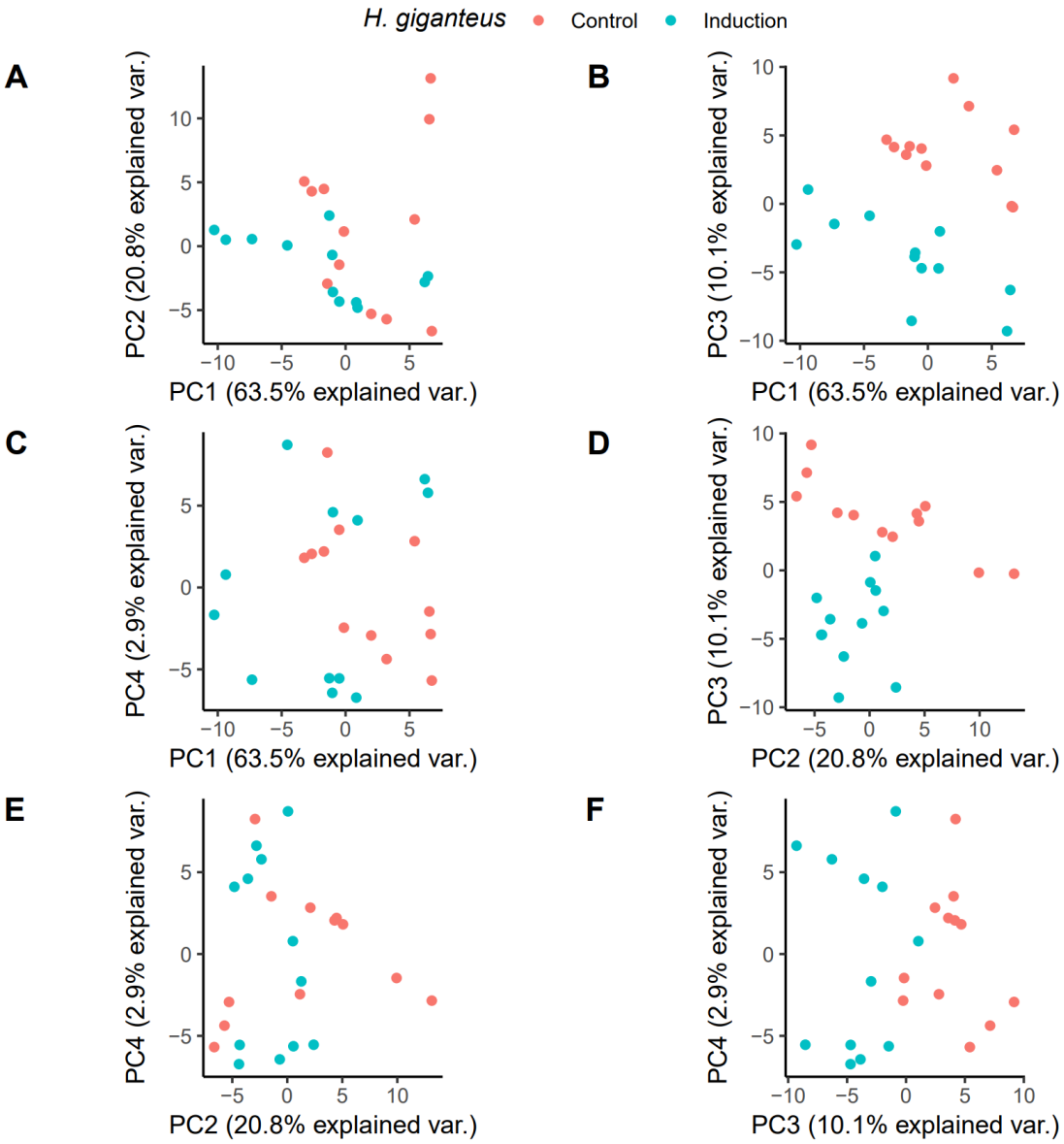

Figure A7: Shows all six PC plots for classifying control versus induced plants in species *H. giganteus*

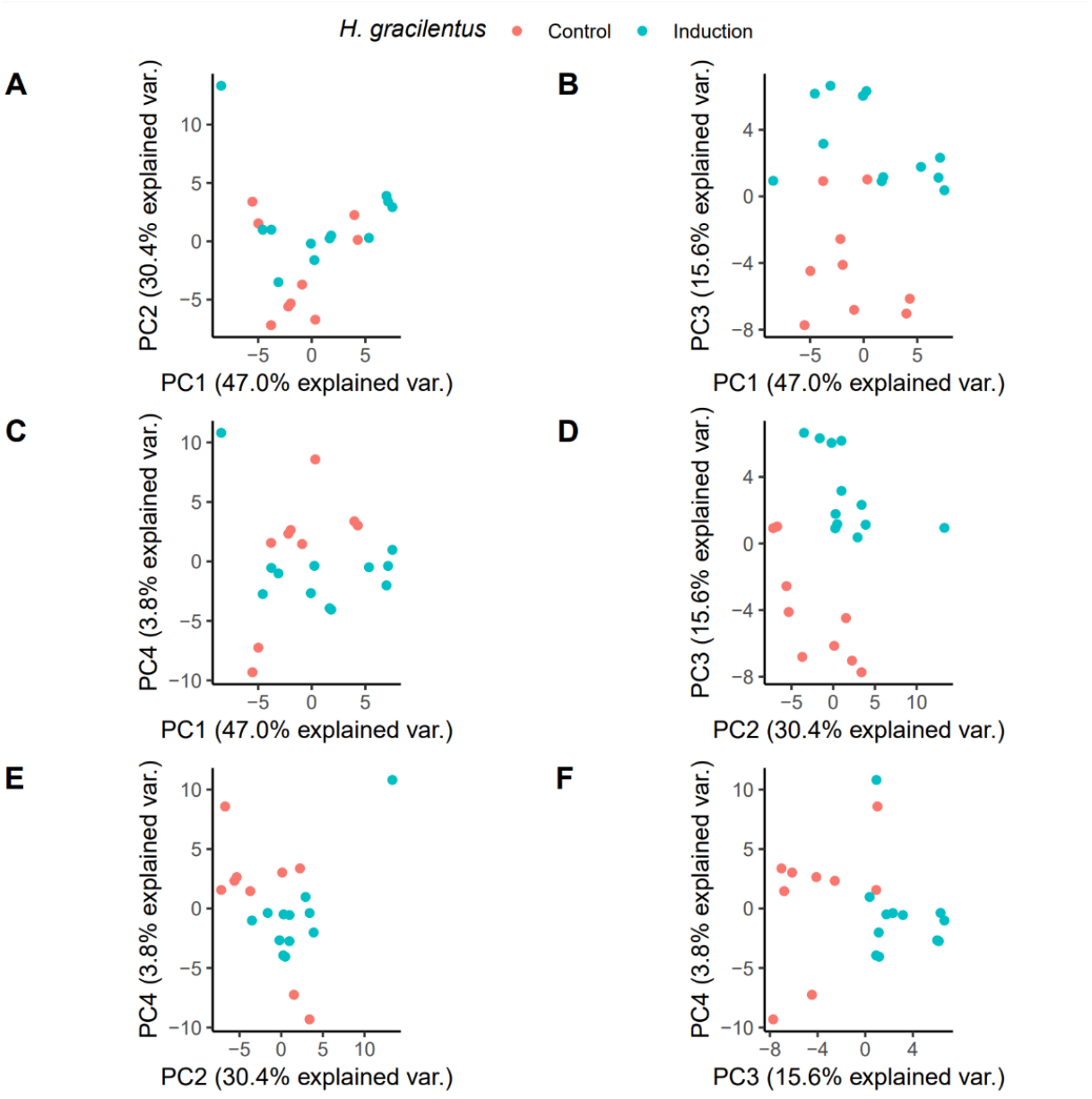

Figure A8: Shows all six PC plots for classifying control versus induced plants in species *H. gracilentus*

