University of Arkansas, Fayetteville
ScholarWorks@UARK

Graduate Theses and Dissertations

12-2012

Taxonomy and DNA Barcoding in the Genus Manfreda (Salisb.) Asparagaceae

William David Ritchie
University of Arkansas, Fayetteville

Follow this and additional works at: https://scholarworks.uark.edu/etd

 Part of the Horticulture Commons, Plant Biology Commons, and the Plant Breeding and Genetics Commons

Citation

Ritchie, W. D. (2012). Taxonomy and DNA Barcoding in the Genus Manfreda (Salisb.) Asparagaceae. *Graduate Theses and Dissertations* Retrieved from https://scholarworks.uark.edu/etd/576

This Thesis is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of ScholarWorks@UARK. For more information, please contact scholar@uark.edu.

TAXONOMY AND DNA BARCODING IN THE GENUS *MANFREDA* SALISB. (ASPARAGACEAE)

TAXONOMY AND DNA BARCODING IN THE GENUS *MANFREDA* SALISB. (ASPARAGACEAE)

A thesis submitted in partial fulfilment
of the requirements for the degree of
Master of Science in Horticulture

By

William David Ritchie
Royal Botanic Garden Edinburgh
Bachelor of Science in Horticulture, 2010

December 2012
University of Arkansas

ABSTRACT

The genus *Manfreda* Salisb. of Asparagaceae is a genus of potential horticultural interest and is currently subject to breeding efforts at the University of Arkansas. A lack of taxonomic clarity however undermines the classification of potential inter - and intrageneric hybrids. The study aims to assess existing species delimitation within the genus *Manfreda* employing morphology while investigating the potential utility of Consortium for the Barcode of Life Plant (CBOL) DNA Barcodes for identification of specific taxa and an External Transcribed Spacer (ETS) - Internal Transcribed Spacer (ITS) DNA barcode for developed hybridized taxa.

Observation of 855 herbarium specimens facilitated phylogenetic and Principal Component Analysis of morphology. Phylogenetic analysis employing Maximum Parsimony and Bayesian techniques of qualitative characters failed to identify any interspecific relationships with sufficient confidence. Principal Component Analysis identified 14 species exhibiting uniformity in categorical characters. The residual seven were subjected to further review employing existing literature, biogeographical and morphological data from herbarium specimens. The study supported specific designations of 19 of 21 species studied. A proposal for *Manfreda pubescens* (Regel & Ortgies) Verh.-Will. ex Espejo & López-Ferr. to be relegated to a varietal rank of *Manfreda maculata* (Mart) Rose was concluded based on a lack of consistent derived characters as well as biogeographical and ecological continuity.

Utility of the CBOL Plant DNA Barcode for identification of *Manfreda* species was investigated employing phylogenetic and nucleotide networking techniques. The CBOL Plant DNA Barcode failed to identify any interspecific relationships via Maximum Parsimony or Bayesian techniques. Sufficient variation however was available for differentiation of each species of *Manfreda* via composition of a nucleotide network map. Results allude to minimal

divergence between species of *Manfreda*, yet sufficient derived characters for functionality of CBOL Plant DNA Barcodes.

Analysis of an ETS - ITS DNA barcode for identification of intergeneric hybridized taxa of *Agave* L., *Manfreda* and *Polianthes* L. could not be conducted due to inefficiencies in DNA amplification techniques. The ETS gene region could not be amplified, a trial of three different amplification parameters was conducted and a lack of appropriate PCR primers was identified as the cause of no amplification. Amplification of the ITS gene region was successfully achieved; however, subsequent analysis of the electropherogram alluded to intra-individual polymorphisms within the genome. Therefore,it was concluded that the utility of the ITS region is negligible for DNA barcoding of the maag 01-07-13, mapo 01-04-07 and mapo 05-04-02 intergeneric hybrids.

This thesis is approved for recommendation

to the Graduate Council.

Thesis Director:

_____

Dr. Jon Lindstrom

Thesis Committee:

_____

Dr. Brad Murphy

_____

Dr. John Clark

_____

Dr. Ashley Dowling

THESIS DUPLICATION RELEASE

I hereby authorize the University of Arkansas Libraries to duplicate this thesis when needed for research and/or scholarship.

Agreed _____
William David Ritchie

Refused _____
William David Ritchie

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Jon T. Lindstrom for his mentorship throughout this project and for allowing me to work under his tutelage. I would also like to thank my committee members, Dr. John R. Clark, Dr. J. Brad Murphy and Dr. Ashley P.G. Dowling for their assistance and guidance. Finally, I would also like to thank Dr. Johnnie L. Gentry Jr. who facilitated study at the University of Arkansas Herbarium.

DEDICATION

This thesis is dedicated to those who continue to explain and explore the world of plants in Arkansas, through botanical gardens, herbaria and academia.

TABLE OF CONTENTS

Page

Introduction....1
Literature Review....8
Morphological Systematic Research....8
Morphological Characters....15
Roots....15
Leaves....15
Stem....20
Inflorescence....20
Flowers....21
Fruit....24
Seeds....24
Principal Component Analysis....25
Molecular Systematic Research....27
Molecular Systematic Study of Asparagales Link. and Asparagaceae Juss....28
Molecular Systematic Study of Agavoideae and *Manfreda* Salisb....31
Phylogenetic Utility of *rbcL* Gene Region....33
Phylogenetic Utility of *MatkK* Gene Region....34
Maximum Parsimony Analysis....35
Bayesian Analysis....35
Nucleotide Network Analysis....38
DNA Barcoding Research....40
Sufficient Size....41
Taxon Discrimination....41
Routine Amplification....42
Universal DNA Barcoding....43
Selection of Appropriate Gene Region(s)....45
Selection of CBOL Plant DNA Barcoding Regions....47

DNA Barcoding for Horticulture 48
Potential DNA Barcoding Regions for Horticultural Application 52
Internal Transcribed Spacer (ITS) 52
External Transcribed Spacer (ETS) 54
Literature Cited 57
Chapter 1: Study of Specific Delimitation for *Manfreda* Salisb. (Asparagaceae) inferred from Principal Component, Phylogenetic Analysis of Morphology and Geographical Data 72
Introduction 72
Materials and Methods 74
Results 82
Discussion 155
Glossary of Terms 168
Literature Cited 170
Chapter 2: Assesment of CBOL Plant DNA Barcodes for Phylogenetic Research and DNA Barcoding in the Genus *Manfreda* Salisb. (Asparagaceae) 172
Introduction 172
Materials and Methods 175
Results 183
Discussion 190
Glossary of Terms 194
Literature Cited 196
Chapter 3: Development of a DNA Barcode for Identification of intergeneric hybrids between *Agave* L., *Manfreda* Salisb. and *Polianthes* L. (Asparagaceae) based on nuclear rDNA Internal and External Transcribed Spacers 199
Introduction 199
Materials and Method 202
Results - ITS 210
Results - ETS 212
Discussion 215
Literature Cited 218
Conclusions 221
Appendix 1 223

Appendix 2.....226
Appendix 3.....230
Appendix 4.....231
Appendix 5.....233

LIST OF TABLES

Page

Table 1. Gene regions surveyed by Dr. Salazar (National Autonomous University of Mexico) for barcoding utility in *Agave* L. and associated taxa including *Manfreda* Salisb. (Salazar, personal communication). 33

Table 2. Summary of analysis conducted for seven potential barcoding regions by the Plant Working Group (2009). 48

Table 3. List of herbaria from Europe, Mexico and the United States from which specimens of *Manfreda* were observed for morphometric analysis. 75

Table 4. List of specific taxonomic units employed in the study and of which herbarium specimens were observed. Taxa denoted with (*) did not have sufficient representative samples to be included in the Principal Component Analysis. 76

Table 5. List of the nine quantitative characters employed for morphometric study via Principal Component Analysis of *Manfreda* herbarium specimens. 77

Table 6. List of categorical characters employed in the phylogenetic analysis via Maximum Parsimony and Bayesian Analysis of herbaria specimens of *Manfreda*. 79

Table 7. The Loading Matrix exhibits the load attributed to each principal component by each of the nine quantitative characters during Principal Component Analysis. 82

Table 8. List of specimens employed in phylogenetic analysis employing Maximum Parsimony and Bayesian Analysis. Nineteen specimens of *Manfreda* and three outgroups were utilized. 176

Table 9. Primers employed in PCR amplification of the *MatK* and *rbcL* plastid gene regions, as suggested by the Consortium for the Barcode of Life's Plant Working Group (Hollingsworth, 2009). 179

Table 10. Cycling conditions employed in the PCR reaction for amplification of the *MatK* and *rbcL* plastid gene regions as suggested by the Consortium for the Barcode of Life's Plant Working Group (Hollingsworth, 2009). 180

Table 11. Settings altered for parameters and priors from default for implementation of Bayesian Analysis via MrBayes 3.1 for phylogenetic analysis. 182

Table 12. Outline of distrubutions, phenology and hybridization of the six taxa allocated to an clade independant of all other taxa of *Manfreda* when subjected to Maximum Parsimony and Bayesian Analysis employing the *MatK* and *rbcL* gene regions. 192

Table 13. Primers employed for amplification of the ITS gene region via PCR for *Manfreda* cultivars; mapo 01-04-07, mapo 05-04-02 and maag 01-07-13. 206

Table 14. PCR amplification parameters for amplification of the ITS gene region for *Manfreda* cultivars; mapo 01-04-07, mapo 05-04-02 and maag 01-07-13. 206

Table 15. List of primers employed in the ETS Amplification Trial for nine parameters employing *Manfreda* cultivars; mapo 01-04-07, mapo 05-04-02 and maag 01-07-13 and *Rudbeckia hirta* 'Prairie Sun'. 209

Table 16. Standard PCR amplification parameters for amplification of the ETS gene region of *Manfreda* cultivars; mapo 01-04-07, mapo 05-04-02 and maag 01-07-13 and *Rudbeckia hirta* 'Prairie Sun'. 209

Table 17. Table of the three different parameters for each of the three experiments included in the ETS Amplification Trial. 210

LIST OF FIGURES

Page

Figure 1. Diagram of states in which *Manfreda* species have been documented adapted from Verhoek-Williams (1975) and Castillejos-Cruz (2009) encompassing El Salvador, Guatemala, Honduras, Mexico and the United States of America. 3

Figure 2. The Loading Plot illustrates the directionality and influence of the nine quantitative characters in the resultant score plot of Principal Component Analysis. 84

Figure 3. The Scree Plot demonstrates the allocation of eigenvalues via Principal Component Analysis to each principal component in a line graph. 85

Figure 4. Eigenvalues are a mathematical description of the amount of variability hosted by each principal component. Eigenvalues are attributed to each principal component by conducting Principal Component Analysis. 86

Figure 5. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda chamelensis*. 88

Figure 6. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda chamelensis* are highlighted by dark blue. 89

Figure 7. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda elongata*. 91

Figure 8. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda elongata* are highlighted by brown. 92

Figure 9. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda guttata*. 94

Figure 10. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda guttata* are highlighted by purple. 95

Figure 11. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda involuta*. 97

Figure 12. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda involuta* are highlighted by turquoise. 98

Figure 13. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda jaliscana.* 100

Figure 14. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda jaliscana* are highlighted by purple. 101

Figure 15. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda littoralis*. 103

Figure 16. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda littoralis* are highlighted by green. 104

Figure 17. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda maculata*. 106

Figure 18. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda maculata* are highlighted by dark green. 107

Figure 19. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda maculosa*. 109

Figure 20. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda maculosa* are highlighted by green. 110

Figure 21. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda nanchititlensis.* 112

Figure 22. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda nanchititlensis* are highlighted by navy blue. 113

Figure 23. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda parva.* 115

Figure 24. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda parva* are highlighted by purple. 116

Figure 25. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda planifolia.* 118

Figure 26. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda planifolia* are highlighted by brown. 119

Figure 27. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda potosina.* 121

Figure 28. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda potosina* are highlighted by dark green. 122

Figure 29. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda pringlei.* 124

Figure 30. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda pringlei* are highlighted by green. 125

Figure 31. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda pubescens.* 127

Figure 32. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda pubescens* are highlighted by purple. 128

Figure 33. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda revoluta.* 130

Figure 34. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda revoluta* are highlighted by purple. 131

Figure 35. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda rubescens*. 133

Figure 36. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda rubescens* are highlighted by brown. 134

Figure 37. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda scabra*. 136

Figure 38. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda scabra* are highlighted by green. 137

Figure 39. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda singuliflora.* 139

Figure 40. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda singuliflora* are highlighted by blue. 140

Figure 41. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda umbrophila.* 142

Figure 42. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda umbrophila* are highlighted by purple. 143

Figure 43. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda variegata.* 145

Figure 44. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda variegata* are highlighted by brown. 146

Figure 45. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda virginica.* 148

Figure 46. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda virginica* are highlighted by blue. 149

Figure 47. Maximum Parsimony cladogram (50% Majority Rule) representing selected quantitative morphological characters of *Manfreda* and selected outgroups. Bootstrap values in red denote scores below 50 and values in back denote scores above 50. 151

Figure 48. Bayesian analysis cladogram representing quantitative morphological characters of *Manfreda* and selected outgroups. Only Posterior Probabilities in excess of 50 are displayed. 153

Figure 49. Distribution of all 855 herbarium specimens observed during the study with associated legend. 154

Figure 50. Cladogram generated employing TPM1uf+1 model of nucleotide evolution for Bayesian Analysis of the *MatK* and *rbcL* gene regions. 184

Figure 51. Cladogram generated employing F81 model of nucleotide evolution for Bayesian Analysis of the *MatK* and *rbcL* gene regions.. 185

Figure 52. Cladogram generated (50% Majority Rule) employing Maximum Parsimony for phylogenetic analysis of the *MatK* and *rbcL* gene regions.. 187

Figure 53. Nucleotide network exhibiting nucleotide divergence between specimens of *Manfreda* and selected outgroups based on *MatK* and *rbcL* gene sequences. 189

Figure 54. Gel electrophoresis of genomic DNA extract for the three intergeneric cultivars: mapo 01-04-07 (Lane 2), mapo 05-04-02 (Lane 3) and maag 01-07-13 (Lane 4). 205

Figure 55. Gel electrophoresis of ITS amplicons for intergeneric cultivars: mapo 01-04-07 (Lane 2), mapo 05-04-02 (Lane 3) and maag 01-07-13 (Lane 4). 207

Figure 56. Example of an electropherogram generated for mapo 01-04-07 demonstrating low signal strength and intra-individual polymorphisms between ITS copies. 211

Figure 57. PCR amplification trial of nine alternative parameters conducted for *Manfreda* cultivars; mapo 01-04-07, mapo 05-04-02 and maag 01-07-13 and *Rudbeckia hirta* 'Prairie Sun' (Continued). 214

INTRODUCTION

The genus *Manfreda* Salisb. of the family Asparagaceae is a taxon within the order Asparagales (APG, 2009; Govaerts, 2011). These monocotyledonous angiosperms are characterized and differentiated from other members of Asparagales by possessing an inferior ovary, subterranean stems and a flexible leaf apex (Verhoek, 2002). The genus was first described by Richard Anthony Salisbury FRS in his publication Genera of plants, a fragment, containing a part of Liriogamæ (1866); the use of *Manfreda* as a generic rank was however not widely employed until Rose (1905).

The modern circumscription for the genus is thought to consist of 26 species according to the most comprehensive account of the genus to date authored by Susan Verhoek-Williams (1975). An alternative and more modern system, that accounts for the discovery of new species, is a 32 species-based treatment authored by Castillejos-Cruz (2009). This treatment however has yet to gain prominence outside of Mexico. The genus has an extensive distribution across Central and North America with southerly populations inhabiting Guatemala, Honduras and El Salvador while northerly populations are found in the northeastern U.S. (Figure 1). The center of species diversity is located in the Mexican states of Chihuahua, Jalisco and San Luis Potosí (Verhoek-Williams, 1975).

Manfreda and allied taxa of the former Agavaceae are endemic to the American continent (Garcia-Mendoza, 2002). Both *Agave* L. and members of tribe Poliantheae have origins located in Central Mexico. As such, approximately 75% of the aforementioned taxa are located in Mexico with as much as 69% endemic (Rose, 1905; Garcia-Mendoza, 2002). The extent of the

native range of *Manfreda* extends far beyond Mexico however, southerly populations have been documented in Guatemala, Honduras and El Salvador while the native range of *M. virginica* (L.) Salisb ex Rose, in particular extends the northern range extensively (Verhoek, 1998; Irish and Irish, 2000). *Manfreda virginica* with its ability to withstand temperatures of -28°C in its dormant period has successfully established along the Atlantic seaboard of the United States. Although other species of *Manfreda* are present in southeastern U.S. such as Texas, Louisiana and Florida, *M. virginica* has populated more northerly states such as Ohio and West Virginia utilizing its ability to withstand severe winter temperatures (Verhoek-Williams, 1975; Verhoek, 1998).

The biogeography of *Manfreda* and sister taxa, *Polianthes* L. and *Prochnyanthes* S.Watson may have played a significant role in speciation (Verhoek-Williams, 1975). Studies into the hybridization potential of the aforementioned species show few barriers to hybridization with intergeneric crosses a commonality in breeding programs (Verhoek, 1975; Lindstrom, 2006). Despite highly conserved and uniform karyology facilitating such crosses, occurrence in the wild is rare (Verhoek-Williams, 1975; Bogler *et al.*, 2006). The key reason is adaption and inheritance of differing environs, leading to the evolution of morphologically distinct populations. To illustrate, some species of *Manfreda* with thin, deciduous leaves inhabit regions with higher rainfall and a cooler climate as its biomass can be reduced significantly via abscission of leaves in the winter for protection. *Agave*, with its characteristic succulence, is better adapted to dry desert conditions, and as an evergreen is more suited to warmer climes as leaf damage at freezing point may be substantial. Although a gradient towards the succulent characteristics of *Agave* exists within *Manfreda* and other members of tribe Poliantheae, distinct populations remain, facilitating further diversification. The exclusivity of breeding populations at

present is effected significantly by physical terrain and geographical distance and thus plays a notable role in the evolution within the Agavoideae complex (Verhoek-Williams, 1975).

Manfreda inhabits a range of ecological conditions from *Pinus* L. spp. and *Quercus* L. spp. - dominated forests to volcanic scree and shrubland (Verhoek, 1998). Members of *Agave* and *Prochnyanthes* have been documented to grow at elevations up to 3400 m but records of altitude for *Manfreda* have been somewhat more conservative (Garcia-Mendoza, 2002). Species from tribe Poliantheae possess the ability to inhabit areas of higher rainfall with populations frequently documented in vegetated woodlands, grasslands, riverbanks and wet meadows (Irish and Irish, 2000).

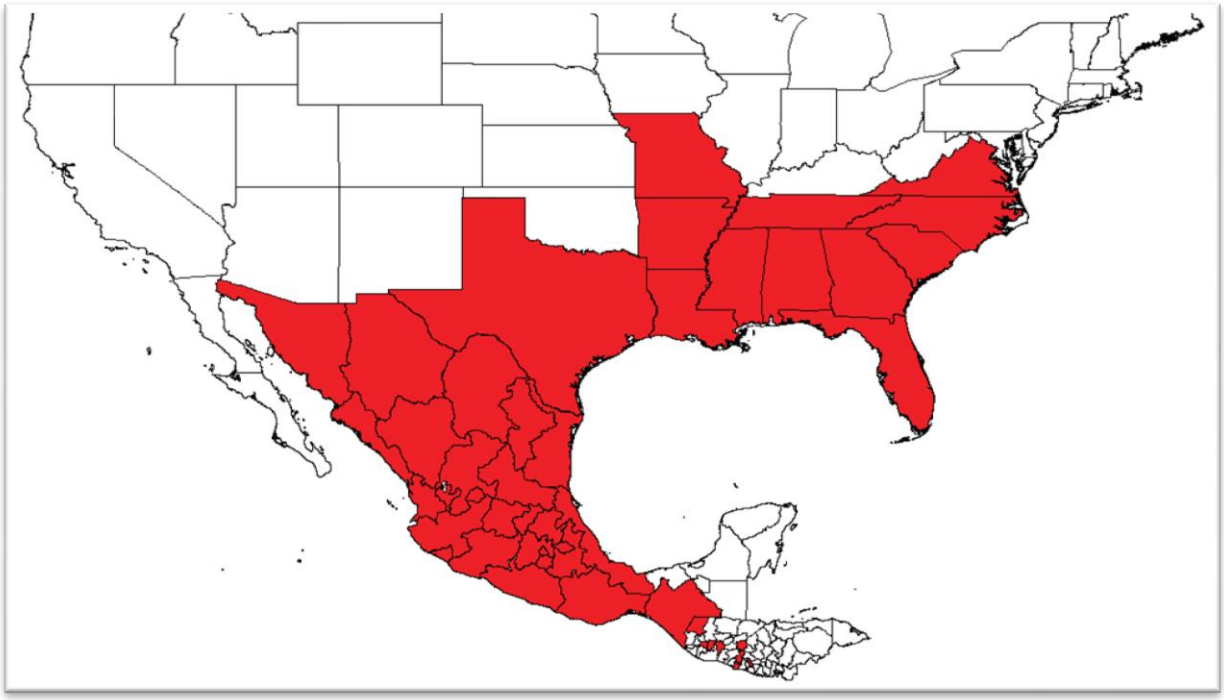


Figure 1. Diagram of states in which *Manfreda* species have been documented adapted from Verhoek-Williams (1975) and Castillejos-Cruz (2009) encompassing El Salvador, Guatemala, Honduras, Mexico and the U.S.

At present, the widest utilization of *Manfreda* in modern culture is in horticulture. The genus comprises a small sector of the market and is propagated as an ornamental crop. Propagation is primarily conducted from seeds, that closely resemble those of other Agavoid species and are thought to retain a period of viability exceeding five years (Irish and Irish, 2000; Cave, 2003). Seeds can be sown in perlite-rich soil mixes to maximize drainage. Irish and Irish (2000) preferred a mix of equal measures of perlite and vermiculite with a small amount of additional organic matter. Optimum growth has been observed in temperatures exceeding 27°C with ontogenesis yielding succulent leaves comprising an obvious rosette form from early juvenility. After the first year, development continues with increasing vigour and flowers typically occur after three years (Irish and Irish, 2000; Howard, 2001).

Specimens of *Manfreda* collected from the wild inhabit localities ranging from deep shade to full sun, demonstrating the versatility of the genus (Verhoek-Williams, 1975). The optimum conditions for cultivated species is thought to be moderate shade (Howard, 2001). Some authors also state that locations in the garden in direct sunlight are applicable but scorching of the leaves can occur (Irish, 2002; Hannon, 2002). Due to the extensive rhizomatic root system, of *Manfreda* lends itself more amenably to landscape than container planting, however vessels such as wine casks have been utilized in an aesthetically pleasing manner (Hannon, 2002). For such plantings a gravel soil dressing has also been recommended as it is thought to emphasize the rosette form and suppress competitive weeds (Ogden, 1994).

In cultivation the growing medium for *Manfreda* should reflect free draining conditions in which they commonly inhabit in the wild, and as such porous soil is essential (Irish and Irish, 2000). *Manfreda* specimens are capable of tolerating a wide range of soil including rich garden loams; the key however is good drainage rather than nutrient availability (Hannon, 2002; Irish,

2002). Authors either suggest for specimens a dry location in the garden or exceptional drainage (Howard, 2001; Irish, 2002). Despite the xeric habits exhibited by *Manfreda,* for substantial growth and development regular watering is required during the active growing season. *Manfreda* species can tolerate extensive watering of four to six summer days out of the week but can also go without water for up to three weeks in moderate summer periods. Fertilizer can also enhance growth but sparse application is necessary. Application of a nitrogen low/potassium high fertilizer twice a year is sufficient to sustain substantial growth (Irish and Irish, 2000). Of garden-grown specimens the pest and disease problems are few and root-knot nematode and deer grazing have been noted as the most significant threats (Hannon, 2002).

Manfreda species are best acclimated to dry and cool winters (Hannon, 2002). Most *Manfreda* specimens available commercially in the United States have a tolerance of below freezing temperatures, incurring damage at -12°C and death at -24°C. The exception is *M. virginica* which, due to the breadth of its native range, can withstand temperatures as low as -34°C (Irish and Irish, 2000). As such, *M. virginica* is included in much of the breeding efforts in progress to increase winter hardiness of hybrids (Lindstrom, 2006). The utility of *Manfreda* for landscape planting is furthered by the ability to uproot the specimen over winter to protect it from freezing conditions and thus extend the possible range for use in a garden (Hannon, 2002).

In the horticultural industry, the varied leaf morphology and garden hardiness of *Manfreda* have evoked an increased interest in the genus for commercial cultivation (Hannon, 2002). The mottling of the leaves has been of particular interest to consumers as well as the sequential changing of floral colors (Howard, 2001). Furthermore, flowering times of *Manfreda* in the wild range throughout the year varying, from only three species flowering in March to 12 species flowering in July, August and September (Rodríguez and Castro, 2006). The limiting

factors for utilization of *Manfreda* for horticultural interest however are attributed to floral aesthetics. A wide diversity of fragrances are present in *Manfreda,* some pleasant, and some less so (Hannon, 2002). The flowers that resemble those of *Agave* have also been attributed as a particular weakness with an aesthetic deemed to have limited commercial value (Lindstrom, University of Arkansas, personal communication). To overcome such problems breeding efforts aimed at intergeneric crosses with *Polianthes,* a close relative with more aesthetically pleasing floral character, have been undertaken (Verhoek-Williams, 1975, Lindstrom, 2006).

Breeding efforts to gain the desirable traits from *Manfreda* and *Polianthes* species have been sparse. The first documentation of a breeding program involving either of the two genera was from an anonymous account dated to 1899, that noted the successful crossing of *Polianthes geminiflora* (Lex) Rose and *Prochnyanthes mexicana* (Zucc.) Rose. This intergeneric cross was followed by Worsely (1911), who achieved a number of crosses within the genus *Polianthes* itself (Verhoek, 1975). Karyological work conducted by McKelvey and Sax (1933) and Sâto (1935) illustrated a highly uniform chromosomal complement in the former family Agavaceae and highlighted an extremely close affinity between the species that are included in the tribe Poliantheae (Bogler *et al.*, 2006).

The close chromosomal complement observed in *Manfreda* suggested a close relationship between the genera and that intergeneric breeding could be feasible. In light of this evidence Verhoek-Williams (1975) conducted an extensive breeding program including intergeneric and intrageneric crosses. She achieved viable crosses were achieved between *Manfreda* and *Polianthes*. One of the crosses produced by Verhoek was between *M. virginica* and *Polianthes tuberosa* L., this cross demonstrated improved aesthetic qualities, increased cold tolerance and as such is discussed favorably by a number of horticultural publications (Irish and Irish, 2000;

Howard, 2001, Lindstrom, 2006). Outside of academia Verhoek-Williams (1975), also noted that cut-flower producers were creating similar crosses, one such example was Mr. F. Meyer of Escondido, Calif. whose family cut-flower business established and utilized a number of *Manfreda* and *Polianthes*-based crosses (Verhoek-Williams, 1975).

Contemporary breeding work regarding *Manfreda* and *Polianthes* began at the University of Arkansas in 2003 led by Dr. J Lindstrom. Utilizing plant materials acquired from Yucca Do Nursery in Giddings, Tex. and from Pine Ridge Nursery in London, Ark., a number of crosses have been made. As of 2006, 15 crosses have been successfully achieved with eight retaining viable seeds (Lindstrom, 2006).

LITERATURE REVIEW

Morphological Systematic Research

The early explorations of the Americas by Sahagun (1571) and Hernandez (1571, 1576) were the first to document *Manfreda* and allied taxa in tribe Poliantheae (Verhoek-Williams, 1975). Despite the rich descriptions and ethnobotanical accounts, little taxonomic work was conducted by European botanists from the samples collected and descriptive accounts documented. The allied taxa *Polianthes* was the first to be introduced to Europe, as it was known in Aztec cultivation for fragrance. Taxonomic descriptions later occured however in herbals and texts such as Anonymous (1601), Bauhin (1623) and Parkinson (1629) based on cultivated species received by Parkinson in 1594 (Trueblood, 1973). Linnaeus, in *Hortus Cliffortianus* (1738), was the first to differentiate the modern *Polianthes* from *Hyacinthus* L. by naming the type species *Polyanthes floribus alternis* L. and thus engaging a greater interest in *Polianthes* and allied specimens such as those that are currently regarded as *Manfreda* (Verhoek-Williams, 1975).

The greater interest surrounding the specimens of *P. tuberosa* also facilitated further enquiry into *Manfreda virginica*, a native of the American Southeast. Grovinus was sent specimens of such by John Clayton during the composition of *Flora virginica, Part 2* (1743). Grovinus assigned the modern *M. virginica* to *Aloe* L., however this classification was shortly superseded. Based on the same specimen sent to Grovinus a decade prior, Linnaeus saw a much closer resemblance to *Agave* and thus reclassified the specimen as *Agave virginica*, a classification still utilized on occasion in modern classification systems (Verhoek-Williams, 1975; Garcia-Moya *et al.*, 2011).

In the latter half of the 18th century, the classification of *A. virginica* was commonly utilized and infrequently challenged. The higher orders of the taxa were changing, however instigating theories and taxonomic investigations into the evolution of the genus itself. The work of De Jussieu (1781) was seminal and remains highly influential today. The order (equivalent of family in contemporary taxonomic systems) Lilia and Bromilae were both employed for members of the modern subfamily Agavoideae. This classification divided the taxon based on the ovary position with *Yucca* L. assigned to Lilia and *Agave* assigned to Bromilae. The principals of this system were observed until Hutchinson (1934), who retained the core species of *Agave* and tribe Poliantheae (Bogler *et al.*, 2006). Enlicher (1841) observed and understood this close relationship; his order Agaveae, the precursor to Agavaceae and latterly subfamily Agavoideae, was centered around these taxa and also included *Furcraea* Vent. and the now defunct genus of *Littaea* Tagl. (Verhoek-Williams, 1975; Bogler *et al.*, 2006).

The first distinction between the modern *Manfreda* species and *Agave* occurred in Brown (1850). His subgeneric system, based on simple or branching inflorescences, naturally grouped *Manfreda* specimens due to their simple spike or raceme form. Similarly research conducted at the Komarov Botanical Institute in St. Petersburg, Russia by Regal (1858) resulted in the publication of a subgeneric system for *Agave* composed of similar groups. The characters employed were poorly documented but within the system, and *Manfreda* specimens were united within one of the subgeneric units. Subsequently, Koch (1860), who disregarded the work of Regal (1858), also composed an eight taxa subgeneric system for *Agave*. All *Manfreda* specimens except one were included in the subfamily Herbaceae predominantly due to their herbaceous habit but also leaf and stem characters. The exception was *A. maculata* Regal. (the contemporary *M. maculata* (Hook.) Rose.), which was placed in the subfamily Canalicultatae, as

the leaves were deemed more herbaceous and differed in their margin and shape (Verhoek-Williams, 1975).

The alignment of herbaceous *Agave* taxa with simple raceme or spike inflorescences in repeated subgeneric systems warranted investigation and possible revision. An English botanist, Richard Anthony Salisbury FRS., undertook such work and utilized the Linnaean-type specimens to do so. Salisbury's contributions to horticulture and botany were significant with substantial works prior to his 1866 publication *Genera of Plants* (Elliott, 2004). It was in the aforementioned text that the genus *Manfreda* was conceived and a description was first published. The genus was named in honor of Manfredus de Monte (born ca. 1335), an Italian writer on issues of botanical and horticultural interest (Paris and Janick, 2008). The circumscription of the genus was based on the type specimen *M. virginica* and the limits of the genus were based on leaf, inflorescence and habit characters in a similar vein to the subgeneric systems proposed previously (Verhoek-Williams, 1975).

The generic system of Salisbury (1866) struggled for recognition and widespread adoption. His work was often undermined by his personal animosity toward contemporaries and bitter rivalries with major figures in European botanical circles (Verhoek-Williams, 1975; Brent, 2005). However, In North America his system suffered from a lack of exposure. Of those that were aware of his work, many questioned the system based limited plant material utilized and perceived limitations to his holistic understanding of Agavoid species. Many prominent systems were proposed later to Salisbury (1866) and most reverted to a generic system that included species of the modern *Manfreda* in *Agave* (Verhoek-Williams, 1975).

Baker (1877) utilized a taxonomic system based on inflorescences and keyed species by leaf texture in composition of a subgeneric system for *Agave*. Recognition for the work of Salisbury (1866) was included in the system, however, a subgeneric rank was assigned and included all specimens of the modern *Manfreda* observed in the study. This study was built upon a lesser known investigation by Engelmann (1859), both utilizing the three subgenera *Euagaveae*, *Littaea* and *Manfreda*. The work of Bentham and Hooker (1883) also accepted the subgeneric system and provided a wider audience for the system. The system of Bentham and Hooker was adapted at the family level, echoing the works of De Jussieu (1789) where position of the ovary was regarded as a highly diagnostic character. As such, *Agave* was placed with its three subgenera in Amarylilideae, while *Yucca* and other hypogynous Agavoid genera were placed in Liliaeae. Further works, such as Engler (1888) and Krause (1930), employed a near identical system (Verhoek-Williams, 1975).

In the years subsequent to Bentham and Hooker (1883), the confines of the subgeneric rank of *Manfreda* began to diversify and alter due to the discovery and description of new species within *Agave*. The description of the genus *Prochnyanthes* by Watson (1887) added substantial diversity to the family and thus the taxon required revision (Verhoek-Williams, 1975; Bogler and Simpson, 1996). Rose, while serving as Assistant Curator of the Smithsonian Institutes herbarium, reintroduced the generic rank of *Manfreda* and brought the work of Salisbury (1866) to the fore of the botanical community (Rose, 1905). Rose's contribution has stood the test of time as it formed the basis for tribe Poliantheae in contemporary taxonomy. During revision of Bentham and Hooker (1883), Rose addressed issues regarding the subgeneric rank of *Manfreda* by assigning full generic status to *Prochnyanthes*, merging *Bravoa* Lex. into *Polianthes* sensu lato and reinstating *Manfreda* to generic status. Rose (1905) went further to

suggest that all herbaceous Agavoid species should be included in the genus *Manfreda*. His confidence in the genus was resolute and confirmed such by stating "I am more strongly convinced than ever that *Manfreda* is generically distinct from *Agave* proper. It differs from *Agave* in its habit, manner of growth, foliage and inflorescence" (Rose, 1905; Verhoek-Williams, 1975).

At the dawn of the 20th century the taxonomy of *Agave*, *Manfreda*, *Polianthes* and *Prochnyanthes* and closely allied taxa remained debated taxonomically. Lotsy (1911) devised the family rank of Agavaceae and was the first usage of such to contain the allied taxa. The group was however still narrowly circumscribed and omitted hypogynous Agavoid species (Bogler *et al.,* 2006). Berger (1915) was one of the first botanists in Europe to address the Agavaceae taxon. He composed 274 monographs to complement previous works and aimed to bring clarity to the generic systems utilized. His system reunited *Agave* and *Manfreda* but did so under much hesitation. His misgivings were exemplified by a note attached to a *M. variegata* (Jacobi) Rose specimen that he observed during his studies. The note read "I again include *Manfreda* under *Agave* under much hesitation. When I was at Kew (Royal Botanic Garden Kew, London) in June 1913 they pressed me to include it under *Agave*" (Verhoek-Williams, 1975; Gentry, 1982).

In the initial half of the 20th century, Agavaceae was also subject to much revision. A substantial reclassification of Agavaceae was devised by John Hutchison, a renowned horticulturist, botanist and taxonomist (Hutchinson, 1934). His 1934 publication *Families o*f *Flowering Plants* devised a novel system for Agavaceae that took its habit into account alongside systematic characters. Characters previously held with the highest regard such as ovary position,

were relegated from apomorphic characters as it was thought evolution of such may have occurred independently throughout the family and that homoplasy was prevalent. Habitat was promoted to a character of high regard as it reflected their life history and evolutionary strategies. As such, *Yucca* was united with *Agave* for the first time within Agavaceae based on a xerophytic habit (Cronquist, 1981; Bogler *et al.*, 2006). The influence of Hutchinson (1934) is still eminent today, although Agavaceae has been relegated to a subfamily rank within Asparagaceae based on APG III the subfamily still comprises the genera proposed by Hutchinson (1934) with minor modifications (APG, 2009).

Hutchinson (1934) brought stability to Agavaceae, and few alternate systems were proposed in subsequent years. The limits of the family were robust and resolute. The internal composition of the family however remained much changed and constantly debated. Novel systems such as that of Shinner (1966) where *Manfreda* was sunk into *Polianthes* sensu lato were proposed yet widely disregarded (Shinner, 1966; Verhoek, 1975). The more prominent argument was whether species of *Manfreda* should be retained as a genus or included once more in *Agave* sensu lato. Work during the 1950s and 1960s was sparse and little resolution or increased confidence in taxonomic systems utilized at generic or family levels was inferred. Advancement of techniques in plant systematics were yielding changes in the higher orders, however. Huber (1969) employed 'microcharacters' such as seed coat morphology, cuticle form and embryonic ontogeny to aggregate the Agavaceae and allied families into one order (Bogler *et al.*, 2006). The Asparagoid order was a precursor to the order Asparagales proposed by Dahlgren *et al.* (1985).

Arthur Cronquist, one of the most influential botanists of the 20th century (New York Times, 1992), revisited the question of the generic composition assigned to Agavaceae in his

1981 work *An Integrated System of Classification of Flowering Plants* (Cronquist, 1981). Cronquist was a staunch advocate of creating taxonomic units that could be taught with ease and were appropriate for workers and students alike. As such, he continued with the work of Hutchinson (1934) and developed a broadly circumscribed family Agavaceae sensu lato as part of the Cronquist system. Cronquist, however, recognized the failings of the system, he noted himself that the system was overly dependent on the xeric habitat and without it, the taxonomic structure would collapse (Cronquist, 1981). The system was in response to the work of Armen Takhtajan (New York Times, 1993). The two debated the composition of Agavaceae. Takhtajan (1980) favored a proliferated system, made up of several smaller families centered around a core Agavaceae including *Agave* sensu lato (Takhtajan, 1980). The polarization between the two works epitomized the segregation of workers as to the composition of Agavaceae. The later work of Dahlgren *et al.* (1985) supported the system of Tahktajan (1980) and was highly influential and no other system in the latter decades of the addressed the issue with greater resolution.

The advent of molecular systematics originally supported Agavaceae as a functional taxonomic group as did early karyological evidence from McKevely and Sax (1933) and Sâto (1935). The original APG system (APG, 1998) did not modify or reclassify Agavaceae from that of Dahlgren *et al.* (1985). Many opponents of the system argued that the algorithms employed favored proliferation and adjustments were made accordingly in subsquent systems (APG, 2003). In the APG II and APG III system evidence suggested that Agavaceae should be merged with neighbouring taxa into a larger family unit, namely Asparagaceae (APG, 2003; APG, 2009). A morphological system to support the reclassification of the Asparagales was devised by Chase (2009) and the system is supported by many of the leading botanical institutes (RBG Kew Press, 2010).

Despite the uncertainty surrounding the family and generic status of the species under study, many of the leading floral works utilized the generic rank of *Manfreda* (i.e. Verhoek and Ness, 2002). Most of this work was based on a comprehensive dissertation composed by Dr. Susan Verhoek (then Verhoek-Williams) and subsequent works. Her 1975 dissertation outlined a categorical system for determination of *Agave* and *Manfreda* as well as allied taxa *Polianthes* and *Prochnyanthes*. Her work is held in high regard by many of the leading botanical institutes today and in the majority of major herbaria worldwide surveyed for this study *Manfreda* is filed independently of *Agave*.

Morphological Characters

Below is a review of the current literature regarding studied anatomical features of species that comprise the genus *Manfreda*.

Roots

The root system of *Manfreda* most commonly consists of fibrous filiform roots that are contractile. The exception to this are *M. hauniensis* (J.B. Petersen) Verh.-Will. and *M. longibracteata* Verh.-Will., that have stout fleshy roots (Verhoek, 1998; Castillejos-Cruz 2009). The diameter of the roots ranges from 0.2 mm to 0.5 mm (Castillejos-Cruz, 2009). This form of root system is common in monocotyledonous species and *Manfreda*, like *Agave*, form such roots in a prompt manner, developing radially and to a shallow depth (Irish and Irish, 2000).

Leaves

The leaves of *Manfreda* hold many key diagnostic characters and are especially important due to prolonged maturation period from seed to flowering (Verhoek-Williams, 1975). *Manfreda* have long, concave leaves, similar to that of *Aloe* (Cave, 2003). However, *Manfreda* possesses

leaves held in a rosette and the leaves themselves are thin, flexible and succulent, semi-succulent or non-succulent (Verhoek-Williams, 1975). The leaf blades are diverse in shape and are often dimorphic. Many forms from linear to oblanceolate have been observed, yet each leaf terminates in an acute non-pungent tip (Verhoek, 1998). Dentate teeth are present on the leaf margin of many species but are very small relative to those of *Agave* (Castillejos-Cruz, 2009). Apart from the obvious characters listed above, taxonomic studies revealed a suite of other characters, some resolute, which aided identification and taxonomy.

The number of leaves held by the rosette is variable between individuals, thus not a viable taxonomic character. The leaf arrangement is consistently spiral or alternate within each species (Verhoek and Ness, 2002).

The base of the leaf is attenuate or cuneate in all species. In *M. littoralis* García-Mend., A.Castañeda and S.Franco, *M. guerrerensis* Matuda. and *M. maculata* (Mart.) Rose, the attenuate form is very narrow and long and is occasionally described as pseudopetiolate (Castillejos-Cruz, 2009). The leaf shape itself is most commonly linear-lanceolate, however variation from this form is extensive. Dimorphism is commonly exhibited in many species and *M. virginica* is noted as being particularly inclined to such development. The leaf forms in species of *Manfreda* differ significantly. *Manfreda nanchititlensis* Matuda. possesses a linear leaf shape with a breadth of 2 mm at most, and at the alternate end of the spectrum *M. planifolia* (S.Watson) Rose possesses leaves that are almost orbicular, evolving from an oblanceolate formation (Verhoek, 1975; Castillejos-Cruz, 2009). Leaf shape, even though inconsistent between specimens, still yields significant taxonomic value, and especially so in the light of the lack of other consistent and informative taxonomic characters. The leaf shape divides the genus into a number of groups but

the evolutionary parameter that drove such diversification is largely unknown (Verhoek-Williams, 1998).

The tip of the leaf is highly diagnostic of the genus, yet highly consistent within the genus and thus offers little taxonomic insight at species level. Unlike *Agave*, *Hesperaloe* Engelm. and *Yucca*, which possess a lignified, pungent tip from early juvenility, *Manfreda* like *Polianthes*, *Prochnyanthes* and *Beschorneria* Kunth., has an acute but soft apex (Verhoek-Williams, 1975). The character is highly consistent and employed commonly to separate *Agave* from *Manfreda.* The only exception to this classification is *M. hauniensis* which has a sharp thickened tip, yet contains no lignin and is technically not spinose. Teeth on the margin are similarly diagnostic, most species in *Agave* possess large lignified teeth. In *Manfreda*, only very small cartilaginous teeth exist. A gradient is evident within *Manfreda* between small teeth and an entire margin, aiding separation of species based on marginal characters. The genus is also highly distinguishable from *Agave* based on this character (Castillejos-Cruz, 2009).

McVaugh (1989) argued that the habit of *Manfreda* and other members of the former Agavaceae were indistinguishable from members of Liliaceae. The solution, according to McVaugh was to include all herbaceous members of Agavaceae in Liliaceae and retain Agavaceae for the succulent and woody taxa. It illustrated the divide within the subfamily Agavoideae between herbaceous members and taxa with persistent succulent leaves. The evolution of succulence is poorly understood but is surmised to relate to adaption for xeric habitats. Although *Manfreda* contains species which exhibit persistent succulent forms, a distinct and significant difference to those of *Agave* are evident. The thickened storage leaves of *Manfreda* lack the volume of fibers that *Agave* possess and retains a flexible thinner leaf (Castillejos-Cruz, 2009).

The succulence associated with species within *Manfreda* alludes to a natural group based around two species that may represent basal lineages. The *M. guttata* (Jacobi and C.D.Bouché) Rose group includes species with semi-succulent or non-succulent leaves. These leaves are deciduous thus senesce and abscise during the dormant season. This adaptive system allows for a reduction of mass during unfavorable conditions, reducing respiration and aiding storage of water and nutrients. The other group aligned around *M. scabra* (Ortega) McVaugh is characterized by succulent persistent leaves that are evergreen and utilize their increased water storage capacity to inhabit areas such as the Rio Grande Valley and the Chihuahua Desert. Fleshy flowers and relatively large teeth are other characteristics associated with this succulent and evergreen group (Verhoek-Williams, 1975).

The coloration of the leaves is variable among species but has not been utilized as a taxonomic character (Verhoek-Williams, 1975; Castillejos-Cruz, 2009). Leaves can harbor dark-green to light-green colors with some authors describing the darkest leaves as having a blue coloration. A maroon mottling of leaves is also common, desired in the horticultural trade and species-dependant (Irish and Irish, 2000). The mottling is not consistent however and many workers have disregarded it as a taxonomic character (Verhoek-Williams, 1975).