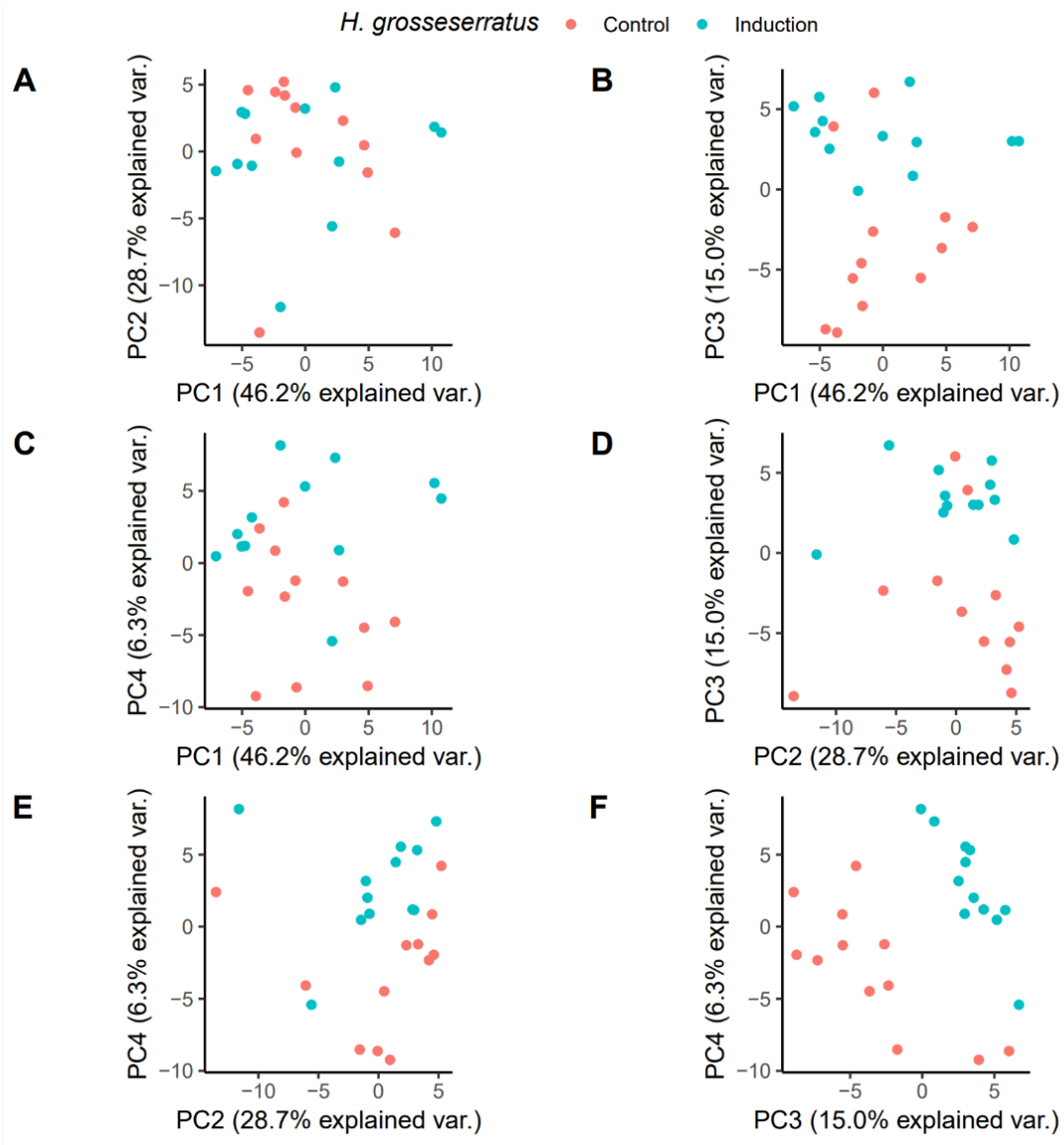

Figure A9: Shows all six PC plots for classifying control versus induced plants in species *H. grosseserratus*

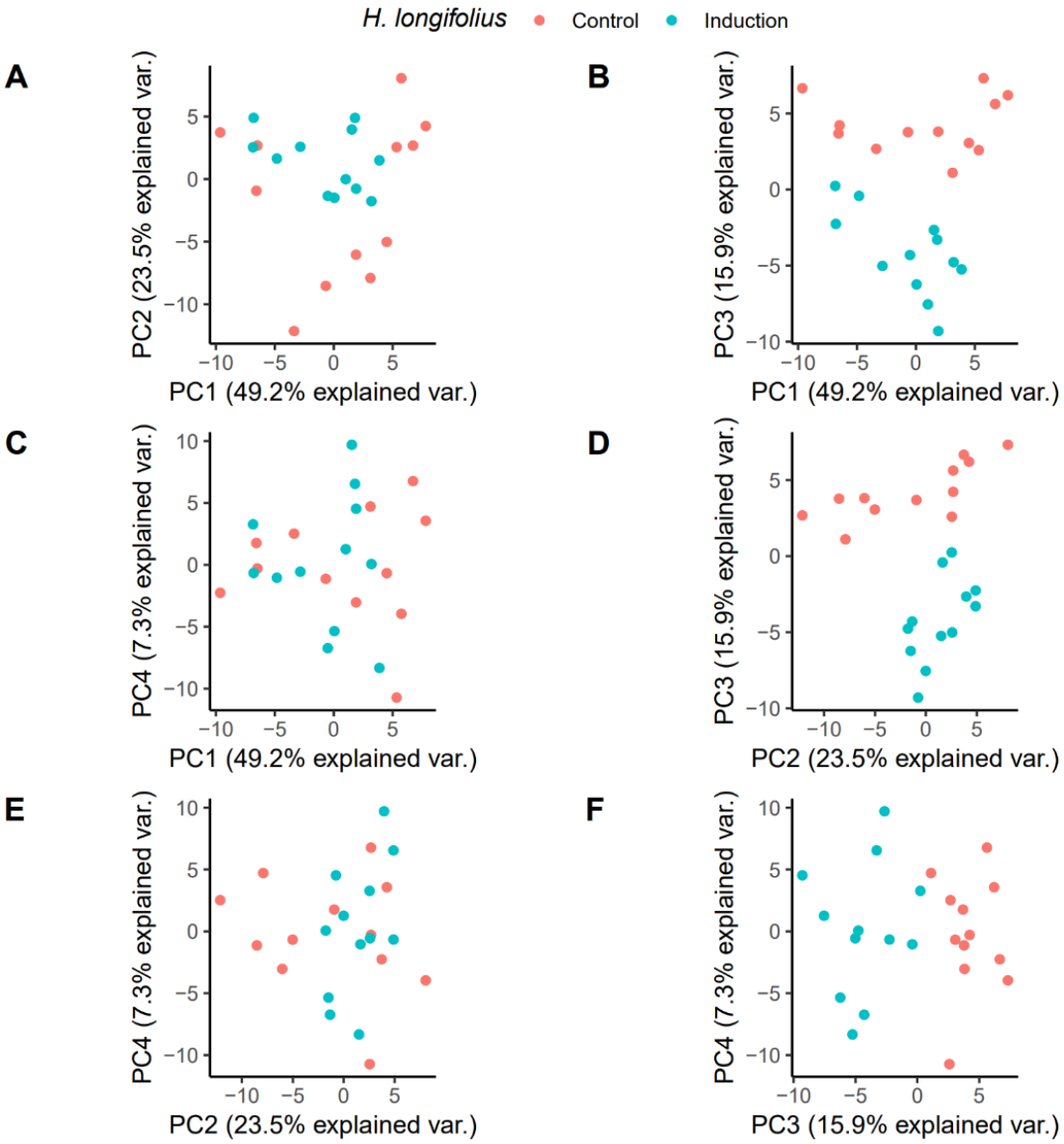

Figure A10: Shows all six PC plots for classifying control versus induced plants in species *H. longifolius*

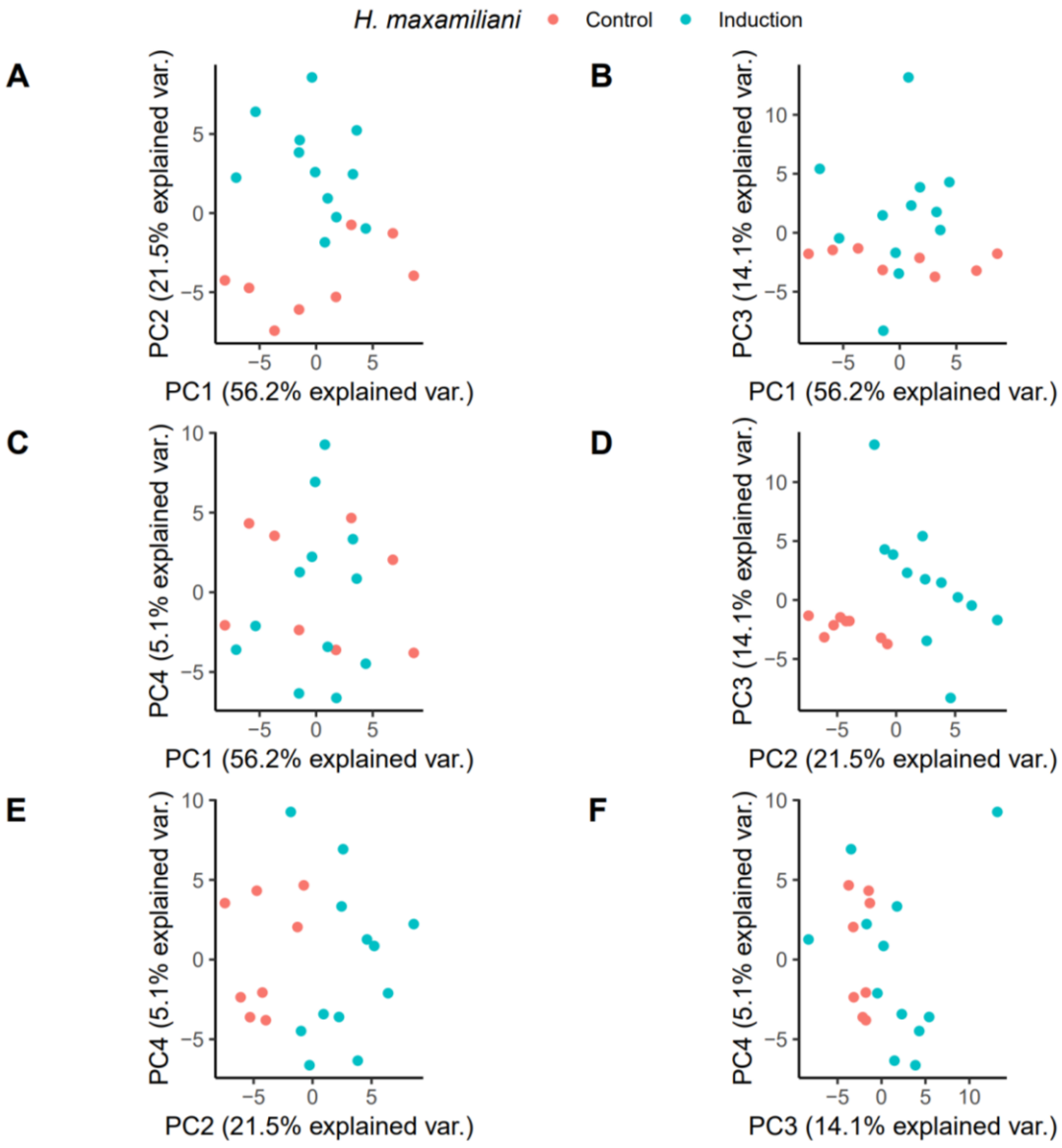

Figure A11: Shows all six PC plots for classifying control versus induced plants in species *H. maximiliani*

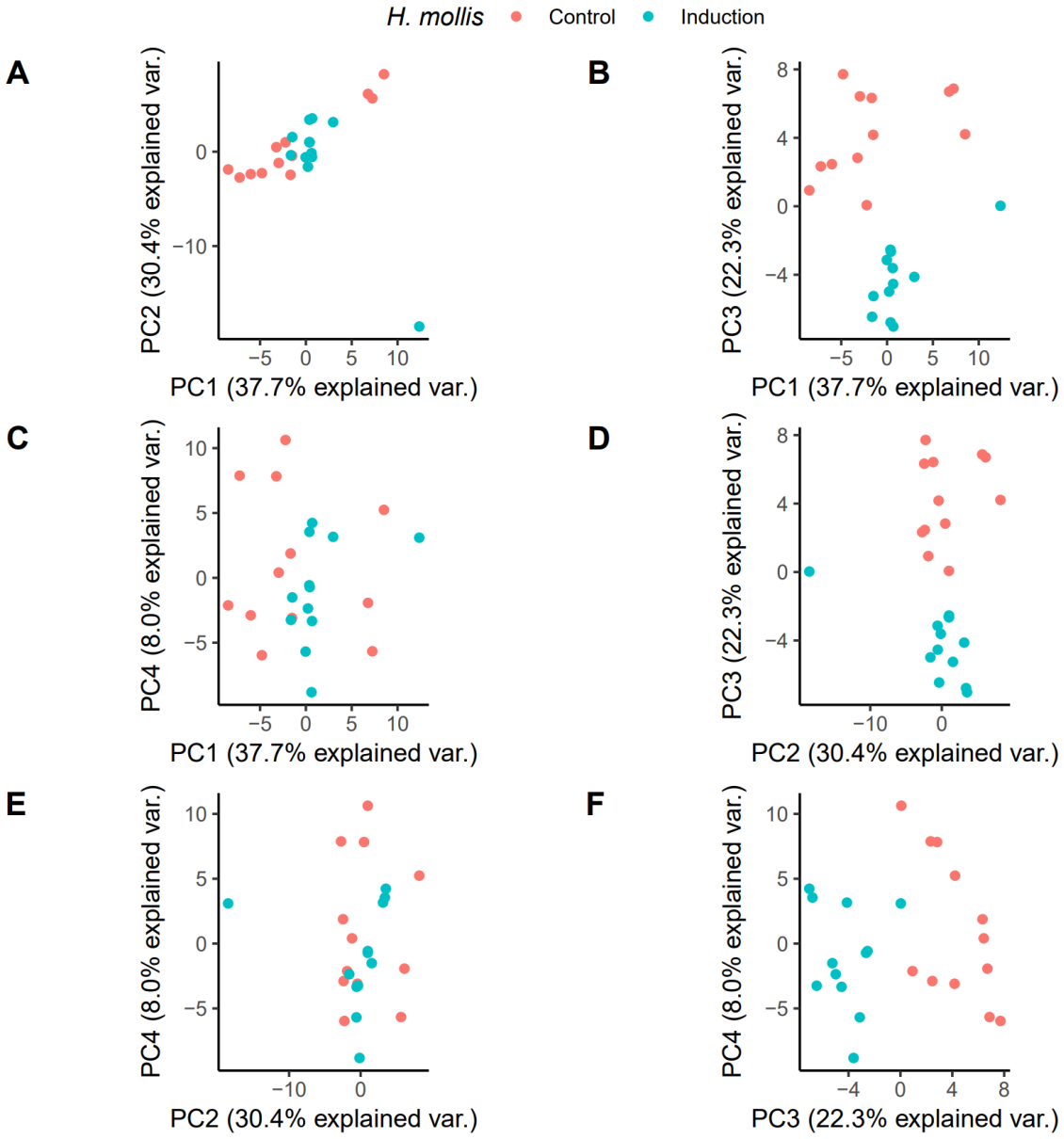

Figure A12: Shows all six PC plots for classifying control versus induced plants in species *H. mollis*