Other, more novel characters, have also been used to assess the taxonomy of *Manfreda*. Verhoek (1998) pointed out a consistent difference in Agavoideae genera based on papillate epidermal cells on the leaf. The survey revealed an unnatural alignment between *Prochnyanthes, Hesperaloe* and *Yucca* section *Hesperoyucca* (Engelm) Trel. exhibited the trait arranged over veins. In contrast, the other species of *Yucca, Furcraea* and *Beschorneria* exhibit non-uniform groups of papillae epidermal cells. *Manfreda* was observed to have a similar arrangement to

Prochnyanthes, *Hesperaloe* and *Yucca* Sect. *Hesperoyucca* with linear formations of papillae cells.

The morphology of the stomata can be employed taxonomically at an intergeneric level. *Manfreda*, along with other members of tribe Poliantheae, are consistently paracytic (with two parallel subsidiary cells). Although this trait offers no insight into the internal relationship with in *Manfreda*, it does serve to separate tribe Poliantheae from *Agave*, which is tetracytic (with four subsidiary cells) (Verhoek, 1998).

Stem

Species of *Manfreda* arise from either bulbs or slender upright rhizomes (Howard, 2001). The correlation between ontogeny and the presence of either a bulb or a slender corm was alluded to by De La Cruz (1998). This theory was supplementary to the earlier work of Gonzalez (1997), who established a similar trait in *Polianthes.* The trait is unique to the tribe Poliantheae and *Yucca elata* (Engelm.) Engelm. as horizontal rhizomes are the archetypal form in the subfamily Agavoideae (Verhoek, 1998). Either the bulb or the upright rhizomes depending on species generally possesses a substantial basal plate for the rosette (Howard, 2001). The stem is described as subterranean and provides little elevation for the leaves (Verhoek, 1998; Verhoek and Ness, 2002). This phenomenon has been cited as an evolutionary adaptation that aids temperature regulation and thus facilitates inhabitance of dry and warm environs (Nobel, 1994).

Inflorescence

The inflorescence of *Manfreda* differs from that of *Agave* and is another diagnostic character to separate the two taxa. *Manfreda* exhibit racemes or spikes, whereas *Agave* possesses paniculate inflorescences (Verhoek, 1998). The exception to this is *Agave* subg. *Littaea* which shares a spicate or rarely racemose inflorescence (Reveal and Hodgson, 2002). The terminal region of the inflorescence bears flowers in dense, lax or intermediate clusters. The number of flowers at each node is a diagnostic character to determine *Manfreda* from *Polianthes* and *Prochnyanthes*, with *Manfreda* most commonly bearing a single flower per node while the other members of tribe Poliantheae bear two (Verhoek-Williams, 1975). The inflorescence, which is extremely tall relative to the plant itself, is highly variable in size and is somewhat taxonomically debated (Irish and Irish, 2000 and Castillejos-Cruz, 2009).

Although commonly referred to as a scape (Verhoek, 1978; Hernándex-Sandoval, 2008), Castillejos-Cruz (2009) made clear that it was not. Building upon previous works of Font-Quer (1979) and Solano (2000), it was stated that for the inflorescence to be a scape it must be devoid of bracts and present flowers at the apex. In this regard, the inflorescence is not a scape as modified bracts are located towards the base of the inflorescence, and flowers are arranged in a raceme or spike and not allocated at the terminus, therefore the term peduncle is preferred (Castillejos-Cruz, 2009). Other key taxonomic characters are aligned to the inflorescence itself. For example, the insertion of bracts and presence of trichomes are important for distinguishing between species, especially the latter, which is unique to *M. maculata* and *M. pubescens* (Regal Ortegies) Verh.-Will. ex Espejo and Lopez-Ferr. (Verhoek-Williams, 1975: Castillejos-Cruz, 2009).

Flowers

The form of the perianth in *Manfreda*, unlike the majority of the order Asparagales, is epigynous along with sister taxa in the Agavoideae subfamily. This trait is thought to have evolved independently on only a few occasions in the group. The lack of other taxa that exhibit this form makes it a diagnostic character for members of subfamily Agavoideae and a strong field character for identification (Bogler *et al.*, 2006; Simpson, 2010). The perianth within the Agavoideae subfamily shares many similar characters. The tepals are connate at the base in *Agave*, *Manfreda*, *Polianthes* and *Prochnyanthes* and differs to that of aligned taxa *Yucca* and *Furcraea*, which exhibit free tepals (Verhoek, 1998). The constituents of the Agavoideae subfamily also share the characters of being biseriate yet homochlamydeous (three outer tepals and three inner tepals), syntepalous and bracteate (Simpson, 2010).

The perianth tube in *Manfreda* exhibits two forms, funnelform or cylindrical with width of the tube and constriction above the ovary the varying factor (Verhoek and Ness, 2002). The limbs are most commonly recurved with the exception of *M. virginica* where erect limbs are apomorphic (Verhoek-Williams, 1975). The variation in the floral tube and the limbs is substantial (Irish and Irish, 2000). The greatest variation was demonstrated by Castillejos-Cruz (2009) who compared *M. bulbulifera* Castillejos and E. Solano and *M. longiflora* (Rose) Verh.-Will., which varied in size from a few millimetres to 6 cm respectively.

Color of the perianth in *Manfreda* is variable and colors are consistent with *Agave* and most species possess a green to yellow color (Verhoek and Ness, 2002). Coloration with maroon is also common however in the form of bands on the tepal (Castillejos-Cruz, 2009). Some species are completely red in color while others are white or pink intermediates (Verhoek, 1998). In *M. variegata*, a brown coloration has been observed (Verhoek-Williams, 1975). The green and white colors are associated with pollination by bats and such relationships have been observed in *Manfreda*. Pollination syndromes of *Manfreda* in relation to color have been poorly studied (Groman and Pellymr, 1999).

The androecium of *Manfreda* consists of six stamens, which is consistent with other members of the Agavoideae subfamily (Verhoek and Ness, 2002; Simpson, 2010). The stamens are inserted at the base of the tepal in the majority of species, some however, have characteristic insertion levels on the tepal however (Verhoek 1998; Castillejos-Cruz, 2009). Insertion is in a uniform single series in all species with the exception of *M. potosina* B.L. Rob & Greenham. Within *M. potosina*, the apomorphic characteristic of having stamens arranged in two series is observed (Verhoek-Williams, 1975; Castillejos-Cruz, 2009).

The color of the filament demonstrates some degree of variation, which is similar to coloration of the tepals where white, green and yellow colors are most common while maroon has also been observed (Castillejos-Cruz, 2009). The filaments of *Manfreda* differ from those of *Polianthes* as they are elongated and mostly exserted (Williams-Verhoek, 1975). The filament that develops from a downward bent position in the bud extend out with the floral tube (Castillejos-Cruz, 2009). The exceptions to the normal condition are *M. guttata* (Jacobi and C.D. Bouché) Rose. and *M. hauniensis*, which have inserted stamens (Williams-Verhoek, 1975). At anthesis, the style is considerably shorter and immature, it can be surmised that this condition helps avoid self-fertilization. The style however, is often of comparable size to the stamens three days after anthesis (Irish and Irish, 2000; Verhoek and Ness, 2002).

The inferior ovary of *Manfreda* shares many characters with those of *Agave*, *Polianthes* and *Prochnyanthes* (Verhoek, 1998). The ovules are anatropous (occasionally campylotropous), bitegmic and aligned into two rows within the ovule (Cronquist, 1988; Simpson, 2010). The placentation of the ovules is axile divided between three locules (Cronquist, 1988). The style, as previously mentioned, is shorter than the stamens during anthesis. The filiform style develops post anthesis and varies considerably in length (Verhoek, 1998). Castillejos-Cruz (2009) illustrated the variation by comparing *M. parva* Aáron Rodr. and *M. longistaminata* Castillejos and E. Solano. which measured 2.6 mm and 12.5 mm respectively. Stigmas of *Manfreda* are trigonous, which is similar to the vast majority of Agavoideae, although *Polianthes* and *Prochnyanthes* exhibit three distinct lobes (Verhoek-Williams, 1975; Verhoek, 1998). The stigma of *Manfreda* is moist when receptive like those of *Polianthes* but differs to that of *Agave* which has a dry stigmatic surface (Verhoek, 1998). The white, yellow and green stigmas of *Manfreda* are also receptive at night and suggest a relationship with night flying animals such as

Sphingidae (Hawk Moth) and Chiroptera (Bats) (Groman and Pellymer, 1999; Irish and Irish, 2000).

Fruit

The Asparagales, was largely characterized based on the studies of Huber (1969). These studies utilized the seed and fruit characters to define the limits of the taxon (Dahlgren *et al.*, 1985). As such, subfamily Agavoideae possess similiar characters in fruit as other taxa in Asparagaceae and the broader Asparagales. The fruit of the subfamily are locuilicidal or indehiscent capsules. The fruit of *Manfreda* are loculicidal, globuse to cylindrical and possess obvious sutures aligned along the locule walls (Verhoek and Ness, 2002; Castillejos-Cruz, 2009: Simpson, 2010).

Seeds

The seeds are highly uniform thoughout the Asparagales (Dahlgren, 1985). Agavaceae was often circumscribed based on the black phytomelan-encrusted seeds that were associated uniformly with all its genera and species (Bogler *et al.*, 2006). In *Manfreda* these seeds are flat and nearly triangular with a radial margin ranging from 0.3-0.6 cm (Verhoek and Ness, 2002). The seeds contained a thin membrane formed by the collapsed inner integument of the seed coat, while the outer epidrmis of the testa is coated with the characteristic phytomelan (Dahlgren, 1985; Castillejos-Cruz, 2009). A characteristic which is found in *Manfreda* and other Asparagales that seperates it from the Liliales is the lack of fat glands that characterise the seeds of liliaceous species. Within the locule, the seeds are aligned in two rows and vary in size depending on their position along these two rows (Castellejos-Cruz, 2009).

Principal Component Analysis

Principal Component Analysis (PCA) is a means of studying multivariate data to identify patterns, recognize affinities within the dataset and highlight significant differences (Smith, 2002). Principal Component Analysis facilitates the exploration of data to discover trends, such as those aforementioned, but it employs different strategies to those of cluster analysis techniques. PCA is less rigorous than techniques of cluster analysis, however PCA can be more informative as it summarizes only the most illuminating relationships within the dataset (Agilent, 2005). The concepts supporting PCA are based on the linear transformation of a source set of correlated variables. The transformation results in a substantially smaller set of uncorrelated variables. The reduction of the dataset to a smaller number of uncorrelated variables aids analysis as data with little or no relevance is discarded (Shlens, 2005). The elimination of data dimensions by employing covariance analysis between factors makes utilization of such analyses for studies of high dimension datasets feasible. By reducing the dimensionality of the dataset, PCA allows for the analysis to be reduced to specified number of principal components with minimal loss of informative data (Smith, 2002).

For botanical studies, PCA is one of the most widely adopted means of statistical analysis of morphological data. The application of PCA to taxonomic investigations has been employed in various taxa. Examples of the exploitation of PCA include Doebley (1989) who employed PCA as a component of a systematic study of *Zea* L., Brunell and Whitkus (1998) who utilized PCA in the examination of sub-specific taxa assigned to *Eriastrum densifolium* (Benth.) H. Mason. and Barrington (2003) who studied potential hybridization between three species of *Polystichum* Roth. When wild populations or cultivated collections are rare and molecular data is limited, PCA of herbarium specimens is commonly employed as a primary means of taxonomic

study, as exemplified by Sears (2008) who conducted a major revision of *Platanthera* Rich. PCA has also been employed in *Prochnyanthes*, a closely related taxon to *Manfreda*. Castro-Castro *et al*. (2010) was able to identify two distinct genotypes of *Prochnyanthes mexicana* (Zucc.) Rose employing the technique.

JMP® statistical software has been elected as the software package of choice to conduct PCA. JMP®, which was originally written in 1989 under the working title the "John MacIntosh Project", possesses in its most recent version (JMP® 10) an arsenal of features that aid academic study and industry related research (Sall, 1996). JMP® represents a more dynamic working environment than many rival statistical packages. JMP® boasts greater flexibility regarding utilization of datasets in many different forms (i.e. text files, Microsoft Excel documents and other SAS® files) and a dynamic interface that allows for real-time modification of data and graphics. These features in combination with multithreading and an internal data storage system make JMP® one of the fastest and capable platforms available at present for processing large statistical datasets (Sall, 1996; SAS Institute, 2010).

Principal Component analysis in JMP® can be utilized in a number of different forms. The concept can be employed in either the Multivariate Analysis or Scatterplot 3D platforms, otherwise the Principal Component analysis platform itself can be utilized (SAS Institute, 2010).

Molecular Systematic Research

The early development of techniques based on molecular evidence in liliaceous genera can be traced to the works of McKelvey and Sax (1933), Whittaker (1934) and Sâto (1935). These early karyological studies had a major bearing on the subfamily Agavoideae of Asparagaceae and the former Agavaceae (Verhoek, 1975; Bogler *et al.*, 2006). These studies identified a bimodal arrangement of five extended chromosomes complimented by 25 reduced chromosomes (Singh, 2004). The karyological arrangement was identified with high uniformity among members of the former Agavaceae and some peripheral taxa such as *Camassia* Lindl., *Hesperocallis* A. Gray. and *Hosta* Tratt., suggesting close affinity (Bogler *et al.*, 2006). The karyological evidence was significant and consistent with the taxonomic system devised by Hutchinson (1934). The evidence increased confidence in the taxon Agavaceae when no clear apomorphic characters for the group were known and only habit characterized the taxonomic limits (Verhoek-Williams, 1975; Cronquist, 1981).

Molecular research regarding species of the Agavoideae subfamily advanced little until Chupov and Kutiavina (1981). Their immunological studies of the Lilioid monocots employed serological techniques and immunoelectrophoresis to the separation and characterization of protein-utilizing immunoglobulins. Their study demonstrated a close affinity between *Yucca* and *Agave*, while also reconfirming the close relations to *Camassia* and *Hosta* (Chupov and Kutiavina, 1981). Their study complemented the works of Hutchinson (1934) and Cronquist (1981), however the study was poorly recognized in North America and Western Europe and was attributed little attention at the time (Bogler *et al.*, 2006). Dahlgren *et al.* (1985) was one of a few who recognized the work of Chupov and Kutiavina (1981) and employed their findings to support their ordinal system, which included the order Asparagales.

Molecular Systematic Study of Asparagales Link. and Asparagaceae Juss.

The study of Lilioid monocots utilizing the *rbcL* gene region by Chase *et al.* (1995) was the first molecular study to address the phylogenetic relationship and limits of the order Asparagales proposed by Dahlgren (1983). The study supported the distinction between Asparagales and Liliales with robust bootstrap values and resolution. With morphological taxonomy and molecular studies both supporting the existence of a monophyletic Asparagales the system was adopted in many contemporary classification systems (Singh, 2004; Simpson, 2010).

At the dawn of the APG system with the publication of APG (1998) the order of Asparagales was resolute due to its well-supported monophyletic lineage. Asparagales was recognized as one of 40 orders and was deemed to include 29 different families. Support for Asparagales has been consistent and resolute since the inception of molecular research concerning the taxon (Bogler *et al.*, 2006). Subsequent APG systems (2003 and 2009) have retained the order and its presence has been rarely challenged, while other research has also continued to support its presence in modern classification systems (Fay *et al.*, 2000b; Soberg, 2012). It is however, the internal circumscription of Asparagales which possesses limited resolution and consensus (Bogler *et al.*, 2006).

Chase *et al.* (1995) presented the original molecular research for the families that comprise Asparagales. Despite the high levels of confidence inferred for the ordinal system based on ITS sequences, no such confidence could be inferred for their system concerning families. With low bootstrap values and resolution in his Maximum Parsimony Analysis, the study inferred that the broad families of Cronquist (1981) were inaccurate and supported, with

limited confidence, the proliferated family systems proposed by Takhtajan (1980) and Dahlgren *et al.* (1985). Limited confidence in the botanical community was attributed, to the research however as under strict consensus (consensus between all fundamental trees proposed) many branches of the proposed phylogeny collapsed (Bogler *et al.*, 2006).

The premise that the Agavaceae sensu lato of Cronquist (1981) was not monophyletic and was a sister clade to the Nolinaceae-Dracaenaceae-Convallariaceae clade proposed by Chase *et al.* (1995) gained some notoriety despite its poor bootstrap values and resolution. The restricted fragement length polymorphism (RFLP) study of Bogler and Simpson (1995) of 110 loci in the chloroplast genome reinvestigated such a hypothesis and yielded similar conclusions with significantly higher confidence levels. The study inferred that indeed the Nolinaceae-Dracaenaceae-Convallariaceae clade was distinct but closely aligned to the clade containing Agavaceae. The study also alluded to some relationships within Agavaceae itself, but was unable to distinguish all. The study demonstrated a close affinity between Agavaceae, *Hesperaloe* and *Hesperoyucca whipplei* (Torr.) Trel.

Bogler and Simpson (1996) readdressed the issue of the taxonomy of Asparagales once more but utilized ITS sequence data to do so. By undertaking such a study, many of the relationships proposed in Bogler and Simpson (1995) improved in resolution and support. Many of the proposed relationships with inadequate support and resolution were re-examined and gained sufficient merit to warrant significant consideration. One such relationship was the close affiliation between *Camassia, Hosta* and Agavaceae.

With the advent of the APG system, speculative grouping based on morphology and habit and paraphyletic taxa were no longer tolerated. The widely utilized system only supported

monophyletic taxa of resolute support and confidence (Bogler *et al.*, 2006; APG, 2003). The system of APG (1998) addressed Asparagales by creating 29 families to account for the high instance of paraphyly. The APG (1998) system was widely critized however for being biased towards proliferation and splitting of taxonomic groups (APG, 2003). As such, the algorithms employed for APG (2003) and APG (2009) were adjusted, resulting in 25 families in Asparagales. In anticipation of the continued scepticism of the rampant proliferation inferred by the system, two alternate systems for Asparagales were proposed. The first suggested a reduction to a core 11 families including Agavaceae sensu lato, while another proposed the radical alternative of two families, Asparagaceae and Alliaceae (APG, 2009).

The proposal for a two-family system, well-defined morphologically with Alliaceae possessing the apomorphy of an umbellate inflorescence as opposed to the raceme, spikes and panicules of Asparagaceae (Chase *et al.,* 2009). By APG (2009) support for a two-family system and a major revision of the family units of Asparagales was supported by the numerous influential authors included in the APG Working Group. The argument centered around the lack of apomorphic characters for the previous families, the staunch molecular support and clearly defined field characters attributed to Alliaceae and Asparagaceae. This solution also supported ease of teaching by keeping the number of family units to a minimum as advocated by Cronquist (1981).

Although the family unit of Asparagaceae is novel and contemporary whereas Agavaceae is well established as well as highly utilized, it is highly likely that the family of Asparagaceae will persevere due to the widescale adoption of the APG system by leading botanical institutes and herbaria (APG, 2009; Chase *et al.*, 2009; RBG Kew Press, 2010). As such, this study will be

one of the first accounts of *Manfreda* that addresses the genus as a member of Asparagaceae and subfamily Agavoideae rather than Agavaceae.

Molecular Systematic Study of Agavoideae and *Manfreda* Salisb.

Many of the greatest advancements in the systematics of Asparagales and lower taxa such as the former Agavaceae/subfamily Agavoideae and *Manfreda* has been established based on DNA sequencing and phylogenetic inference. Important studies such as that of Eguiarte *et al.* (2000), Chase *et al.* (1995), Bogler and Simpson (1995, 1996) and APG (1998, 2003 and 2009) have all employed DNA sequencing and furthered our understanding of the morphologically complex group. Different research groups have investigated alternative DNA regions to varied success.

The study of Chase *et al.* (1995) employed the *rbcL* which is a large single-copy region of the chloroplast. The function of the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (RUBISCO) is predominantly involved with fixation of carbon dioxide in dark reactions. The DNA sequence of the gene is highly variable and used extensively in plant systematic research for DNA sequencing. The research concluded there was a clear distinction between Liliales and Asparagales at an ordinal level. Conclusions drawn from the study supported the classification system of Dahlgren (1983) and were the foundation for further research into the Lilioid monocots and the order Asparagales. The *rbcL* region was used to study the family level systematics within Asparagales by Duvall (1993), Eguiate (1994) and Chase *et al.* (1995). All three studies supported the proliferation of the former Agavaceae and the expulsion of *Dracaena* Vand. ex L., *Nolina* Michx. and *Dasylirion* Zucc. into adjacent families.

The Internal Transcribed Spacer (ITS) region located between the 18S and 26S of nuclear ribosomal is one of the most widely used region for DNA sequence analysis in plant systematics and studies into the Asparagales order (Chase *et al.*, 2007; Simpson, 2010). The region, which contains multiple DNA copies, high variability and conserved flanking regions, has made it a highly applicable and utilized tool in modern systematics (Kress *et al.*, 2005; Simpson, 2010). The region was first employed for the former Agavaceae by Bogler and Simpson (1996), subsequent to Bogler and Simpson (1995) a study of the same taxon using chloroplast DNA restriction site analysis. The utilization of the ITS region greatly increased resolution and identified a monophyletic Agavaceae with a high bootstrap value. The results also alluded to an expansion of Agavaceae sensu lato to include affiliated taxa *Camassia* and *Hosta.*

Recent studies by Bogler *et al.* (2006) have attempted to build upon the success of Bogler and Simpson (1996) by exploring other regions. Regions such as *ndhF* have been identified as bearing a similar level of resolution as ITS but without the ease of amplification. The *ndhF* for example is four times as long as ITS and does not possess as well-defined flanking regions which facilitates the development of universal primers. The search for other regions of utility continues however as ITS has been proven to lack the ability at generic and species level to differentiate all taxa (Bogler and Simpson, 2006; G. Salazar, National Autonomous University of Mexico, Personal Communications). The insufficient variability and evolution within the group has prompted researchers to find either a complimentary or new gene region for phylogenetic research within Asparagales at genus and species level (Bogler *et al.*, 2006).

The only research conducted at the genus level and below concerning *Manfreda* has been the research of Dr. Gerardo Salazar as part of his efforts to find a universal barcode for land plants with the Consortium for the Barcoding of Life's Plant Working Group. The surveys of

gene regions applicable for the resolution of genera and species within the Asparagales order however differs somewhat to finding a suitable region for phylogenetic study. Dr. Salazar requires in a universal barcode a highly variable region bearing conserved flanking sites for the development of universal primers. Due to the rigorous requirements of a universal barcode, none of the regions which he surveyed were suitable or studied on a sufficient sample of *Manfreda* species and individuals.

Table 1. Gene regions surveyed by Dr. Salazar (National Autonomous University of Mexico) for barcoding utility in *Agave* L. and associated taxa including *Manfreda* Salisb. (Salazar, National Autonomous University of Mexico, personal communications).

Gene Region	Genome	Approximate Size
*rpo*B	Plastid	410bp
*rpo*C1	Plastid	400bp
*ndh*J	Plastid	380bp
*acc*D	Plastid	900bp
*rbc*L	Plastid	456bp
*Mat*K	Plastid	1,500bp
*psb*A-*trn*H	Nuclear	450bp
ITS	Nuclear	460bp
*trn*G1	Nuclear	600bp
*psb*K-*psb*L	Nuclear	400bp

Phylogenetic Utility of *rbcL* Gene Region

The Ribulose-Bisphosphate Carboxylase/Oxygenase *(rbcL*) gene region, a constituent region of the chloroplast genome, has been utilized extensively in molecular systematics in plants (Simpson, 2010). In 1993, *rbcL* was selected for the first major taxonomic assessment of angiosperms employing a molecular technique. The study, conducted by 42 systematists lead by Dr. Mark Chase of the Royal Botanic Garden, Kew opted for *rbcL* as the orthologous region of choice due to its ease of amplification and universality (Chase *et al.*, 1993 and Savolainen, 2000). The *rbcL* has successfully inferred novel phylogenetic relationships in a host of taxa such

as Cucurbitaceae, Orchidaceae, Rhamnaceae and Araucariaceae employing its measured rate of synonymous nucleotide substitution in comparison to the highly variable nuclear regions (Reddy, 2009). Trials conducted by Erickson *et al.* (2008) noted that *rbcL* was able to successfully discriminate nearly all of 239 species selected from across the plant kingdom. Yet in contrast, a lack of divergence has also been alluded to within other taxa. The slower rate of nucleotide substitution in certain plant groups has failed to provide substantial nucleotide varation for phylogenetic inference, especially at specifc and subspecific taxonomic levels (Kress and Erickson, 2007). Despite the region identified shortcomings it is still one of the best performing gene region avaliable for phylogenetic analysis with highly successful primers and protocols further enhancing utility (Plant Working Group, 2009).

Phylogenetic Utility of *MatkK* Gene Region

Maturase K *(MatK)* is a single-copy gene of the chloroplast located within the *trnK* intron and neighboring the *psbA* gene region (Hilu and Liang, 2007). Both regions flanking *Mat*K are highly conserved and host a series of primers designed for the amplifcation of the highly variable *MatK* region (Hilu *et al.*, 2003). The region persists in the vast majority of plant taxa due to functional constraints. To illustrate, the parasitic genus *Epifagus* (L.) W. Bartram. only retained 45% of the chloroplast genome of allied taxa, however the residual portion included *MatK* in its entirety despite the deletion of introns flanking the region and substantial lose of contiguous portions of the genome (Hilu and Liang, 2007).

Utilization of *Mat*K as a phylogenetic marker arose between 1994-1996 with a number of seminal papers inferring successful phylogenies (Hilu *et al.*, 2003). The utility of *MatK* is attributed to its rate of nucleotide substitution, three times higher than in *rbcL*, high rate of

nonsynonymous substitution and insertions/deletions as well as a capacity for phylogenetic signal in even taxa previously unresolved by other prominent markers (Hilu *et al.,* 2003; Barthet and Hilu, 2007). The high rate of substitution associated with *MatK* is due to a comparable rate of substitution at each codon in contrast to a third codon bias in most protein-coding regions. The utility of *MatK* has been repeatedly proven in marker surveys for phylogenetic and DNA barcoding efforts in that it retains the greatest discriminatory ability. The drawbacks of *Mat*K however has been cited as a lack of universality in primers, especially in gymnosperms (Hilu *et al.*, 2003). A number of new primers have been developed in recent times, increasing universal amplification to levels akin to other commonly employed regions such as *rbcL* and ITS.

Maximum Parsimony Analysis

Maximum Parsimony Analysis is a technique commonly employed for the inference of phylogeny (Simmons and Ochoterena, 2000). Maximum Parsimony employs the minimum number of evolutionary steps to infer phylogenetic relationships; in doing so the method negates poorly supported proliferation (speciation) or misrepresentation of homoplasy (non-derived analogous characters). The non-parametric technique utilizes discrete characters in the generation of an array of cladograms, each of which is subjected to an explicit optimality criterion (best scoring) for the selection of the optimal tree or compilation of trees (Doyle and Davis, 1998; Page and Holmes, 1998 and Swofford, 1993; Kolaczkowski *et al.,* 2004). The selected cladogram(s) can either be rooted by outgroups or unrooted, while nodes harboring speciation are scored using the Bootstrap Resampling Technique (Efron, 1982) where a numerical value between 0 and 1 is assigned to represent the percentage of iterations in which the speciation event occurred (Swofford, 1993).

Maximum parsimony has been a premier technique for phylogenetic inference since the advent of computational cladistics. The simplistic model of evolution employed by maximum parsimony was convenient for processing capabilities of the 1980s and 1990s and thus gained prominence (Steel, 2005). With modern processing proficiency, maximum parsimony is rivalled by likelihood and Bayesian methods of phylogenetic inference. The superiority of maximum parsimony has been championed over such techniques by many researchers (Farris, 1973; Sober, 1988). Kolaczkowski and Thornton (2004) epitomized the strength of maximum parsimony in datasets with high heterogeneity while proponents of the technique continued to compose taxonomic systems of considerable merit utilizing the method (i.e. APG, 1998; *Centaurea* L. - Garcia-Jacas *et al.*, 2000; *Acacia* Mill. - Lucklow, 2003).

PAUP* 4.0 is a widely employed program for the inference of evolutionary trees developed at Florida State University by Dr. David Swofford. PAUP* 4.0 utilizes maximum parsimony to infer phylogenies from discrete character data, the algorithms employed calculate the single most parsimonious tree or group of trees (Swofford, 2003). The program encompasses a range of utilities such as alternative phylogenetic methods, algorithms and parameters allowing for substantial customization of the analysis. Version 4.0 includes and expanded array of analyses from version 3.1 with the inclusion of maximum likelihood capabilities and improvements to branch and bound algorithms further diversifying and increasing the utility of the program (Wilgenbusch and Swofford, 2003).

The limitations of PAUP 4.0* are few relative to comparable phylogenetic analysis platforms. Although PAUP currently hosts a number of analysis techniques such as maximum parsimony, maximum likelihood, neighbor joining and UMPGA but it does not host the highly popular Bayesian analysis. Furthermore, internal platforms for tree editing and sequence

alignment are not included and thus accessory programs must be employed (i.e. TreeView, MacClade and BioEdit) (Swofford, 2003).

Bayesian Analysis

Bayesian Analysis has been established as a mainstream analytical technique for phylogenetic research (Ronquist and Huelsenbeck, 2003). In conjunction with the Markov Chain Monte Carlo (MCMC) method, Bayesian Analysis possesses substantial capabilities for the inference of phylogenies. The method employs a Markov Chain constructed from a probability distribution, the chain is modified with random variables (0,1) and contrasted between accepted probabilities (Andrieu *et al.,* 2003). The iterations can be specified for a desired confidence and speciation events are scored via Posterior Probability, a statistic pertaining to the likelihood of accuracy based on available data (Ronquist and Huelsenbeck, 2003).

Exponential growth in the application of Bayesian Analysis to phylogenetic inference was observed from the 1990s (Congdon, 2003). Bayesian Analysis was seen as an alternative to Maximum Parsimony Analysis, viewed as oversimplified and inadaptable to external parameters (Congdon, 2003 and Weising, 2005). In plant systematics, Bayesian analysis has been employed extensively (Kim *et al*., 2004; Neinhaus *et al.,* 2005 and Smith *et al*., 2008). The consensus within the botanical community is that both Maximum Parsimony Analysis and Bayesian Analysis have merits and limitations, thus both techniques are also employed in tandem (Simmons and Miya, 2003).

MrBayes 3.1.2 is a program that employs Bayesian Analysis for phylogeny estimation. Employing the Monte Carlo Markov Chains (MCMC) method, posterior probabilities (PP) are sampled utilizing Metropolis-coupling. The technique relies upon three 'heated' (increasing PP)

chains and one 'cold' chain (decreasing PP), systematic exchange of parameter values between the 'heated' and 'cold' chains aids progression of 'melting', a flattening of peaks due to application of heated chains and prevents chains from becoming embedded between peaks. After a specified number of generations are run and further optional optimization such as 'burn-in' are completed, an optimal tree is devised (Ronquist and Huelsenbeck, 2003).

Unlike PAUP 4.0*, MrBayes 3.1.2 only possesses the utility to perform Bayesian Analysis. The program is limited to the specific task of analysis and peripheral applications are limited. To conduct model testing or tree modification others programs must be employed (i.e. JModelTest or TreeView). The greatest limiting factor of MrBayes 3.1.2 is that the Bayesian algorithms employed have a much greater complexity than other contemporary analyses such as maximum likelihood or maximum parsimony, thus require significantly more computing power and time to conduct such analysis (Matzke, 2011).

Nucleotide Network Analysis

Nucleotide networks are commonly employed in population genetics for the visualization of alternative haplotypic forms between sequence data. For analysis of DNA barcodes, nucleotide networks possess a number of advantages as opposed to phylogenetic inference such as documentation of relationships between single nucleotide polymorphisms, algorithms designed to decipher between highly aligned sequences and a greater capacity for the incorporation of evolutionary events (i.e. recombination). Commonly employed techniques for composition of nucleotide networks are Templeton-Crandell-Sing Method via TCS 3.2.1 and Forced Directed Method via HapStar 0.7 (Clement *et al.*, 2000; Teacher and Griffith, 2011).

TCS 3.2.1 was the original program available for nucleotide network composition and was employed across a breadth of studies (Blaxter *et al.*, 2005; Pons *et al.*, 2006; Roe *et al.*, 2010; Zhou *et al.*, 2012). Since the release of HapStar 0.5 in 2011 the program has been increasingly utilized. Due to the relatively recent advent of HapStar software platforms, research employing this technique have been limited. The performance of HapStar 0.5 and 0.7 has shown substantial merit and utility in studies for which it has been employed (Odour, 2011; Guiraldelli and Rocha, 2011; Sun *et al.*, 2012). Although limited data is available as to the use of HapStar in the analysis of DNA barcodes, TCS analysis has been employed to examine the performance of DNA barcodes in studies of various taxonomic groups (Wong *et al.*, 2009; Hart and Sunday, 2011; Monaghan *et al.*, 2012).

HapStar 0.7 was favored to TCS 3.2.1, despite prior examples of TCS 3.2.1 being applied to DNA barcode analysis, due to the advanced algorithms and facilities found in HapStar 0.7. HapStar was composed in Python 3.3.0 and processes input files in the Arlequin Results File format compatible with Arlequin 1.1 onwards. The technqiue employs the Force Directed Method algorithm, iterations (repetition of the analysis) test for optimal assemblage of nodes and branches via a series of nodal transfers. The resulting nucleotide network is graphically represented by a series of nodes and branches, each branch representing a nucleotide variation, intermediate nodes representing ancestral or missing OTUs and terminal nodes representing the taxa understudy. Modification of the resultant nucleotide network can be conducted on any platform configured to the Scalable Vector Graphics file format (Teacher and Griffith, 2011).

DNA Barcoding Research

The technique of DNA barcoding was originally derived from systematic studies. Akin to DNA barcoding, DNA sequencing utilizing orthologous DNA sequences which are compared to uniquely identify taxonomic units while estimating relationships. The origin of the DNA barcoding concept, where libraries of standardized DNA sequences were available in the public domain, was developed at the University Of Guelph, Ontario by Dr. Paul Herbert in 2003 (barcodeoflife.org, 2012). The advent of DNA barcoding coincided with an increased awareness within the taxonomic community of escalating biodiversity loss as well as the lethargic and insufficient attributes of contemporary taxonomic techniques (Godfray, 2002; Blaxter and Floyd, 2003 and Newmaster *et al*., 2006). Traditional descriptive and morphology-based taxonomic systems had been the principal technique for species documentation since the advent of the Linnean system (1753) and prior with over 1.7 million species having been described in such a way. With estimates of diversity of life ranging from 10-100 million species, DNA barcoding gained many early proponents as the most appropriate method for cataloguing the vast array of life on earth (United Nations, 1992; Newmaster *et al.,* 2006).

DNA barcoding, with an estimated cost between 2.50 to 5.00 USD per sample and protocols consistent with common laboratory techniques for extraction, amplification and sequencing facilitated a quick and early adoption of the technique by the taxonomic community (Cameron *et al.,* 2006). The technique allowed for study under a variety of circumstances and thus was advocated as a tool of great utility for taxonomists. DNA samples were taken from wild, dried herbarium or archaeological samples for taxa which where extinct (Savolainen *et al.,* 1995; Kress *et al.*, 2005). Such analysis could never have been conducted via morphological based taxonomy. Furthermore, classical taxonomy could only be utilized to its fullest when all

organs of the plant were present; this attribute can be difficult in many angiosperms if flowering is infrequent (Chase *et al.*, 2005). Moreover, DNA barcoding could be conducted via batch analysis of many species during one procedure while traditional morphological taxonomy required considerable input into every taxon (Plant Working Group, 2009).

The criteria cited for selection of an appropriate region or augmented series of DNA regions for DNA barcoding practice differs depending on the dynamics of the study. In studies seeking a universal region the utility and characters of the region must be subjected to strict performance levels (Fazekas, 2009; Plant Working Group and Janzen, 2009). In research concerning smaller taxonomic groups however the gene selection paradigm differs and is often less restricted allowing different regions to be employed (Pryer, 2010). As a generalization however the following criteria have been repeatedly cited for selection of a gene region for DNA barcoding:

Sufficient Size

It has been suggested that between 400-800 bp is the ideal length for a DNA sequence to be employed in DNA barcoding efforts (Kress and Erickson, 2008). This number of suggested base pairs is short enough for the vast majority of PCR procedures to amplify such a region efficiently from even partially degraded samples in single pass sequencing (Kress and Erikson, 2008). The suggested size is also thought to be large enough to permit sufficient sequence variation to represent divergence at a species or subspecies level (Kress, 2005; Ford *et al.*, 2009).

Taxon Discrimination

Size is a significant factor in the level of discriminatory power in many gene regions but increased length is redundant unless a sufficient rate of nucleotide substitution is observed (Kress

and Erikson, 2008). The variability within a desirable gene region would ideally have the discriminatory prowess at all taxonomic level for all species, however the search for such a region has yielded little indication that such a region exists (Hollingsworth, 2009). Thus, the search for a gene region of adequate discriminatory power are generally graded against a 90% species discrimination interval for regions of universal application and customized levels of confidence are applied to taxa specific DNA barcoding operations (Fazekas, 2009).

Routine Amplification

Routine amplification is paramount to establish a universal DNA barcode. For a DNA barcode to be widely applicable, taxonomists should be able to follow an established protocol to amplify the region of interest (Kress, 2005; Kress and Erikson, 2008; Ford *et al.*, 2009). The requirement for routine amplification contend somewhat with the required attributes for variability. Regions must be variable for species discrimination, yet they must also contain conserved flanking sites to allow complimentary primers to be developed (Plant Working Group, 2009). The oxymoronic requirements of DNA barcoding regions to be variable internally yet conserved at both terminal portions is perhaps the most limiting factor in identifying DNA barcodes and restricts the most variable regions from being employed (Hollingsworth, 2009).

The technical challenges faced in addressing documentation of speciation and taxonomic units are being addressed continually (Hollingsworth, 2009). The fundamental challenges remain and are perplexing; the lack of a clearly defined species concept has been cited as an example of such and little has been undertaken to address the issue due to the enormity of the task (Spooner, 2009). The nature of nucleotide substitution varies and the phenotypic expression varies even more. The complexity of introgression and hybridization also facilities diverse

genomes with little indication of epigenetic or phenotypic inference (Mortiz and Cicero, 2004; Fazekas, 2009). The species complex as regards to DNA barcodes would be difficult to categorize based on rate of nucleotide substitution or the presence of autapomorphic traits and as such allocation of taxa to specific taxonomic groups will remain somewhat subjective. Moreover, DNA barcoding will always be in need of traditional taxonomy to quantify inferred systems against morphological traits on which classical taxonomic systems were proposed (Will *et al.*, 2005; Spooner, 2009).

Universal DNA Barcoding

The use of DNA barcoding in plants rose to prominence after the successful utilization of the *CO1* gene extracted from mitochondrial DNA (mtDNA) for identification of species in the animal kingdom. The region was highly applicable for such as the rate of nucleotide substitution in the mtDNA genome is high and the region possessed conserved flanking sites for ease of amplification (Chase *et al.*, 2007). For the plant kingdom the utility of the *CO1* gene was problematic. The gene only exhibited a handful of base alterations across 1.4 kb and genome structure evolved too rapidly in plants for the presence of conserved flanking regions. In animals, the wide adoption of *CO1* has been criticized due to the maternal inheritance patterns and lack of functionality in all taxonomic groups (Kress *et al.*, 2005).

In search of an applicable region for utilization in land plants the Plant Working Group of the Consortium for Barcoding of Life (CBOL) was devised and was funded by the Alfred P. Sloan and Betty Moore Foundation to research 100 plastid regions for utility as a plant barcode (Chase *et al.*, 2007).

The plant working group, led by Dr. Peter Hollingsworth of the Royal Botanic Garden Edinburgh, consisted of 52 researchers from 24 institutions worldwide, based their survey of three primary principals which were universality, sequence quality in addition to coverage and discrimination to conduct the foundation work required to further study genes which were of high utility (Levin, 2009). None of the resultant seven loci had the same utility as the *CO1* for animals, each had their advantages and deficiencies. The most challenging compromise faced was between universality and conserved flanking regions. The most appropriate regions demonstrated only 70% universality while *CO1* in animals exhibited a range upwards of 90%. A multiregion approach may be the most appropriate (CBOL, 2009). Many authors have supported the tiered approach to barcoding while critics of such are prevalent (Fazekas *et al.*, 2009).

The protocol for conducting DNA barcoding is still in its infancy, typically five to ten samples are sequenced for each species to provide confidence in the sequence obtained. A statistical technique known as the Probability of Correct Identification is then utilized to assess the taxa under study. The technique looks for monophyly in a small phylogenetic analysis to infer a confidence level to the uniqueness of the DNA sequence the particular specimen studied. PCI is still being developed and taxonomic weighting and scaling are still to be refined for efficient analysis (Erickson *et al.*, 2008).

The establishment of a minimum level of divergence for a DNA barcode has not been devised to date. Although the need to define taxonomic limits is apparent, skepticism surrounds such work however as many feel limits would be artificial without a holistic species concept being devised first, which is perhaps one of the greatest challenges to modern biology (Erickson *et al.*, 2008).

Selection of Appropriate Gene Region(s)

The argument for and against particular genes in DNA barcoding is further exacerbated by the lack of consensus on whether a singular, multiple regions or a tiered approach is employed. The exploits of animal biologists who have opted for the singular gene approach have yielded significant results by employing the *CO1* gene exclusively. The ease of amplification and routine practice involved have allowed for adoption of the technique universally and for mass barcoding efforts to be conducted (Chase *et al.*, 2007). Approximately 90% of species studied have conformed to such DNA barcoding efforts indicating a high degree of universality and utility for future employment (Fazekas *et al.*, 2009). The challenges of plant systematics are perhaps more diverse with constant hybridization and introgression as well as slower plastid evolution than in animal mitochondria (Hollingsworth *et al.*, 2011). As such a singular region from the mitochondrial genome or chloroplast genome fails to represent parental lineages as they are uniparental and possesses considerably less nucleotide substitution than the *CO1* region in animals (Chase *et al.*, 2007).

In plants, in light of the observed lack of utility of single regions as standardized DNA barcodes, multiple region systems have often been proposed. Erickson and Kress (2007) estimated that from their studies that utility of two DNA barcoding regions rather than one observed a 9% rise in species discrimination from 79% to 88% on average. Chase *et al.* (2007) was the first major study to advocate this system and devised two combinations which exhibited taxonomic utility. The combinations of *rpo*C1, *rpo*B, *Mat*K and *rpo*C1 as well as *Mat*K demonstrated discriminatory prowess, although neither exceeded or matched *CO1* utilized in animal studies and all required considerably more laboratory time and effort to employ.

A third approach, proposed by Newmaster *et al.* (2006), utilizes multiple regions at different phylogenetic levels in a tiered system. The system would avoid simply adding another sequence of bases onto the terminus of the barcode by creating multiple levels and using the appropriate gene region or gene regions to discriminate at each taxonomic level (i.e. order, family and genus). The system is more complex and require a revaluation of how we think about DNA barcoding, but a noteworthy proposition nonetheless. The system utilizes gene regions appropriately and increase the utility of the barcode.

The selection criteria for the choice of gene regions for DNA barcodes for plants were devised by the CBOL Plant Working Group (2009). The following three objectives are utilized to assess the feasibility of potential barcoding regions for employment in future studies.

(1) Universality: The chosen loci must have application across a breadth of land plants if not all are to be deemed suitable for universal application (CBOL, 2009). For studies of specific taxonomic groups, greater flexibility in gene selection can be exploited due to the lesser constraints of universality.
(2) Quality and Coverage: Robust sequences should be gained with ease of amplification with few ambiguous base calls for comparability to all species studied. This utility is particularly reliant on conserved flanking sites for designation of universal primers (Chase *et al.*, 2007; CBOL, 2009).
(3) Discrimination: The selected loci should be able to discriminate at all taxonomic levels or via the tiered approach of Newmaster *et al*. (2006) and have substantial discriminatory power at each of the levels designated. Discrimination is the crux of species determination and DNA barcodes which fail to discriminate the vast majority of species studied have little merit in such studies (CBOL, 2009).

The frailties of DNA barcoding in plants is evident, however. Spooner (2009) cited a number of studies in which DNA barcoding efforts have categorically failed. The weakness of the system were noted by Chase *et al.* (2007) and particular emphasis was placed on the lack of discrimination observed in plastid regions due to measured evolution in comparison to animal mitochondrial genomes. It was also stated that universality of a nuclear region was not feasible due to substantial variation in size and nucleotides in selected spacer regions. Furthermore, species boundaries are less well defined in plants than in animals. Fazekas *et al.* (2009) concluded that gaps between taxonomic units were significantly less in plants and introgression and hybridization occurred at a substantially greater rate complicating molecular systematics in plants.

Selection of CBOL Plant DNA Barcoding Regions

Indecision as to an appropriate universal barcoding standard for land plants hindered early barcoding efforts. Selection of a gene region or augmented series of gene regions proved to be difficult as no one region or combination of regions met the desired criteria (Plant Working Group, 2009; Chase *et al.*, 2007). The CBOL Plant Working Group (2009) conducted the largest survey as to potential DNA barcodes. The study, which incorporated 907 samples representing 445 angiosperms, 38 gymnosperms and 67 cryptogamic species, tested seven potential loci against CBOL data standards (Table 2). The study concluded that *rbcL* offered the greatest universality yet modest discriminatory prowess. *MatK* and *trnH-psbA* demonstrated similar levels of discrimination, superior to the other candidate region. Both regions however were impaired. *MatK* had been disregarded in previous studies due to lack of primer universality, yet demonstrated 90% amplification in the study. *trnH-psbA* had a high instance of mononucleotide repeats and non-consistent bidirectional sequencing which impeded alignment. The consensus

between the CBOL Plant Working Group was that *rbcL* and *MatK* in tandem was the most viable barcode for land plants.