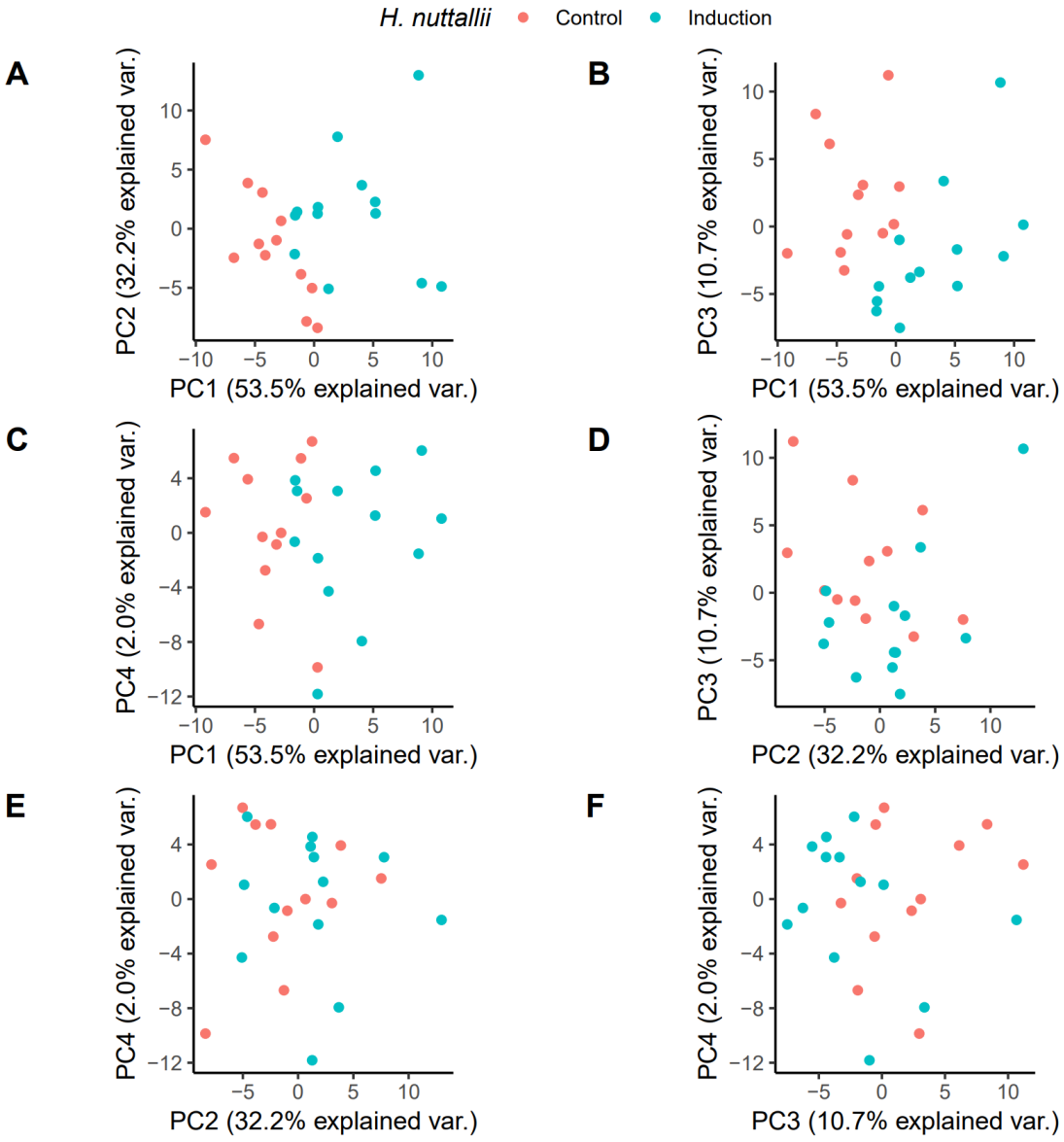

Figure A13: Shows all six PC plots for classifying control versus induced plants in species *H. nuttallii*

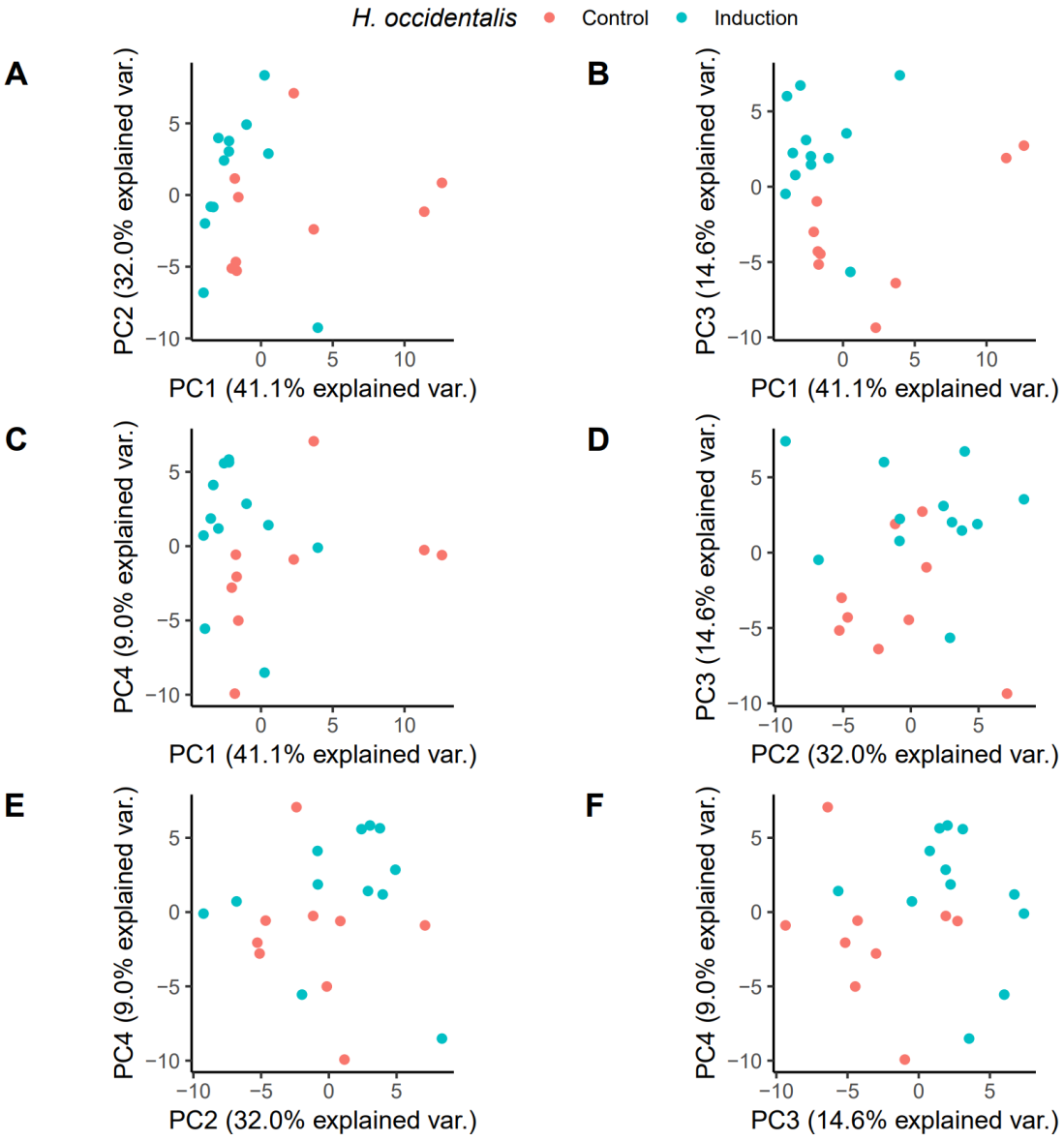

Figure A14: Shows all six PC plots for classifying control versus induced plants in species *H. occidentalis*

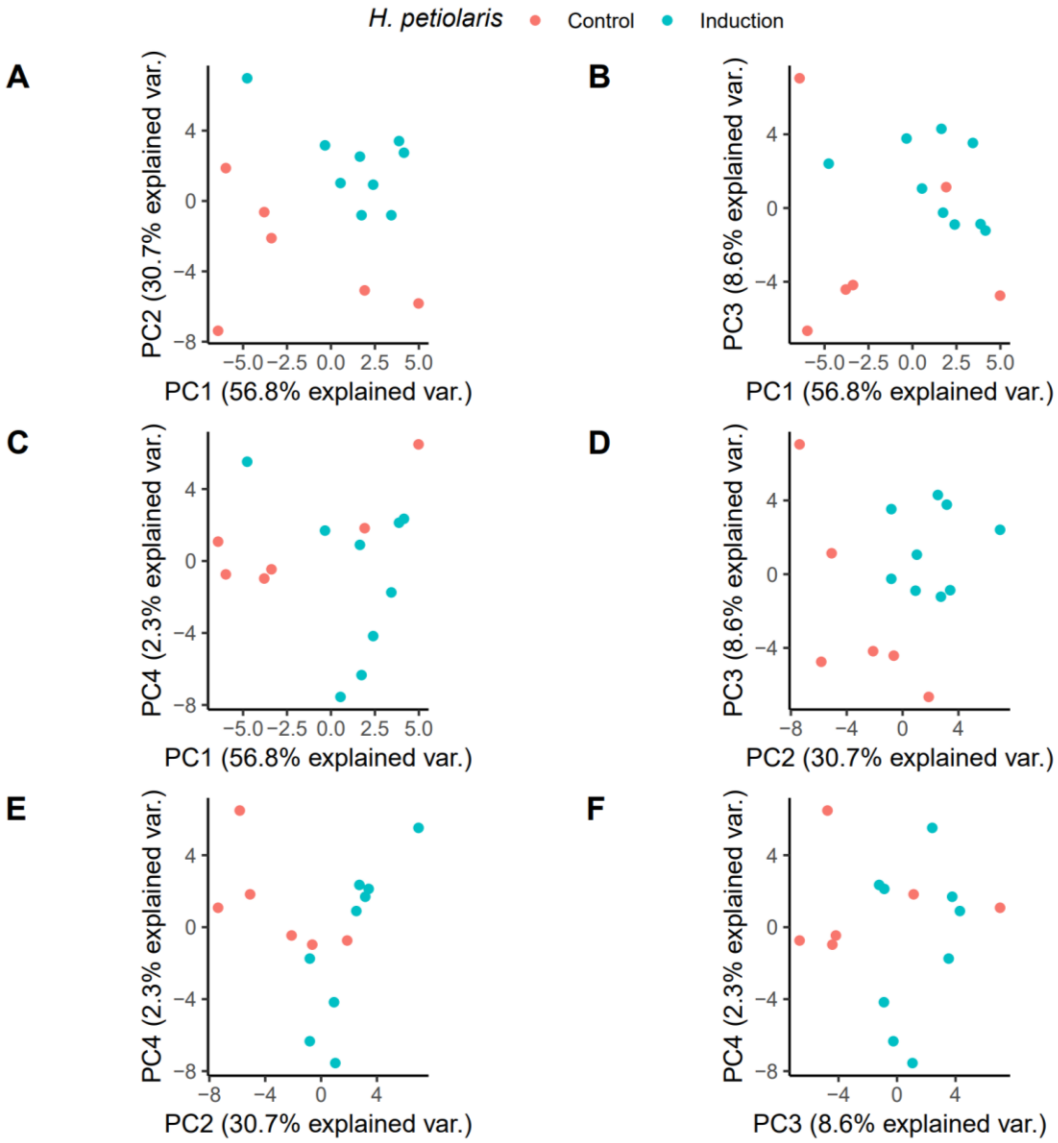

Figure A15: Shows all six PC plots for classifying control versus induced plants in species *H. petiolaris*

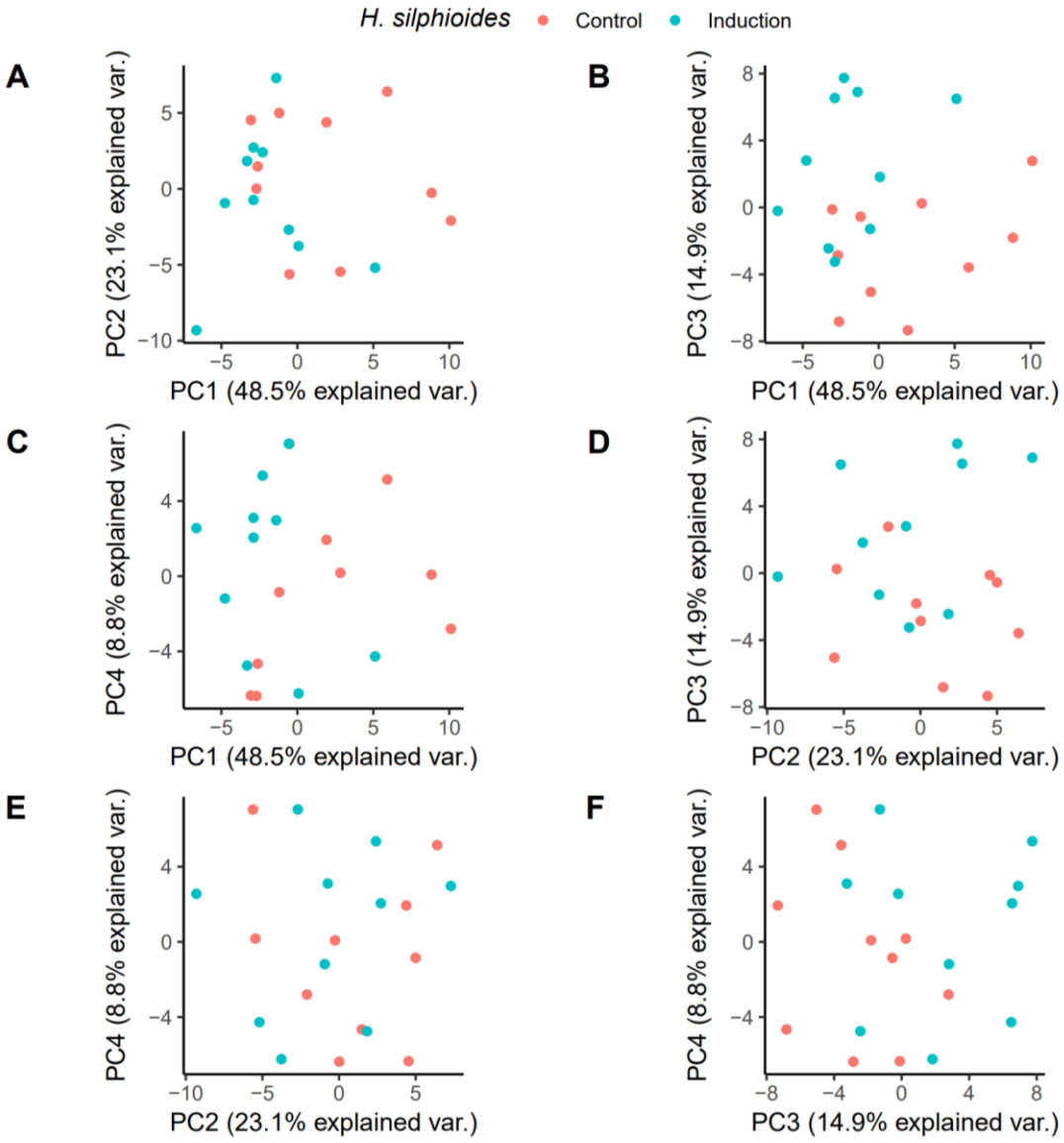

Figure A16: Shows all six PC plots for classifying control versus induced plants in species *H. silphioides*