Table 2. Summary of analysis conducted for seven potential barcoding regions by the Plant Working Group (2009).

Gene region	Function	Universality	Sequence quality	Species discrimination
matK	coding	90%>	moderate	66%
rbcL	coding	90%>	high	61%
rpoB	coding	90%>	high	40%
rpoC1	coding	90%>	high	43%
atpF-atpH	noncoding	90%>	moderate	50%
trnH-psbA	noncoding	90%>	moderate	69%
psbK-psbl	noncoding	77%	moderate	68%

The CBOL Executive committee appointed an ad hoc panel of three independent review groups to evaluate the merits of the proposed *rbcL* and *MatK* barcode of CBOL Plant Working Group (2009) in conjunction with reviewing a three locus barcoding option consisting of *rbcL*, *MatK* and *trnH-psbA* championed by Kress *et al.* (2009). The panel concluded that the advantages of augmenting *trnH-psbA* against time and cost were negligible and therefore supported the two locus barcode. The CBOL Executive Committee therefore approved *rbcL* and *MatK* as the barcode for land plants issuing a declaration statement on November 16, 2009. Primers, protocols and data guidelines were devised and deposited on barcoding.si.edu and kew.org/barcoding to initiate barcoding efforts (CBOL Executive Committee, 2009).

DNA Barcoding for Horticulture

Within the nursery trade one of the greatest challenges is the correct identification of specimens sold during multiple forms and growth stages such as seed, corms/bulbs and vegetative growth. The ontogenesis of a plant can have a number of polymorphic phases of

growth which can be misleading for even the most knowledgeable horticulturalists. The limitations of morphological identification have often lead to the mislabelling and incorrect sale of species and cultivated taxa (Pryer, 2010).

A recent example was highlighted by the sale of *Cheilanthes* Sw. in a nursery franchise located in California, North Carolina and Texas. The nursery was selling *C. wrightii* Hook., a native of Arizona, New Mexico and Texas, and claiming provenance of the plant material to the aforementioned regions (Science Daily, 2010). A study by Pyrer (2010) sampled specimens from the nursery and sequenced the *rbcL*, *AtpA* and *trnG-R* gene regions to be employed as an identifying barcode. The samples were cross-checked against the Pterophyte Barcode Library at Duke University, North Carolina which contained type barcodes for a range of *Cheilanthes* species. With comprehensive evidence the study concluded that the species being sold was in fact *C. distans* Mett., a native of Australia, New Caledonia and New Zealand. The nursery implicated most likely mislabelled the species in error but many issues can arise with mis-sale such as plant patent and variety rights, associated commission and illegal sale of banned or protected species.

The protection of intellectual property in plant breeding is integral for fostering incentives for breeding and continuing development (Rimmer, 2003). The protection of new cultivars allows for the developer to reap financial rewards for such efforts over a prolonged period in the form of sales commissions and thus such regulations are a catalyst for continued development of plant materials for the horticultural industry (Kesan and Janis, 2002). Legislation was first introduced via the Plant Patent Act (1930) in the United States. The act stemmed the encroachment of the unlawful sale of plant cultivars protected under the act via asexual propagation but did not cover sexually propagated nor tuber-propagated materials (Chen, 2006).

A more comprehensive act was enlisted in the 1970s under compliance with the International Union for Protection of Plant of New Varieties of Plants (UPOV). The UPOV convention of 1961 required compliance and the generation of conforming legislation from each of the member states, as such the Plant Variety Protection Act (1970) was established in the U.S. The act, which also constituted part of the United States compliance with the establishment of the World Intellectual Property Organization (1967), was established allowing for the first time the protection of cultivars propagated via sexual means. Under this legislation breeders are awarded 20 years (or 25 years for trees and vines) of exclusive sale or licensing to a vendor who are liable for royalties (Thomas, 2002; Rimmer, 2003; USDA, 2006).

Enforcement of the Plant Patent Act (1930), Plant Variety Protect Act (1970) and the Convention on International Trade in Endangered Species of Wild Fauna and Flora (1973) have all encountered problems with prosecution in recent years. Defense attorneys have learned to challenge the species status or distinct entity of the plants and without comprehensive forms of evidence have often succeed in overturning prosecutions placed against their clients (San Diego Zoological Society, 2006). The subjective manner of morphological identification and taxonomic determination cannot hold significant bearing in court and thus DNA barcoding has been seen as one of few lines of comprehensive evidence which could be employed (Kress and Erickson, 2008; Levin, 2009).

Although the concept of DNA barcoding for species has been well established, difficulties lie with recent hybridization and introgression (Cowan *et al.*, 2006). The plastid gene regions employed for global barcoding efforts have often been criticized for poorly documenting hybridization due to the region's uniparental heritage (Cowan *et al.*, 2006; SDZS, 2006; Newmaster *et al.*, 2006). For use of barcodes for cultivated taxa, the identification of

hybridization and identification of parent lineages is critical (Mortiz and Cicero, 2004). The challenges of establishing barcodes for hybridized material has been overshadowed by species determination and global barcoding efforts and thus has been little studied. The work of Dr. Kevin Burgess at Columbus State University, Ga. is one of few studies investigating identification of hybridized taxa with a horticultural emphasis. His research has focused on the identification of the promiscuous genus of *Narcissus* L. that is known to have a diverse life history strife with introgression (Burgess, 2010). With no data published as of yet the field remains fraught with challenges but yet laden with potential for research.

The need for accommodations of hybridization in DNA barcodes extends beyond identification of cultivated taxa. A study by Fazekas (2009) studied 12 genera and identified the need for a better understanding of hybridization events in DNA barcodes for species determination. The study highlighted the contribution to discontinuity in genomes of plants and that taxa with a disposition towards hybridization have smaller genetic divergence between species. Skepticisms associated with DNA barcoding can also often be traced back to its insufficiencies with identification of taxa with recent hybridization within their lineages (Mortiz and Cicero, 2004).

The reliance of DNA barcoding systems restricted to few gene regions, or solely *CO1* in studies of animals has been cited as often failing to account for recent hybridizations (Newmaster *et al.*, 2006). The problem is only exacerbated if the region is uniparental such as many of the plastid regions favored by the CBOL Plant Working group (Cowan *et al.*, 2006 and Plant Working Group, 2009). The inheritance factor is not improved via the utilization of multiple regions if all are uniparental. In construction of a DNA barcode, which accommodates

introgression and hybridization, the plastid and mitochondrial genome would provide little insight (Cowan *et al.*, 2006).

In studies conducted by Spooner (2009) regarding *Solanum* L. spp. the fragility of plastid regions was illustrated with the lack of utility of the chloroplast regions *psbA-trnH* and *MatK*. The ITS region of the nuclear genome exhibited considerably more variation and was more informative than the aforementioned plastid regions. The ITS region did not yield a clear understanding of the hybrid origin of many of the studied taxa as infraspecific variation was substantial and few evolutionary patterns were revealed. The utility of the nuclear region which is biparental for DNA barcodes and accommodate recent introgression and hybridization is understood by a number of authors, the region of choice is still debated and novel regions have been suggested (Kress *et al.*, 2005).

Potential DNA Barcoding Regions for Horticultural Application

Internal Transcribed Spacer (ITS)

The Internal Transcribed Spacer (ITS) gene region, located between the 18S-26S of the nuclear ribosome (nrDNA), refers to the combination of two independently evolved sub- regions, ITS1 and ITS2 separated by the 5.8S nrDNA (Simpson, 2010). The region is evolutionarily conserved, small and thus has a reduced number of restriction sites. In flowering plants the region varies little in size, the ITS1 region fluctuates from 187 to 298 bp while the ITS2 region varies from 187 to 252 bp depending on species (Hershkovitz, 1999: Simpson, 2010). The function of the ITS regions is thought to be linked to the development of the mature 18S, 5.8S and 26S rRNAs; however, through deletion exercises conducted by Van Der Sande *et al* (1992), it was demonstrated that the affiliation is only partial and placed little evolutionary constraints on

either ITS1 or ITS2 (Baldwin, 1992). The overriding interest in contemporary molecular genetics is purely for utility in molecular ecology and systematics. The negligible function allows for substantial variation with a high GC content within a conserved region facilitating high discriminatory powers between taxa at low phylogenetic levels (i.e. genus and species) (Hershkovitz, 1999). The gene differs from many other studied gene regions with sizeable variation however in possessing conserved flanking regions, which facilitate primer development and universality of primers developed (Bena, 1998b; Linder *et al.*, 2000). The high copy number of rRNA genes allows also for ease of amplification even from small quantities of DNA extracted from degraded specimens or even aged herbarium specimens (Simpson, 2010).

Utilization of the ITS region was first explored with floral studies by Baldwin (1992) who was the first plant systematist to utilize the region, based on employment in research conducted on apes and humans (Gonzalez *et al.*, 1990b). A number of other authors then began to utilize the region due to the considerable ease of processing (Baldwin, 1993; Suh *et al.*, 1993; Wojciechowski, 1993; Baldwin, 1995). Since then the popularity and usage of the region for phylogenetic studies has increased exponentially, making ITS one of the most utilized gene regions in molecular ecology, with numerous primers developed for different taxa and over 741,000 sequences registered with GenBank (Bogler and Simpson, 1996; NCBI, 2011).

The ITS region possesses some characteristics which limit its utility. The crux of the limitations is related to a lack of discriminatory power within certain lineages and taxonomic groups (Bena, 1998b). Although the rate of evolution is high within both the ITS1 and ITS2, each only contain at most 298 bp and 252 bp, respectively (Simpson, 2010). The limited number of base pairs impedes substantial variation in certain taxa due to slower rates of evolution, new or recently diverged lineages (Bena, 1998b).

The ITS region has been seen as unviable for a universal DNA plant barcoding region by most of the leading research (Chase *et al.*, 2007). Problems cited include lack of utility in some taxonomic groups and presence of divergent paralogous ITS copies (Kress *et al.*, 2005). Although single-copy plastid regions are employed for universal DNA barcoding, the ITS still harbors great potential for DNA barcoding in horticultural research. The ITS region has demonstrated considerable utility in the vast majority of plant groups, including Agavoideae (Bogler *et al.*, 2006). The region possesses three to four times greater nucleotide variability than plastid markers and as a nuclear gene region, is biparental facilitating detection of hybrid taxa (Chase *et al.*, 2007).

External Transcribed Spacer (ETS)

The External Transcribed Spacer (ETS) region, which is located adjacent to the ITS and the Non-Transcribed Spacer (NTS), is a segment of the IGS region between 18S and 26S of the nuclear ribosomal DNA. This is a portion of the genome with much taxonomic interest (Simpson, 2010). This region, like ITS, is involved in nrDNA maturation but has minimal functional constraints and is similarly variable (Hershkovitz, 1999). During the 1980s and 1990s, the ETS region was profitably exploited in Restriction Fragment Length Polymorphism (RFLP) analysis of the IGS region. Due to technical issues associated with primer development, only the ITS with conserved flanking sites was later utilized in early DNA sequencing studies (Baldwin and Markos, 1998; Hershkovitz, 1999).

Nuclear loci such as the *H3* intron, *pgiC*, *ncpGS* and *PISTILLATA* intron 1 have been comprehensively investigated with little reward in search of a compliment to ITS (Doyle *et al.*, 1996; Gottlieb and Ford, 1996; Emshiwiller and Doyle, 1999 ; Bailey and Doyle, 1999). Such

research has found that most nuclear regions are poorly characterized, overly conserved or have associated amplification difficulties (Starr *et al.*, 2003).

Restriction site studies conducted by Systma and Schaal (1985) and Kim and Mabry (1991) in the ETS region have shown comparable variability to that of ITS. These early studies provided Bruce Baldwin and Staci Markos of the University of California, Berkeley with substantial evidence for further investigation and saw the ETS region as a potential supplement to the ITS to increase resolution (Markos and Baldwin, 2001). A long-distance PCR technique allowed Baldwin and Markos (1998), to amplify the IGS region to develop internal primers for amplification of the ETS region, showing remarkable promise for the 3' end in particular. The technique of amplifying the larger IGS region with primers set in the 18S and 26S regions via long distance PCR and the design of internal primers was the breakthrough in the use of ETS for systematic studies (Markos and Baldwin, 2001). From the seminal work of Markos and Baldwin (1998) the use of the ETS region to augment the ITS region increased significantly and yielded impressive results (Markos and Baldwin, 2001; Chew *et al.*, 2010; Logacheva *et al.*, 2010).

The advantages gleaned from utilizing the ETS region are widely considered to be the greater variability and size of the region in comparison to the ITS gene region (Volkov *et al.*, 1996; Linder *et al.*, 2000). The combined optimal ITS size of ca. 800 bp is greatly exceeded by that of ETS with results ranging to an upper limit of 3kb (Borisjuk *et al.*, 1997; Hershkovitz, 1999). The ETS region is more informative in particular segments towards the 5' end. Repetitive non-informative DNA is common with much of the variation in size attributed to a tandem repeat sequence. Harbored in the 3' end however is a region found across taxa of approximately 500-600 bp which is substantially variable and highly informative (Hershkovitz, 1999).

Research aimed at circumnavigating the amplification difficulties associated with the ETS region have been centered around amplification of the aforementioned IGS region. The IGS region is ca. 3-6 kb long and constructed with repeating motifs (Hershkovitz, 1999). The size of the gene region has made amplification challenging, especially considering the amount of intergeneric polymorphism contained within its boundaries. The spacer itself is also too long and poses difficulties in analysis due to the large volume of repeated elements but contains three subregions of significant interest, the ITS, ETS and some even suggest that the NTS may harbor an ability for phylogenetic resolution (Persson, 2000; Becerra, 2002). The high variability of the NTS at the 5' prime end of ETS region is one of the limiting factors for ETS primer design (Linder *et al.*, 2000). With use of primers developed for the 18S region downstream of the 3' end of the ETS and a second designed primer in a conserved segment towards the 5' end Baldwin and Markos (1998) were the first to amplify and sequence the ETS region of 700 bp for *Calycadenia* DC. The Ast-1 primer, showed considerable utility across the Asteraceae, a sizeable family of approximately 1620 genera (Simpson, 2010). Similarly, Bena *et al.* (1998a;1998b) devised a primer specific to Fabaceae, Andersen & Baldwin (2001) developed one for Malvaceae and Becerra (2002) designed a primer for *Bursera* Jacq. ex. L. spp. yet did not test the wider utility.

The complex procedure involved in designing primers for the ETS region and the substantial variation in composition and size of the region negate any potential for use as a universal DNA barcode. Such challenges are specific to universal barcoding efforts and can be overcome in smaller taxonomic groups such as Agavoideae or *Manfreda*.

Literature Cited

Agilents Technologies. 2005. Principal component analysis. Agilents Technologies, Santa Clara. 13 July 2012. <http://www.chem.agilent.com/cag/bsp/sig/downloads/pdf/pca.pdf.>.

Andreasen, K., and B.G. Baldwin. 2001. Unequal evolutionary rates between annual and perennial lineages of Checker Mallows (*Sidalcea*, Malvaceae): Evidence from 18s–26s rDNA internal and external transcribed spacers. Mol. Biol. and Evolution 18: 936-944.

Andrieu, C., N. De Freitas, A. Doucet, and M.I. Jordan. 2003. An introduction to MCMC for machine learning. Machine Learning 50: 5-43.

Angiosperm Phylogeny Group. 1998. An ordinal classification for the families of flowering plants. Ann.Miss. Bot. Gard. 85: 531-553.

Angiosperm Phylogeny Group. 2003. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG II. Bot. J. Linnean Soc. 141: 399-436.

Angiosperm Phylogeny Group. 2009. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. Bot. J. Linnean Soc. 161: 105-121.

Bailey, C. D., and J.J. Doyle. 1999. Potential phylogenetic utility of the low copy nuclear gene *PISTILLATA* in dicotyledonous plants: comparison to nrDNA ITS and *trnL* intron in *Sphaerocardamum* and other Brassicaceae. Mol. Phylogenetics Evolution 13(1): 20-30.

Baldwin, B. G. 1992. Phylogenetic utility of the Internal Transcribed Spacers of nuclear ribosomal DNA in plants: An example from the Compositae. Mol. Phylogenetics Evolution 1(1): 3-16.

Baldwin, B. G., and S. Markos. 1998. Phylogenetic utility of the External Transcribed Spacer (ETS) of 18s–26s rDNA: Congruence of ETS and ITS trees of *Calycadenia* (Compositae). Mol. Phylogenetics Evolution 10(3): 449-463.

Baldwin, B. G., M.J. Sanderson, J.M. Porter, M.F. Wojciechowski, C.S. Campbell, and M.J. Donoghue. 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. Ann. Miss. Bot. Gard. 82: 247-277.

Barrington, D. S. 2003. *Polystichum lilianae* sp. nov. (Dryopteridaceae) and its relationships to *P.fournieri* and *P.turrialbae*. Brittonia 55(4): 317-325.

Barthet, M. M., and K.W. Hilu. 2007. Expression of *MatK*: Functional and evolutionary implications. Amer. J. Bot. 98(8): 1402-1402.

Becerra, J. X. 2002. Evolution of Mexican *Bursera* (Burseraceae) inferred from ITS, ETS, and 5s nuclear ribosomal DNA sequences. Mol. Phylogenetics Evolution 26: 300-309.

Bena, G., M.F. Jubier, I. Olivieri, and B. Lejeune. 1998b. Ribosomal External and Internal Transcribed Spacers: Combined use in the phylogenetic analysis of *Medicago* (Leguminosae). J. Mol. Evolution 46: 299-306.

Blaxter, M. L., and R. Floyd. 2003. Molecular taxonomics for biodiversity surveys: already a reality. Trends Ecol. Evolution 18: 268-269.

Blaxter, M., J. Mann, T. Chapman, F. Thomas, C. Whitton, R. Floyd, and E. Abebe. 2005. Defining operational taxonomic units using DNA barcode data. Philosophical Trans. Royal. Soc. 360(1462): 1935-1943.

Bogler, D. J. and B.B. Simpson. 1995. A chloroplast DNA study of the Agavaceae. Systematic Bot. 20(2): 191-205.

Bogler, D. J. and B.B. Simpson. 1996. Phylogeny of Agavaceae based on its rDNA sequence variation. Amer. J. Bot. 83(9): 1225-1235.

Bogler, D. J., J.C. Pires, and J. Francisco-Ortega. 2006. Phylogeny of Agavaceae based on *ndhF*, *rbcL* and ITS sequences: Implications of molecular data for classifications. Aliso 22: 313-328.

Borisjuk, N. V., Y.M. Davidjuk, S.S. Kostishin, G.P. Miroshnichenko, R. Velasco, V. Hemlebem, and R.A. Volkov. 1997. Structural analysis of rDNA in the genus *Nicotiana*. Plant. Mol Biol. 35: 655-660.

Brunell, R. S. and R. Whitkus. 1998. Assessment of morphological variation in *Eriastrum densifolium* (Polemoniaceae): Implications for subspecific delimitation and conservation. Systematic Bot. 23(3): 351-368.

Burgess, K., A.J. Fazekas, P.R. Kesanakurti, S.W. Graham, B.C. Husband, S.G. Newmaster, D.G. Percy, M. Hajibabaei, and S.C. Barrett. 2011. Discriminating plant species in a local temperate flora using the *rbcL MatK* DNA barcode. Methods Ecol. Evolution 2: 333-340.

Cameron, S., D. Rubinoff, and K. Will. (2006). Who will actually use DNA barcoding and what will it cost? Systematic Bot. 55(5): 844-847.

Castillejos-Cruz, C. 2009. Sistemático del genero Manfreda. National Autonomous University of Mexico, Mexico City, Mexico, PhD Diss.

Castro-Castro, A., A. Rodriguez, G. Vargas-Amado, and R. Ramirez-Delgadillo. 2010. Variacion morfologica del genero *Prochnyanthes* (Agavaceae). ACTA Bot. Mexicana 92: 29-49.

Cave, Y. 2003. Succulents for the contemporary garden. 1st ed. Timber Press, Portland, OR.

CBOL Executive Committee. 2009. CBOL approves *MatK* and *rbcL* as the barcode regions for land plants. Consortium for the Barcode of Life., Washington, D.C.

Chase, M. W. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbcL*. Ann. Miss. Bot. Gard. 80: 528-580.

Chase, M. W. 2007. A proposal for a standardised protocol to barcode all land plants. Taxon 56(2): 295-299.

Chase, M. W., J.L. Reveal, and M.F. Fay. 2009. A subfamilial classification for the expanded asparagalean families Amaryllidaceae, Asparagaceae and Xanthorrhoeaceae. Bot. J. Linnean Soc. 161: 132-136.

Chase, M., M. Duvall, H. Hills, J. Conran, A. Cox, L. Eguarite, J. Hartwell, M. Fay, L. Caddick, K. Cameron, and S. Hoot. 1995. Phylogenetics in Lilianae. In: P. Rudall (eds.). Monocotyledons: Systematics and Evolution. Royal Bot. Gard. Kew Press, Richmond, England.

Chen, J. 2006. The parable of the seeds: Interpreting the plant variety protection act in furtherance of innovation policy. Notre Dame Law Rev. 81(4): 1-51.

Chew, T., E. De Luna, and D. Gonzalez. 2010. Phylogenetic relationships of the pseudobulbous *Tillandsia* species (Bromeliaceae) inferred from cladistic analyses of ITS 2, 5.8S ribosomal RNA gene, and ETS sequences. Systematic Bot. 35(1): 86-95.

Chupov, V. S. and N.G. Kutiavina. 1981. The comparative immunoelectrophoretic investigation of seed proteins of Liliaceae. Bot. Zhurnal Leningrad 63: 473-493.

Clement, M., D. Posada, and K.A. Crandell. 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. 9: 1657-1659.

Congdon, P. 2003. Applied bayesian modelling. 1st ed. John Wiley and Sons Ltd, Chicester, England.

Cowan, R. S., M.W. Chase, W.J. Kress, and V. Savolainen. 2006. 300,000 species to identify: problems, progress, and prospects in DNA barcoding of land plants. Taxon 55(3): 611-616.

Cronquist, A. 1981. An integrated system of classification of flowering plants. Columbia Univ. Press, New York, NY.

Cronquist, A. 1988. The evolution and classification of flowering plants. 2nd ed. New York Bot. Gard. Press, New York, NY.

Dahlgren, R. 1983. General aspects of angiosperm evolution and macrosystematics. Nordic J. Bot. 3: 119–149.

Dahlgren, R.M., H.T. Clifford, and P.F. Yeo. 1985. The families of the monocotyledons: structure, evolution and taxonomy. Springer-Verlag, Berlin, Germany.

Doebley, J. 1990. Molecular systematics of *Zea* (Gramineae). Maydica 35: 143-150.

Doyle, J. J. and J.J. Davis. 1998. Homology in molecular phylogenetics, p. 101-131. In: Soltis D., P. Soltis and J. Doyle (eds.). Molecular systematics of plants. Kluwer, Boston, MA.

Doyle, J. J., V. Janazin, and R.C. Shoemaker. 1996. Phylogenetic utility of histone *h3* intron sequences in the perennial relatives of soybean (glycine: Leguminosae). Mol. Phylogenetics Evolution 6: 438-447.

Duvall, M. R., M.T. Clegg, M.W. Chase, W.D. Clark, W.J. Kress, H.G. Hills, L.E. Eguiarte, J.F. Smith, B.S. Guat, E.A. Zimmer, and G.H. Learn. 1993. Phylogenetic hypotheses for the monocotyledons constructed from *rbcL* sequence data. Ann. Miss. Bot. Gard. 80(3): 607-619.

Efron, B. 1982. The jackknife, the bootstrap, and other re- sampling plans. SIAM Monographs 38.

Eguiarte, L. E., V. Souza, and A. Silva-Montellano. 2000. Evolucion de la familia Agavaceae: filogenia, ecologıa evolutiva de la reproduccion y genetica de poblaciones. Bul. Soc. Bot. Mexico 66: 131-150.

Elliott, B. 2004. The birth of the organisation that was to become the RHS was on 7 March 1804. The Gard. 129(2): 92-95.

Emshwiller, E. and J.J. Doyle. 1999. Chloroplast-expressed glutamine synthetase (*ncpGS*): Potential utility for phylogenetic studies with an example from *Oxalis* (Oxalidaceae). Mol. Phylogenetics Evolution 12: 310-319.

Erickson, D. L., J. Spouge, A. Resch, L.A. Weigt, and W.J. Kress. 2009. DNA barcoding in land plants: Developing standards to quantify and maximize success. Taxon 57(4): 1304-1316.

Farris, J.S. 1973 A probability model for inferring evolutionary trees. Systematic Zoology 22: 250-256.

Fay, M. F. 2000b. Phylogenetic studies of Asparagales based on four plastid DNA regions. In: Wilson, K.L., and D.A. Morrison (eds.). Monocots: Systematics and Evolution : Volume I of the proceedings of Monocots II, the Second International Conference on the Comparative Biology of the Monocotyledons. CSPIRO Publishing, New York, NY.

Fazekas, A. J. 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. Public Library of Sci. 3(7): 1-12.

Font-Quer, P. 1979. Diccionario de botánica. Editoral Labor, Barcelona.

Ford, C. S., K.J. Ayres, N. Toomey, N. Haider, J. Van Alphan Stahl, L.J. Kelly, N. Wikström, P.M. Hollingsworth, R.J. Duff, S.B. Hoot, R.S. Cowan, M.W. Chase and M.J. Wilkinson. 2009. Selection of candidate coding DNA barcoding regions for use on land plants. Bot. J. Linnean Soc. 159: 1-11.

Garcia-Jacas, N., A.S. Mozzaffarian, and R. Illarslan. 2000. The natural delimitation of *Centaurea* (Asteraceae: Cardueae). Plant Systematics Evolution 223(3-4): 185-199.

Garcia-Mendoza, A. 2002. Distribution of *Agave* (Agavaceae) in Mexico. Cactus Succulent J. 74: 177-187.

Garcia-Moya, E., A.R. Manzanares, and P.S. Nobel. 2011. Highlights for *Agave* productivity. Global Change Biol. Bioenergy 3: 4-14.

Gentry, H.S. 1982. Agaves of continental North America. The University of Arizona Press, Tucson, AZ.

Godfray, H. C. 2002. Challenges for taxonomy. Nature 417: 17-19.

Gonzalez, D. 1998. Marcadores moleculares para los estudios comparativos de la variacion en ecologia y sistematica. Rev. Mexicana Mycol. 14: 1-21.

Gonzalez, I.L., C. Chamers, J.L. Gorski, D. Stambolian, R.D. Schmickel, and J.E. Sylvester. 1990b. Sequence and structure correlation of human ribosomal transcribed spacers. J. Mol. Biol. 212(5): 27-35.

Gottlieb, L. and V. Ford. 1996. Phylogenetic relationships among the sections of *Clarkia* (Ongraceae) inferred from the nucleotide sequences of *PgiC*. Systematic Biol. 21: 45-62.

Govaerts, R. 2011. World checklist of Asparagaceae. World Checklist of Selected Plant Families Facilitated by the Royal Botanic Garden Kew. 09 July 2012. <http:// http://apps.kew.org/wcsp/asparagaceae>.

Groman, J. D., and O. Pellmyr. 1999. The pollination biology of *Manfreda virginica* (Agavaceae): Relative contribution of diurnal and nocturnal visitors. Oikos 87(2): 373-381.

Guiraldelli, R. H., and R.L. Rocha. 2011. Adaptive unification of graphs applied to haplotype networks. IEEE Latin America Trans. 9(6): 950-955.

Hajibabaei, M., G.A. Singer, P.D. Hebert, and D.A. Hickey. 2007. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. Trends Genet. 23(4): 167-171.

Hannon, D. P. 2002. *Manfreda* and its allies. Cactus Succulent J. 74: 245-251.

Hart, M. W., and J. Sunday. 2011. Things fall apart: biological species form unconnected parsimony networks. Statistical Parsimony Methods 4: 43-49.

Hernández-Sandoval, L., R. Orellana, and G. Carnevali. 2008. Two new species of *Manfreda* Salisb. (Agavaceae) from the Yucatán Peninsula, Mexico. J. Torrey Bot. Soc. 135(2): 168-177.

Hershkovitz, M. A., E.A. Zimmer, and W.J. Hahn. 1999. Ribosomal DNA sequences, p. 268-326. In: P. Hollingsworth, R. Bateman and R. Gornall (eds.), Molecular Systematics. Taylor and Francis, London, England.

Hilu, K. W., T. Borsch, K. Müller, D.E. Soltis, P.S. Soltis, V. Savolainen, M.W. Chase, M.P. Powell, L.A. Alice, R. Evans, H. Sauquet, C. Neinhaus, T.A. Slotta, J.G. Rohwer, C.S. Campbell, and L.W. Chatrou. 2003. Angiosperm phylogeny based on *MatK* sequence information. Amer. J. Bot. 90(12): 1758-1776.

Hilu, K. W. and H. Liang. 1997. The *MatK* gene: Sequence variation and application in plant systematics. Amer. J. Bot. 84(6): 830-839.

Hollingsworth, P. M. 2008. DNA barcoding in biodiversity hotspots: Progress and outstanding questions. Heredity 101: 1-2.

Howard, T.M. 2001. Bulbs for warmer climates. University of Texas Press, Austin, TX.

Huelsenbeck, J. P. and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics Applications Notes 17(8): 754-755.

Hutchinson, J. 1934. The families of flowering plants: Dicotyledons. Claredon Press, London, England.

Irish, M. and G. Irish. 2001. Arizona gardener's guide. Cool Spring Press, Brentwood, TN.

Irish, M. 2002. Growing Agaves in cold climates. Cactus Succulent J. 74(4): 165-173.

Kesan, J. P. and M.D. Janis. 2002. U.S. plant variety protection: Sound and fury..?. Houston Law Rev. 39: 272-275.

Kim, K. J., and T.J. Mabry. 1991. Phylogenetic and evolutionary implications of nuclear ribosomal DNA variation in dwarf Dandelions (*Krigia*, Lactuceae, Asteraceae). Plant Systematics Evolution 177: 53-69.

Kim, Y. D., S.H. Kim and L.R. Landrum. 2004. Taxonomic and phytogeographic implications from its phylogeny in *Berberis*. J. Plant Res. 117: 175-182.

Kolaczkowski, B. and J.W. Thornton. 2004. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. Nature 431: 980-984.

Kress, W. J. and D.L. Erickson. 2007. A two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. Public Library of Sci. e508(6), 1-10.

Kress, W. J. and D.L. Erickson. 2008. DNA barcodes: Genes, genomics, and bioinformatics. Proc. Natl. Acad. Sci. U.S. 105(8): 2761-2762.

Kress, W. J., D.L. Erickson, F.A. Jones, N.G. Swenson, R. Perez, O. Sanjur, and E. Bermingham. 2009. Plant DNA barcodes and a community phylogeny. Proc. Natl. Acad. Sci. U.S. 106(44): 18621–18626.

Kress, W. J., K.J. Wurdack, E.A. Zimmer, L.A. Weigt, and D.H. Janzen. 2005. Use of DNA barcodes to identify flowering plants. Proc. Natl. Acad. Sci. U.S. 102(23), 8369-8374.

Levin, A. 2009. A barcode for botany. The Botanics 39: 4-7.

Lin, W., R. Zhou, L. Porter, J. Chen, and Y. Wu. 2010. Evolution of *Sousa chinensis*: A scenario based on mitochondrial DNA study. Mol. Phylogenetics Evolution 57(2): 907-911.

Linder, R., L.R. Goertzen, B.V. Heuval, J. Francisco-Ortega, and R.K. Jansen. 2000. The complete external transcribed spacer of 18S-26S rDNA: Amplification and phylogenetic utility at low taxonomic levels in asteraceae and closely allied families. Mol. Phylogenetics Evolution 14(2): 285-303.

Lindstrom, J.T. 2006. Intergeneric hybrids between *Polianthes* and *Manfreda*. Southern Nursery Assn. Res. Conf. 51: 599-601.

Logacheva, M. D. 2010. A comparison of nrDNA ITS and ETS loci for phylogenetic inference in the Umbelliferae: An example from tribe Tordylieae. Mol. Phylogenetics Evolution 57: 471-476.

Lucklow, M. A., M. Miller, J.T. Murphy, and T. Livshultz. 2003. A phylogenetic analysis of the Mimosoideae (Leguminosae) based on chloroplast DNA sequence data. Adv. Legume 10: 197-220.

Markos, S. and B.G. Baldwin. 2001. Higher-level relationships and major lineages of *Lessingia* (Compositae, Astereae) based on nuclear rDNA Internal and External Transcribed Spacer (ITS and ETS) sequences. Systematic Bot. 26(1): 168-183.

Matzke, N. 2011. Lab 11: MrBayes lab - principals of phylogenetics. University of California, Berkeley. 13 July 2012. <http://ib.berkeley.edu/courses/ib200a/labs/ ib200a_lab11_mrbayes.pdf>.

McKelvey, S. D. and K. Sax. 1933. Taxonomic and cytological relationships of *Yucca* and *Agave*. J. Arnold Arboretum 15: 76-80.

McVaugh, R. 1989. Bromeliaceae to Dioscoreaceae, p.1-398. In: W. Anderson (eds.). Flora Novo-Galiciana. University of Michigan Press, Ann Arbor, MI.

Monaghan, M. T., M. Balke, J. Pons, and A. Vogler. 2012. Beyond barcodes: complex DNA taxonomy of a south pacific island radiation. Proc. Royal Society 273: 887-893.

Mortiz, C. and C. Cicero. 2004. DNA barcoding: Promise and pitfalls. Public Library Sci. 10(2): 1529-1531.

NCBI. 2012. Genbank search: Internal Transcribed Spacer. National Center for Biotechnology Information. 09 July 2012. <http://www.ncbi.nlm.nih.gov/nuccore/?term =internal tracribed spacer/>.

Neinhaus, C., S. Wanke, K.W. Hilu, K. Muller, and T. Borsh. 2005. Phylogeny of Aristolochiaceae based on parsimony, likelihood, and bayesian analyses of *trnL-trnF* sequences. Plant Systematics Evolution 250: 7-26.

New York Times. 1992. Arthur Cronquist, 73, a specialist in classifying the plant kingdom. New York Times. 09 July 2012.<http://www.nytimes.com/1992/03/26/nyregion/arthur-cronquist-73-a-specialist-in-classifying-the-plant-kingdom.html>.

New York Times. 1993. Scientist at work: Armen Takhtajan; botanist plans survey of world's flowers. New York Times. 09 July 2012. <http://www.nytimes.com/1993/04/06/science/scientist-at-work-armen-takhtajan-botanist-plans-survey-of-world-s-flowers.html>.

Newmaster, S. G., A.J. Fazekas, and S. Ragupathy. 2006. DNA barcoding in land plants: Evaluation of the *rbcL* in multigene tiered approach. Can. J. Bot. 84: 335-341.

Nobel, P.S. 1994. Remarkable Agaves and succulents. Oxford University Press (USA), New York, NY.

Odour, A. M. 2011. Dissecting ecological and evolutionary mechanisms underlying invasiveness of plants: Using *Brassica nigra* as a model plant. University of Granada, PhD Diss.

Ogden, S. 1994. Garden bulbs for the south. Timber Press, Portland, OR.

Page, R. D. and E.R. Holmes. 1998. Molecular evolution: A phylogenetic approach. Blackwell, Oxford, England.

Paris, H. S. and J. Janick. 2008. What the Roman emperor Tiberius grew in his greenhouses. Proc. IXth EUCARPIA Mtg. Genet. Breeding Cucurbitaceae 33-40.

Persson, C. 2000. Phylogeny of the neotropical *Alibertia* group. Amer. J. Bot. 87: 1018–1028.

Plant Working Group., and D.H. Janzen. 2009. A DNA barcode for land plants. Proc. Natl. Acad. Sci. U.S. 106(31): 12794-12797.

Pons, J. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Systematic Biol. 55: 595-609.

Pyrer, K. M., E. Schuettpeltz, L. Huiet, A.L. Gusz, C.J. Rothfels, T. Avent, D. Swartz, and M.D. Windham. 2010. DNA barcoding exposes a case of mistaken identity. Mol. Ecol. Resources 10: 979-985.

Reddy, B. U. 2009. Molecular phylogeny of angiospermic plant families using *rbcL* gene sequences. Intl. J. Bioinformatics Res. 1(2): 27-36.

Reveal, J.L. and W.C. Hodgson. 2002. Flora of North America - *Manfreda* Salisb., p. 323. In: Flora of North America Editorial Commitee (eds.). Flora of North America, Volume 26: Lilidae: Liliales and Orchidales. Oxford University Press (USA), New York, NY.

Rimmer, M. 2003. Franklin Barley: Patent and plant breeders' rights. Murdoch University J. Law, 10(4): 1-45.

Rodríguez, A., and A. Castro-Castro. 2006. Potencial ornamental de los "Amoles" (*Manfreda*, Agavaceae) en Mexico. Ibungana 15(1): 3-11.

Roe, A. D., A.V. Rice, S.E. Bromilow, J.E. Cooke, and F. Sperling. 2010. Multilocus species identification and fungal DNA barcoding: insights from blue stain fungal symbionts of the Mountain Pine Beetle. Mol. Ecol. Resources 10(6): 946-959.

Ronquist, F. and J.P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics Applications Notes 19(12): 2003.

Rose, J.N. 1905. Contributions for the United States National Herbarium, volume VIII: Mexican, Central American and Porto Rican plants. Government Press Office, Washington, D.C.

Royal Botanic Garden Kew Press Office. 2009. As easy as APG III - scientists revise the system of classifying flowering plants. 09 July 2012. <http://www.kew.org/about-kew/press-media/press-releases-kew/easy-as-apg3/index.htm>

Sall, J. 1996. JMPer cables. 09 July 2012. <http://www.jmp.com/about/newsletters/jmpercable/pdf>.

SAS Institute. 2010. JMP release 8: User guide, second edition. SAS Institute, Cary. 12 July 2012. <http://www.jmp.com/support/downloads/pdf/jmp8/jmp_user_guide.pdf>.

Sāto, D. 1935. Analysis of the karyotype in *Yucca*, *Agave* and related genera with special reference to the phylogenetic significance. Jpn. J. of Genet. 11: 272-278.

Savolainen, V. 2000. Phylogenetics of flowering plants based on combined analysis of plastid *atpB* and *rbcL* gene sequences. Systematic Biol. 49(2): 306-362.

Savolainen, V., P. Cuenoud, R. Spichiger, M. Martinez, M. Crevecoeur, and J. Mamen. 1995. The use of herbarium specimens in DNA phylogenetics: Evaluation and improvement. Plant Systematics Evolution 197(1-4): 87-89.

Science Daily. 2010. DNA barcoding exposes fake ferns in international plant trade. Science Daily. 09 July 2012.<www.sciencedaily.com/1246643.htm>

San Diego Zoological Society. 2006. Barcoding: Developing tools for species identification as part of international efforts for maintaining sustainable wildlife populations. Annu. Mtg., San Diego, 13-14 Jan. 2006.

Sears, C. J. 2008. Morphological discrimination of *Platanthera aquilonis*, *P. huronensis*, and *P. dilatata* (Orchidaceae) herbarium specimens. Rhodora 110(94): 389-405.

Seberg, O., G. Petersen, J.I. Davis., J.C. Pires, D.W. Stevenson, M.W. Chase, M.F. Fay, D.S. Devey, T. Jørgensen, K.J. Sytsma, and Y. Pillon. 2012. Phylogeny of the Asparagales based on three plastid and two mitochondrial genes. Amer. J. Bot. 99(5): 875-889.

Shinners, L. H. 1966. Texas *Polianthes*, including *Manfreda* (*Agave* subgenus *Manfreda*) and *Runyonia* (Agavaceae). Sida 2: 333-338.

Shlens, J. 2005. A tutorial on principal component analysis. Center for Neural Science, New York University, New York. 09 July 2012. <http://www.snl.salk.edu/~shlens/pca.pdf.>.

Simmons, M. P. and M. Miya. 2003. Efficiently resolving the basal clades of a phylogenetic tree using bayesian and parsimony approaches: a case study using mitogenomic data from 100 higher teleost fishes. Mol. Phylogenetics Evolution 31: 351-362.

Simmons, M.P. and H. Ochoterena. 2000. Gaps as characters in sequence based phylogenetic analyses. Systematic Biol. 49: 369-381.

Simpson, M. G. 2010. Plant systematics. 2 ed. Elsevier, London, England.

Singh, G. 2004. Plant systematics: An integrated approach. Science Publishers Inc., Enfield, NH.

Smith, L. I. 2002. A tutorial on principal component analysis. University of Auckland, Auckland, New Zealand. 12 July 2012. <http: http://www.sccg.sk/~haladova/ principal_components.pdf>.

Smith, L.J. and I.N. Donoghue. 2008. Rates of molecular evolution are linked to life history in flowering plants. Science 322(5898): 86-89.

Sober, E. 1988. Reconstructing the past: Parsimony,evolution and inference. MIT Press, Cambridge, MA.

Spooner, D.M. 2009 DNA barcoding will frequently fail in complicated groups: An example in wild potatoes. Amer. J. Bot. 96(6): 1177-1189.

Starr, J. R., S.A. Harris., and D.A. Simpson. 2003. Potential of the 5' and 3' ends of the intergenic spacer (IGS) of rDNA in the Cyperaceae: New sequences for lower-level phylogenies in sedges with an example from *Uncinia*. Intl. J. Plant Sci., 164(2): 213-227.

Steel, M. 2005. Should phylogenetic models be trying to 'fit an elephant'. Trends Genet. 21(6): 307-309.

Suh, Y., L.B. Thein, H.E. Reeve, and E.A. Zimmer. 1993. Molecular evolution and phylogenetic implications of Internal Transcribed Spacer Sequences of ribsomal DNA in Winteraceae. Amer. J. of Bot. 80(9): 1042-1055.

Sun, Y., X. Wen, and H. Huang. 2012. Genetic diversity and differentiation of *Michelia maudiae* (Magnoliaceae) revealed by nuclear and chloroplast microsatellite markers. Genetica 139: 1439-1447.

Swofford, D. L. 1993. PAUP: a computer program for phylogenetic inference using maximum parsimony. J. Gen. Physiol. 102: A9.

Systma, K. J. and B.A. Schaal. 1985. Phylogenetics of the *Lisianthus skinneri* (Gentianaceae) species complex in Panama utilizing DNA restriction fragment analysis. Evolution 39: 594-608.

Takhtajan, A. 1997. Diversity and classification of flowering plants. Columbia University Press, New York, NY.

Teacher, A. G. and D.J. Griffiths. 2011. Hapstar: automated haplotype network layout and visualization. Mol. Ecol. Res. 11: 151-153.

Trueblood, E. W. 1973. "Omixochitl" - the Tuberose (*Polianthes tuberosa*). Econ. Bot. 27: 157-173.

United Nations. 1992. Convention on biological diversity. United Nations. 09 July 2012. <http://www.cbd.int/doc/legal/cbd-en.pdf.>

USDA. 2006. Plant variety protection and regulations and rules of practice. U.S. Department of Agriculture, Plant Variety Protection Office. Beltsville, MD.

Van Der Sande, C. A., M.M. Kwa, R.W. Van Nues, H. Heerikhuizen, H.A. Raue, and R.J. Planta. 1992. Functional analysis of Internal Transcribed Spacer 2 of *Saccharomyces*. J. Mol. Biol. 223: 899-910.

Verhoek, S.E. and W.J. Ness. 2002. Flora of North America - *Manfreda* Salisb., p. 462. In: Flora of North America Editorial Commitee (eds.). Flora of North America, Volume 26: Lilidae: Liliales and Orchidales. Oxford University Press (USA), New York, NY.

Verhoek, S.E. 1998. Agavaceae, p 60-70. In: K. Kubitzki (eds.). The families and genera of vascular plants: Monocotyledons: Lilianae (except Orchidaceae). Springer, Berlin, Germany.

Verhoek, S.E. 1978. Two new species and a new combination in *Manfreda* Agavaceae. Brittonia 30(2): 165-171.

Verhoek-Williams, S.E. 1975. A study of the tribe Poliantheae (including *Manfreda*) and revision of *Manfreda* and *Prochnyanthes* (Agavaveae). Cornell Univ., Ithaca, NY, PhD Diss.

Volkov, R. A., N.Y. Komarov, I.I. Panchuk, and V. Hemlebem. 2003. Molecular evolution of rDNA External Transcribed Spacer and phylogeny of sect. *Petota* (genus *Solanum*). Mol. Phylogenetics Evolution 29: 187-202.

Weising, K., H. Nybom, K. Wolf, and G. Kahl. 2005. DNA fingerprinting in plants. (2 ed.). CRC Press - Taylor Francis Group, London, England.

Whittaker, T. W. 1934. Chromosome constitution in certain monocotyledons. J. Arnold Arboretum 15: 135-143.

Wilgenbusch., J.C. and D.L. Swofford. 2003. Inferring evolutionary trees with PAUP. Current Protocols Bioinformatics 10: 1-28.

Will, K. W., B.D. Mishler, and B.Q. Wheeler. 2005. The perils of DNA barcoding and the need for integrative taxonomy. Systematic Biology 54(5): 844-851.

Wojciechowski, M. F., M.J. Sanderson, B.G. Baldwin, and M.J. Donoghue. 1993. Monophyly of aneuploid *Astragalus* (Fabaceae): Evidence from nuclear ribosomal DNA. Amer. J. Bot. 80(6): 711-722.

Wong, E. H., M.S. Shivji, and R. Hanner. 2009. Identifying sharks with DNA barcodes: assessing the utility of a nucleotide diagnostic approach. Mol. Ecol. Resources 9(1): 243-256.

Chapter 1

STUDY OF SPECIFIC DELIMITATION FOR *MANFREDA* SALISB. (ASPARAGACEAE) INFERRED FROM PRINCIPAL COMPONENT, PHYLOGENETIC ANALYSIS OF MORPHOLOGY AND GEOGRAPHICAL DATA

Introduction

Manfreda Salisb. is a genus of the family Asparagaceae, in the tribe Poliantheae due to possessing an inferior ovary, subterranean stems and the lack of a distal leaf spine (Verhoek-Williams, 1975; APG, 2009; Chase *et al.*, 2009). Within the tribe Poliantheae, *Manfreda* can be distinguished from *Polianthes* L. and *Prochnyanthes* S. Watson. via the commonality of flowers paired at nodes, cryptantherous stamens and style in addition to a trigonous stigma (Verhoek-Williams, 1975).

Manfreda consists of between 26-32 species dispersed across North and Central America, inhabiting a climatically diverse range between West Virginia of the American North and northern Honduras, El Salvador as well as limited documentation in Guatamala. Constituent species of *Manfreda* inhabit an equally diverse array of ecological niches from desert to pine-oak forest and high altitude chaparral, isolating populations and imposing diversification (Verhoek-Williams, 1975; Castillejos-Cruz, 2009).

Manfreda sensu Verhoek-Williams (1975) was the first comprehensive taxonomic system devised for the genus since Rose (1905). The system utilized 26 specific taxa, in addition to two subspecific taxa for the species *M. variegata* (Jacobi) Rose, devised solely upon morphological observations. The system was devised utilizing predominantly leaf characters, floral form and tepal curvature to segregate divisions within the taxon. The latter system of Castillejos-Cruz

(2009) was derived from Verhoek-Williams (1975) and employed 32 specific taxa, accounting for the documentation of new species. The system utilized both morphological observations and analysis of variance. Although Castillejos-Cruz (2009) adjusted little as to the taxonomic composition of *Manfreda,* the characters utilized for his key differed substantially. As well as leaf characters, plant size and bract form were promoted to premier diagnostic characters with further utilization of floral and fruit forms employed for interspecific identification.

Phylogenetic inference of specific delimitation in the genus *Manfreda* is limited to few affiliated studies. Bogler and Simpson (1995) utilized chloroplast DNA restriction site analysis to investigate generic relationships in the former family Agavaceae. Similarly, Bogler and Simpson (1996) re-examined their research applying Maximum Parsimony Analysis to the Intergeneric Transcribed Spacer (ITS) gene region. Both studies utilized only *M. scabra* (Ortega) McVaugh and *M. virginica* (L.) Salisb. ex Rose yet demonstrated monophyly of the genus and differentiation between the two species resulting from orthologous substitutions shared between *Agave lechuguilla* Torr. and *M. virginica.* Paraphyly was however observed by Good Avila *et al* (2006) who utilized *M. hauniensis* (J.B. Petersen) Verh.-Will., *M. nanchititlensis* Matuda and *M. potosina* (B.L. Rob. & Greenm.) Rose in the analysis of speciation in *Agave* employing Maximum Likelihood and the *trnL* and *trnL-trnF* gene regions. All three taxa were interspersed between two separate clades, however the associated bootstrap values were low.

Defined diagnostic characters differentiating species of *Manfreda* are diverse, yet limited. The high instance of phenotypic plasticity and possible cytonuclear disequilibrium due to low hybridization barriers exhibited by the genus, as demonstrated by horticultural breeding efforts and a highly analogous karyology, have also made delimitation challenging (McKelvey & Sax, 1933; Lindstrom, 2006). The low number of herbarium specimens, documentation of wild

populations and supporting taxonomic studies collectively infer a low confidence in existing taxonomic systems (Oldfield, 1997; Castillejos-Cruz, 2009).

The aim of this study is to review existing species designations via means of morphometric, phylogenetic and geographical studies. In doing so, the goals are to either increase confidence in existing specific taxonomic units or further scrutinize inadequately supported designations. The overall objective is to aid composition of a system of species delimitation for *Manfreda* which is well supported and associated with consistent field characters to simplify identification and thus increase documentation, cultivation and breeding efforts.

Materials and Methods

Character selection. Characters employed in the study were adapted from the Complete Morphological Character List of Simpson (2010). The selected 58 characters included 14 quantitative and 36 categorical variables appropriate to taxa within Asparagaceae, including all apomorphic characters utilized for taxonomic discrimination in Verhoek-Williams (1975) and Castillejos-Cruz (2009) (Appendix 1).

Data collection. Plant samples identified to *Manfreda* from 22 herbaria were observed. Collections of *Manfreda* from ten herbaria were loaned to the University of Arkansas Herbarium and specimens of eight herbaria were viewed online via JSTOR Plant Science. Furthermore, the collections of both the University of Guadalajara and National Autonomous University of Mexico, Mexico, in addition to the Royal Botanic Garden Edinburgh, Scotland were examined at respective herbaria (Table 3). A total of 855 samples were recorded including 27 specific taxa designated to the genus *Manfreda* (Table 4). All quantitative characters were measured in millimeters.

Table 3. List of herbaria from Europe, Mexico and the United States from which specimens of *Manfreda* were observed for morphometric analysis.

Herbarium name	Location	Herbarium code	Number of ppecimens
Arizona State University	Tempe, AZ	ASU	58
Conservatory and Botanical Gardens of the City of Geneva	Geneva, Switzerland	G	7
Desert Botanic Garden	Phoenix, AZ	DES	25
Field Museum of Natural History	Chicago, IL	F	1
Friedrich Schiller University	Jena, Germany	JE	2
Harvard University	Cambridge, MA	A	26
Missouri Botanical Garden	Saint Louis, MO	MO	130
National Autonomous University of Mexico	Mexico City, Mexico	MEXU	183
National Botanic Garden of Belgium	Meise, Belgium	BR	3
National Museum of Natural History	Paris, France	P	5
Royal Botanic Garden Edinburgh	Edinburgh, Scotland	E	8
Royal Botanic Garden Kew	Richmond, England	K	6
Royal Botanic Garden Madrid	Madrid, Spain	MA	2
Russian Academy of Science	Moscow, Russia	MHA	2
Smithsonian Institute Museum of Natural History	Washington, DC	US	13
Texas A & M University	College Station, TX	TAMU	3
The Natural History Museum	London, England	BM	3
University of Arizona	Tucson, AZ	ARIZ	104
University of Arkansas	Fayetteville, AR	UARK	64
University of Guadalajara	Zapopan, Mexico	IBUG	141
University of Michigan	Ann Arbor, MI	MICH	15

Table 4. List of specific taxonomic units employed in the study and of which herbarium specimens were observed. Taxa denoted with (*) did not have sufficient representative samples to be included in the Principal Component Analysis.

Specfic epithet	Author	Publication
Manfreda brunnea*	(S. Watson) Rose	Contr. U. S. Natl. Herb. 8: 21 1903
Manfreda chamelensis	E.J.Lott & Verh.-Will.	Phytologia 70: 366 1991
Manfreda elongata	Rose	Contr. U. S. Natl. Herb. 8: 21 1903
Manfreda fusca*	(Ravenna) Thiede & Eggli	Herbertia 43: 17 1987
Manfreda guttata	(Jacobi & C.D.Bouché) Rose	Contr. U. S. Natl. Herb. 8: 21 1903
Manfreda hauniensis*	(J.B. Petersen) Verh.-Will.	Brittonia 30: 165 1978
Manfreda involuta	McVaugh	Fl. Novo-Galiciana 15: 231 1989
Manfreda jaliscana	Rose	Contr. U. S. Natl. Herb. 8: 21 1903
Manfreda littoralis	García-Mend., A.Castañeda & S.Franco	Acta Bot. Mex. 50: 39 2000.
Manfreda longibracteata*	Verh.-Will.	Brittonia 30: 166 1978
Manfreda longiflora*	(Rose) Verh.-Will.	Baileya 19: 163 1975
Manfreda maculata	(Mart) Rose	Contr. U. S. Natl. Herb. 8: 21 1903
Manfreda maculosa	(Hook.) Rose	Contr. U. S. Natl. Herb. 8: 21 1903
Manfreda nanchititlensis	Matuda	Anales Inst. Biol. Univ. Nac. Autón. México, Bot. 43: 54 1972
Manfreda parva	Aarón Rodr.	Acta Bot. Mex. 88: 2 2009
Manfreda planifolia	(S. Watson) Rose	Contr. U. S. Natl. Herb. 8: 21 1903
Manfreda potosina	(B.L.Rob. & Greenm.) Rose	Contr. U. S. Natl. Herb. 8: 21 1903
Manfreda pringlei	Rose	Contr. U. S. Natl. Herb. 8: 21 1903
Manfreda pubescens	(Regel & Ortgies) Verh.-Will. ex Espejo & López-Ferr.	Monocot. Mexic. Sinopsis Flor. 1(1): 35 1993
Manfreda revoluta	(Klotzsch) Rose	Contr. U. S. Natl. Herb. 8: 21 1903
Manfreda rubescens	Rose	Contr. U. S. Natl. Herb. 8: 21 1903
Manfreda scabra	(Ortega) McVaugh	Fl. Novo-Galiciana 15: 234 1989
Manfreda sileri*	Verh.-Will.	Brittonia 30: 168 1978
Manfreda singuliflora	(S. Watson) Rose	Contr. U. S. Natl. Herb. 8: 21 1903
Manfreda umbrophila	Garcia-Mend.	Rev. Mex. Bio. 82: 747 2011
Manfreda variegata	(Jacobi) Rose	Contr. U. S. Natl. Herb. 8: 21 1903
Manfreda virginica	(L.) Salisb. ex Rose	Contr. U. S. Natl. Herb. 8: 21 1903

Quantitative Characters Analysis. Of the 14 quantitative variables, only nine were selected for morphometric analysis, due to the inconsistent presence of available date for the residual five variables (Table 5). All sample records with missing data were excluded from the analysis and all taxonomic units with less than three complete sample records were also excluded. Utilizing JMP® 9.0, descriptive statistics including mean and standard deviation were generated for each of the residual 145 complete sample records via distribution function. Dot plots were also produced by correlating taxon names and each variable individually, demonstrating both intra and interspecific variation employing the graph builder function.

Table 5. List of the nine quantitative characters employed for morphometric study via Principal Component Analysis of *Manfreda* herbarium specimens.

Character name	Character description
Plant height	Measure from ground level to apex of longest leaf
Leaf length	Mean of the three longest leaves, measured from rosette to leaf apex
Leaf width	Mean of the three widest leaves, measured at the widest portion
Inflorescence length	Measure from base of the rosette to the peduncle apex
Flower width	Mean of three widest flowers, measured at the widest portion of the perianth
Calyx length	Mean of three longest tepals, measure from the base of the receptacle attachment to the apex of the tepal
Filament length	Mean of three longest filaments, measured from the receptacle to the point of anther attachment
Ovary length	Measure from apex of receptacle to base of style
Style length	Measure from apex of ovary to base of stigma

For Principal Component Analysis, all measures were standardized to z-scores to prevent bias towards anatomical proportions. Employing the Principal Component Analysis platform in JMP® 9.0, eigenvectors, eigenvalues and a Scree Plot were initially synthesized. The first three principal components, representing 71% of variation within the dataset, were selected for construction of the principal components. A triphasic score plot, plotting each permutation of bifurcate plots, was created to represent the data.

Categorical Characters Analysis. Phylogenetic analysis was conducted under Maximum Parsimony in PAUP* 4.0 and Bayesian Analysis via MrBayes 3.1. The analysis consisted of 29 taxa and 36 categorical characters transformed to numeric values ranging from 1-3 (Table 6). Consensus sequences were generated for each taxonomic unit from all 855 samples by calculating the mode for each character. Of the 1037 character states, nine were missing and denoted "?" indicating gaps. Outgroups included in the analysis were *Polianthes tuberosa* L. due to the sister taxa status attributed to *Polianthes* and *Agave americana* L. as an outgroup of greater evolutionary distance, yet pertaining close affinity within Asparagaceae and subfamily Agavoideae. Both outgroups were designated and were represented as a monophyletic sister group to the monophyletic ingroup.

Table 6. List of categorical characters employed in the phylogenetic analysis via Maximum Parsimony and Bayesian Analysis of herbaria specimens of *Manfreda*.

Character name	Character state 1	Character state 2	Character state 3
Root type	Adventious	Other	
Underground stem	Bulb	Other	
Arial stem type	Rosette	Other	
Leaf type	Simple	Compound	
Leaf succulence	Succulent	Semi-succulent	Non-succulent
Leaf attachment	Sessile	Petiolate	
Leaf arrangement	Rosette	Other	
Leaf blade shape	Linear	Narrowly elliptical	
Leaf blade margin	Entire	Serrate	
Leaf blade apex	Acute	Spinose	
Leaf pubescence	Pubescent	Glabrous	
Inflorescence type	Scapose	Other	
Flower arrangement	Spicate with one flower per node	Spicate with two flowers per node	Paniculate
Floral symmetry	Actinomorpic	Zygomorphic	
Floral attachment	Pedicellate	Sessile	
Perianth cycly	Uniseriate	Biseriate	
Perianth type	Homochylamydeous	Dichlamydeous	
Tepal fusion	Synsepalous	Aposepalous	
Perianth symmetry	Actinomorpic	Aposepalous	
Stamen type	Filamentous	Laminar	
Stamen fusion	Apostemonous	Other	
Stamen insertion	Single series	Two series	
Anther attachement	Dorsifixed	Basifixed	
Anther dehiscence	Longitudinal	Other	
Anther type	Dithecal	Monothecal	
Ovary position	Inferior	Superior	
Ovary shape	Globose	Intermediate	Elliptical
Style number per pistil	One	Two	
Style position	Exserted	Inserted	
Number of stigma	One	Two	
Stigma position	Terminal	Other	
Number of carpels	One	Two	Three

A Pairwise Distance Matrix was generated exhibiting pairwise morphological distances between each combination of paired sequences calculated from the sum of all base pair differences. Default parameters were employed for substitutions and gap penalties (Appendix 2).

Maximum Parsimony. A heuristic search was selected for Maximum Parsimony Analysis due to the number of taxa and characters employed in the analysis. Data type was elected as standard due to its applicablility to numerical datasets, Bootstrap replicates were set at 100, all characters were unordered and equally weighted while gaps were treated as missing. The heuristic search was supplemented by Stepwise Addition with a singular tree held at each step. Additional branch swapping was conducted via the default Tree-Bisection-Reconnection (TBR) utility. The Bootstrap 50% Majority Rule was implemented in the selection of a consensus tree.

Bayesian Analysis. Analysis was conducted on a Mac Pro "Quad Core" 3.1 using the MrBayes 3.1.2. program (Huelsenbeck & Ronquist, 2001). Four Markov Monte Carlo chains were employed, three warm and one cool with default parameters. Generations were set at 5,000,000 repetitions with sampling every 1000 replicates. Burn-in was conducted on 500,000 trees (10%). Default settings were utilized for the analysis.

Tree editing was conducted in TreeView 1.6.6., tree style was selected as rectangular cladogram and tree order was defined as ladderside right. Trees were saved to an enhanced metafile format.

Geographical Distribution Analysis. Of the 855 samples observed, full geographical coordinates were printed on 453 samples and the residual 402 coordinates were estimated employing geographical locations included in the footnote. Estimation of coordinate was attributed a 50 mile margin of error. Coordinates were analyzed employing JMP® 9.0,

coordinates were separated into latitude and longitude then formatted to the Longitude DMS and Latitude DMS formats, respectively. Latitude and longitude for each taxon were plotted utilizing the Graph Builder platform against the World Countries Map. Taxa were differentiated by point coloration and denoted in the association legend.

Results

Loading Matrix and Plot. The Loading Matrix (Figure 2) illustrates the weighting applied by each character on each principal component. The associated Loading Plot (Figure 3) demonstrates such 'factor loading' via line lengths while direction aids navigation of the principal components by illustrating the directionality imposed on each specimen by each character.

Loading Matrix

	Prin1	Prin2	Prin3	Prin4	Prin5	Prin6	Prin7	Prin8	Prin9
Std Plant Height	0.80047	-0.52091	0.14492	0.03563	0.14015	-0.04126	0.20270	-0.01673	0.05403
Std Leaf Length	0.80083	-0.51545	0.14316	0.03740	0.14166	-0.03740	0.21572	-0.01503	-0.05351
Std Leaf Width	0.36986	0.30574	-0.67944	0.51696	-0.02899	0.02003	0.19892	-0.00529	0.00053
Std Inflorescence Length	0.71295	-0.32068	-0.04754	0.32309	0.07262	0.18765	-0.49128	0.01935	-0.00079
Std Calyx Width	0.07665	0.51659	0.62691	0.54754	-0.00079	-0.17922	0.03536	0.03290	0.00012
Std Calyx Length	0.55971	0.43766	0.19032	-0.13640	-0.12032	0.64209	0.11577	-0.01385	0.00044
Std Filament Length	0.81189	0.31283	-0.00707	-0.17177	-0.28365	-0.26287	-0.10159	-0.23138	-0.00045
Std Ovary Length	0.40725	0.57562	-0.08447	-0.23426	0.65542	-0.08772	-0.05917	0.00550	0.00004
Std Style Length	0.84240	0.20458	-0.07111	-0.25021	-0.28005	-0.20740	-0.02286	0.24268	0.00007

Table 7. The Loading Matrix exhibits the load attributed to each principal component by each of the nine quantitative characters during Principal Component Analysis.

Examination of the Loading Matrix shows that Principal Components 1 is substantially influenced by six characters ranging from 0.56 for Standardized (Std) Calyx Length to Std Style Length at 0.84. The other characters were Std Leaf Width at 0.36 and Std Calyx Width with a score of 0.08, these character had little influence on Principal Component 1.

Five characters imposed considerable influence on Principal Component 2 ranging from -0.52 for Std Plant Height to 0.57 for Std Ovary Length. The characters which possessed scores of between -0.50 to 0.50, Std Filament Length at 0.31, Std Leaf Width at 0.30, Std Style Length and Std Inflorescence Length at -0.32, imposed minimal influence on Principal Component 2.

Principal Component 3 exhibited only two characters with scores in excess of -0.50 and 0.50. Std Leaf Width, which possessed minimal bearing in the first two Principals Components, scored -0.67 and Std Calyx Length at 0.63.

The Loading Plot, incorporating Principal Components 1-3, demonstrated that standardized Leaf Length and Std Plant Height were the two most influential characters and that Std Leaf Width had the least influence.

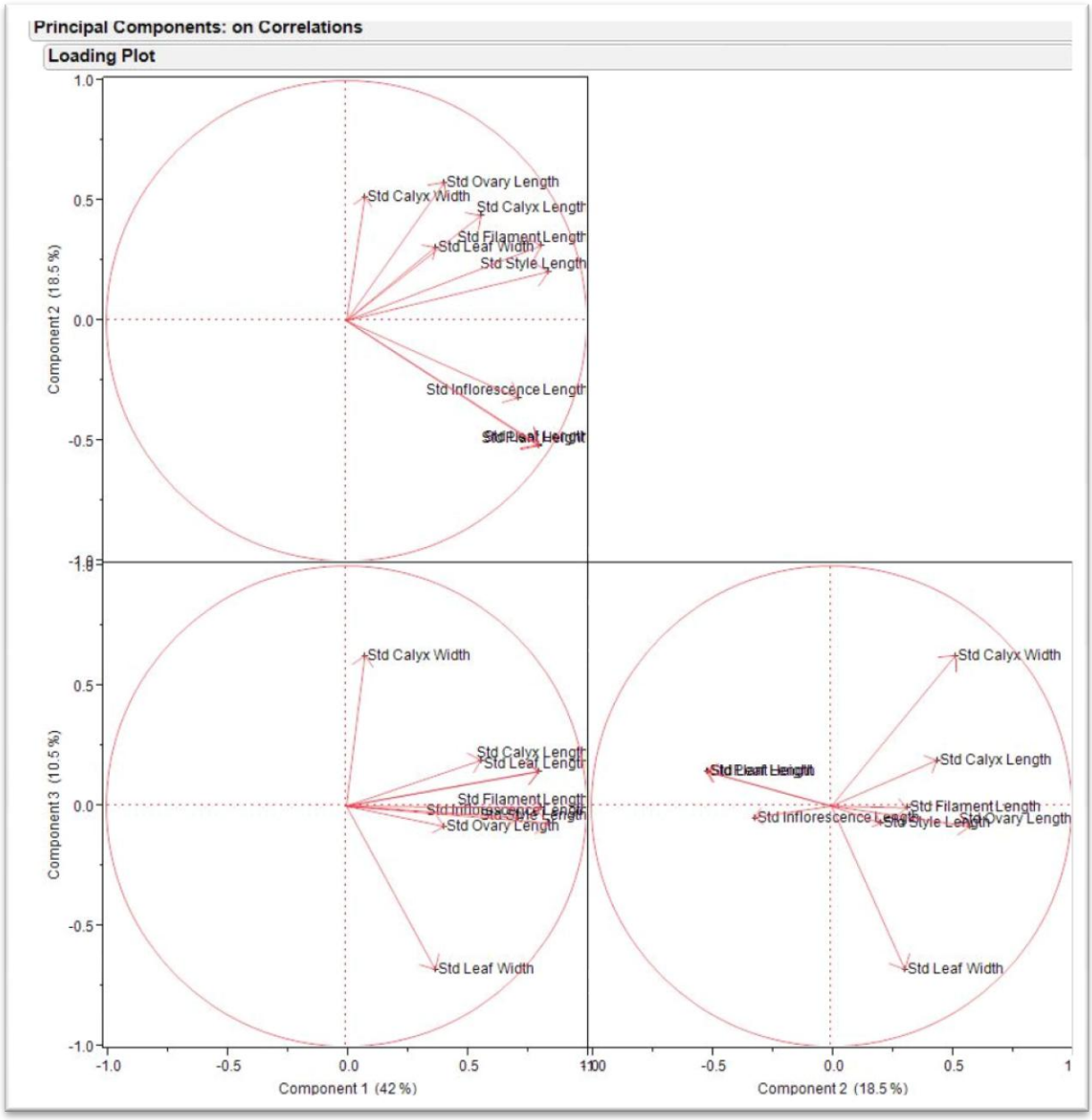


Figure 2. The Loading Plot illustrates the directionality and influence of the nine quantitative characters in the resultant score plot of Principal Component Analysis.

Scree Plot and Eigenvalues. Variation within the multivariate sample is represented by the Scree Plot (Figure 4) and eigenvalues (Figure 5). All variation within the sample is partitioned between eigenvalues associated to each principal component.

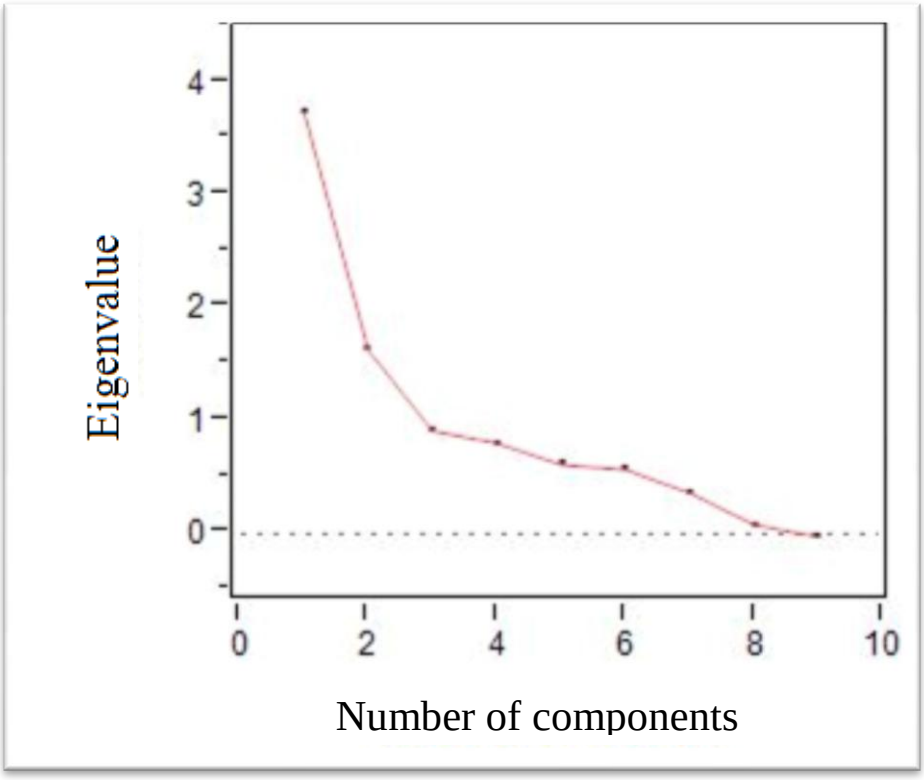


Figure 3. The Scree Plot demonstrates the allocation of eigenvalues via Principal Component Analysis to each principal component in a line graph.

Eigenvalues

Number	Eigenvalue	Percent	20 40 60 80	Cum Percent	ChiSquare	DF	Prob>ChiSq
1	3.7810	42.011		42.011	1090.15	44.000	<.0001*
2	1.6628	18.476		60.487	789.611	35.000	<.0001*
3	0.9469	10.521		71.007	672.259	27.000	<.0001*
4	0.8397	9.330		80.337	627.611	20.000	<.0001*
5	0.6488	7.208		87.546	574.037	14.000	<.0001*
6	0.6029	6.699		94.245	526.794	9.000	<.0001*
7	0.3975	4.417		98.662	427.919	5.000	<.0001*
8	0.1146	1.274		99.936	244.633	2.000	<.0001*
9	0.0058	0.064		100.000	0.000	0.000	.

Figure 4. Eigenvalues are a mathematical description of the amount of variability hosted by each principal component. Eigenvalues are attributed to each principal component by conducting Principal Component Analysis.

Principal Component 1 encompassed 42% of the variation within the sample. The correlated eigenvalue is 3.7810 with 44.000 degrees of freedom. Principal Component 2 represented 18.48% of the variation cumulated to 60.49% when combined with Principal Component 1. Principal Component 3 possesses 10.52% of the variation and in combination with Principal Component 1 and 2, the three largest principal components accounted for 71.01% of all variation within the sample. Principal Component 3 has an eigenvalue of 0.94 and 17 degrees of freedom.

Score Plots. The score plots representing each permutation of Principal Components 1 to 3 represents 71.007% of the variation within the dataset. Analysis for each individual taxon are subsequently outlined:

Manfreda chamelensis - Dot plot analysis exhibited limited variation, of which Standardized (Std) Leaf Width demonstrated the greatest variation and Std Filament Length demonstrated the least (Figure 6). In score plots 1-3 all four specimens were located in close proximity, demonstrating close correlation between specimens. Scores for Principal Component 1 ranged from 0.5298 to 0.9596, scores for Principal Component 2 range from -1.29 to -2.00 and scores for Principal Components 3 range from -0.96 to 0.33. No evident single character was influential in defining the cluster (Figure 7).

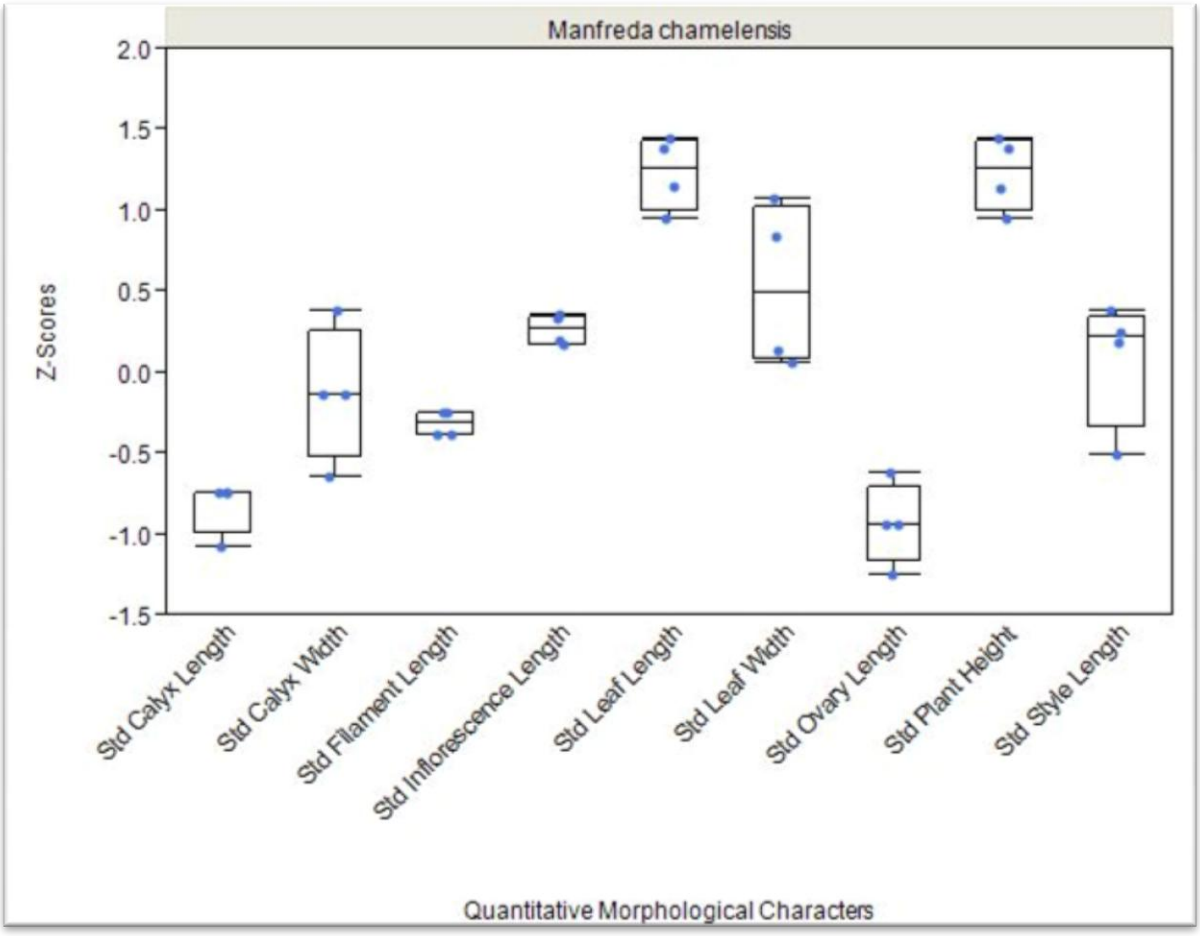


Figure 5. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda chamelensis.*

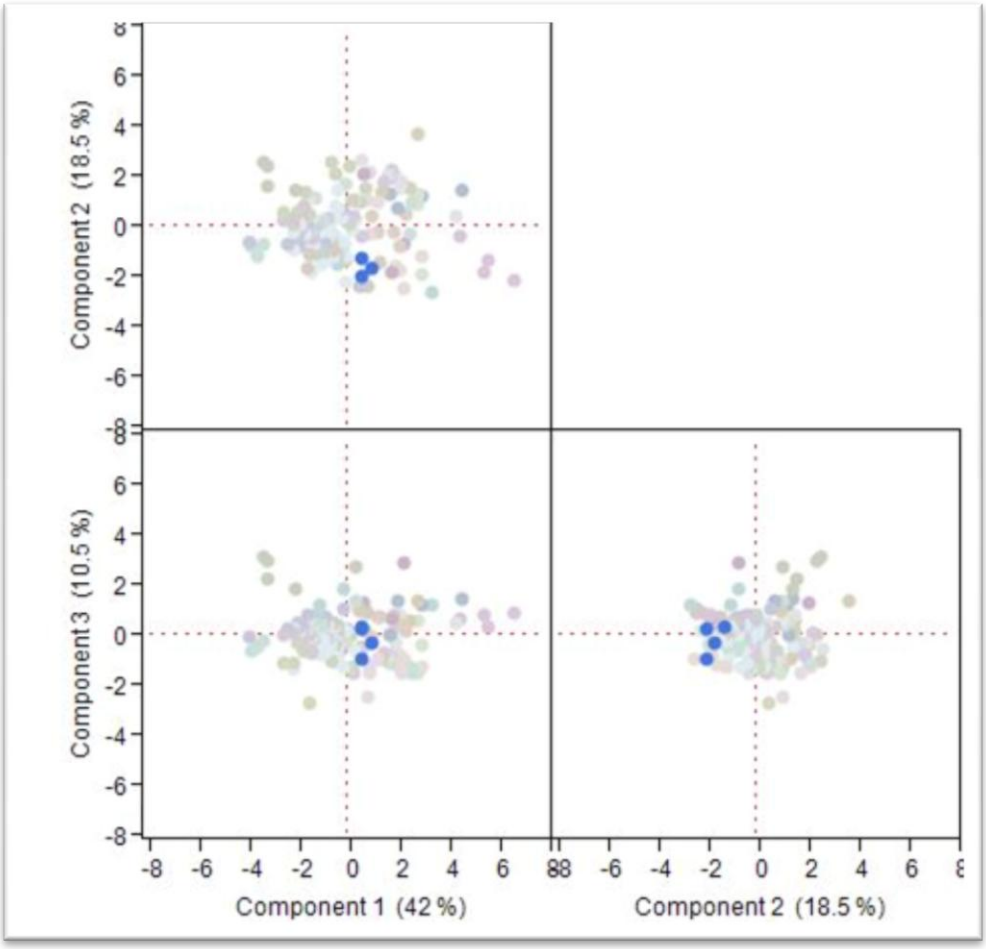


Figure 6. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda chamelensis* are highlighted by dark blue.

Manfreda elongata - The characters with the greatest variation observed in the dot plot were Std Style Length and Std Calyx Width, while the least variation was attributed to Std Ovary Length (Figure 8). Score plots 1 and 2 demonstrated close alignment, yet in score plot 3 the distribution of the taxon was disjunct. Scores ranged from 0.9288 to 2.3160, -0.8043 to 0.5041 and -0.1022 to 0.7346 for Principal Components 1-3 respectively. The disjunct nature of Principal Component 3 exhibited two clusters predominantly influenced by Std Leaf Width and Std Calyx Width, the two prominent variables of Principal Component 3 (Figure 9).

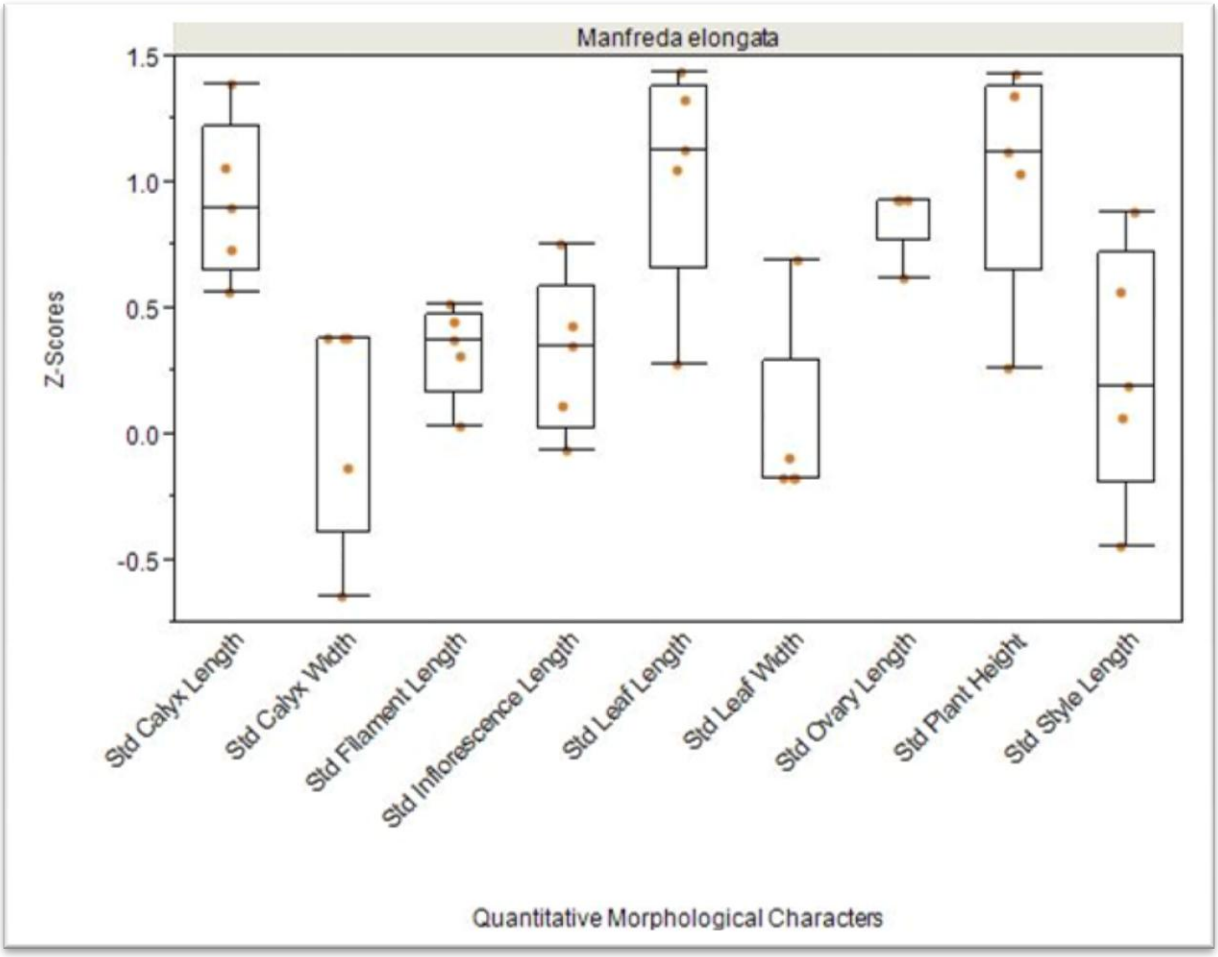


Figure 7. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda elongata.*

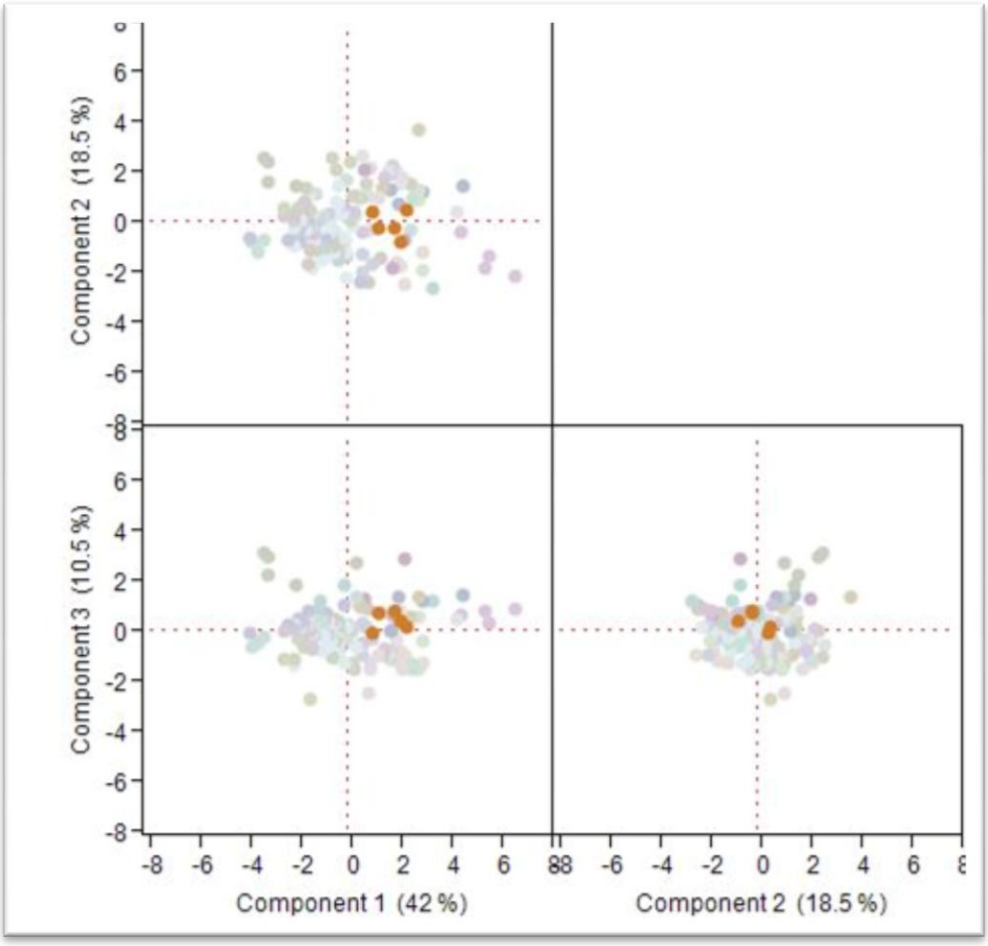


Figure 8. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda elongata* are highlighted by brown.

Manfreda guttata - Substantial variation was exhibited in the dot plot with Std Leaf Length and Std Plant Height demonstrating the greatest uniformity (Figure 10). Score plots 1-3 exhibited both a disjunct and broad distribution. Scores ranged for the 12 specimens from -1.8014 to 1.7337 for Principal Component 1, -0.4459 to 2.2570 for Principal Component 2 and -1.5522 to 0.2330 for Principal Component 3. In score plot 1, four clusters were apparent, Std Ovary Length separated the apical cluster while the other three clusters are not differentiated by any single character. Score plot 2 demonstrated an ambiguous and sparse distribution, while score plot 3 contained a central cluster with outliers allocated to the right and no clear separation by a single character (Figure 11).

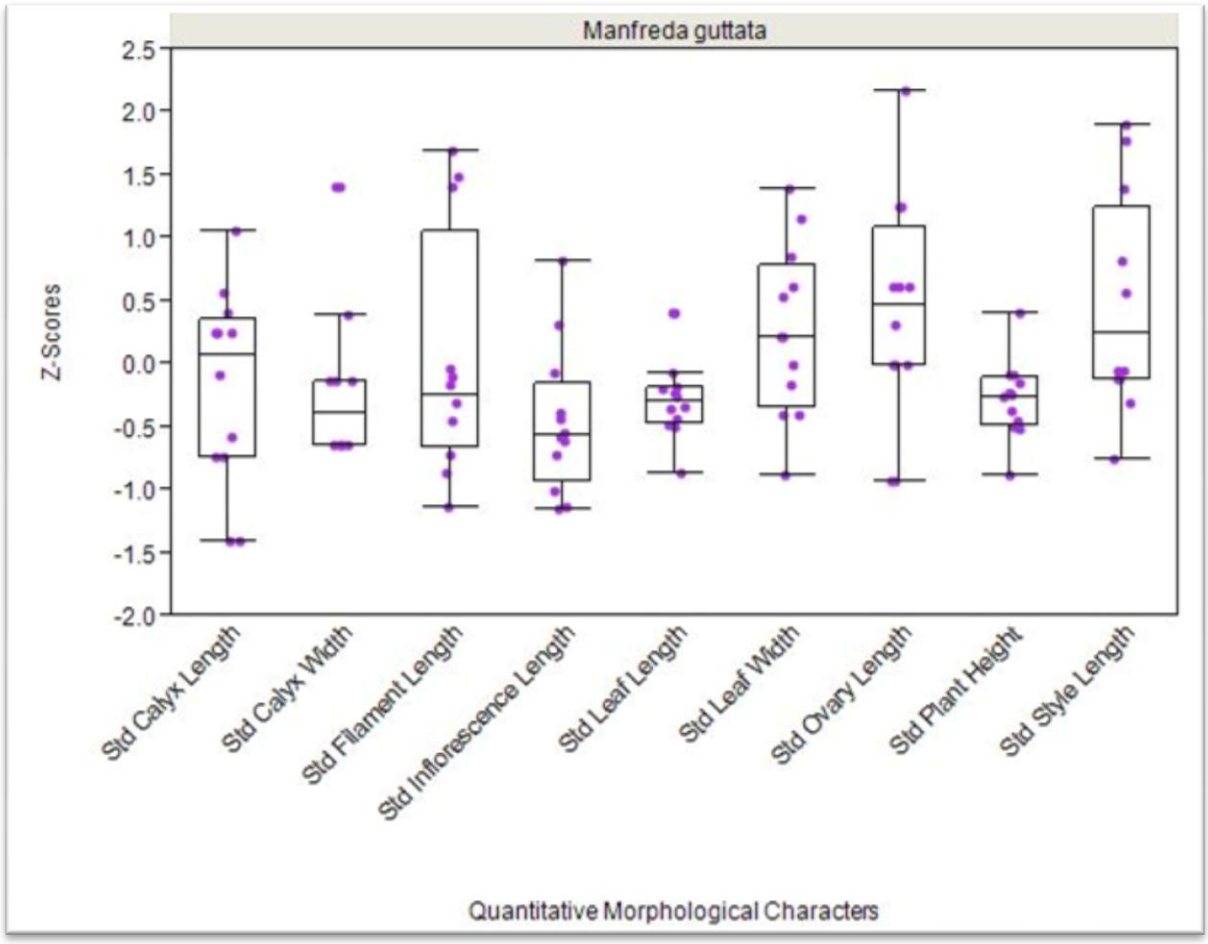


Figure 9. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda guttata.*

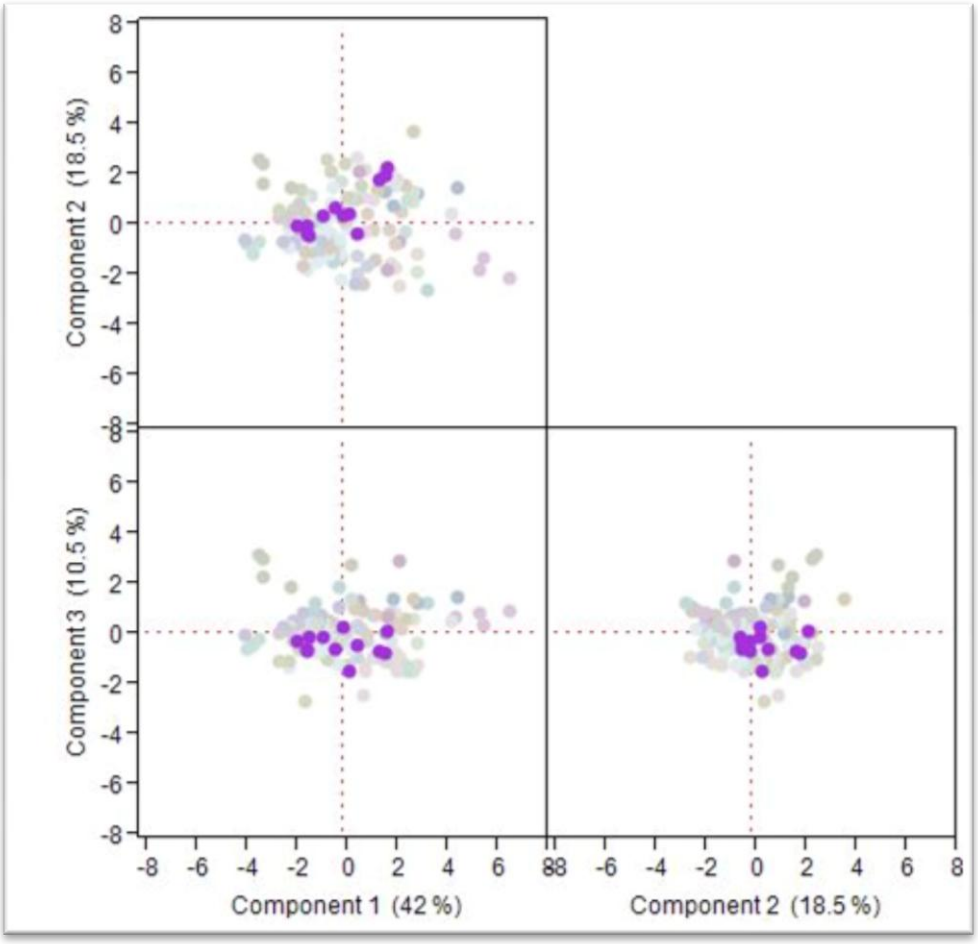


Figure 10. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda guttata* are highlighted by purple.