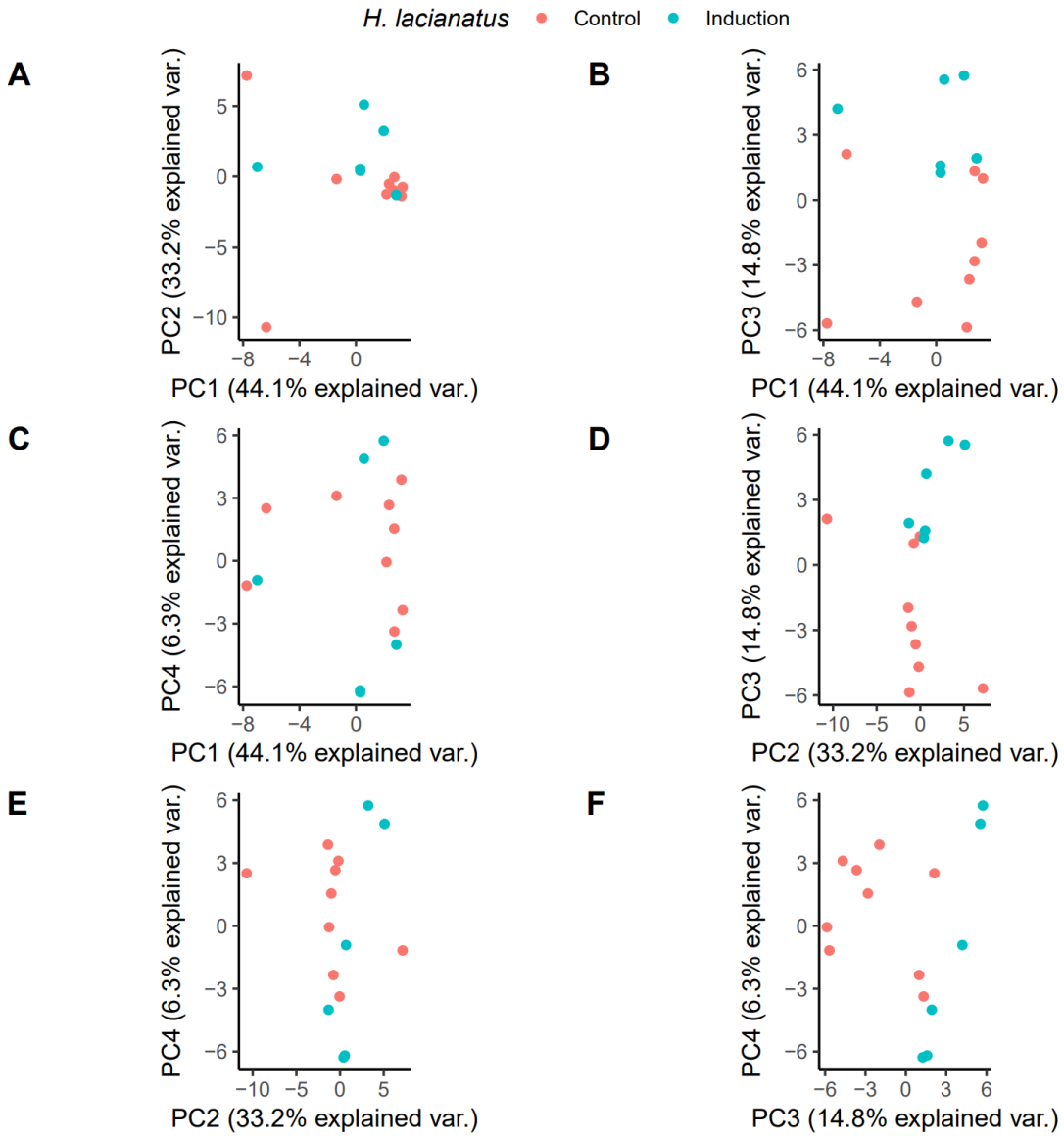

Figure A17: Shows all six PC plots for classifying control versus induced plants in species *H. lacianatus*

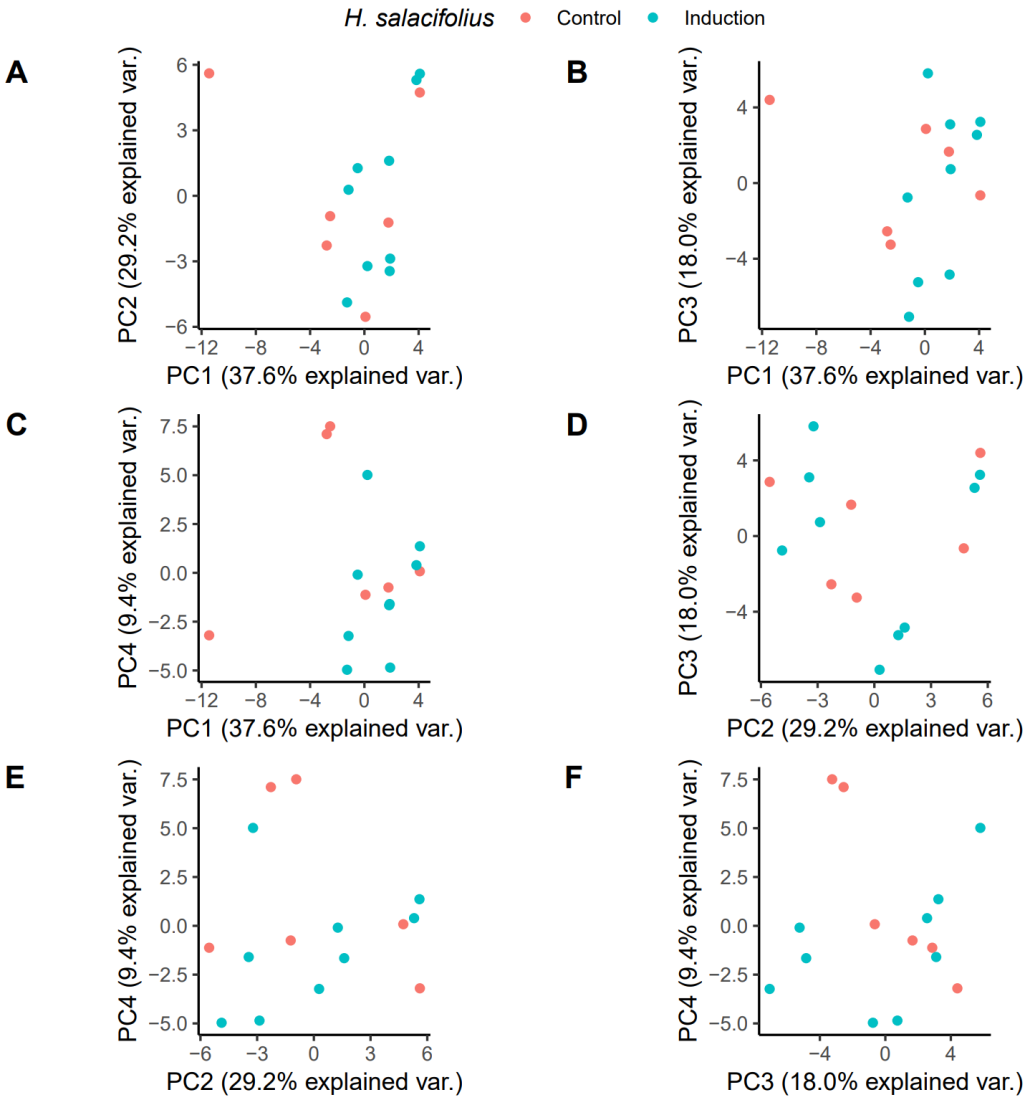

Figure A18: Shows all six PC plots for classifying control versus induced plants in species *H. salicifolius*

Helianthus

Models:					
PC 1-2			Predicted		
			Control	Induction	Classification Error
	Observed	Control	136	46	0.33
		Induction	61	169	0.27
					OOB: 0.26
PC 1-3			Predicted		
			Control	Induction	Classification Error
	Observed	Control	178	25	0.12
		Induction	19	190	0.10
					OOB: 0.11
PC 1-4			Predicted		
			Control	Induction	Classification Error
	Observed	Control	185	16	0.09
		Induction	12	199	0.06
					OOB: 0.07

Table A1: Shows OOB for the SVC using PCs 1-2, 1-3 and 1-4 for all of the *Helianthus* species

H. maximiliani

Models:					
PC 1-2			Predicted		
			Control	Induction	Classification Error
	Observed	Control	7	0	0.00
		Induction	1	12	0.08
					OOB: 0.05
PC 1-3			Predicted		
			Control	Induction	Classification Error
	Observed	Control	8	0	0.00
		Induction	0	12	0.00
					OOB: 0.00
PC 1-4			Predicted		
			Control	Induction	Classification Error
	Observed	Control	8	0	0.00
		Induction	0	12	0.00
					OOB: 0.00

Table A2: Shows OOB for the SVC using PCs 1-2, 1-3 and 1-4 for *H. maximiliani*

H. atrorubens

Models:					
PC 1-2			Predicted		
			Control	Induction	Classification Error
	Observed	Control	3	1	0.33
		Induction	6	11	0.55
					OOB: 0.33
PC 1-3			Predicted		
			Control	Induction	Classification Error
	Observed	Control	12	0	0.00
		Induction	0	9	0.00
					OOB: 0.00
PC 1-4			Predicted		
			Control	Induction	Classification Error
	Observed	Control	12	0	0.00
		Induction	0	9	0.00
					OOB: 0.00

Table A3: Shows OOB for the SVC using PCs 1-2, 1-3 and 1-4 for *H. atrorubens*

H. giganteus

Models:					
PC 1-2			Predicted		
			Control	Induction	Classification Error
	Observed	Control	12	0	0.00
		Induction	0	12	0.00
					OOB: 0.00
PC 1-3			Predicted		
			Control	Induction	Classification Error
	Observed	Control	12	0	0.00
		Induction	0	12	0.00
					OOB: 0.00
PC 1-4			Predicted		
			Control	Induction	Classification Error
	Observed	Control	12	0	0.00
		Induction	0	12	0.00
					OOB: 0.00

Table A4: Shows OOB for the SVC using PCs 1-2, 1-3 and 1-4 for *H. giganteus*

H. mollis

Models:					
PC 1-2			Predicted		
			Control	Induction	Classification Error
	Observed	Control	12	1	0.08
		Induction	0	11	0.00
					OOB: 0.04
PC 1-3			Predicted		
			Control	Induction	Classification Error
	Observed	Control	12	0	0.00
		Induction	0	12	0.00
					OOB: 0.00
PC 1-4			Predicted		
			Control	Induction	Classification Error
	Observed	Control	12	0	0.00
		Induction	0	12	0.00
					OOB: 0.00

Table A5: Shows OOB for the SVC using PCs 1-2, 1-3 and 1-4 for *H. mollis*

H. longifolius

Models:					
PC 1-2			Predicted		
			Control	Induction	Classification Error
	Observed	Control	6	3	0.5
		Induction	6	9	0.67
					OOB: 0.38
PC 1-3			Predicted		
			Control	Induction	Classification Error
	Observed	Control	12	0	0.00
		Induction	0	12	0.00
					OOB: 0.00
PC 1-4			Predicted		
			Control	Induction	Classification Error
	Observed	Control	12	0	0.00
		Induction	0	12	0.00
					OOB: 0.00

Table A6: Shows OOB for the SVC using PCs 1-2, 1-3 and 1-4 for *H. longifolius*

H. occidentalis

Models:					
PC 1-2			Predicted		
			Control	Induction	Classification Error
	Observed	Control	8	1	0.13
		Induction	1	11	0.09
					OOB: 0.10
PC 1-3			Predicted		
			Control	Induction	Classification Error
	Observed	Control	9	1	0.11
		Induction	0	11	0.00
					OOB: 0.05
PC 1-4			Predicted		
			Control	Induction	Classification Error
	Observed	Control	9	0	0.00
		Induction	0	12	0.00
					OOB: 0.00

Table A7: Shows OOB for the SVC using PCs 1-2, 1-3 and 1-4 for *H. occidentalis*

H. agrophyllus

Models:					
PC 1-2			Predicted		
			Control	Induction	Classification Error
	Observed	Control	6	1	0.17
		Induction	0	10	0.00
					OOB: 0.06
PC 1-3			Predicted		
			Control	Induction	Classification Error
	Observed	Control	6	0	0.00
		Induction	0	11	0.00
					OOB: 0.00
PC 1-4			Predicted		
			Control	Induction	Classification Error
	Observed	Control	6	0	0.00
		Induction	0	11	0.00
					OOB: 0.00

Table A8: Shows OOB for the SVC using PCs 1-2, 1-3 and 1-4 for *H. agrophyllus*

H. silphioides

Models:					
PC 1-2			Predicted		
			Control	Induction	Classification Error
	Observed	Control	4	2	0.50
		Induction	6	8	0.75
					OOB: 0.40
PC 1-3			Predicted		
			Control	Induction	Classification Error
	Observed	Control	9	3	0.33
		Induction	1	7	0.14
					OOB: 0.20
PC 1-4			Predicted		
			Control	Induction	Classification Error
	Observed	Control	8	1	0.13
		Induction	2	9	0.22
					OOB: 0.15

Table A9: Shows OOB for the SVC using PCs 1-2, 1-3 and 1-4 for *H. silphioides*

H. debilis

Models:					
PC 1-2			Predicted		
			Control	Induction	Classification Error
	Observed	Control	9	2	0.22
		Induction	2	8	0.25
					OOB: 0.19
PC 1-3			Predicted		
			Control	Induction	Classification Error
	Observed	Control	10	0	0.00
		Induction	1	10	0.10
					OOB: 0.05
PC 1-4			Predicted		
			Control	Induction	Classification Error
	Observed	Control	11	0	0.00
		Induction	0	10	0.00
					OOB: 0.00

Table A10: Shows OOB for the SVC using PCs 1-2, 1-3 and 1-4 for *H. debilis*

H. angustifolius

Models:					
PC 1-2			Predicted		
			Control	Induction	Classification Error
	Observed	Control	7	0	0.00
		Induction	2	9	0.22
					OOB: 0.11
PC 1-3			Predicted		
			Control	Induction	Classification Error
	Observed	Control	9	0	0.00
		Induction	0	9	0.00
					OOB: 0.00
PC 1-4			Predicted		
			Control	Induction	Classification Error
	Observed	Control	9	0	0.00
		Induction	0	9	0.00
					OOB: 0.00

Table A11: Shows OOB for the SVC using PCs 1-2, 1-3 and 1-4 for *H. angustifolius*

H. divericatus

Models:					
PC 1-2			Predicted		
			Control	Induction	Classification Error
	Observed	Control	12	1	0.08
		Induction	0	11	0.00
					OOB: 0.04
PC 1-3			Predicted		
			Control	Induction	Classification Error
	Observed	Control	12	1	0.08
		Induction	0	11	0.00
					OOB: 0.04
PC 1-4			Predicted		
			Control	Induction	Classification Error
	Observed	Control	12	0	0.00
		Induction	0	12	0.00
					OOB: 0.00

Table A12: Shows OOB for the SVC using PCs 1-2, 1-3 and 1-4 for *H. divericatus*

H. gracilentus

Models:					
PC 1-2			Predicted		
			Control	Induction	Classification Error
	Observed	Control	5	1	0.20
		Induction	4	11	0.36
					OOB: 0.23
PC 1-3			Predicted		
			Control	Induction	Classification Error
	Observed	Control	9	0	0.00
		Induction	0	12	0.00
					OOB: 0.00
PC 1-4			Predicted		
			Control	Induction	Classification Error
	Observed	Control	9	0	0.00
		Induction	0	12	0.00
					OOB: 0.00