Manfreda involuta - The greatest variation observed in the dot plot was attributed to the characters Std Inflorescence Length, Std Leaf Length and Std Plant Height. However, in each of the aforementioned characters, a substantial degree of the variation could be attributed to a single outlier (Figure 12). The distribution of the five *M. involuta* specimens in score plots 1-3 consist of one close cluster and one distinct outlier (Figure 13). Scores ranged from 1.2763 to 3.3922, 2.6399 to 0.1511 and 0.4281 to 1.8012 for Principal Components 1-3, respectively. The central cluster was well defined yet the singular outlier was distinct and disjunct in relation to the rest of the samples. From score plot 1 the outlying specimen showed substantial differentiation based on Std Inflorescence Length, in score plot 2 the affinity was closer related to Std Style Length and in score plot 3 no evident character affinity was distinguishable.

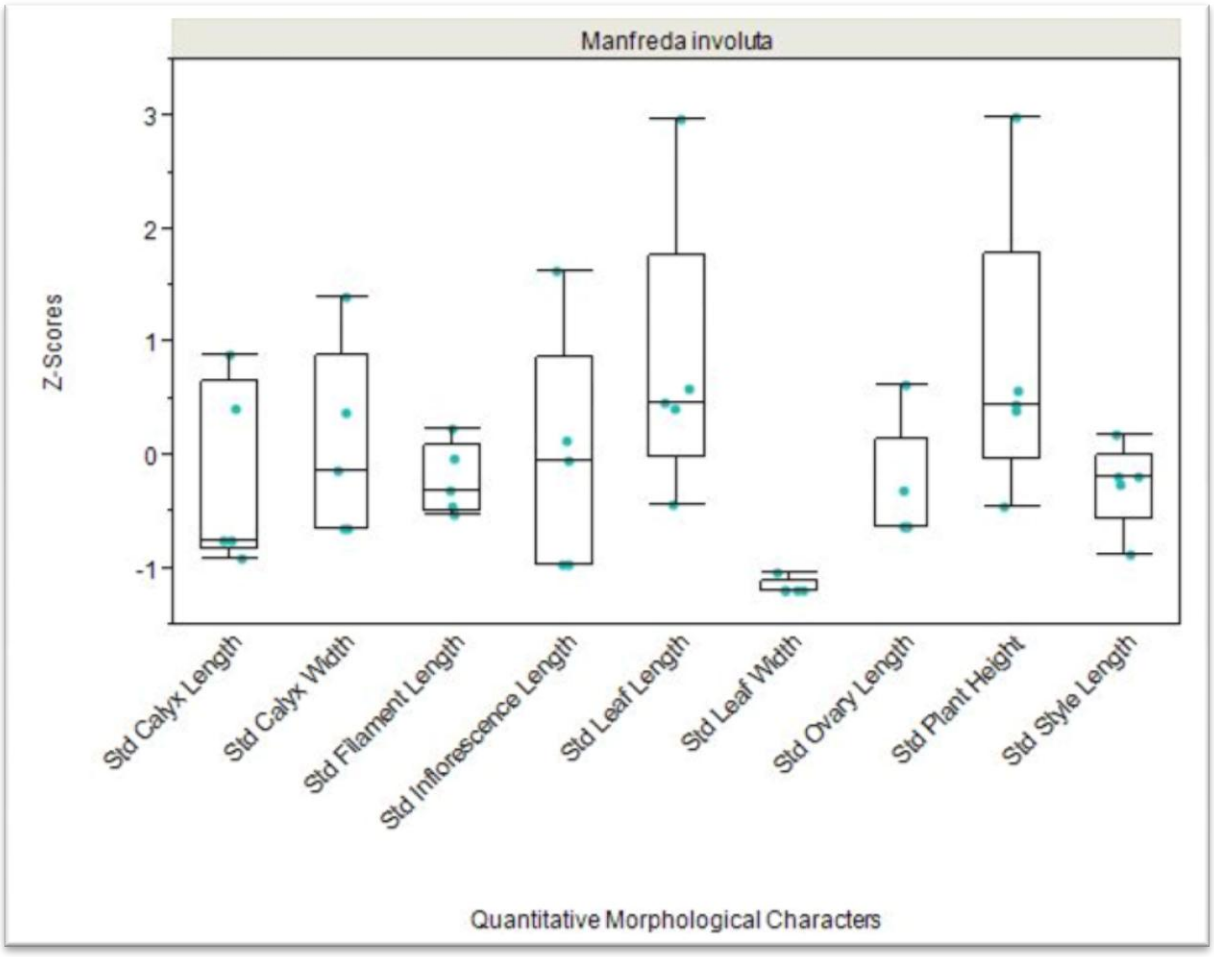


Figure 11. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda involuta.*

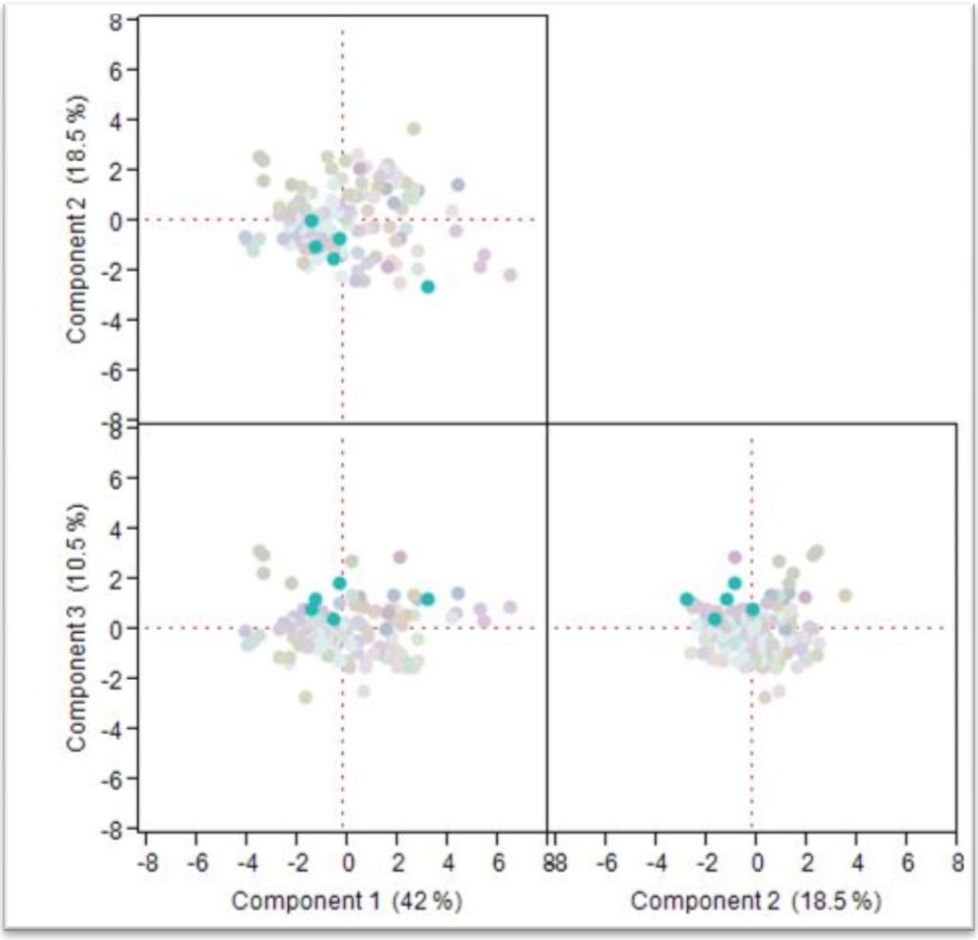


Figure 12. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda involuta* are highlighted by turquoise.

Manfreda jaliscana - Dot plot analysis demonstrated substantial uniformity between characters examplified by Std Filament Length and Std Ovary Length, while other characters exhibited sizeable variation such as Std Leaf Length and Std Plant Height (Figure 14). The distribution of the four *M. jaliscana* specimens was distinct, well defined and in score plot 1 and score plot 2, far removed from all *Manfreda* specimens. Scores for Principal Components 1-3 range from 4.45 to 6.63, -2.19 to 0.44 and 0.2937 to 0.89, respectively. Data points in score plot 1 seem to be highly influenced by Std Inflorescence Length, in score plot 2 Std Style Length was more influential and in score plot 3 no clear affiliation with any character was obvious (Figure 15).

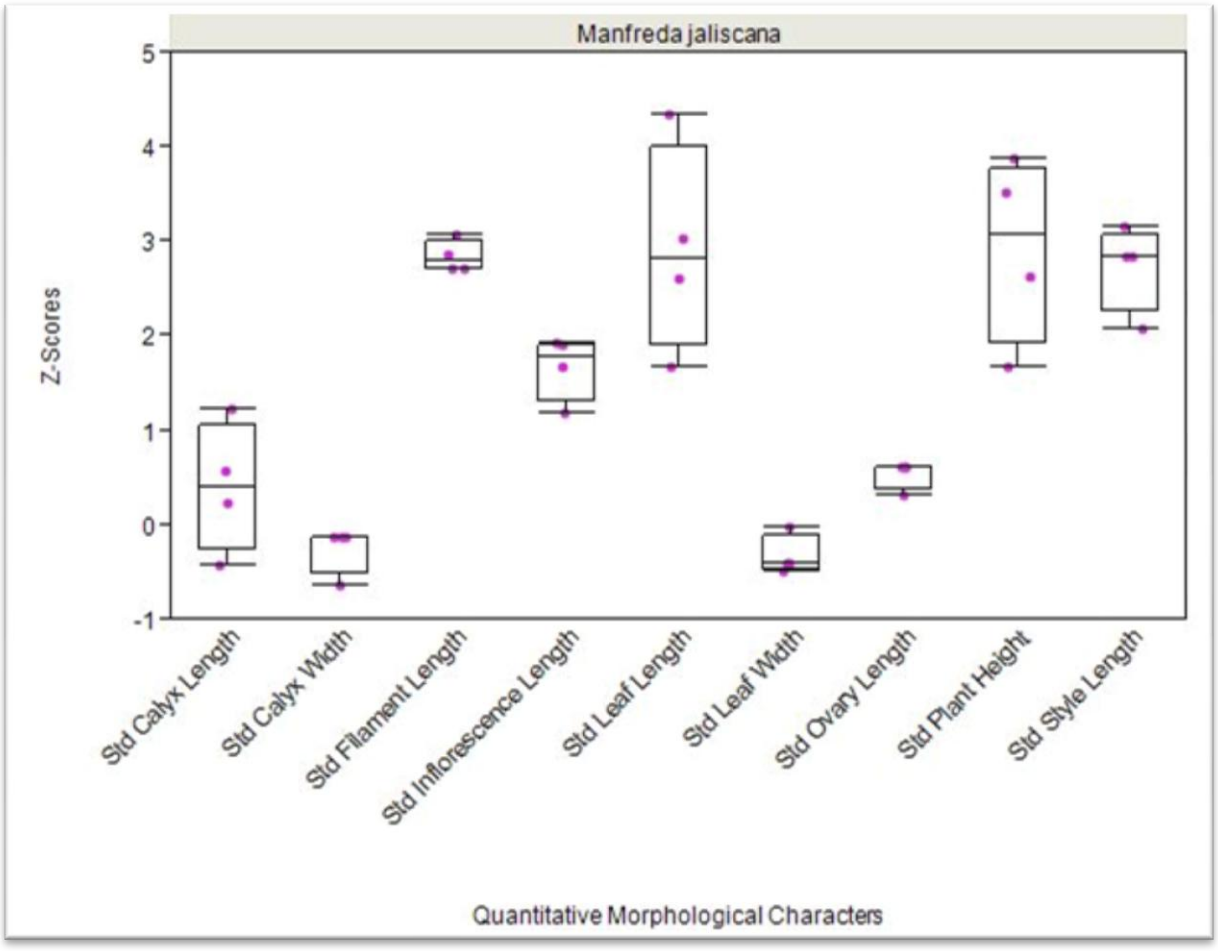


Figure 13. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda jaliscana.*

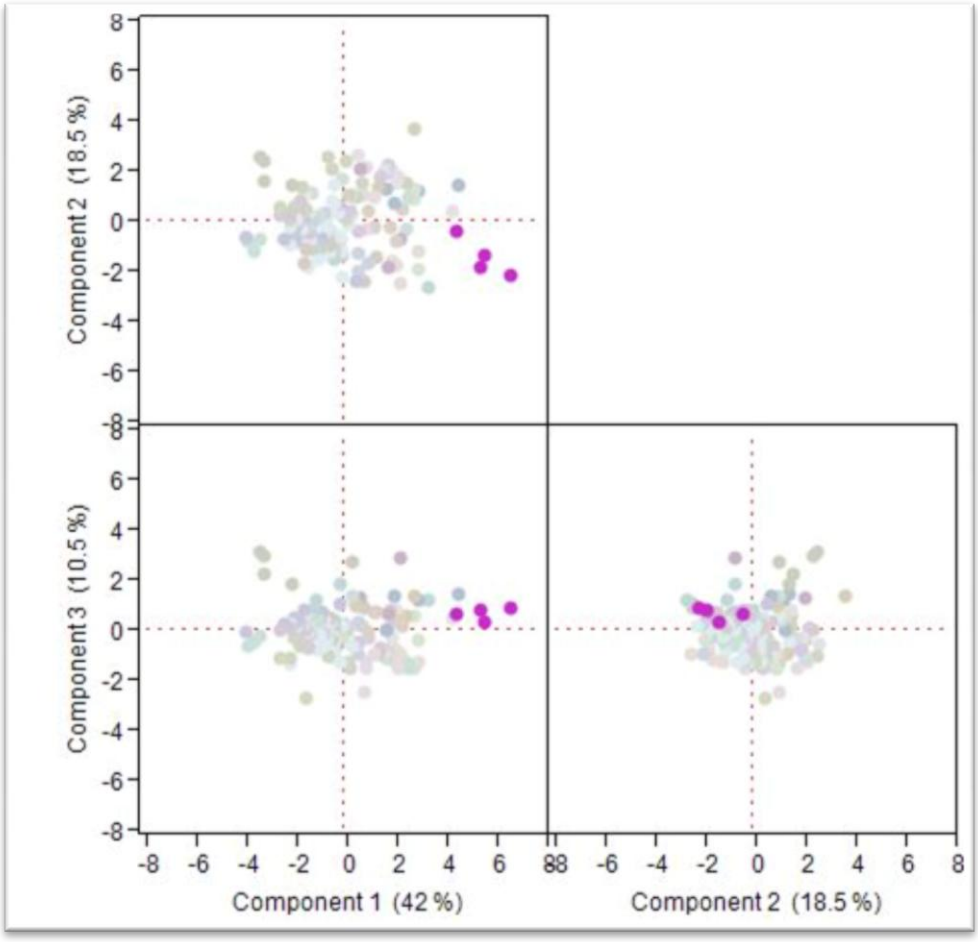


Figure 14. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda jaliscana* are highlighted by purple.

Manfreda littoralis - The dot plot exhibited a diverse range for each character with the greatest variation observed among Std Leaf Width and the least variation among Std Style Lengths (Figure 16). The limited three samples of *M. littoralis* had an ambiguous distribution pattern in all three score plots. Scores ranged from -2.57 to 0.06 for Principal Component 1, 0.06 to 0.55 for Principal Component 2 and -2.71 to 1.14 for Principal Component 3. No clustering or obvious strong correlations with a single character was evident, possibly attributed to small sample size (Figure 17).

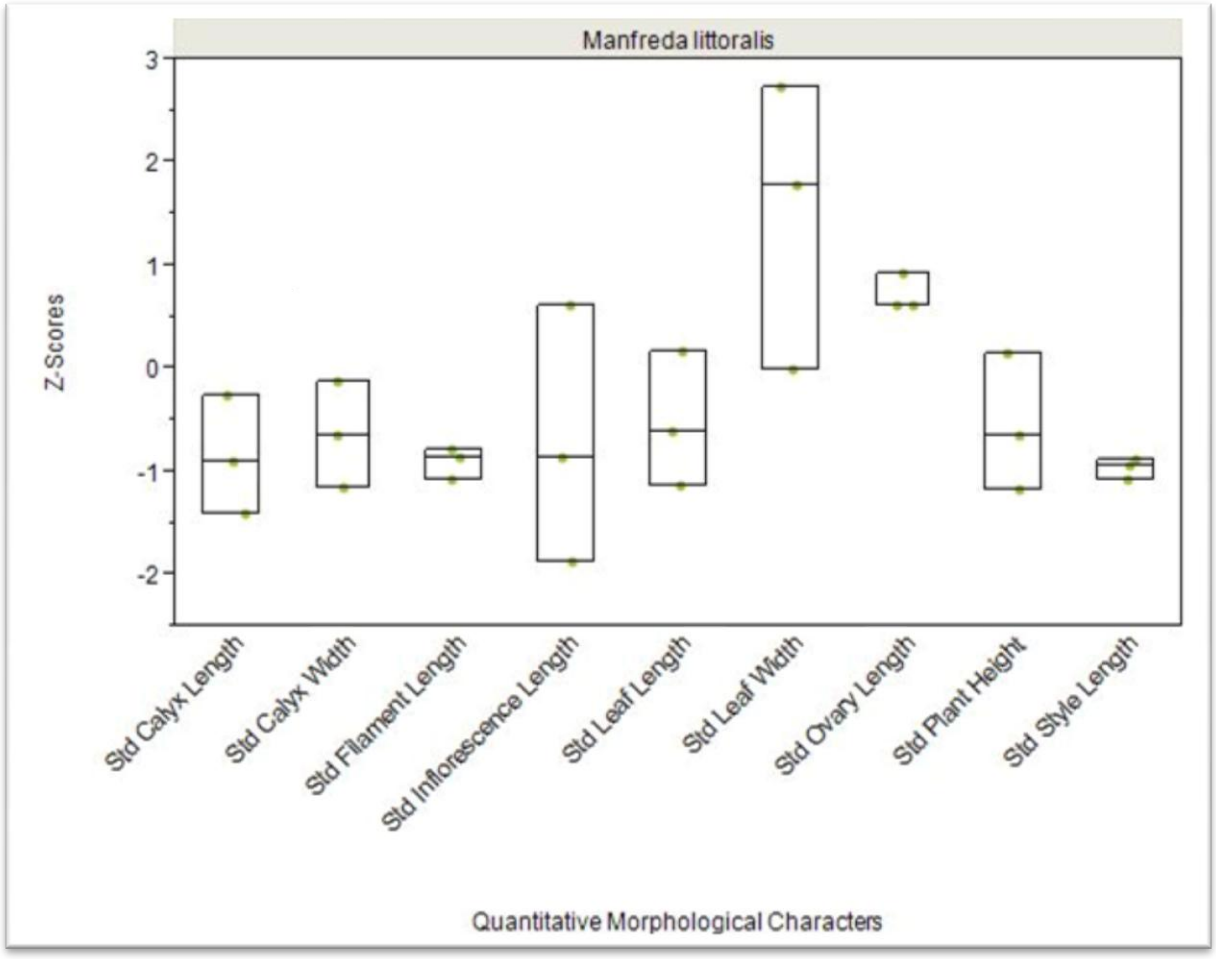


Figure 15. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda littoralis*.

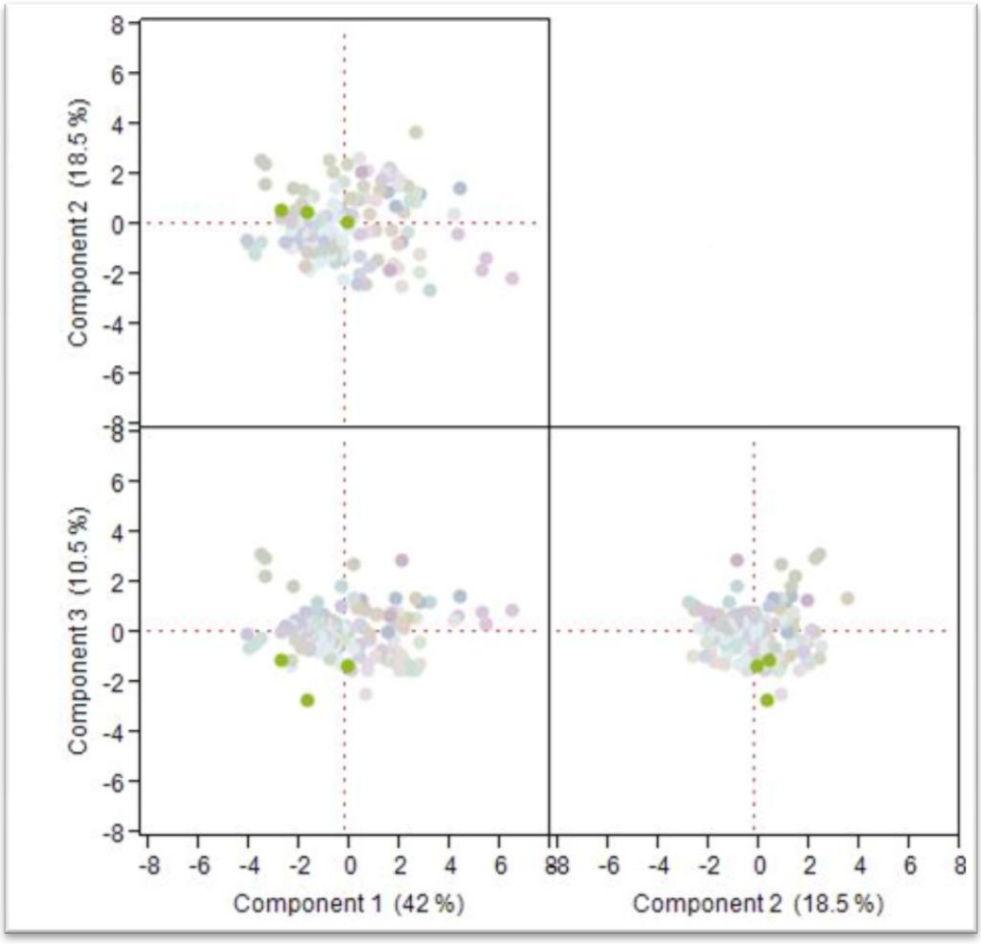


Figure 16. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda littoralis* are highlighted by green.

Manfreda maculata - Of the dot plot, the greatest uniformity was observed in Std Style Length, while the greatest variation iwas observed in Std Ovary Length due predominetly to one exceptional outlier (Figure 18). Of the nine specimens of *M. maculata*, a disjunct distribution was observed in score plot 1 and score plot 2 as well as to a lesser extent in score plot 3. Scores ranged from -2.16 to 0.75, 0.07 to 2.54 and -1.28 to 0.35 in Principal Components 1-3, respectively. In score plot 1-2, two distinct cluster were observed consisting of four and five specimens. score plot 3 could either form two distinct cluster once more or a singular clusters, more specimens would be required to gain further insight into the distribution (Figure 19).

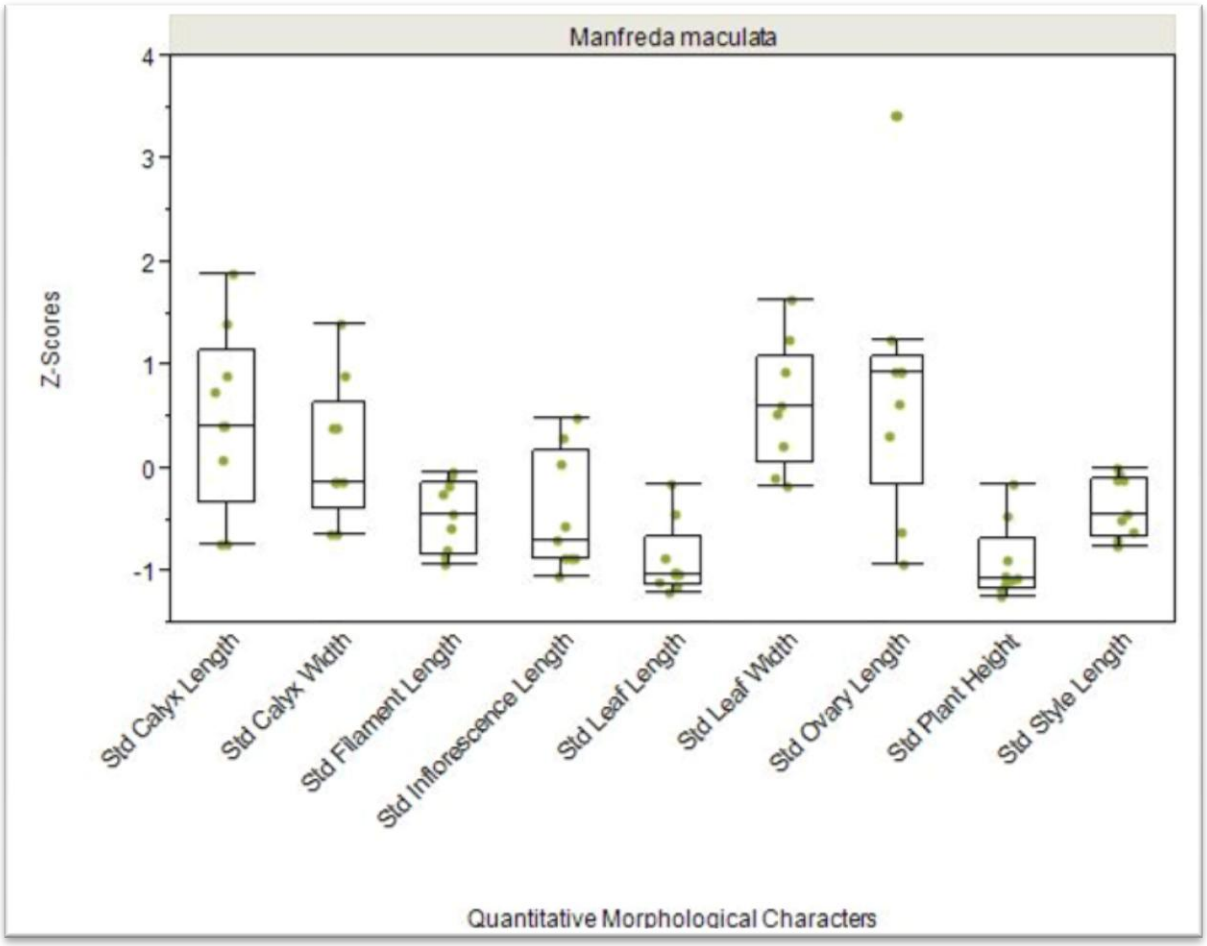


Figure 17. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda maculata*.

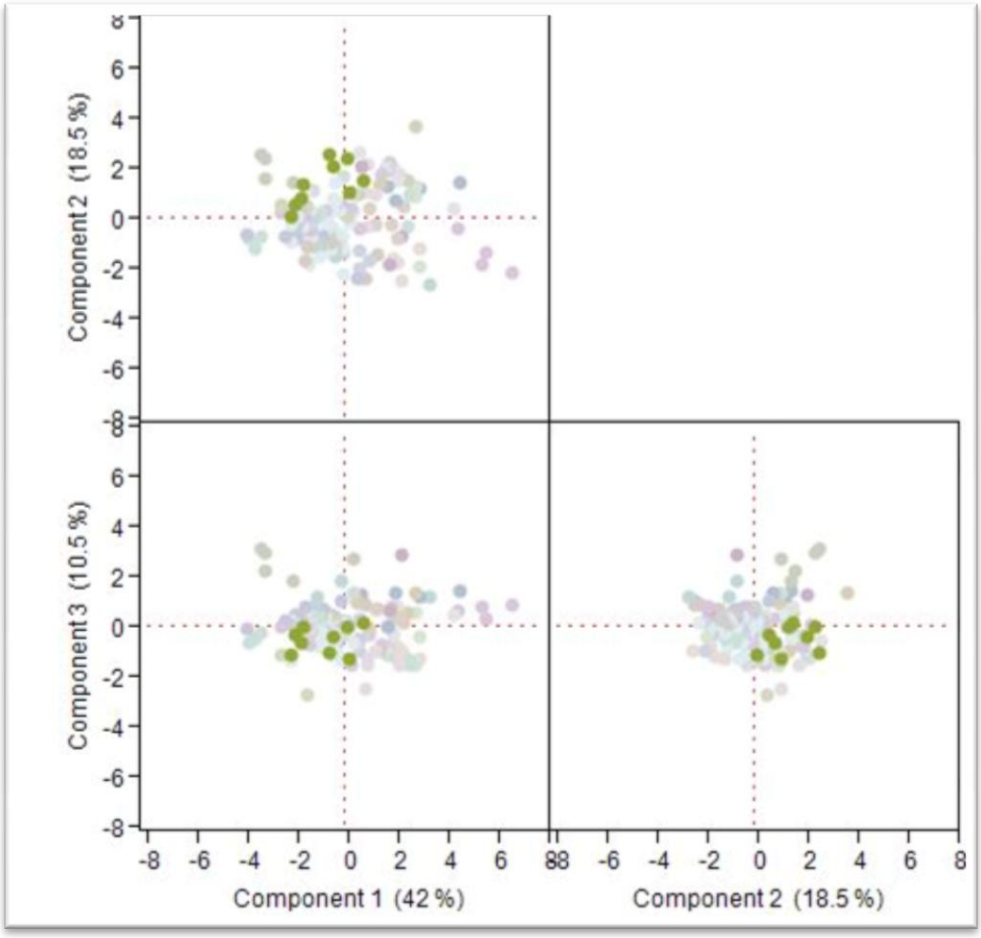


Figure 18. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda maculata* are highlighted by dark green.

Manfreda maculosa - Dot plot analysis demonstrated that the greatest variation was attributed to Std Calyx Length, while Std Leaf Length and Std Plant Height contributed the least variation (Figure 20). No distinct cluster or pattern was ascertained from the three *M. maculosa* specimens. Scores range from -1.38 to -0.11, -0.38 to 1.65 and 0.28 to 0.44 for Principal Components 1-3, respectively. A combination of an uninformative dispersal and only three specimens has caused the interpretation of the score plots to be too challenging to identify any relationships with substantial confidence (Figure 21).

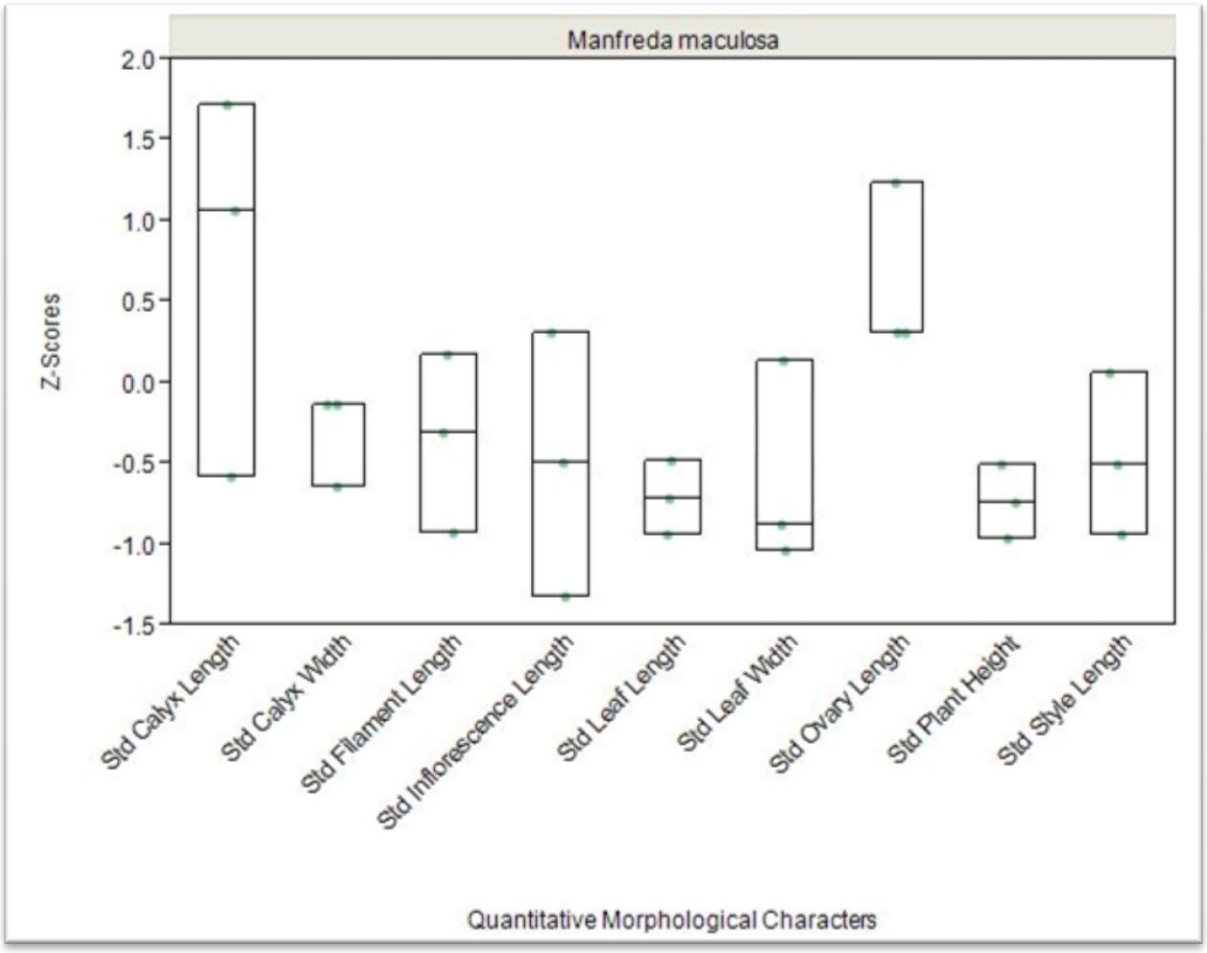


Figure 19. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda maculosa*..

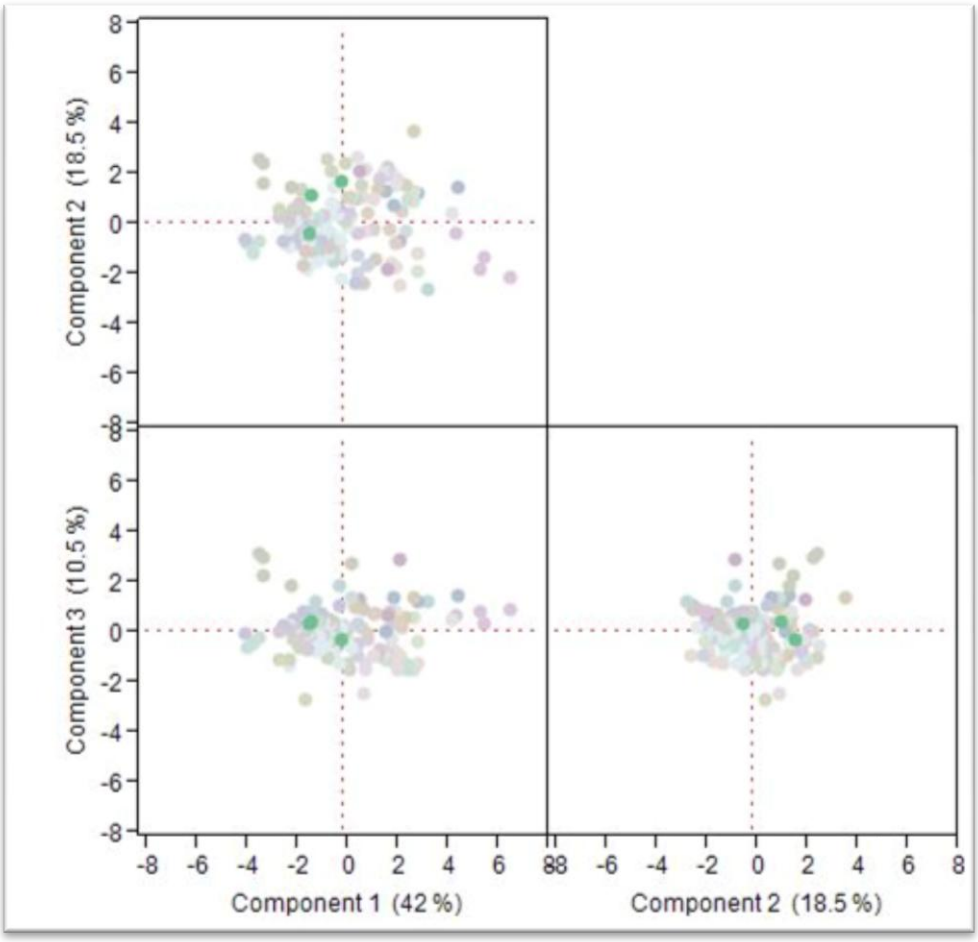


Figure 20. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda maculosa* are highlighted by green.

Manfreda nanchititlensis - Dot plot analysis exhbited substantial uniformity in Std Leaf Width, while Std Filament Length contributed the greatest variation (Figure 22). The distribution of *M. nanchititlensis* data points in each score plot demonstrated a relationship with moderate to high confidence due to a close affinity. Scores for *M. nanchititlensis* ranged from 1.6485 to 4.5945 for Principal Component 1, 0.7 to 1.47 for Principal Component 2, 0.00 to 1.44 for Principal Component 3. Score plot 1 and 2 demonstrated a strong affinity to Std Filament Length and Std Style Length. Score plot 3 showed no evident correlation to any single character (Figure 23).

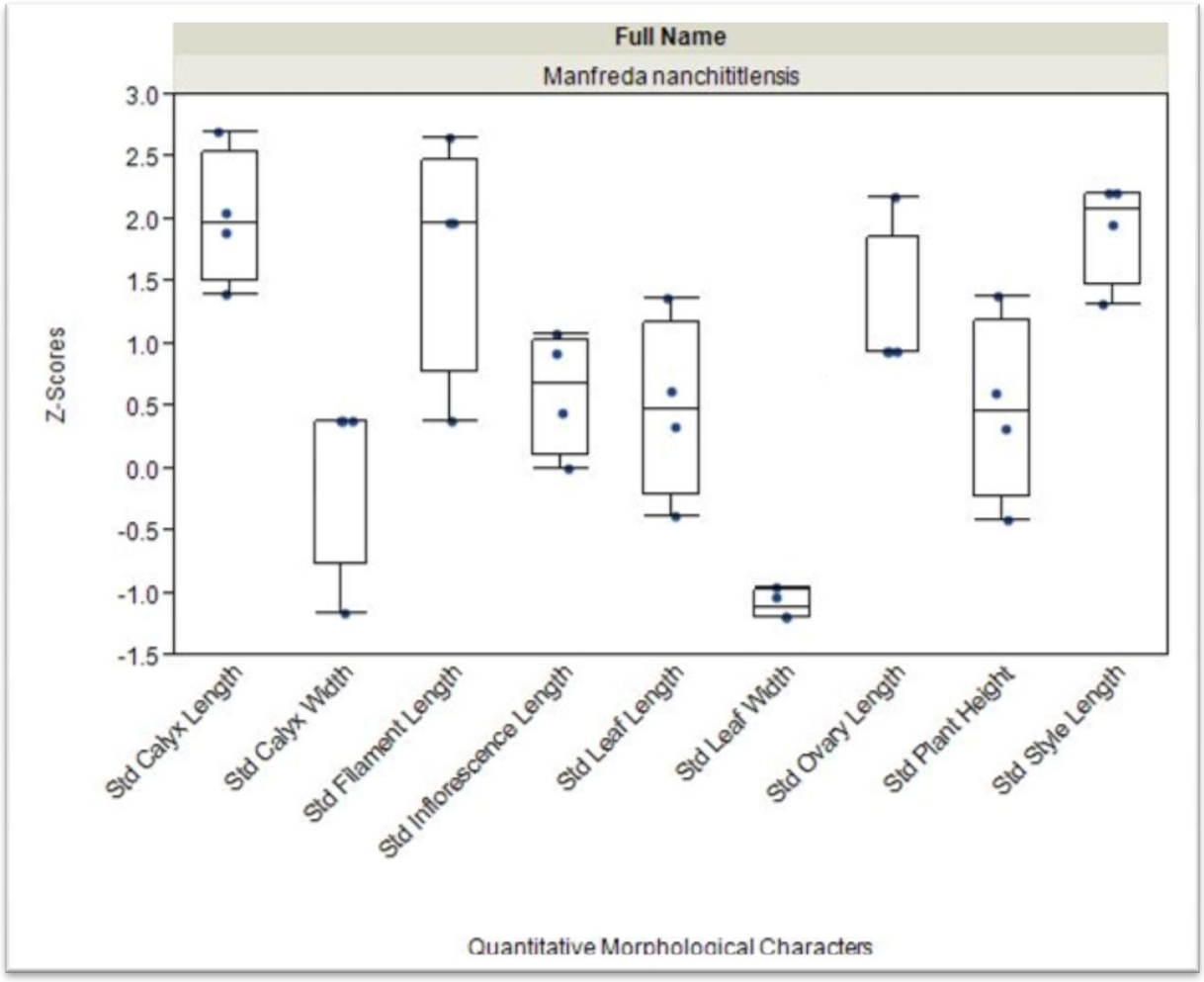


Figure 21. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda nanchititlensis*.

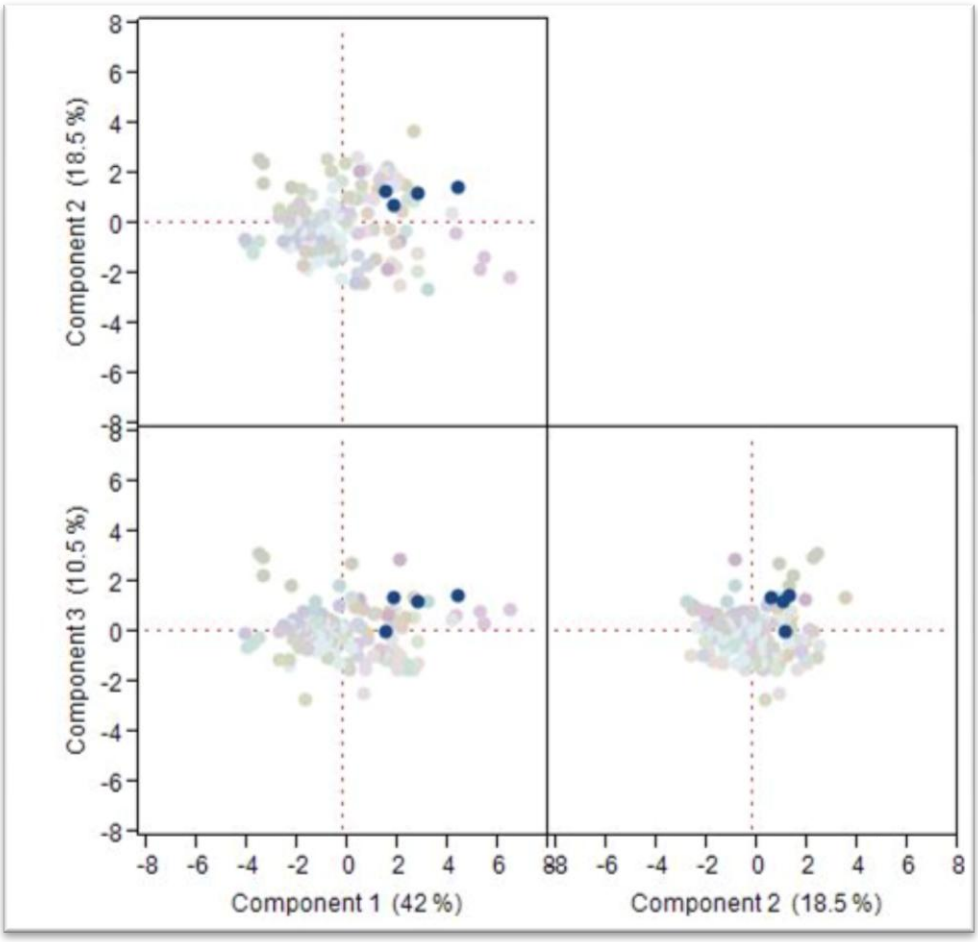


Figure 22. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda nanchititlensis* are highlighted by navy blue.

Manfreda parva - The distrubution of Std Calyx Length demonstrated the greatest variation in the dot plot analysis, while Std Calyx Width exhibited no variation (Figure 24). The distribution of *M. parva* has an unambiguous clustering pattern in all three principal components. Principal Components 1-3 ranged from -2.55 to 1.78, 0.2339 to 0.71 and 0.23 to 0.71, respectively. *M. parva* demonstrated consistent clustering across all three score plots for all four specimens with no clear correlation to any single direction (Figure 25).

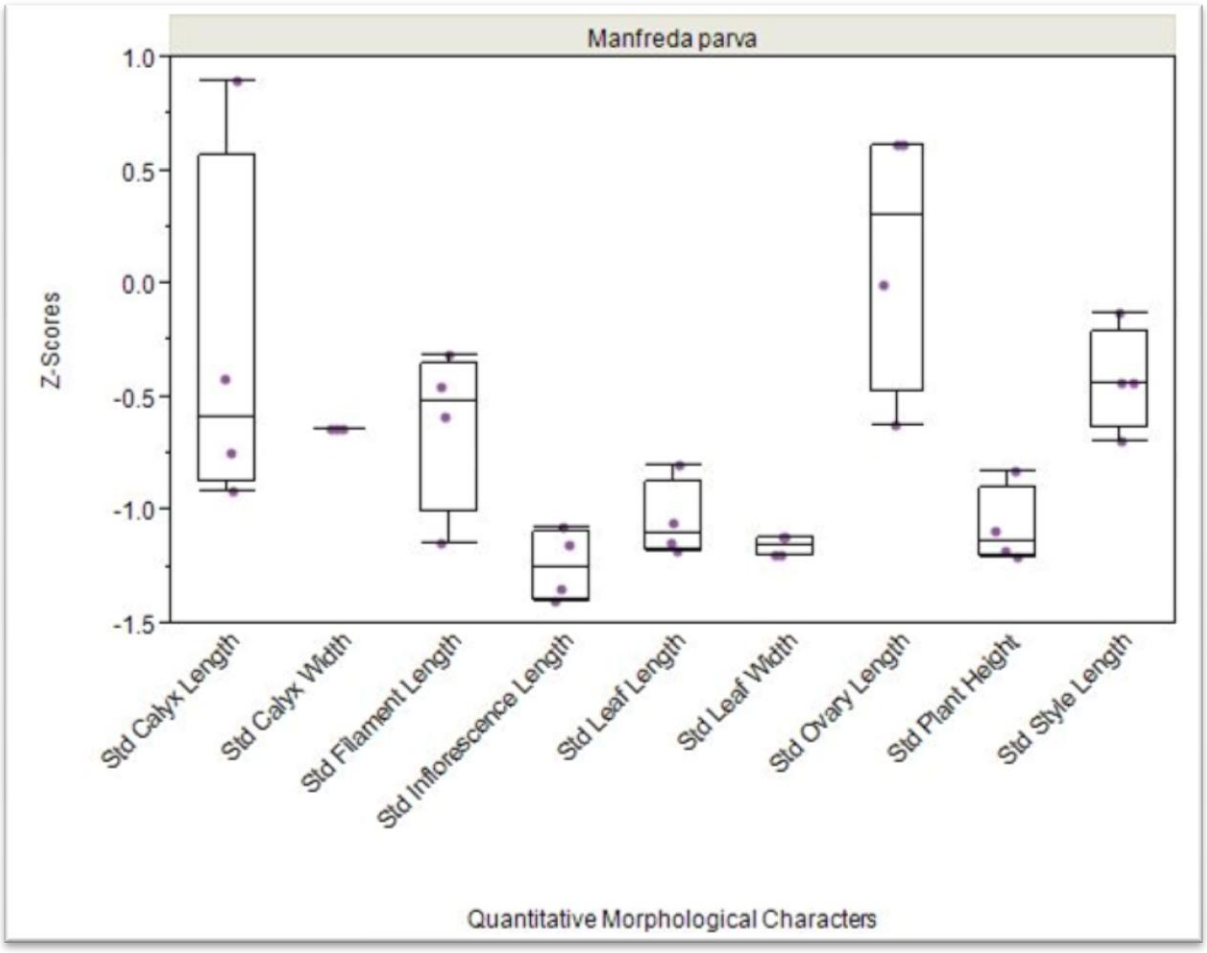


Figure 23. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda parva.*

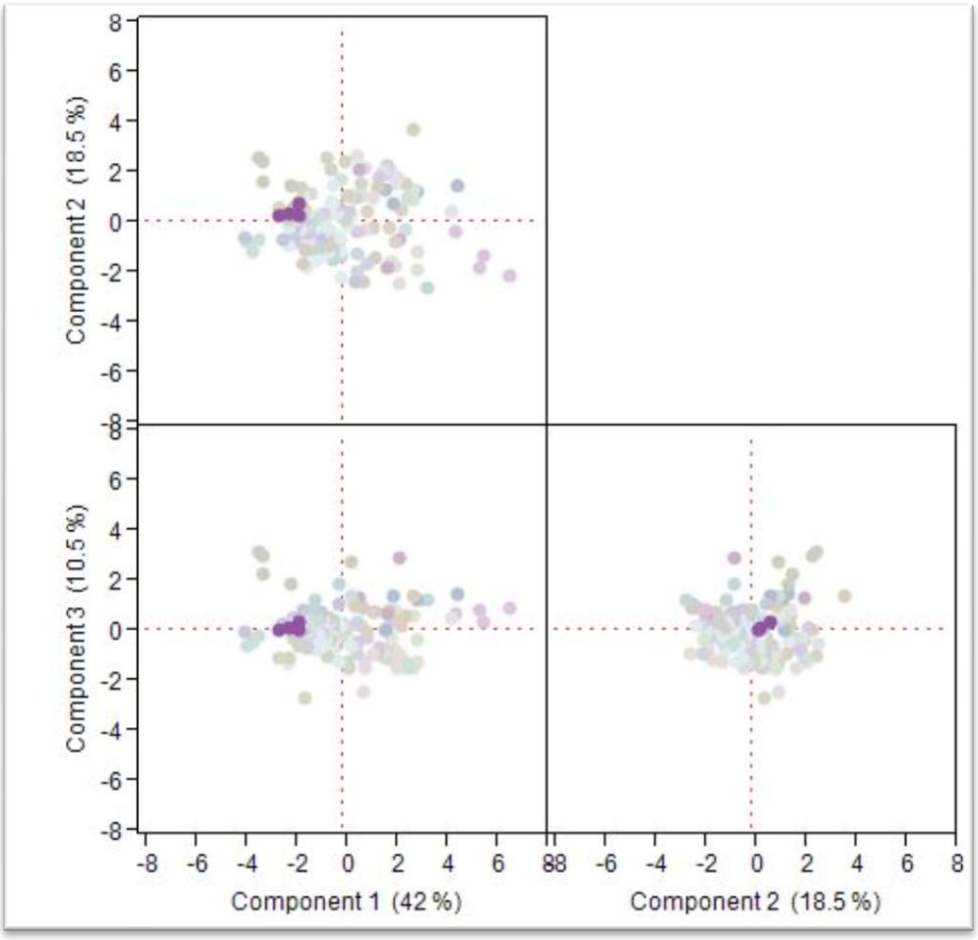


Figure 24. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda parva* are highlighted by purple.

Manfreda planifolia - Dot plot analysis exhibited substantial uniformity for all characters, Std Calyx Width possessed no variation between specimens. However, Std Calyx Length demonstrates the greatest variation (Figure 26). The distribution of *M. planifolia* exhibited a close affinity and evident clustering. Scores ranged from 1.95 to 2.97, -2.35 to 1.79 and -1.29 to 0.83 for Principal Components 1and 3, respectively. In score plots 1 and 2 a correlation with Std Inflorescence Length can be identified, for score plot 3 no clear character correlation was apparent (Figure 27).

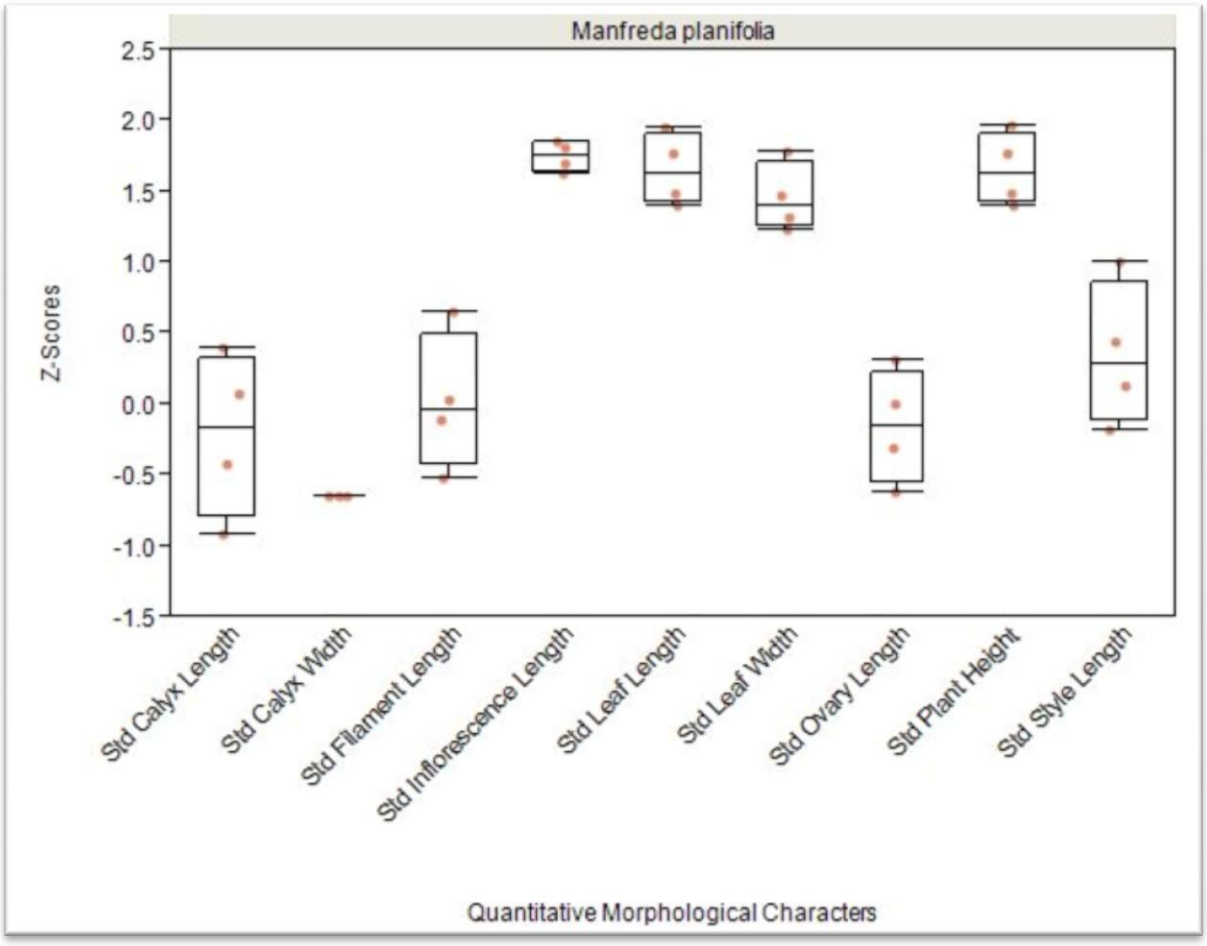


Figure 25. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda planifolia.*

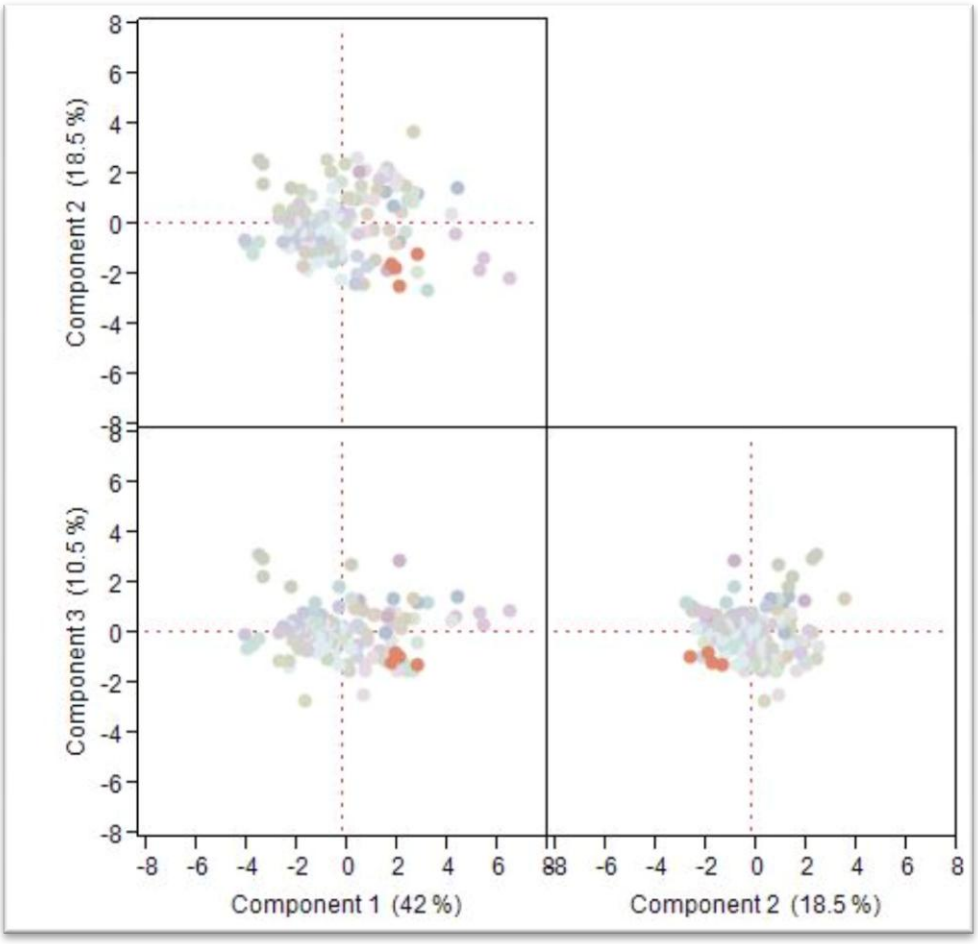


Figure 26. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda planifolia* are highlighted by brown.

Manfreda potosina - Dot plot analysis revealed that Std Inflorescence Length exhibited the greatest variation, while minimal variation was observed for Std Style Length. Minimal diversity was also observed Std Filament Length with the exception of a single outlier (Figure 28). Although no cluster formation was observed in *M. potosina* specimens, all data point were external to any other *Manfreda* specimen in score plots 2 and 3. Scores ranged from 3.33 to 0.31 in Principal Component 1, 1.00 to 2.59 in Principal Component 2 and 1.81 to 3.13 in Principal Component 3. No correlation with any single character can be observed in score plot 1-2, in score plot 3 however affinity was exhibited with Std Ovary Length (Figure 29).

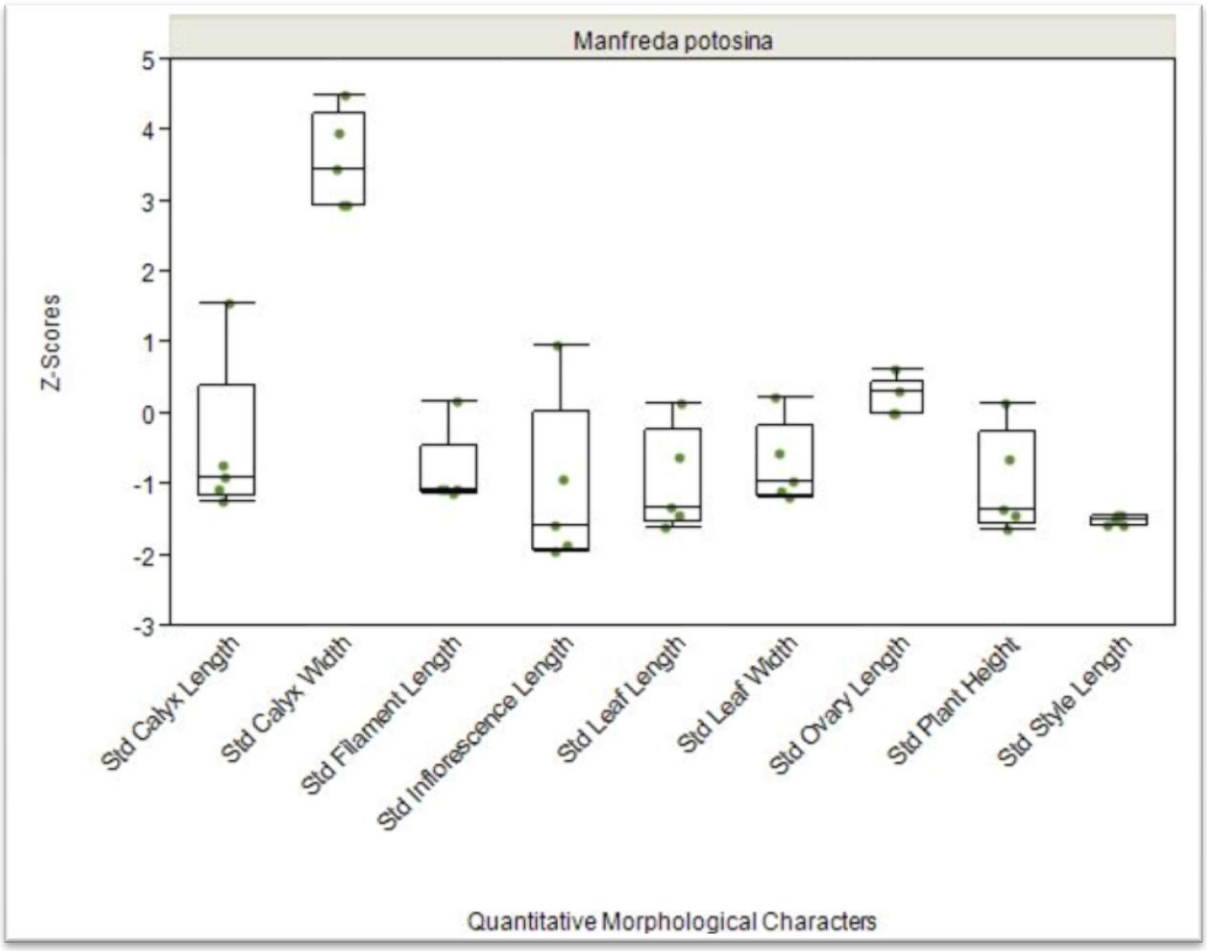


Figure 27. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda potosina*.

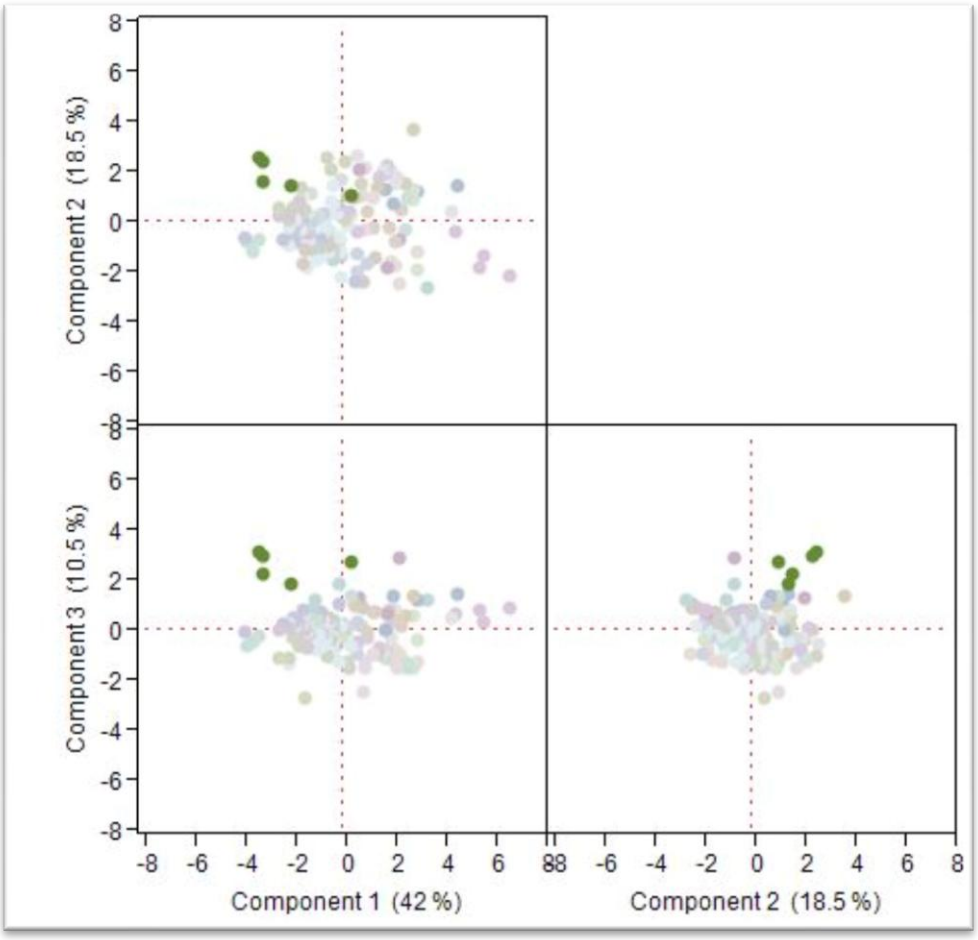


Figure 28. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda potosina* are highlighted by dark green.

Manfreda pringlei - Substantial variation was observed in the dot plot among all characters (Figure 30). The distribution of *M. pringlei* has no clear clustering pattern and score plots 1-2 exhibited multiple disjunct groupings. Scores ranged from -3.8509 to 2.5497, -1.1997 to 2.0885 and -1.5560 to 1.3921 for Principal Components 1-3 respectively. The distribution of data points observed no correlation to any single character (Figure 31).

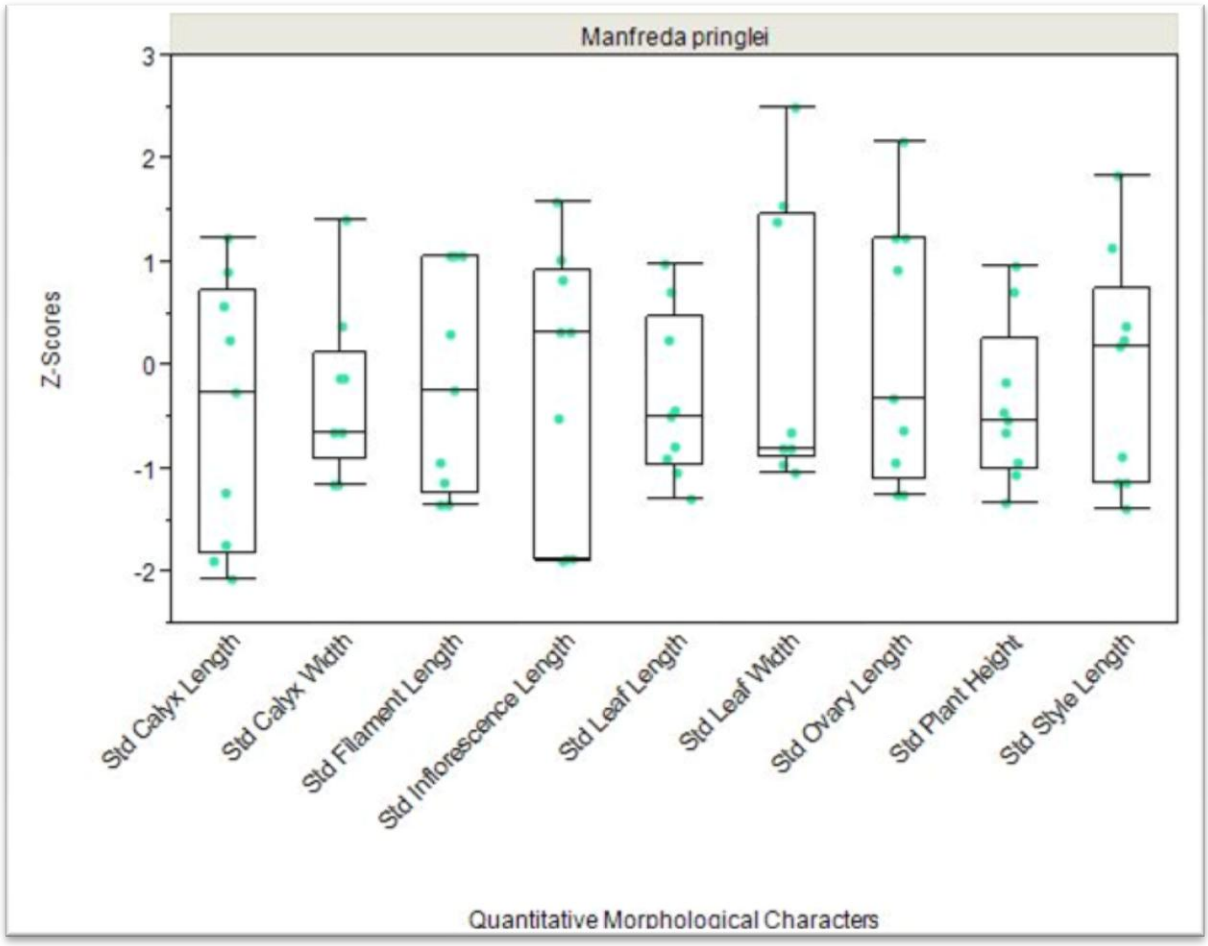


Figure 29. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda pringlei.*

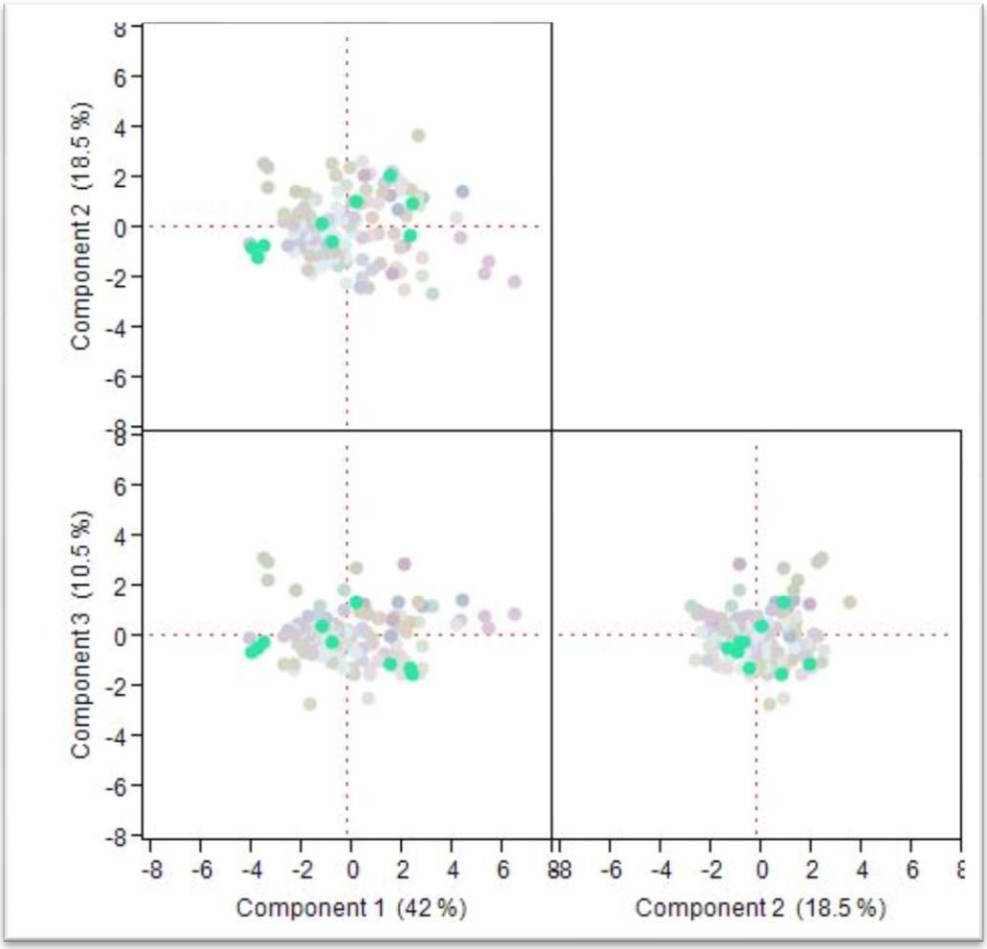


Figure 30. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda pringlei* are highlighted by green.

Manfreda pubescens - Dot plot analysis demonstrates that Std Leaf Length and Std Plant Height contributed the greatest variation, while Std Leaf Width, Std Ovary Length and Std Style Length posssessed equivelently low variation (Figure 32). Specimens of *M. pubescens* demonstrated a collective directionality in all three score plots, however some data points are distantly dispersed. For Principal Components 1-3, scores ranged from 0.56 to 4.33, 0.39 to 2.62 and -0.88 to 0.79. Two clear central clusters were evident in score plots 1 and 3, however peripheral data points were distant and score plot 2 demonstrated no clear cluster. No clear correlation with any single character was dound (Figure 33).

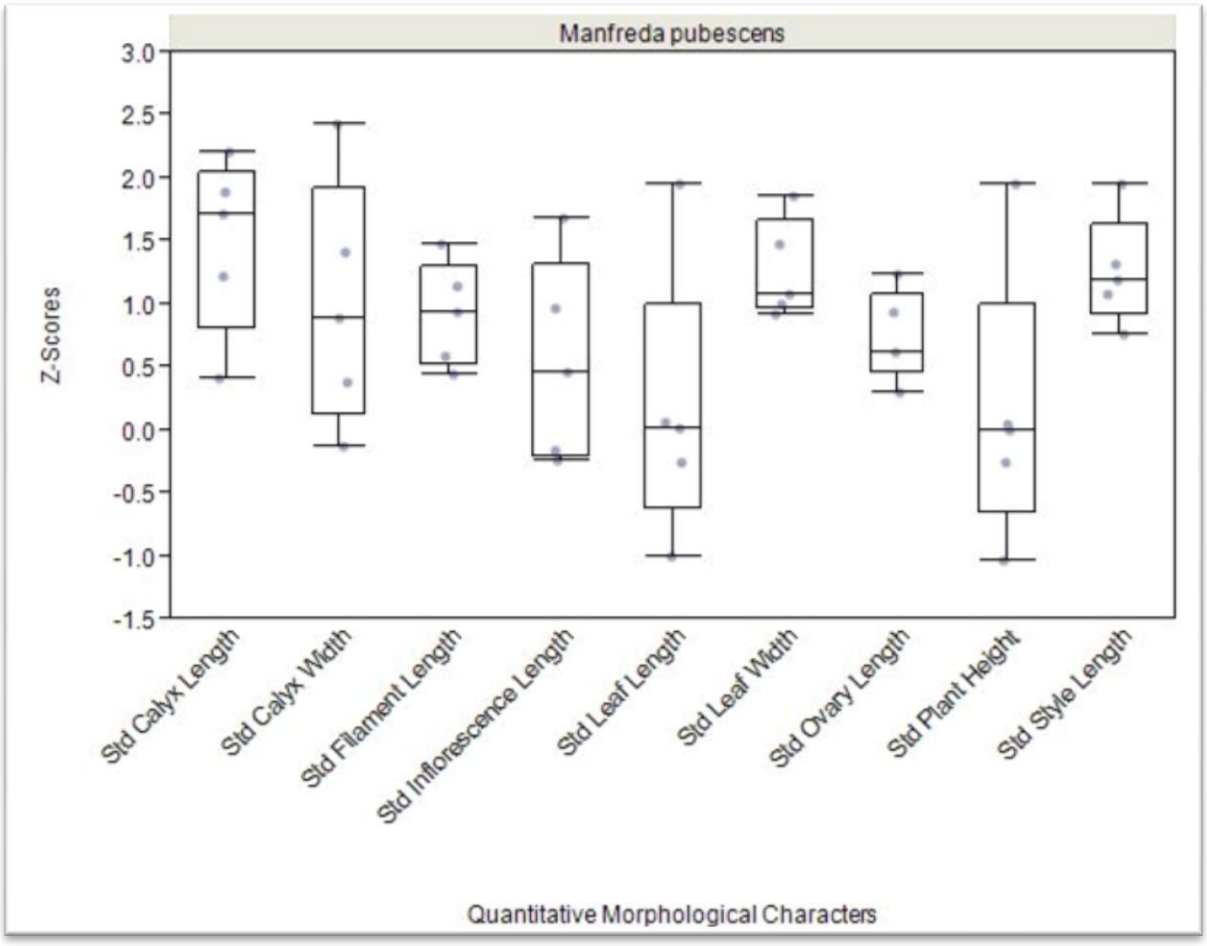


Figure 31. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda pubescens.*

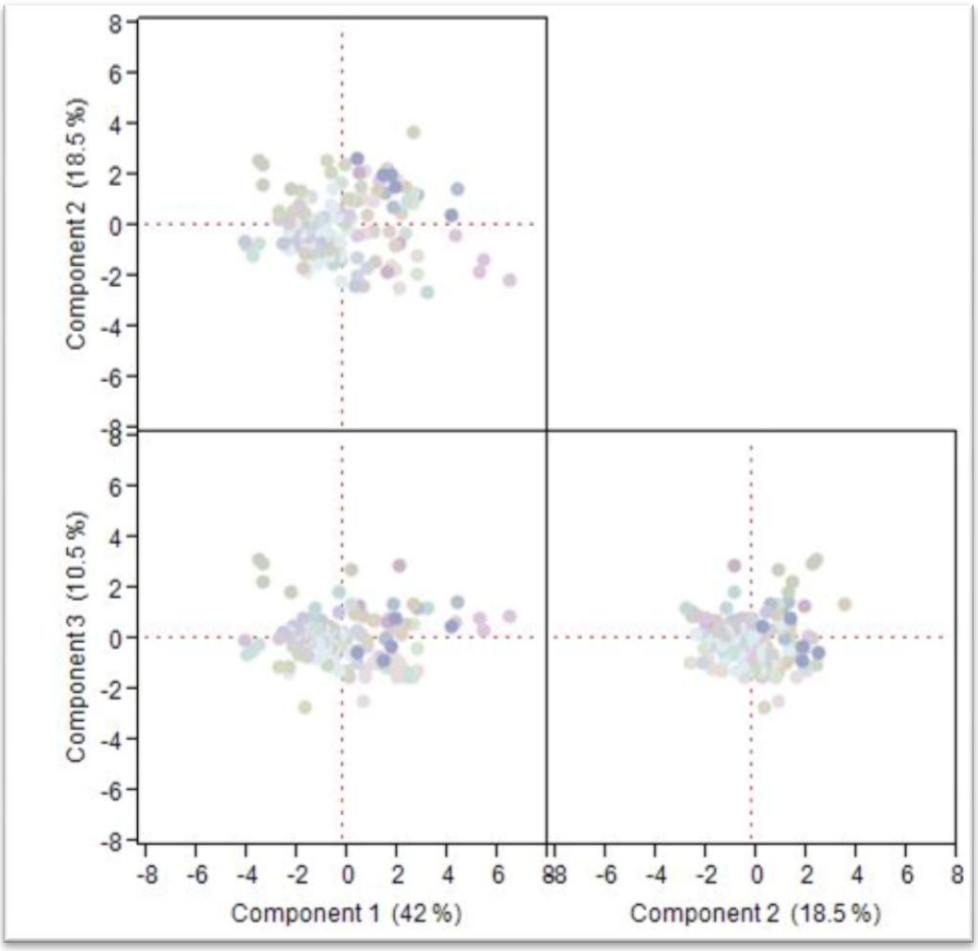


Figure 32. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda pubescens* are highlighted by purple.

Manfreda revoluta - Of the dot plot analysis, Std Filament Length exhibited the least variation, while Std Ovary Length demonstrated the greatest variation (Figure 34). The distribution of each of the three specimens of *M. revoluta* possessed no clear clustering pattern and no correlation was surmised. Scores range from 0.6354 to 2.2303 for Principal Component 1, 0.6354 to 2.2303 for Principal Component 2 and 0.6290 to 2.9101 for Principal Component 3. The relatively few samples and diverse distribution did not facilitate inference of a clear correlation to one another or any defining characters (Figure 35).

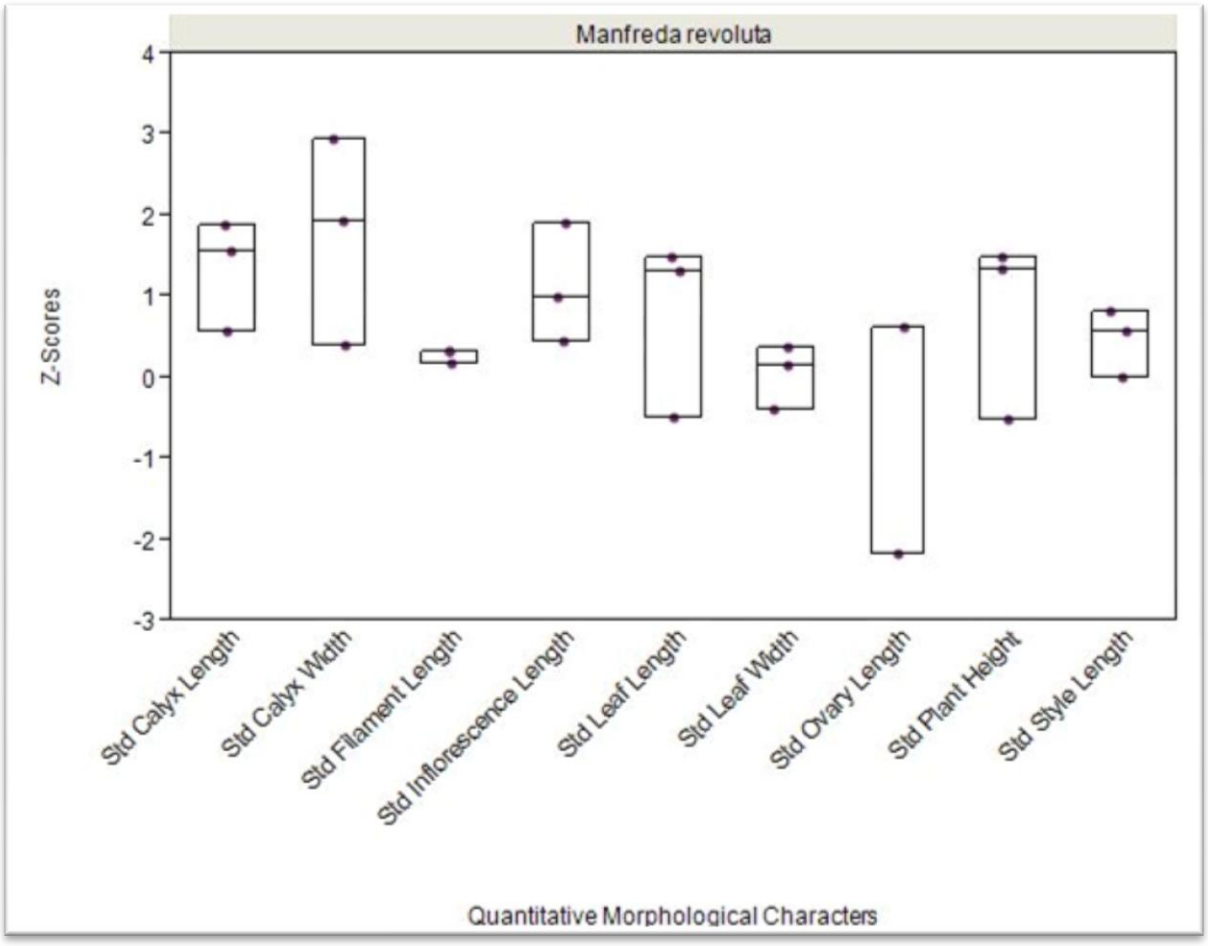


Figure 33. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda revoluta.*

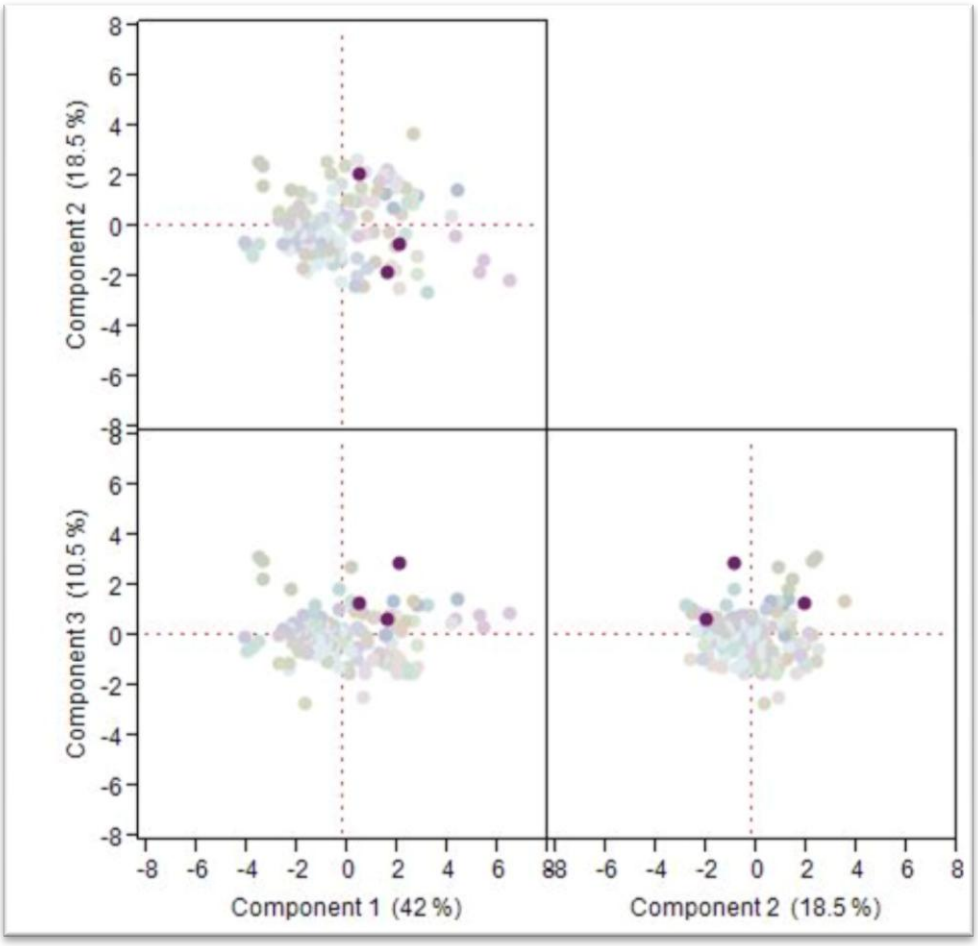


Figure 34. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda revoluta* are highlighted by purple.

Manfreda rubescens - Dot plot analysis revealed that Std Inflorescence Length contributed the greatest degree of variation, while Std Leaf Width exhibited the least variation (Figure 36) The 11 specimens of *M. rubescens* demonstrated a clear clustering pattern in all three score plots with few outlying specimens. Scores range from -1.5859 to -0.3913 for Principal Component 1, -2.40 to 0.72 for Principal Component 2 and -0.03 to 0.85 for Principal Component 3. In score plot 1 and 2 two taxa were peripheral to the central cluster and in score plot 3 only one. No clear correlation with any character was observed (Figure 36).