Table A13: Shows OOB for the SVC using PCs 1-2, 1-3 and 1-4 for *H. gracilentus*

H. grosseserratus

Models:					
PC 1-2			Predicted		
			Control	Induction	Classification Error
	Observed	Control	12	0	0.00
		Induction	0	12	0.00
					OOB: 0.00
PC 1-3			Predicted		
			Control	Induction	Classification Error
	Observed	Control	12	0	0.00
		Induction	0	12	0.00
					OOB: 0.00
PC 1-4			Predicted		
			Control	Induction	Classification Error
	Observed	Control	12	0	0.00
		Induction	0	12	0.00
					OOB: 0.00

Table A14: Shows OOB for the SVC using PCs 1-2, 1-3 and 1-4 for *H. grosseserratus*

H. nuttallii

Models:					
PC 1-2			Predicted		
			Control	Induction	Classification Error
	Observed	Control	12	1	0.08
		Induction	0	12	0.00
					OOB: 0.04
PC 1-3			Predicted		
			Control	Induction	Classification Error
	Observed	Control	12	0	0.00
		Induction	0	13	0.00
					OOB: 0.00
PC 1-4			Predicted		
			Control	Induction	Classification Error
	Observed	Control	12	0	0.00
		Induction	0	13	0.00
					OOB: 0.00

Table A15: Shows OOB for the SVC using PCs 1-2, 1-3 and 1-4 for *H. nuttallii*

H. petiolaris

Models:					
PC 1-2			Predicted		
			Control	Induction	Classification Error
	Observed	Control	6	1	0.17
		Induction	0	9	0.00
					OOB: 0.06
PC 1-3			Predicted		
			Control	Induction	Classification Error
	Observed	Control	6	0	0.00
		Induction	0	9	0.00
					OOB: 0.00
PC 1-4			Predicted		
			Control	Induction	Classification Error
	Observed	Control	6	0	0.00
		Induction	0	9	0.00
					OOB: 0.00

Table A16: Shows OOB for the SVC using PCs 1-2, 1-3 and 1-4 for *H. petiolaris*

H. salicifolius

Models:					
PC 1-2			Predicted		
			Control	Induction	Classification Error
	Observed	Control	1	0	0.00
		Induction	5	9	0.55
					OOB: 0.33
PC 1-3			Predicted		
			Control	Induction	Classification Error
	Observed	Control	1	0	0.00
		Induction	5	9	0.55
					OOB: 0.33
PC 1-4			Predicted		
			Control	Induction	Classification Error
	Observed	Control	3	0	0.00
		Induction	3	9	0.33
					OOB: 0.20

Table A17: Shows OOB for the SVC using PCs 1-2, 1-3 and 1-4 for *H. salicifolius*

H. lacianatus

Models:					
PC 1-2			Predicted		
			Control	Induction	Classification Error
	Observed	Control	9	1	0.11
		Induction	0	5	0.00
					OOB: 0.07
PC 1-3			Predicted		
			Control	Induction	Classification Error
	Observed	Control	9	0	0.00
		Induction	0	6	0.00
					OOB: 0.00
PC 1-4			Predicted		
			Control	Induction	Classification Error
	Observed	Control	9	0	0.00
		Induction	0	6	0.00
					OOB: 0.00

Table A18: Shows OOB for the SVC using PCs 1-2, 1-3 and 1-4 for *H. laciantaus*

LITERATURE CITED

Adler LS, Schmitt J, Bowers MD. 1995. Genetic variation in defensive chemistry in *Plantago lanceolate(*Plantaginaceae) and its effect on the specialist herbivore *Junionia coenia*(Nymphalidae).*Oecologia* 101: 75-85.

Bebber PD, Holmes T, Smith D, Gurr SJ. 2014.Economic and physical determinants of the global distributions of crop pests and pathogens. *New Phytologist* 202(3): 901-910

Challis GL and Hopwood DA. 2003. Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by *Streptomyces* species. *Proceedings of the NationalAcademy of Sciences of the United States of America* 100: 14555-14561.

Choudhary DK, Prakash A, Johri BN. 2007. Induced systemic resistance (ISR) in plants: mechanism of action. *Indian Journal of Microbiology* 47: 289-297.

Constans A. 2000. Taking It Higher. *The Scientist.https://www.the-scientist.com/technology profile/taking-it-higher-55417*

Couture JJ, Singh A , Rubert-Nason KF , Serbin SP, Lindroth RL, Townsend PA. 2016. Spectroscopic determination of ecologically relevant plant secondary metabolites. *Methods in Ecology and Evolution* 7: 1402–1412.

Croteau R, Kutchan TM, Lewis NG. 2000. Natural products (secondary metabolites*). Biochemistry andMoleculary Biology of Plants* 1250-1318.

Dixon RA and Paiva NL. 1995. Stress-Induced Phenylpropanoid Metabolism. *The Plant Cell* 7: 1085-1097.

Emwas AM. 2015. The Strengths and Weaknesses of NMR Spectroscopy and Mass Spectrometry with Particular Focus on Metabolomics Research.*Metabonomics*161-193.

Forseth RR and Schroeder FC. 2010. NMR-spectroscopic analysis of mixtures: from structure to function. *Current Opinion in Chemical Biology*15(1): 38–47.

Grebeand SK and Singh RJ. 2011. LC-MS/MS in the Clinical Laboratory – Where to From Here?*The Clinical Biochemist Reviews*32(1): 5–31.

Hartmann T. 2007. From waste products to ecochemicals: Fifty years research of plant secondary metabolism. *Phytochemistry*68: 2831-2846.

Jones CG and Firn RD. 1991. On the evolution of plant secondary chemical diversity. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 333: 273-28.

Jolliffe T and Cadima J. 2016. Principal component analysis: a review and recent developments. *Philosophical Transactions A*374(2065): 20150202.

Lange MB, Rujan T, Martin W, Croteau R. 2000. Isoprenoid Biosynthesis: The evolution of two ancient and distinct pathways across genomes. *Proceedings of the National Academy of Sciences of theUnited States of America* 97(24): 13172–13177.

Mason CM, Bowsher AW, Crowell BL, Celoy RM, Tsai C, Donovan LA. 2015. Macroevolution of leaf defenses and secondary metabolites across the genus Helianthus. *New Phytologist* 209: 1720-1733

Moore BD, Andrew RL, Kulheim C, Foley WJ. 2013. Explaining intraspecific diversity in plant secondary metabolites in an ecological context. *The New Phytologist* 201: 733-750.

Richardson AD, Duigan SP, Berlyn GP. 2002. An evaluation of noninvasive methods to estimate foliar chlorophyll content. *New Phytologist*1(153): 185-194.

Road K and Dun D. 1989.Spectral reflectance characteristics of vegetation and their use in estimating productive potential. *Plant Science.* 99(1): 59-81

Shihong Y, Ping L,Peiyi H. 2003. SVM classification: Its contents and challenges. *Applied Mathematics-A Journal of Chinese Universities*18(3): 332–34.

Skoric D. 2016. Sunflower breeding for resistance to abiotic and biotic stresses. *InTech* 585-635.

Stamp N. 2003. Out of the quagmire of plant defense hypotheses. *The Quarterly Review of Biology*78:23-50.

Speed MP, Fenton A, Jones MG, Ruxton GD, Brockhurst MA. 2015. Coevultion can explain defensive secondary metabolite diversity in plants. *New Phytologist* 208: 1251-1263.

Tankeu S, Senguttuvan J, Paulsamy S, KarthikaK. 2014.Phytochemical analysis and evaluation of leaf and root parts of the medicinal herb, Hypochaerisradicata L. for in vitro antioxidant activities. *Phytochemical Analysis.* 25(1): 81-84

Theis N, Lerdau M. 2003. The evolution of function in plant secondary metabolites. *International Journal of Plant Sciences* 164: S93-S102.

War A, Paulraj MG, Ahmad T, Buhroo AA, Hussain B, Ignacimuthu S, Sharma HC.2012. Mechanisms of plant defense against insect herbivores. *Plant Signaling and Behavior.* 7(10): 1306-1320

Wink M. 1987. Plant Breeding: importance of secondary metabolites for protection against pathogens and herbivores. *Theoretical an Applied Genetics* 75: 225-233.