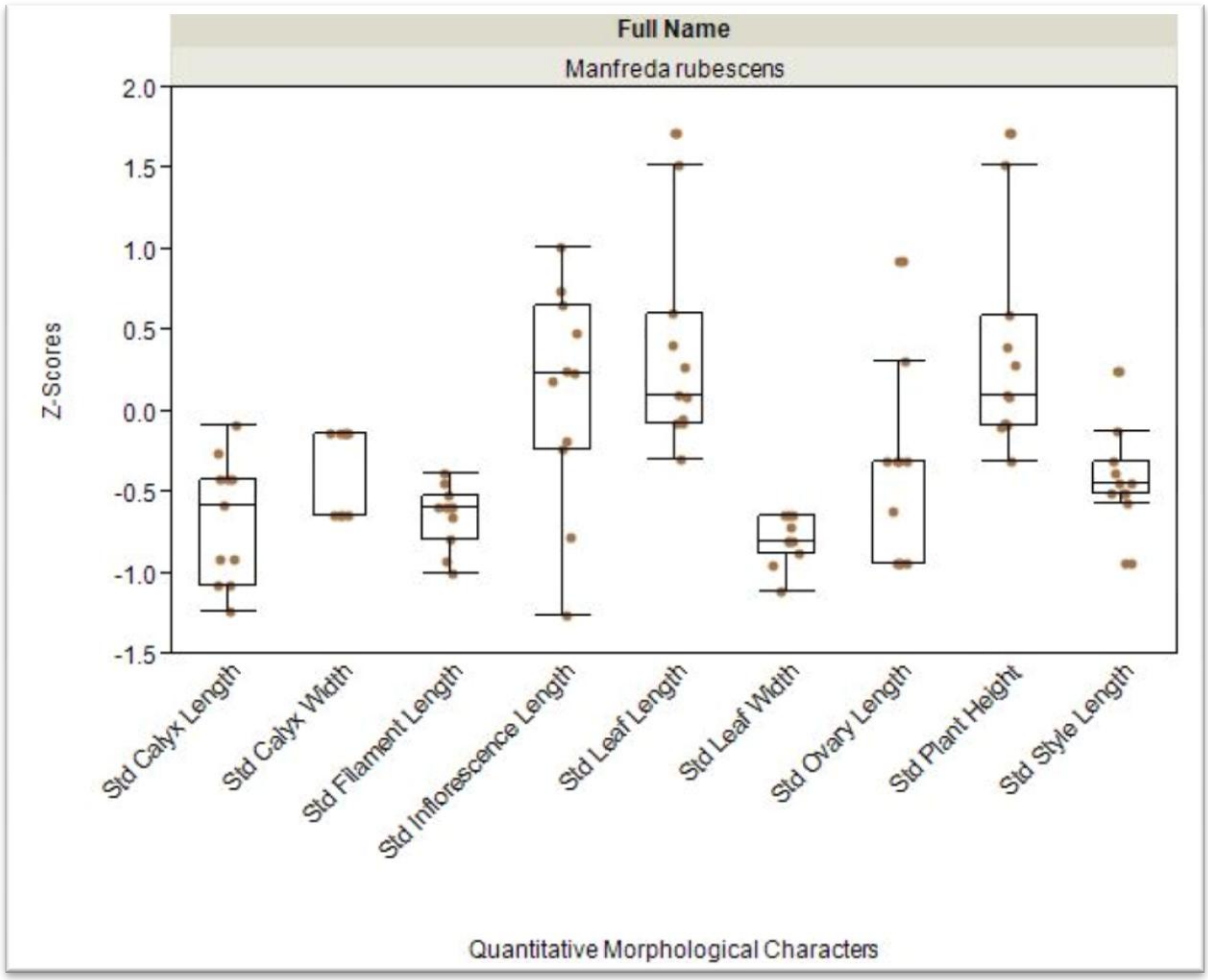


Figure 35. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda rubescens*.

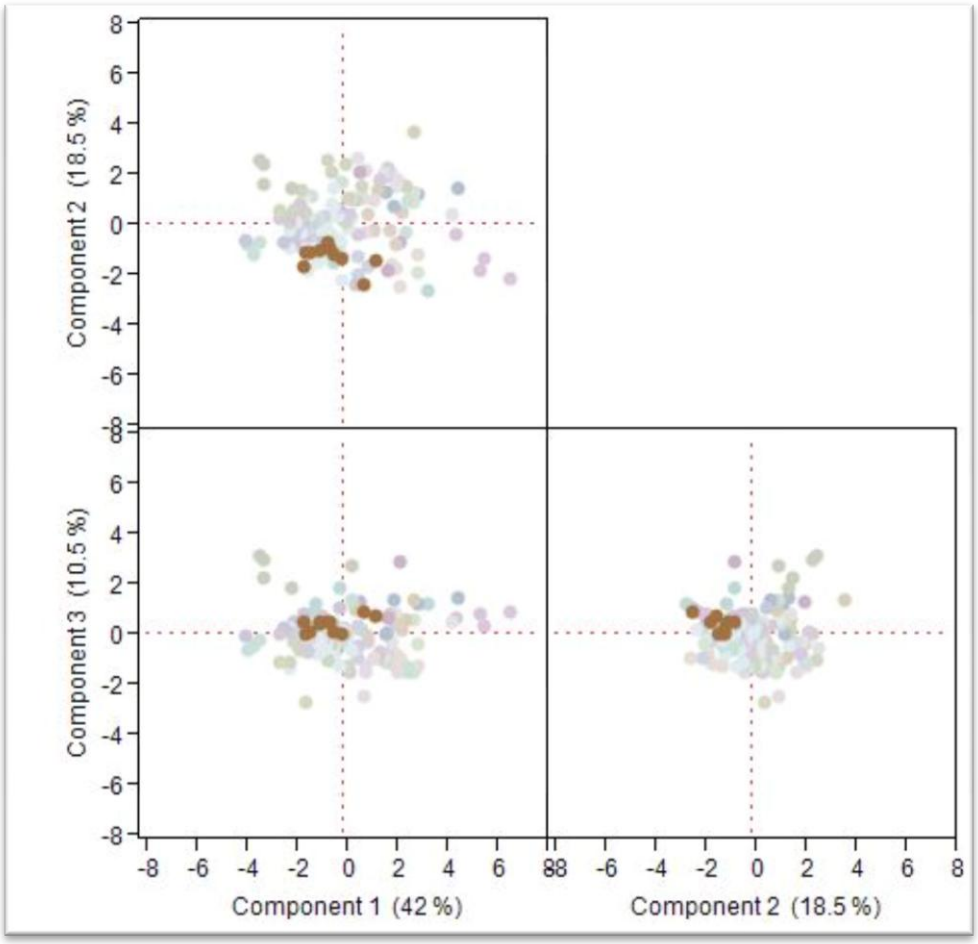


Figure 36. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda rubescens* are highlighted by brown.

Manfreda scabra - Of the dot plot analysis, Std Calyx Length demonstrated the greatest variation, in contrast Std Filament Length exhibited minimal variation (Figure 38). Specimen of *M. scabra* exhibit a varied pattern of distribution. In score plot 1, two specimens were highly aligned and one was disjunct. In score plot 2 the three data points showed a close affinity while in score plot 3 data points were distant. For Principal Components 1-3, scores range from 2.79 to 2.98, -1.90 to 1.31 and -0.38 to 0.55, respectively. Despite substantial uniformity in Principal Component 1 the variation within Principal Component 2-3 has resulted in a broad dispersal of data points (Figure 39).

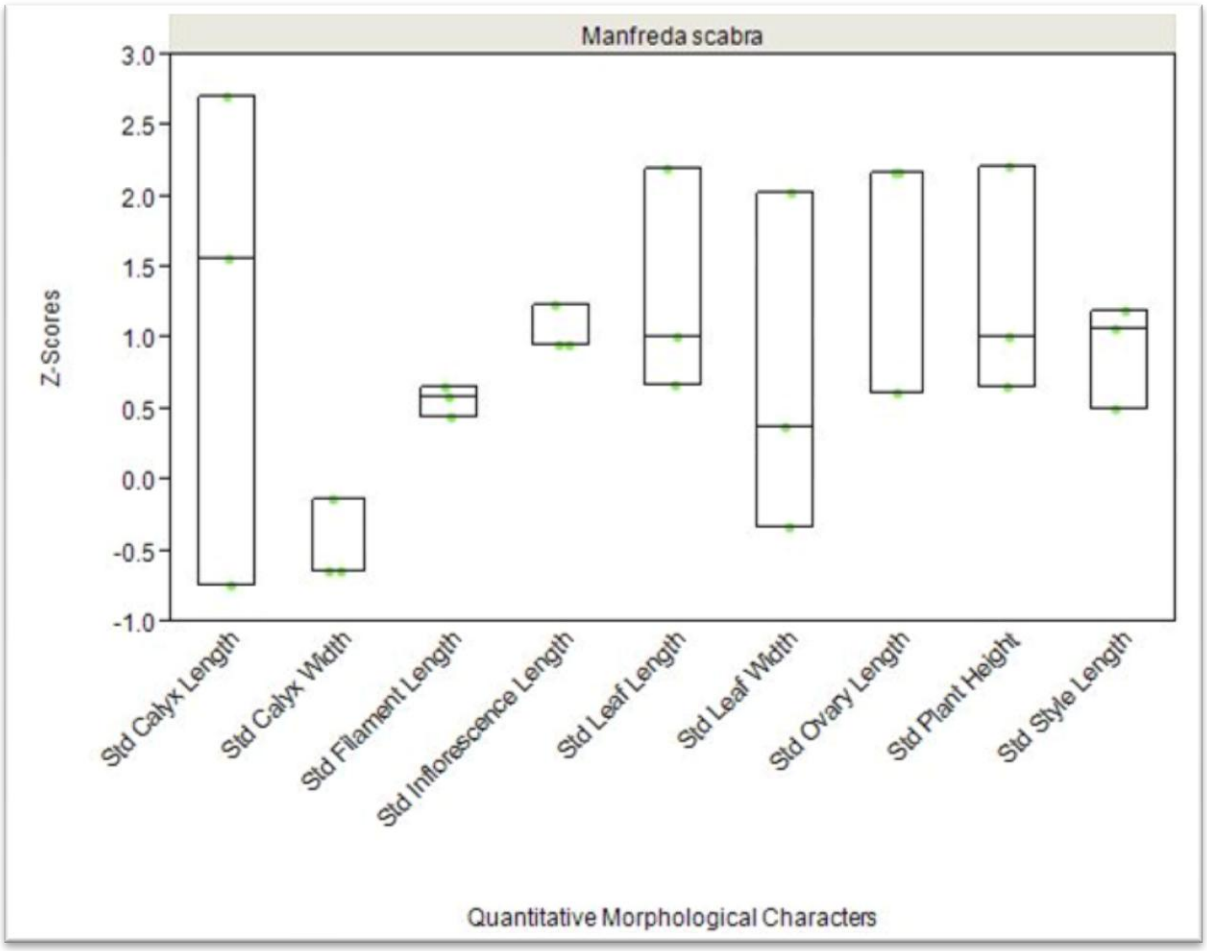


Figure 37. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda scabra*.

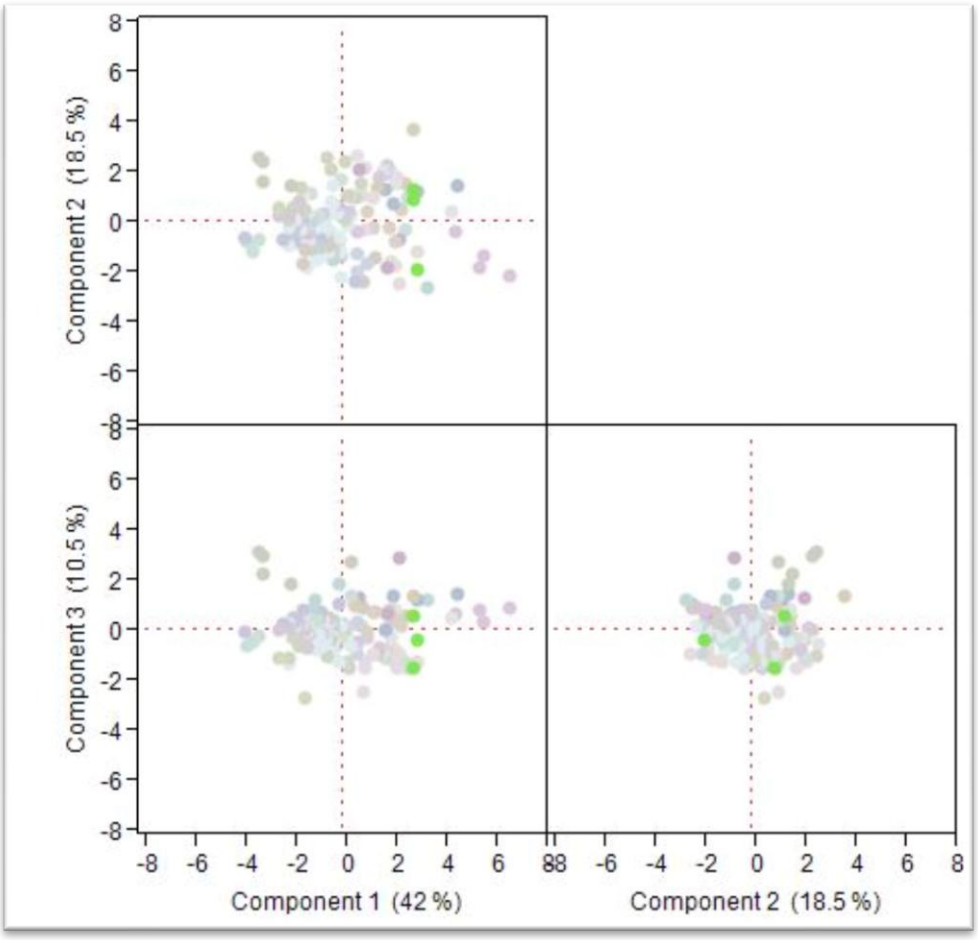


Figure 38. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda scabra* are highlighted by green.

Manfreda singuliflora - Of the dot plot analysis, the greatest variation was attributed to Std Calyx Length, while Std Leaf Width and Std Ovary length possessed the least variation (Figure 40). The distribution of *M. singuliflora* demonstrated both clustering and disjunct data points. In score plot 1, a central cluster was augmented by three peripheral specimens. In score plot 2 data point exhibited a linear correlation with one outlying data point, and similarly in score plot 3 a single data point is an outlier to a well defined cluster. Scores ranged from -3.89 to 0.49 for Principal Component 1, -2.40 to 0.53 for Principal Component 2 and -0.09 to 1.08 for Principal Component 3. Despite occasional outliers, a strong clustering pattern was observable in score plots 1 and 3 in particular. No strong relationship with any single character was exhibited (Figure 41).

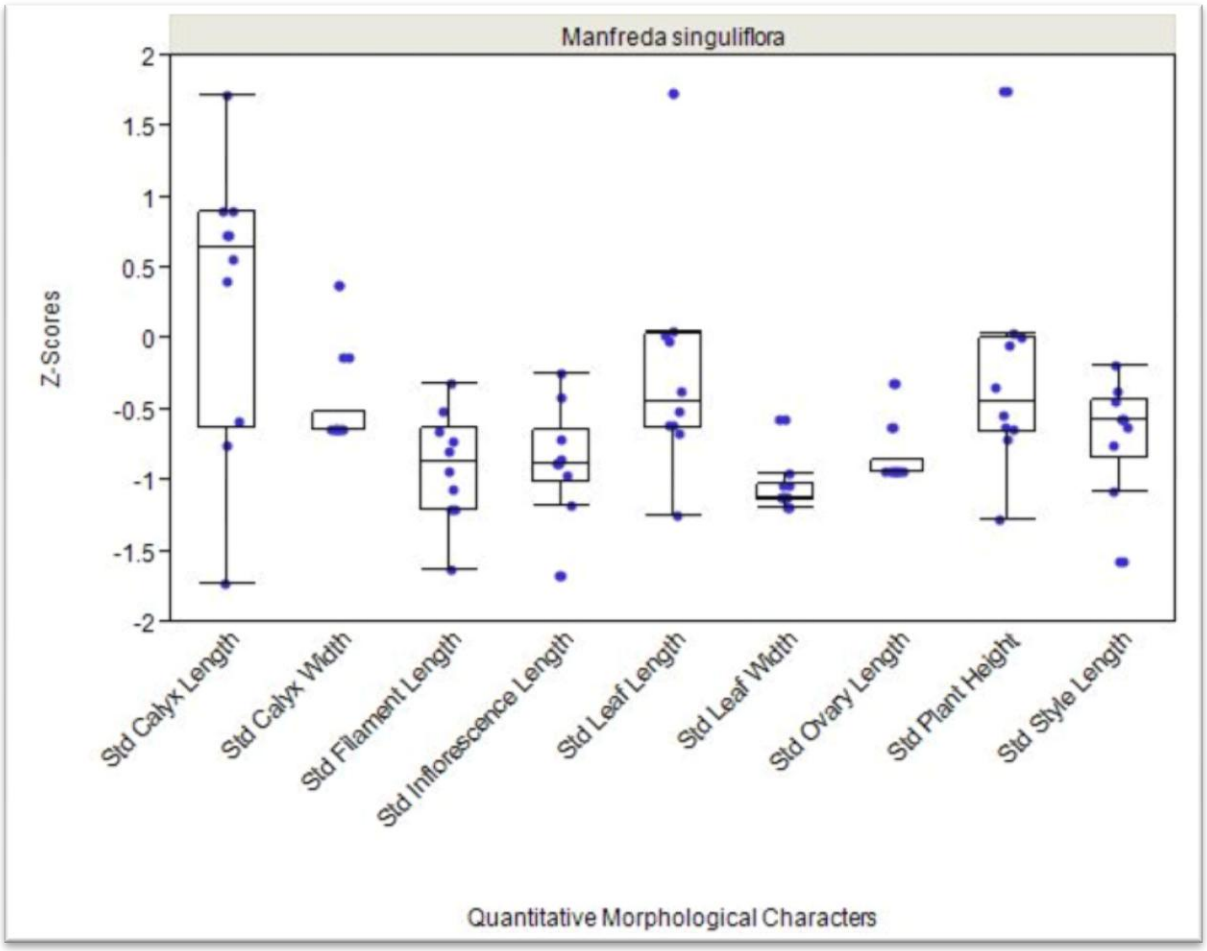


Figure 39. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda singuliflora*.

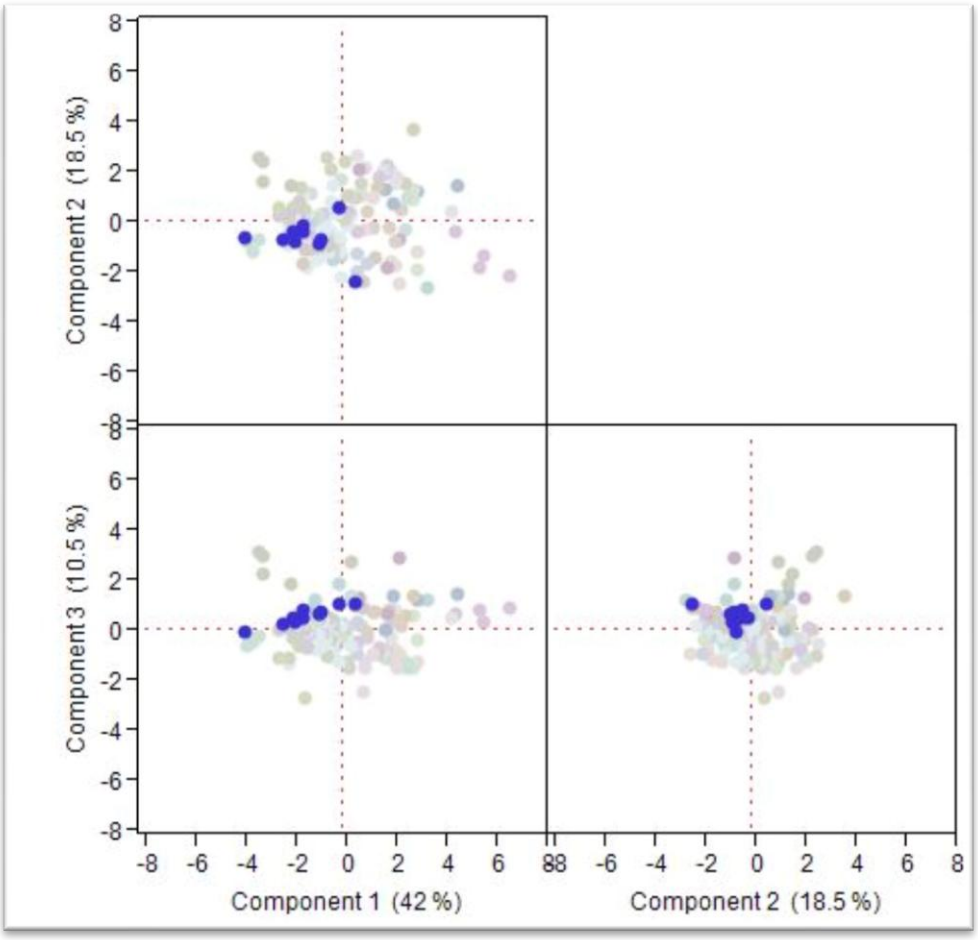


Figure 40. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda singuliflora* are highlighted by blue.

Manfreda umbrophila - Dot plot analysis revealed that the greatest variation was attributed to Std Filament Length and Std Style Length. The least diversity was observed in Std Calyx Length (Figure 42). The distribution of *M. umbrophila* specimens was close, however with sufficient distance to prevent distinct clustering. For Principal Components 1-3, scores ranged from 0.83 to 2.18, -0.30 to 2.19 and -2.50 to 0.99 respectively. Specimens of *M. umbrophila* in score plot 1 exhibited a strong correlation with Std Calyx Width, however in score plot 2 and 3 the relationship was strong with either Std Plant Height, Std Leaf Length or Std Inflorescence Length (Figure 43).

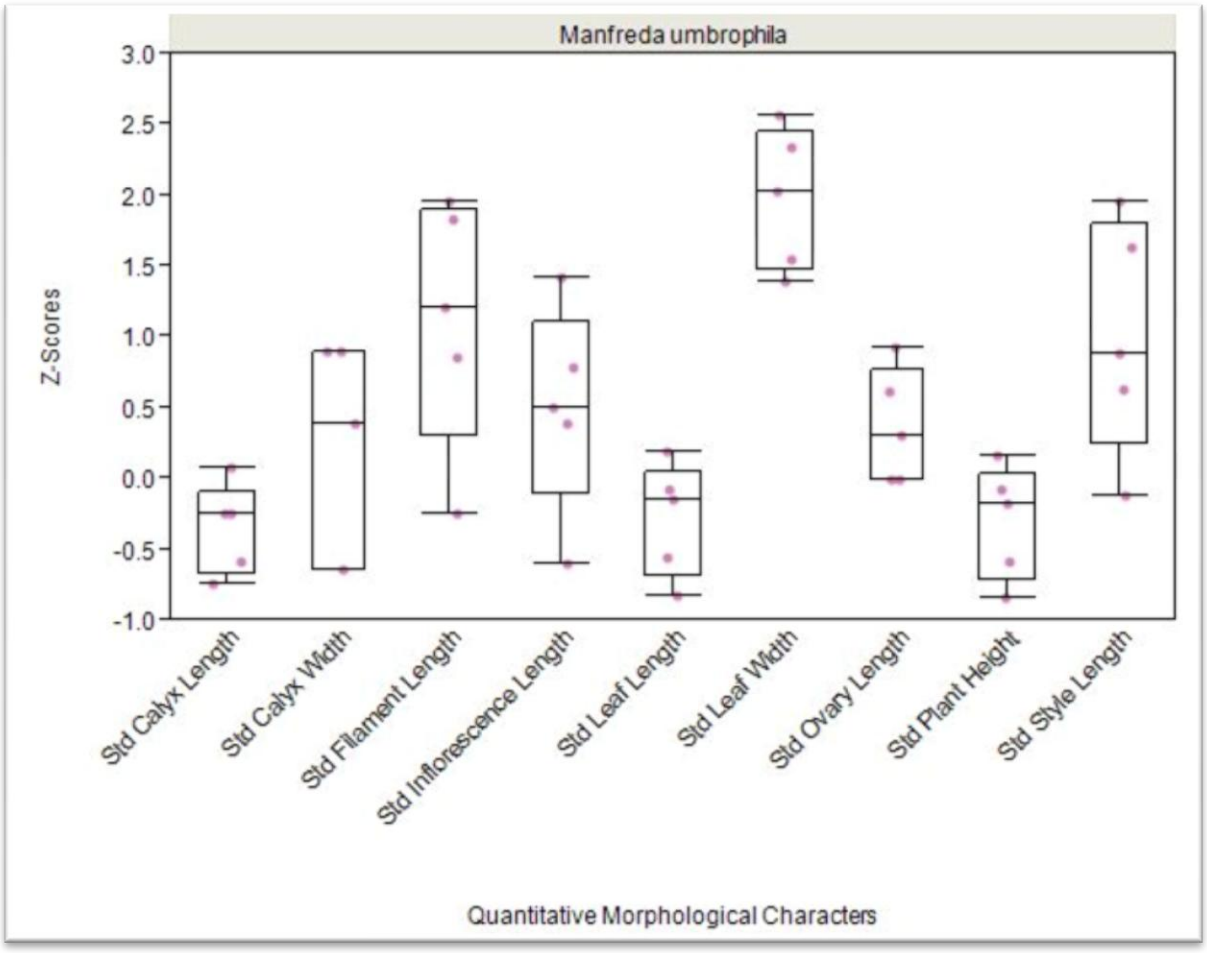


Figure 41. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda umbrophila.*

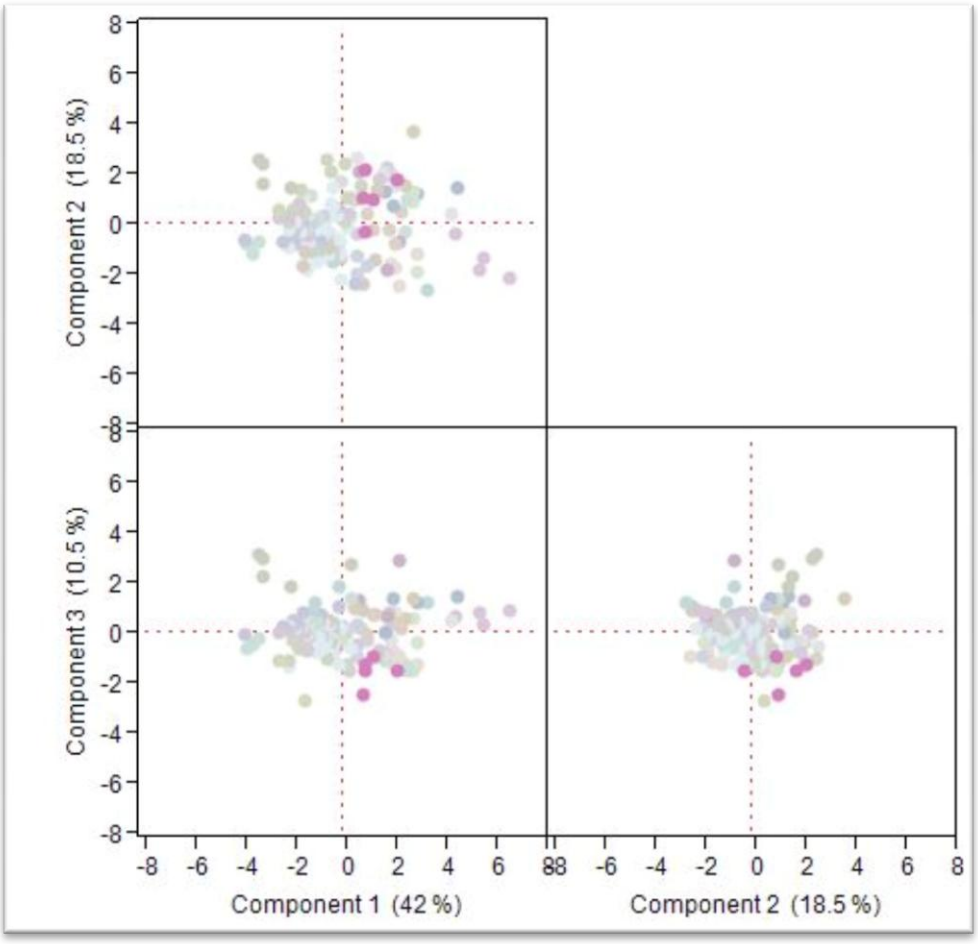


Figure 42. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda umbrophila* are highlighted by purple.

Manfreda variegata - Dot plot analysis demonstrated that Std Filament Length contributes the greatest variation, while Std Leaf Width contributed the least (Figure 44). The distribution of *M. variegata* specimens was distant yet congruent with a single outlier. Scores for Principal Components 1-3 ranged from 0.46 to 2.51, 0.92 to 3.67 and 0.97 to 1.33, respectively. Score plot 1 exhibited a strong relation to Std Leaf Width for three out of four specimens, no strong correlation to any single character was observed in score plot 2 or 3, however (Figure 45).

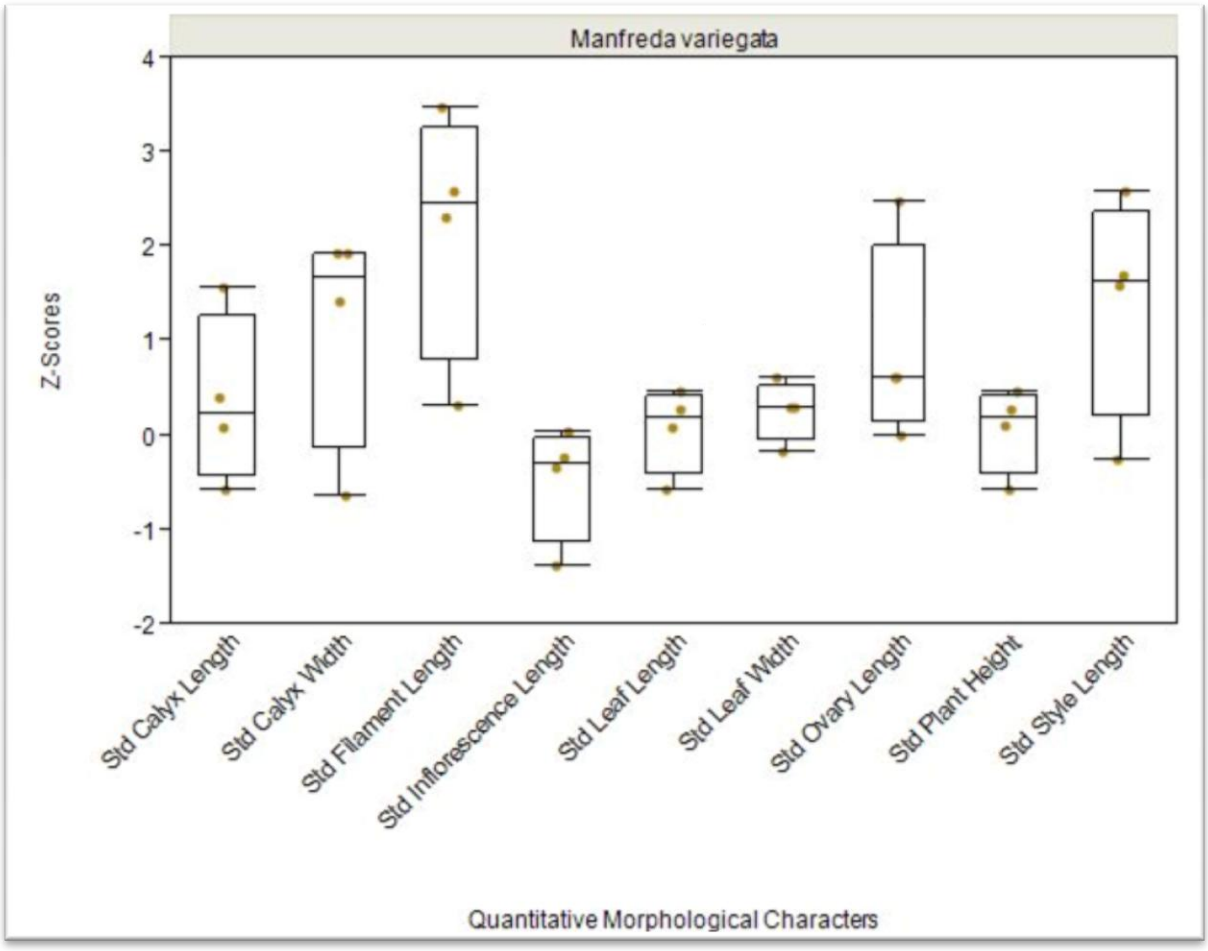


Figure 43. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda variegata.*

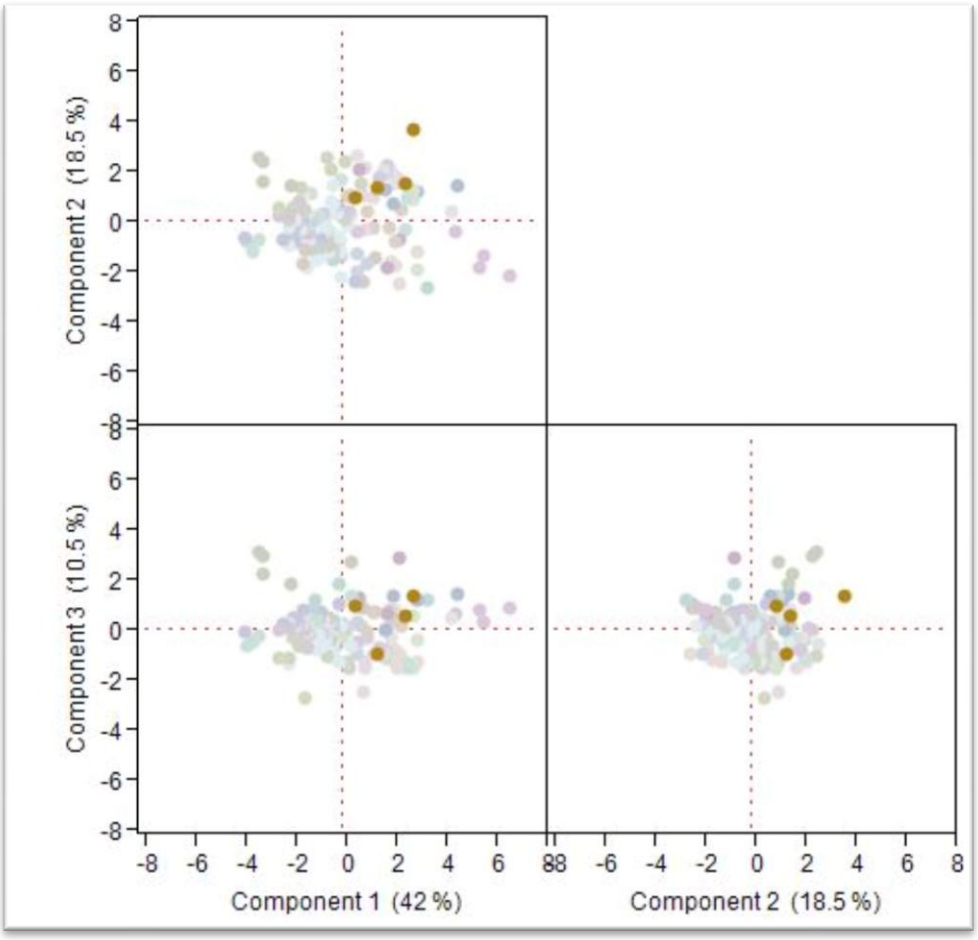


Figure 44. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda variegata* are highlighted by brown.

Manfreda virginica - Of the dot plot analysis, Std Ovary Length demonstrated the greatest variation due to distant outliers, however the least variation between the first and third quartile wasattributed to the same character. Std Leaf Length possessed the least overall variation (Figure 46). With 33 specimens the distribution of *M. virginica* was well defined as a cluster in all three score plots. Scores range from -2.17 to 0.31 for Principal Component 1, -2.27 to 1.45 for Principal Component 2 and -1.26 to 0.56 for Principal Component 3. The well formed cluster had minimal outliers and no correlation with any individual character (Figure 47).

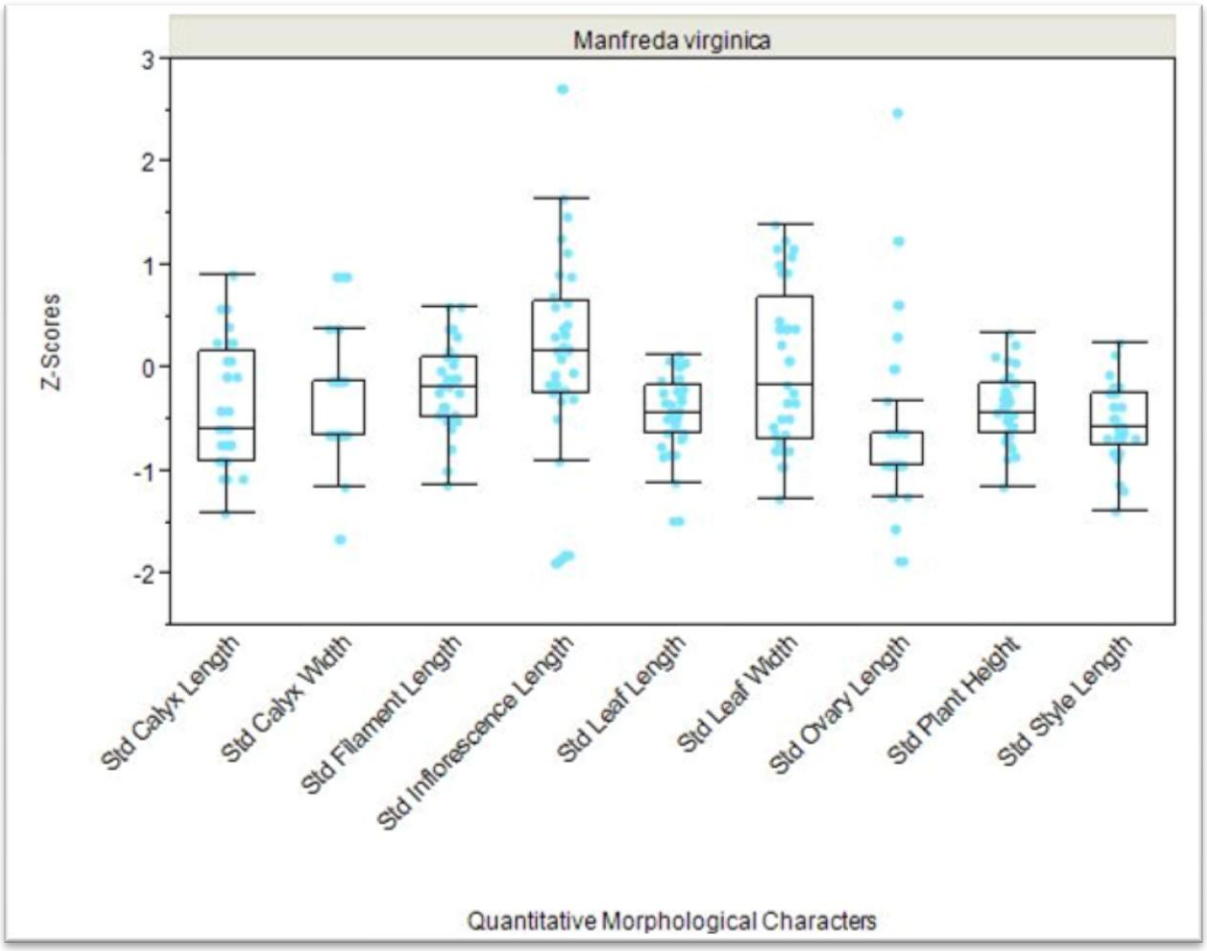


Figure 45. Box plots depicting variation represented by z-scores (standardized values) between nine quantitative morphological characters for *Manfreda virginica.*

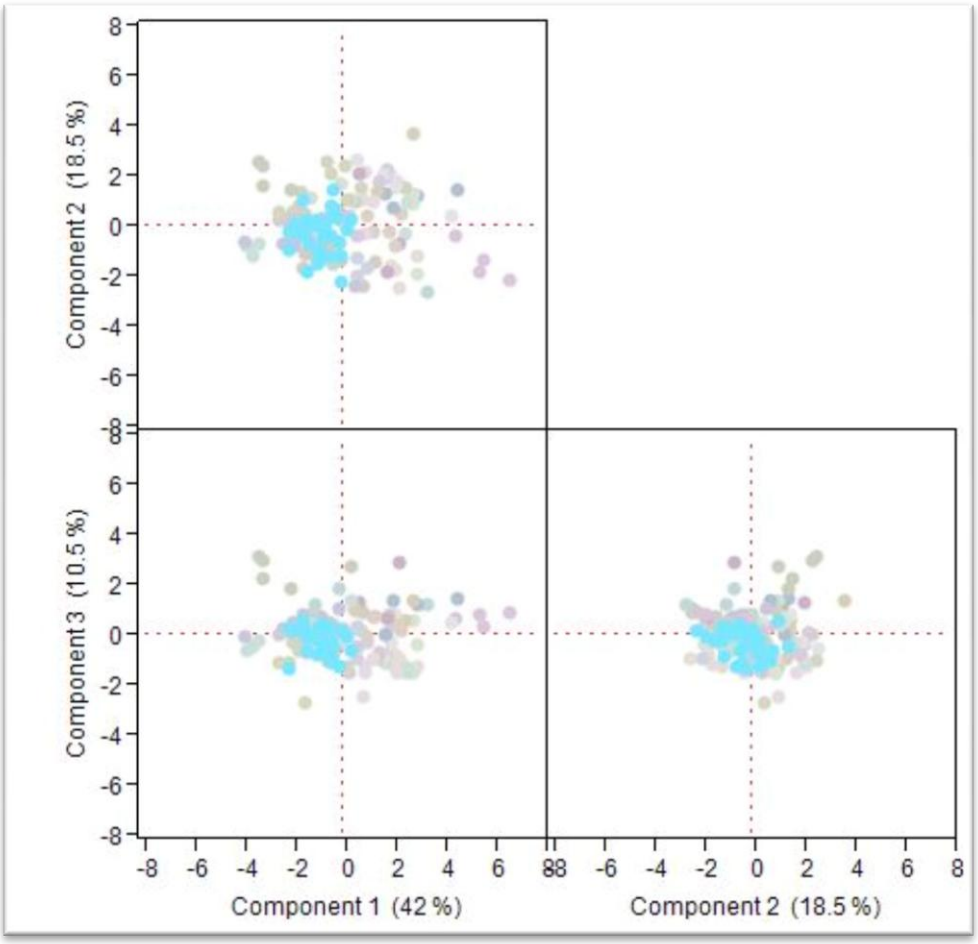


Figure 46. Resultant plot of Principal Components 1, 2 and 3 from Principal Component Analysis of herbarium specimens representing *Manfreda*. Specimens representing *Manfreda virginica* are highlighted by blue.

Pairwise Distance Matrix Analysis. Pairwise distances calculated between each of the 29 taxa ranged from 0.0000 between six pairwise relationships and between *Agave americana* and *M. nanchititlensis* at 0.3243. The greatest morphological distance observed between *Manfreda* taxa was between a complex of *M. parva, M. potosina, M. pringlei and M. rubescens* at 0.2500. The mean morphological distance was 0.1306 (Appendix 2).

Maximum Parsimony Analysis. Of the 36 characters employed in the Maximum Parsimony Analysis, 21 were constant, four were parsimony uninformative and 12 residual parsimony informative characters. Proliferation was extensive throughout the cladogram, although many speciation events were poorly supported. Bifurcation ranging from 3 to 49 (denoted in red) represented 21 proposed speciation events. Only the clades containing *M. longiflora* and *M. potosina* as well as *M. maculosa* and *M. sileri* were supported by bootstrap values in excess of 50 (denoted in black). Both clades were attributed a bootstrap value of 51, inferring minimal confidence in the proposed clades (Figure 48).

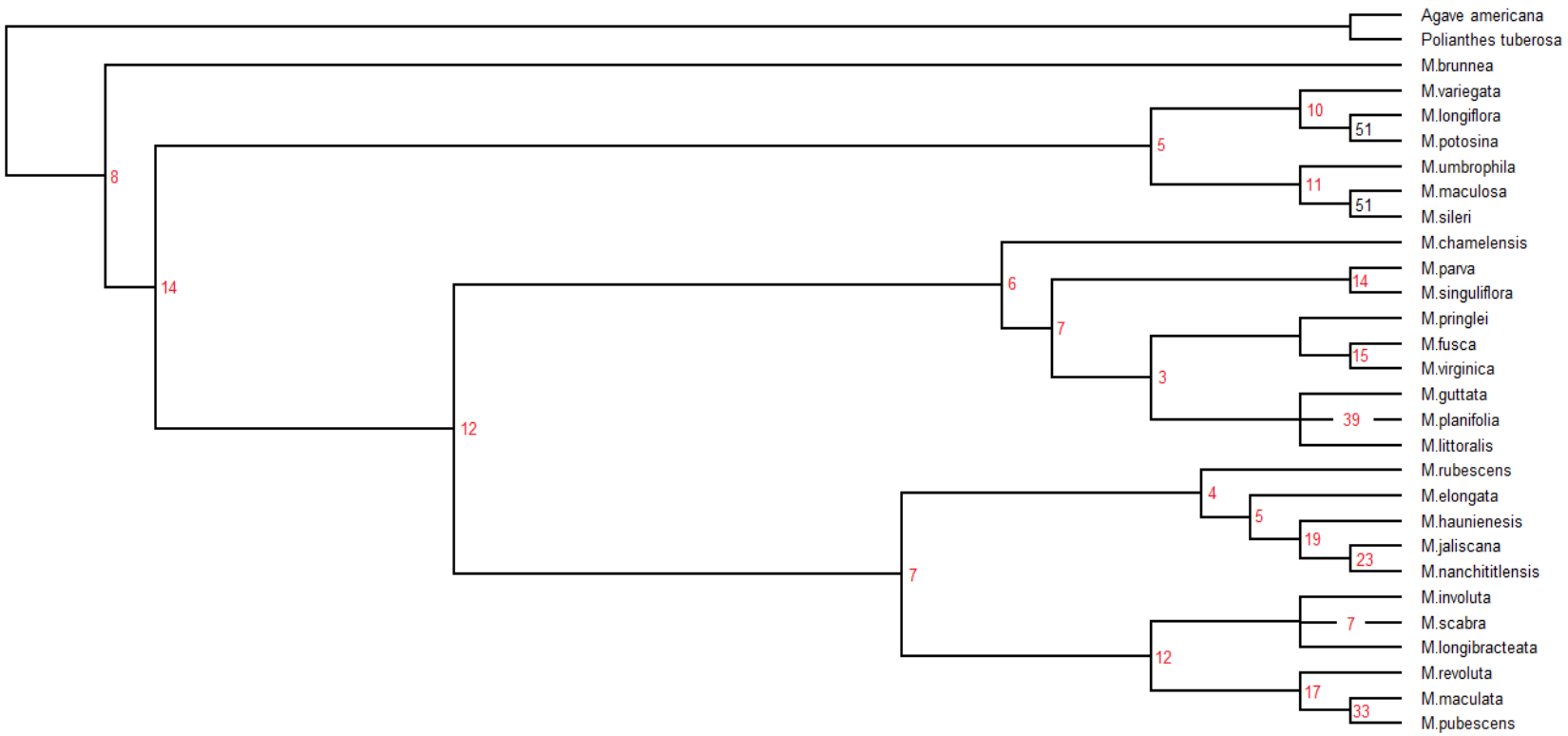


Figure 47: Maximum Parsimony cladogram (50% Majority Rule) representing selected quantitative morphological characters of *Manfreda* and selected outgroups. Bootstrap values in red denote scores below 50 and values in back denote scores above 50.

Bayesian Analysis. Of the phylogenetic analysis of 29 taxa with 37 categorical characters a piece under Bayesian Analysis methods, minimal inference of evolution and interspecific relationships could be ascertained. Only three proposed clades possessed associated posterior probabilities (PP) in excess of 0.50. A clade containing *M. maculata* and *M. pubescens* was attributed a PP value of exactly 0.50 while a tritypic clade containing *M. guttata*, *M. littoralis* and *M. planifolia* was attributed an associated PP of 0.54. Both clades possessed posterior probability values too low for inference of support for either clade. The clade containing *M. longiflora* and *M. potosina* was the best supported clade with a PP of 0.82. Despite a PP greatly in excess of the other two clades, once more the PP value was too low for supporting an inference of a relationship between *M. longiflora* and *M. potosina.* All other taxa were unresolved (Figure 49).

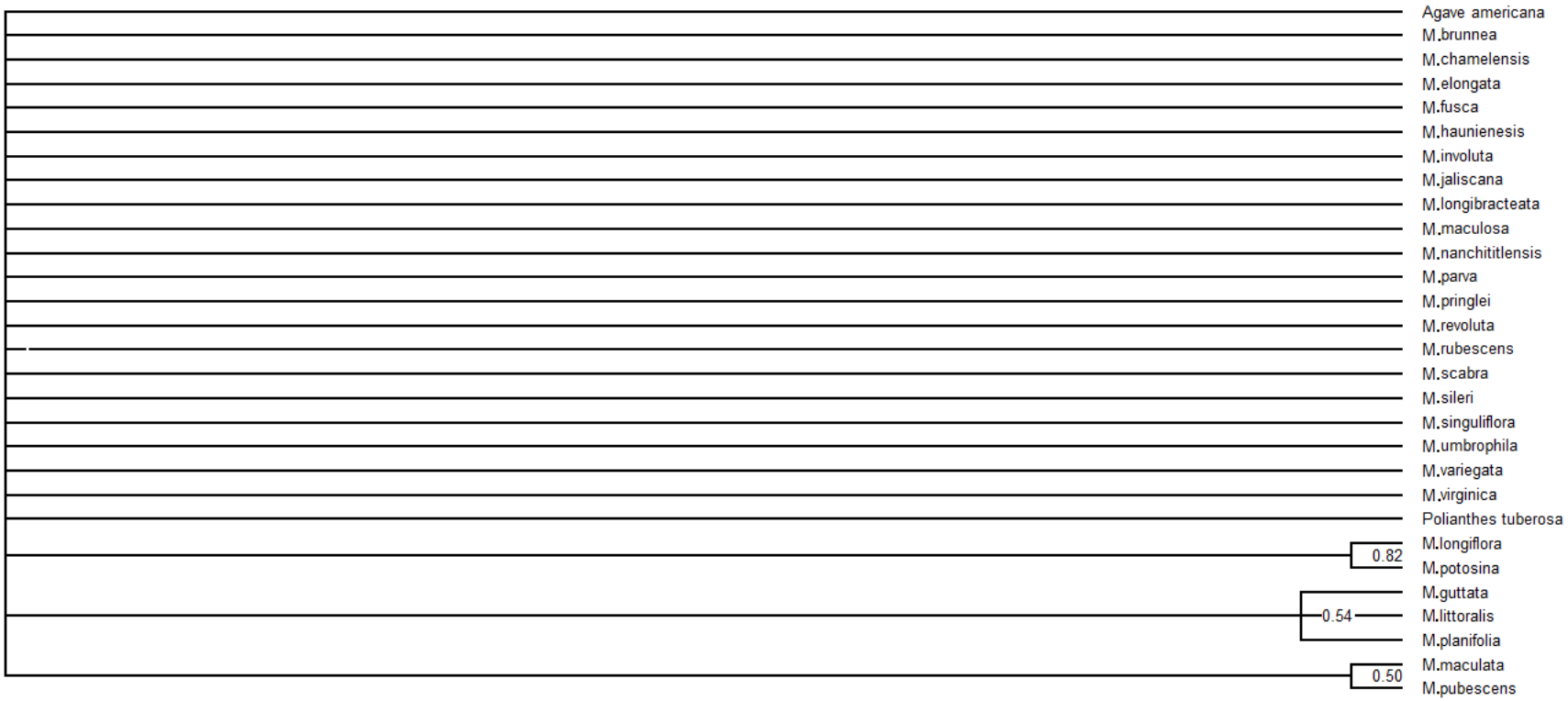


Figure 48: Bayesian analysis cladogram representing quantitative morphological characters of *Manfreda* and selected outgroups. Only Posterior Probabilities in excess of 50 are displayed.

Geographical Distribution Analysis. Of the 855 taxa studied, a geographical distribution was compiled encompassing the southeastern United States and Mexico. All taxa exhibited contiguous patterns of distribution across varied ranges (Figure 50).

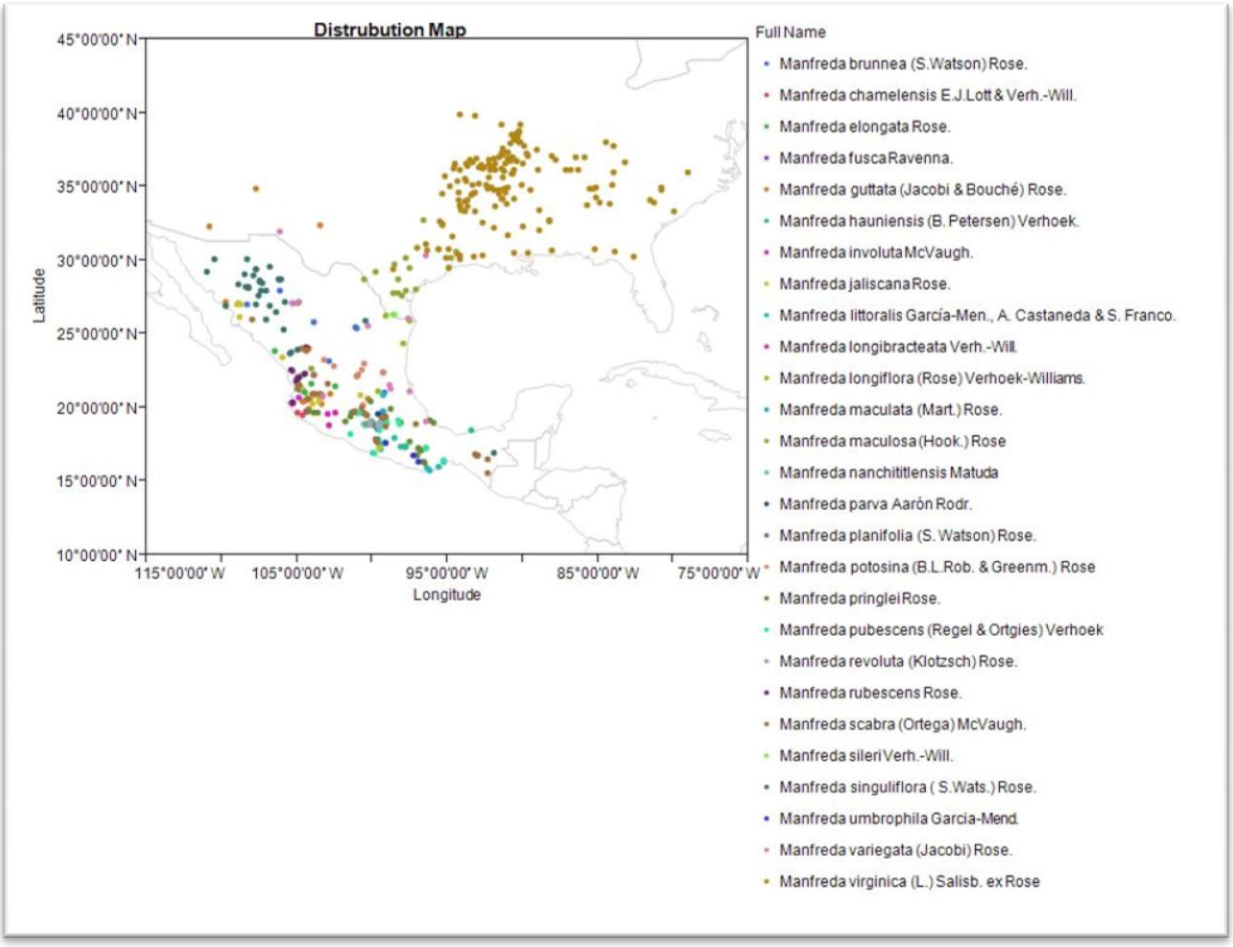


Figure 49. Distribution of all 855 herbarium specimens observed during the study with associated legend.

Discussion

The analysis of morphological characters has varied success in accord with the techniques employed. Of the 21 specific taxa subjected to Principal Component Analysis, 12 possessed sufficient uniformity to demonstrate clustering and close morphological affinity, while two lacked enough data or discernible pattern for any form of meaningful systematic inference. The residual seven taxa exhibited substantial variation in morphological features inciting query as to their specific status. Both the Maximum Parsimony and Bayesian phylogenetic techniques inferred minimal confidence in taxonomic relationships, with poor support for all relationships proposed. The Geographical Distribution Analysis identified 21 of 21 contiguous populations and highlighted potential imbricate distributions of morphologically similar taxa. The aforementioned analyses identified the following taxa as poorly supported taxonomic entities and in need of further revision:

Manfreda guttata. Principal Component Analysis of 12 specimens of *M. guttata* resulted in a broad distribution of data points. Moreover, the Pairwise Distance Matrix exhibited uniformity between *M. guttata, M. littoralis* and *M. planifolia* (all 0.0000). Such a close affinity was supported by both Maximum Parsimony Analysis and Bayesian Analysis, however a bootstrap value of 39 and a PP of 0.54 inferred limited confidence in the proposed clade.

Variability within the dataset was attributed to Filament Length, Ovary Length and Style Length. The standard deviation of each of the standardized characters aforementioned was 0.96, 0.89 and 0.86 with ranges of -1.14 to 1.68, -0.94 to 2.17 and -0.76 to 1.89 respectively.

Manfreda guttata possessed close morphological affinity to three other taxa, *Manfreda riosramirii* Solano & Castillejos, which was not studied due to limited herbarium specimens and

the aforementioned *M. littoralis* and *M. planifolia*. The specific status of *Manfreda riosramerii* was defined by leaves with widths between 0.7 cm to 1.2 cm which are glabrous on both the ventral and dorsal surfaces in contrast to the presence of verrucation on 1.5 cm to 2.5 cm wide leaves in *M. guttata*. Similarly, *M. planifolia* was differentiated due to elliptical leaves and a minutely denticulate margin in contrast to linear-lanceolate leaves with entire margins. Although the analysis failed to differentiate between *M. guttata* and *M. littoralis*, morphological distinction was apparent as *M. littoralis* is a small plant with a short peduncle of between 60-90 cm and a floral tube less than 0.57 mm in length (Verhoek-Williams, 1975; Castillejos-Cruz, 2009).

Geographical isolation of breeding populations is also highly likely; of morphological similar species only *M. planifolia* has an overlapping distrubution with *M. guttata*. Similarly, only *M. planifolia* exhibited congruent phenology, also flowering in June and July (Verhoek-Williams, 1975; Castillejos-Cruz, 2009).

Manfreda involuta. Of the five specimens employed in the Principal Component Analysis, clustering was observed in four with one distinct outlier. The Pairwise Distance Matrix was unable to differentiate between *M. involuta, M. longibracteata* and *M. nanchititlensis* (all 0.00). Maximum Parsimony analysis constructed a terminal clade containing *M.involuta*, *M. longibracteata* and *M. scabra* with an extremely low bootstrap value of seven; Bayesian Analysis was unable to resolve any relations to *M. involuta.*

The greatest variation was observed in the characters Plant Height, Leaf Length and Inflorescence Length. Standardized scores ranged from -0.4596 to 2.9920 with a standard deviation of 1.2973 for Plant Height, -0.4374 to 2.9648 with a standard deviation of 1.2778 for Leaf Length and -0.9763 to 1.6293 with a standard deviation of 1.0673 for Inflorescence Length.

The observed affinity with *M. longibracteata* and *M. scabra* was separated by Castillejos-Cruz (2009) by means of *M. involuta* possessing primary bracts of less than 0.6 cm in length as opposed to both of the other taxa having bracts far in excess of 0.6 cm. The closest related taxon morphologically to *M. involuta* is *M. nanchititlensis*. Differentiation between both taxa is defined by *M. involuta* possessing involute leaves and a cylindrical perianth (Verhoek-Williams, 1975; Castillejos-Cruz, 2009).

Manfreda scabra is the sole taxon with morphologically similar characters that inhabits both Jalisco and Michoacán states in Mexico, to which *M. involuta* is native. Overlap in flowering times does occur, both *M. longibracteata* and *M. scabra* flower from July to September while *M. involuta* flowers between March and July (Verhoek-Williams, 1975; Castillejos-Cruz, 2009).

Manfreda littoralis. The Principal Component Analysis of three *M. littoralis* specimens was sparse, and with such a limited sample minimal confidence could be attributed to such analysis. As previously discussed *M. guttata, M. planifolia and M. scabra* (all 0.00) could not be differentiated from *M. littoralis* and formed a clade under Maximum Parsimony (39) and Bayesian Analysis (0.54) (*see M. guttata*).

The premier source of variation was from Leaf Width, Inflorescence Length and Plant Height. Standardized scores for each ranged from 0.02 to 2.72, -1.87 to 0.62 and -1.17 to 0.15 with standard deviations of 1.39, 1.25 and 0.66, respectively.

Manfreda littoralis is differentiated from most species of *Manfreda*, including those aforementioned due to a small plant size, floral tube less than 0.57 cm in length and a peduncle of only 60-90 cm. The only other species to be characterized by these features is *M. bulbulifera*

Castillejos & E. Solano, a species of limited collections, and thus could not be sourced for inclusion in the analysis. Morphological characters used to differentiate *M. littoralis* from *M. bulbulifera* are floral; linear oblong tepals of 0.8 to 1.1 cm in length which are erect at anthesis differ to the linear 2.5-3.6 cm tepals which are reflexed at anthesis of *M. bulbulifera* (Castillejos-Cruz, 2009).

Manfreda scabra and *M. bulbulifera* are the only two species of close morphological affinity which inhabit a similar range, *M. scabra* inhabits both the states of Oaxaca and Guerrero in Mexico to which *M. littoralis* is native. *M. bulbulifera* has only been documented in the state of Guerrero. *M. scabra* is the sole morphologically similar species to flower in tandem with *M. littoralis*, with *M. scabra* flowering from July to September and *M. littoralis* flowering from August to October (Castillejos-Cruz, 2009).

Manfreda maculata. Of the nine specimens employed for *M. maculata*, the Principal Component Analysis calculated a disjunct distribution. The Pairwise Distance Matrix alluded to *M. pubescens* (0.0540) being the closest related taxon. Similarly, both Maximum Parsimony Analysis and Bayesian analysis grouped *M. maculata* and *M. pubescens* into a terminal clade with a bootstrap value of 33 and posterior probability of 0.50.

Substantial variation was observed in Ovary Length, Calyx Length and Calyx Width. Standardized scores ranged from -0.9388 to 3.4121, -0.7509 to 1.8805 and -0.6469 to 1.4035 with standard deviations of 1.2441, 0.8822 and 0.6934, respectively.

Manfreda maculata is differentiated from *M. pubescens* by bearing an inflorescence between 40 to 50 cm, tepals 0.5 to 1.2 cm and filaments 2.0 to 2.2 cm in length. In contrast *M.*

pubescens bears an inflorescence of 63 to 185 cm, tepals of 0.9 to 1.3 cm and filaments of 2.3-4.0 cm in length (Verhoek-Williams, 1975; Castillejos-Cruz, 2009).

The distribution of both *M. maculata* and *M. pubescens* bisect in the state of Guerrero and the phenology of both taxa are aligned as *M. maculata* flowers between July and September while *M. pubescens* has been documented as flowering during August (Verhoek-Williams, 1975; Castillejos-Cruz, 2009).

Manfreda pringlei. Principal Component Analysis of nine specimens for *M. pringlei* revealed a wide dispersal of data points and disjunct clusters. The Pairwise Distance Matrix revealed three taxa of equally close alignment, *M. guttata, M. parva* and *M. planifolia* (all 0.0273). Bayesian Analysis was unable to resolve the relationships to allied taxa, Maximum Parsimony aligned *M. pringlei* in proximity to a clade containing *M. scabra* and *M. virginica* supported by a 15 bootstrap value.

Variation within the dataset was high. The three characters with the largest variations were Leaf Width, Inflorescence Length and Ovary Length, standardized values ranged from -1.8881 to 2.3890, -1.8881 to 1.5881 and -1.2495 to 2.1690 with standard deviations of 1.3635, 1.3608 and 1.2729 respectively.

The differentiation of *M. pringlei* to morphologically similar taxa *M. guttata*, *M. parva* and *M. planifolia* utilizes attributes of the bract, inflorescence and fruit. *M. parva* has a primary bract less than 0.6 cm in length and both *M. guttata* and *M. planifolia* possesses fruit which are ellipsoid as well as a dense fertile portions of the peduncle in contrast to the cylindrical fruit and loose fertile portion of peduncle observed in *M. pringlei* (Verhoek-Williams, 1975; Castillejos-Cruz, 2009).

Both *M. pringlei* and *M. guttata* have considerable overlap in their distributions across central Mexico, both *M. parva* and *M. planifolia* inhabit ranges within Guerrero and Oaxaca in Mexico outside the range of *M. pringlei,* however. The phenology of the aforementioned taxa are aligned with *M.guttata, M. parva, M. planifolia* and *M. pringlei,* all flowering in May (Verhoek-Williams, 1975; Castillejos-Cruz, 2009).

Manfreda revoluta. Principal Component Analysis revealed wide variation between the three *M. revoluta* samples. Although samples were limited, it was surmised that the scale of dissimilarity warranted further investigation. The Pairwise Distance Matrix resulted in the two closest species, in regard to the analysis, *M. involuta* and *M. longibracteata* (both 0.0270). Maximum Parsimony inferred that *M. revoluta* had the strongest affinity with a clade containing *M. maculata* and *M. pubescens* with a bootstrap value of 33. Bayesian Analysis was unable to resolve the phylogenetic position of *M. revoluta*.

The substantial variation can be attributed primarily to the Ovary Length, Calyx Width and Plant Height with standardized scores ranging between -2.18 to 0.82, 0.38 to 2.94 and -0.52 to 1.48 with standard deviations of 161, 1.29, 1.11, respectively.

Manfreda revoluta most closely morphologically resembles *M. elongata*. They differ in that *M. revoluta* possesses a loose fertile portion of the peduncle, the entire inflorescence is 12 to 60 cm long, flowers are erect at anthesis and leaves are 12 to 20 cm long, 1.5-2 cm wide. In contrast, *M. elongata* has a dense fertile portion of the inflorescence, an entire inflorescence of between 44 to 96 cm, reflexed flowers at anthesis and leaves 35 to 46.5 cm long and 2.8 to 3.9 cm wide. *M. involuta* and *M. longibracteata* are separated by Castillejos-Cruz (2009) based on

M. revoluta having bracts smaller than 0.6 cm and filaments inserted in the uppermost quarter of the perianth tube, respectively (Verhoek-Williams, 1975; Castillejos-Cruz, 2009).

Manfreda revoluta is endemic to Mexico State, in which no documentation of *M. elongata*, *M. involuta* or *M. longibracteata* have been recorded. However, all three aforementioned taxa do have aligned phenology with *M. revoluta*, as all flower in July and August (Verhoek-Williams, 1975; Castillejos-Cruz, 2009).

Manfreda scabra. Principal Component Analysis of three specimens resulted in a varied pattern of dispersal. The Pairwise Distance Matrix identified *M. hauniensis* (0.0000) as the most closely related species in morphology. Despite this, Maximum Parsimony analysis placed *M. scabra* in a terminal clade with two other constituents, *M. involuta* and *M. longibracteata* supported by a very low bootstrap value of 7. Bayesian Analysis was unable to resolve any relationship involving *M. scabra.*

Ovary Length, Plant Height and Leaf Length contributed the greatest variation to the dataset with ranges between 0.6151 to 2.1690, 0.6557 to 2.1690 and 0.6640 to 2.1939 with standard deviations of 1.6510, 1.2919 and 1.2882, respectively.

Despite close affinities between morphological data, distinct phenotypic attributes employed by Castillejos-Cruz (2009) differentiated between *M. scabra* and affiliated taxa. *M. hauniensis* possesses succulent leaves, *M. involuta* has primary bracts which are smaller than 0.6 cm and *M. longibracteata* has cylindrical fruit and a loose fertile portion of the inflorescence.

Manfreda scabra has the widest distribution of any species of *Manfreda* in Mexico, and thus overlaps with all other morphologically similar species. With flowering periods between

June and February, phenology also overlaps all aforementioned taxa (Verhoek-Williams, 1975; Castillejos-Cruz, 2009).

The classification systems proposed by Verhoek-Williams (1975) and Castillejos-Cruz (2009) when subjected to analysis, performed admirably, with the majority of Opertional Taxonomic Units (OTUs) employed fairing well. Two taxa however, post analysis and after further review were still subject to low confidence due to hypervariablity in morphometric measures, negating the ability for classification via Principal Component Analysis. The relationship between *M. maculata* and *M. pubescens* were separated morphologically by only proportions of anatomical features, and yet in distribution and phenology there is much overlap.

Observation of herbarium specimens for both taxa exhibited considerable variation in the morphological characters on which they are separated. Of the 26 inflorescence lengths observed for *M. maculata*, a range of 35 cm and 162 cm was recorded, greatly exceeding the 40 cm to 50 cm range defined by Castillejos-Cruz (2009). Of the 15 inflorescence lengths observed for *M. pubescens*, a range of 85 cm to 184 cm was recorded. Although differences in range were observed the majority of specimen exhibit inflorescences lengths within both ranges and thus differentiating between both OTUs based on such a character would be challenging.

Castillejos-Cruz (2010) also employed tepal length to separate *M. maculata* with a range between 5 mm to 12 mm from *M. pubescens* with a range of 9 mm to 13 mm. Of the 15 tepal lengths recorded for *M. maculata*, a range of between 10 mm and 31 mm was observed and for the 19 examples observed in *M. pubescens* a range between 12 mm to 36 mm was recorded. Once more, the majority of specimens would be intermediates of both ranges and thus identification of either OTU would be difficult based on this character.

Filament length was the third derived character employed by Castillejos-Cruz (2009) to differentiate between both OTUs. With a range of 19 mm to 32 mm based on 11 specimens and 23 mm to 82 mm based on 25 specimens for *M. maculata* and *M. pubescens*, respectively. Filament length performed better as a field character and both the means for observed specimens of *M. maculata* (25.7 mm) and *M. pubescens* (48.85 mm) lie within their respective ranges.

The sole other character utilized to differentiate between both OTUs was leaf base form utilized by Verhoek-Williams (1975). Of herbarium specimens observed, the form of the base was consistent to both specimens of *M. maculata* and *M. pubescens*. The basal portion of *M. maculata* leaves were consistently attenuate with an approximately straight or biconcave form creating an acute angle of 45° or less. In contrast, *M. pubescens* specimens possessed a broader basal leaf portion, with no concave margins and an angle greater than 45°. It was therefore deemed that the attribute was of sufficient merit to be utilized as a field character for differentiation between the two OTUs.

The limited derived characters consistently observed between *M. maculata* and *M.pubescens*, phylogenetic affinity, albeit poorly supported, as well as corresponding biogeography and phenology, are indicative of taxa of limited diversification. No records of hybridization between the two taxa has been documented, with the acquisition of such however it can only be surmised that support for two independent taxonomic entities is limited.

Early taxonomic investigations of J.N Rose and J.M. Greenman utilized a single taxonomic entity, *M. maculata*, for taxa characterized by pubescent leaves, filaments of equal length and moderate curvature of the perianth tube (Rose, 1905). Consideration however must be taken for

the derived characters identified by Verhoek-Williams (1975) and Castillejos-Cruz (2009), although only leaf base and filament length were deemed consistent and viable field characters.

Therefore, in accord with Hamilton & Reichard (1992) and the International Code of Botanical Nomenclature (ICBN Editorial Committee, 2005) the following varietal system for *M. maculata* is proposed.

Key to varieties of *Manfreda maculata* (Mart) Rose. sensu Ritchie

1a. Leaves attenuate at base, filaments 19 mm to 32 mm in length.....

.....*Manfreda maculata* var. *maculata*

1b. Leaves cuneate at base, filaments 23 mm to 82 mm in length .....

.....*Manfreda maculata* var. *pubescens*

1a. Manfreda maculata* (Mart) Rose. var. *maculata, stat. nov. - TYPE: Oaxaca. Valles Calientes de Oaxaca.

Ab omnibus subspeciebus characteribus combinatis differt: attenuatis basi foliis, filamentis 19 mm ad 32 mm in longitudine. Leaves attenuate at base; filaments 19 mm to 32 mm in length (mean 26 mm, n = 11).

Phenology. Flowering from July to September.

Distrubution. Occuring in *Pinus* spp.-*Quercus* spp. or deciduous woodland in Guerrero, Oaxaca and Mexico State, Mexico.

***1b. Manfreda maculata* (Mart.) Rose. var. *pubescens* (Verh.-Will.) Ritch.**, stat. nov. - TYPE: ex horto bot. petro-politano, 73.4, Collector unknown.

Ab omnibus subspeciebus characteribus combinatis differt: cuneatae basi foliis. filamentis 23 mm ad 82 mm in longitudine. Leaves attenuate at base; filaments 23 mm to 82 mm in length (mean 49 mm, n = 25).

Phenology. Flowering during August.

Distrubution. Occuring in *Pinus* spp. and *Quercus spp.* woodland or on Scree Slopes in Guererro, Morelos and Oaxaca (Occasionally Michoacán and Puebla) Mexico.

Specimens examined. Manfreda maculata - MEXICO. State of Mexico: Tejupilec, 1954, E. Matuda (MEXU). Tejupilec, 1954, E. Matuda (MEXU). Tejupilec, 1954, E. Matuda (MEXU). Tejupilec, 1961, H.S. Gentry (ASU). Tejupilec, 1961, H.S. Gentry (ASU). Tejupilec, 1961, H.S. Gentry (ASU). Tejupilec, 1961, H.S. Gentry (ASU). Tejupilec, 1961, H.S. Gentry (ASU). Tejupilec, 1961, H.S. Gentry (DES). Tejupilec, 1961, H.S. Gentry (MEXU).Temascaltepec, 1990, A. Garcia-Mendoza (ASU). Temascaltepec, 1990, A. Garcia-Mendoza (ASU). Temascaltepec, 1990, A. Garcia-Mendoza (DES). Temascaltepec, 1990, A. Garcia-Mendoza (DES). 7km N Zacualpan, 1990. A. Garcia-Mendoza (MEXU). 7km N Zacualpan, 1995. A. Garcia-Mendoza (DES). 7km N Zacualpan, 1995. A. Garcia-Mendoza (DES). 7km N Zacualpan, 1995. A. Garcia-Mendoza (MEXU). 7km N Zacualpan, 1995, A. Garcia-Mendoza (MEXU). A. Garcia-Mendoza (MEXU). A. Garcia-Mendoza (MEXU). 6km SW Temascaltepec, 2004, A. Rodríguez (MEXU). Temascaltepec, 2004, A. Rodríguez (IBUG). Temascaltepec, 2007, A. Rodríguez (IBUG). Temascaltepec, 2007, A. Rodríguez (IBUG). Oaxaca: 2.5km W Temascal, 1987. Guererro: 1km E Chipla, 1982. R. Torres (MEXU) Chilpa, 1986, C.Catalan (MEXU). Tujupilco, 2007, A. Rodríguez (IBUG). Temascaltepec, 2007, A. Rodríguez (IBUG). Temascaltepec, 2007, A. Rodríguez (IBUG). Temascaltepec, 2009, A. Rodríguez (IBUG).

Manfreda pubescens. MEXICO. Cultivation: UNAM Jardin Botanico, 1992, A. Garcia-Mendoza (MEXU). Guererro: Chilapa, 1983, I.P. Lujan (MEXU). Chilapa, 1988, I.P. Lujan (MEXU). De Los Amates, 1997, A. Garcia-Mendoza (MEXU). De Los Amates, 1998, A. Garcia-Mendoza (ASU). De Los Amates, 1998, A. Garcia-Mendoza (ASU). De Los Amates, 1998, A. Garcia-Mendoza (ASU). 2km N De Los Amates, 1998, A. Garcia-Mendoza (ASU). 1km N De Los Amates, 1998, A. Garcia-Mendoza (MEXU). De Los Amates, 1998, A. Garcia-Mendoza (MEXU). De Los Amates, 1998, A. Garcia-Mendoza (MEXU). De Los Amates, 1998, A.

Garcia-Mendoza (MEXU). Pilcaya, 1998, A. Garcia-Mendoza (IBUG). Zirandaro. Michoacán: Tzitizio, 1992, A. Garcia-Mendoza (MEXU). Morelos: Tepoztlan, 1995, A. Garcia-Mendoza (MEXU). San Jose de Los Laureles, 1998, A. Garcia-Mendoza (ASU). Oaxaca: Ixtlan, 1952, H.S. Gentry (ARIZ). 18km NW Temascaltepec, 1986, C. Martinez (MEXU). Temascaltepec, 1992, A. Garcia-Mendoza (MEXU). Tehuantepec, 1998, A. Garcia-Mendoza (ASU). Tehuantepec, 1998, A. Garcia-Mendoza (ASU). Tehuantepec, 1998, A. Garcia-Mendoza (ASU). Puebla: 2km W San Jose de Jaluca, 2008, A. Garcia-Mendoza (IBUG).

Glossary of Terms

Z-Score. A standardized score indicating how many standard deviations a value is above or below the mean.

Eigenvalue. A mathematical description of the amount of variability assigned to each principal component.

50% Majority Rule. A procedure in consensus tree construction were only speciation events documented in greater than 50% of trees generated are retained.

Bootstrap Value. A technique utilized in Maximum Parsimony analyses to assign an estimated degree of confidence to speciation events.

Branch Swapping. A technique employing a series of branch rearrangements of an initially generated tree to test for greater parsimony.

Heuristic Search. An abbreviated search technique for the most parsimonious trees utilizing a series of branch rearrangements.

Ingroup. The group studied by the investigator in phylogenetic analyses.

Monophyletic. A group of taxa which includes an ancestral taxon and all descendants.

Outgroup. A group of taxa employ in phylogenetic analysis for comparative purposed and not directly under study by the investigator.

Parsimony Informative Character. A nucleotide position with a minimum of two alternate states, each of which is represented by at least two taxa understudy.

Parsimony Uninformative Character. A nucleotide position with a minimum of two alternate states, yet represented by a single taxon.

Score Plot. A two dimensional plot comprising of two principal components. The scores are the intersection between the first and second principal components.

Scree Plot. A line graph of eigenvalues visually illustrating the variation assigned to each principal component.

Tree-Bisectional-Reconnection (TBR). A basic branch swapping technique which employs 'pruning' of tree sections and reattachment to survey for greater parsimony.

Literature Cited

Angiosperm Phylogeny Group. 2009. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. Bot. J. Linnean Soc. 161: 105-121.

Bogler, D. J. and B.B. Simpson. 1995. A chloroplast DNA study of the Agavaceae. Systematic Bot. 20(2): 191-205.

Bogler, D. J. and B. B. Simpson. 1996. Phylogeny of Agavaceae based on its rDNA sequence variation. Amer. J. Bot. 83(9): 1225-1235.

Castillejos-Cruz, C. 2009. Sistemático del genero *Manfreda*. Natl. Autonomous Univ. Mexico, Mexico City, PhD Diss.

Chase, M. W., J. L. Reveal and M. F. Fay. 2009. A subfamilial classification for the expanded asparagalean families Amaryllidaceae, Asparagaceae and Xanthorrhoeaceae. Bot. J. Linnean Soc. 161: 132-136.

Good-Avila, S. V., V. Souza, B. S. Gaut and L.E. Eguiarte. 2006. Timing and rate of speciation in *Agave* (Agavaceae). Proc. Natl. Acad Sci. U.S. 103(24): 9124-9129.

Hamilton, C. W. and S. H. Reichard. 1992. Current practice in the use of subspecies, variety, and forma in the classification of wild plants. Taxon 41(3): 485-498.

McNeill, J., F.R. Barrie, H.M. Burdet, V. Demoulin, D.L. Hawksworth, K. Marhold, D.H. Nicolson, J. Prado, P.C. Silva, J.E. Skog, J.H. Wiersema and N.J. Turland. 2006. 146: International Code of Botanical Nomenclature (Vienna Code) adopted by the Seventeenth International Botanical Congress Vienna, Austria, July 2005. Gantner Verlag, Ruggell, Liechtenstein.

Lindstrom, J.T. 2006. Intergeneric hybrids between *Polianthes* and *Manfreda*. Southern Nursery Assn. Res. Conf. 51: 599-601.

McKelvey, S. D. and K. Sax. 1933. Taxonomic and cytological relationships of *Yucca* and *Agave*. J. Arnold Arboretum 15: 76-80.

Rose, J.N. 1905. Contributions for the United States National Herbarium, volume VIII: Mexican, Central American and Porto Rican plants. Government Press Office, Washington, D.C.

Simpson, M. G. 2010. Plant systematics. 2 ed. Elsevier, London, England.

Verhoek-Williams, S.E. 1975. A study of the tribe Poliantheae (including *Manfreda*) and revision of *Manfreda* and *Prochnyanthes* (Agavaveae). Ithaca, NY: PhD Thesis, Cornell University.

Chapter 2

ASSESSMENT OF CBOL PLANT DNA BARCODES FOR PHYLOGENETIC RESEARCH AND DNA BARCODING IN THE GENUS *MANFREDA* SALISB. (ASPARAGACEAE)

Introduction

The technique known as DNA Barcoding, as stated by the Consortium for the Barcode of Life, is "a technique for characterizing species of organisms using a short DNA sequence from a standard and agreed-upon position in the genome." The name DNA barcode employs the metaphor of barcode to illustrate the ability of the technique to assign uniquely identifying sequences to taxonomic entities (Barcodeoflife.org, 2012). The comparison can be misconstrued, because unlike barcodes, a degree of variation can occur between populations and the gene regions employed for DNA barcoding will be constantly subject to evolution and base alteration (Moritz & Cicero, 2004).

DNA barcoding rose to prominence as a molecular method of much merit and interest from the taxonomic community, due predominately to its application to large scale batch processing. Advocates of the technique favor DNA barcoding over traditional descriptive taxonomy, via which 1.7 million species have been described, to undertake or assist in the estimated next 10-100 million species still to be documented (Newmaster *et al*, 2006).

The task of identifying suitable DNA barcoding regions for the plant kingdom was assigned to an international collaboration of 52 plant scientists from 24 institutions, known as the Plant Working Group (Levin, 2009). Collective research concluded that the most viable DNA barcode for plants would consist of two plastid regions, *MatK* and *rbcL*. The newly proposed

DNA barcode, despite demonstrating discriminatory ability in only 72% of species, was verified and certified by the CBOL Executive Committee as the official DNA barcode of land plants (CBOL Executive Committee, 2009; Plant Working Group, 2009).

Application of the CBOL DNA Barcode to studies of taxonomic groups within the plant kingdom has been subject to varied success. Studies such as that of De Vere (2009) and Roy (2010), investigating species delimitation in the order Rosales Perleb. and *Berberis* L. respectively, noted failure of the DNA barcode to discriminate between all species. The majority of studies demonstrated utility of the technique, however. The CBOL Plant DNA Barcode has been employed in a range of taxonomic groups and purposes such as local biodiversity inventories, forensic identification and indentification in the horticultural nursery trade, demonstrating the wide array of applications for the technique if successful discrimination can be achieved (De Vere, 2009; Lou *et al*, 2010; Burgess *et al*, 2011; Khew & Chia, 2011).

Manfreda Salisb., is a genus of complex taxonomy and life history. The 26 - 32 constituent species of the genus inhabit a diverse array of localities and conditions, predominantly in the southeastern United States and Mexico, but have also been documented in El Salvador, Honduras and Guatemala (Castillejos-Cruz, 2009). The high likelihood of cytonuclear disequilibrium and reticulation being a commonality within populations in *Manfreda* is high due to an observed low resistance to hybridization (Verhoek-Williams, 1975). Similarly, phenotypic plasticity and hypervariability of anatomical proportions have made identification and classification of species within the genus challenging.

The genus *Manfreda* was described from specimens of a single species, *Manfreda virginica* (L.) Salisb. ex Rose by Richard Salisbury FRS in *The genera of plants - A fragment*

containing part of Liriogamae (1866). The newly conceived genus was not widely adopted until Rose (1905). During travels to Mexico, Joseph Nelson Rose was convinced of the distinction between *Manfreda* and *Agave* L. and subsequently composed a 17-species system for the genus. Verhoek-Williams (1975) and Castillejos-Cruz (2009) are the only two other systems employed for the genus consisting of 26 and 32 species, respectively. All three multi-specific systems possess limitations: all three systems are highly dependent on inflorescence and floral characters, rely upon arbitrary measures and are based on phenetic principals with no accommodation for homoplasy, reticulation or plesiomorphic characters. The aforementioned factors in combination with inadequate collections of many species for study, permit only a limited degree of confidence to be inferred for existing taxonomic systems based on morphology (Oldfield, 1997).

Molecular systematic studies concerning *Manfreda* have been predominantly focused on ordinal and familiar classifications. Early karyological works of McKelvey & Sax (1933), Whittaker (1934) and Sāto (1935) united the former Agavaceae by documenting a near identical bimodal chromosomal complement of five large and 25 small pairs at meiotic division. Immunological studies by Chupov and Kutiavena (1981) and later restriction site analysis of chloroplast DNA by Bogler and Simpson (1995) inferred inter and intra familiar relationships, throughout which *Manfreda* remained stable as a core species of the former Agavaceae.

Of phylogenetic studies, only two prominent publications have incorporated species of *Manfreda*. Bogler and Simpson (1996) utilized the Intergeneric Spacer Region (ITS) to study the former Agavaceae employing *M. scabra* (Ortega) McVaugh. and *M. virginica*. Good-Avila *et al* (2006) utilized *M. hauniensis* (J.B. Petersen) Verh.-Will., *M. nanchititlensis* Matuda. and *M. potosina* (B.L.Rob. & Greenm) Rose in their study of speciation in Agavaceae employing the *trnL* intron and *trnL-trnF* intergenic spacer (*trnL* and *trnL-trnF*). Both studies were unable to

infer a monophyletic *Manfreda*, separated by intermittent species of *Agave* and *Polianthes* L. Moreover, limited specimens were employed and both studies trialed only four phylogenetic markers between them, thus minimal meaningful inference as to the specific composition of *Manfreda* has been accomplished to-date.

The aim of this study was to assess the utility of the CBOL Plant DNA Barcode for both phylogenetic studies and identification through DNA barcoding efforts. Testing the ability of the employed gene regions to infer taxonomic relationships between species of *Manfreda* will facilitate a greater understanding of the potential of *MatK* and *rbcL* to convey greater confidence in the phylogeny of *Manfreda*. Study of the utility of the CBOL Plant DNA Barcode, within the genus *Manfreda*, will add further evidence to the debated universality of the CBOL Plant DNA Barcode. Furthermore, such investigations will aid our understanding as to whether DNA barcoding can be applied to studies such as population genetics, conservation and horticulture.

Materials and Methods

Specimen Collection. Specimens were harvested from living collections, trimmed to 5 mm x 5 mm, placed in a tea bags which were submerged in a Ziploc® bag of silica gel. All specimens were stored at room temperature. Specimens were sourced predominantly from living collections of Dr. Jon T. Lindstrom at the University of Arkansas and Dr. Aáron Rodríguez at the University Center for Biological and Agricultural Sciences, University of Guadalajara, Mexico. Leaf material was also harvested on request and sent to the University of Arkansas from the Huntington Library, Art Collection and Botanic Garden in San Marino, California and the Royal Botanic Garden Edinburgh, Scotland (Table 7).

Table 8. List of specimens employed in phylogenetic analysis employing Maximum Parsimony and Bayesian Analysis. Nineteen specimens of *Manfreda* and three outgroups were utilized.

Name	Author	Institute of origin	Determination
Agave sisalana [Outgroup]	Perrine.	GenBank: GU135234 (*rbcL*) FR717534 (*MatK*)	J. Abbott
Camassia cusickii [Outgroup]	S. Watson.	GenBank: HM640479 (*rbcL*) HM640593 (*MatK*)	D. Kim
Manfreda brunnea	(S.Watson) Rose.	University of Guadalajara	A. Rodríguez
Manfreda guttata	(Jacobi & C.D.Bouché) Rose.	University of Guadalajara	A. Rodríguez
Manfreda hauniensis	(J.B.Peterse n) Verh.-Will.	University of Guadalajara	A. Rodríguez
Manfreda jaliscana	Rose.	University of Guadalajara	A. Rodríguez
Manfreda longibracteata	Verh.-Will.	University of Guadalajara	A. Rodríguez
Manfreda longiflora	(Rose) Verh.-Will.	University of Arkansas	W. Ritchie
Manfreda maculata	(Mart.) Rose.	University of Guadalajara	A. Rodríguez
Manfreda maculosa	(Hook.) Rose.	The Huntington	W. Ritchie
Manfreda nanchititlensis	Matuda.	University of Guadalajara	A. Rodríguez
Manfreda potosina	(B.L.Rob. & Greenm.) Rose.	The Huntington	W. Ritchie
Manfreda pringlei	Rose.	University of Guadalajara	A. Rodríguez
Manfreda pubescens	(Regel & Ortgies).	Royal Botanic Garden Edinburgh	W. Ritchie
Manfreda rubescens	Rose.	University of Guadalajara	A. Rodríguez
Manfreda scabra	(Ortega) McVaugh	University of Arkansas	W. Ritchie
Manfreda singuliflora	(S.Watson) Rose.	University of Guadalajara	A. Rodríguez
Manfreda undulata	(Klotzsch) Rose.	University of Arkansas	W. Ritchie
Manfreda variegata	(Jacobi) Rose.	The Huntington	W. Ritchie
Manfreda virginica	(L.) Salisb. ex Rose.	The Huntington	W. Ritchie
Yucca gigantea [Outgroup]	Lem.	GenBank: JQ590093 (*rbcL*) JQ586436 (*MatK*)	M. Hajibabaei

DNA Extraction. The following protocol was adapted from Keb-Llanes *et al*. (2002), Tapia-Tussel *et al*. (2005) and personal communication from Dr. Gerardo Salazar, National Autonomous University of Mexico, Mexico and optimized for the attributes of *Manfreda* leaf tissue. (CTAB and sodium borate buffer solutions composed are detailed in Appendix 3).

1. Hexadecyltrimethylammonium bromide (CTAB) solution was warmed to 65°C on a Thermolyne® 17600 Dri-Bath

2. 5 mm x 5 mm leaf segments were submerged in liquid nitrogen and pulverized using a mortar and pestle

3. 2µl of β-mercaptoethanol was added to 500 µl of pre-warmed CTAB solution and mixed on a Vortex-Genie® 2 vortex mixer

4. Pulverized leaf material was added to the CTAB/β-mercaptoethanol solution in an 2ml eppendorf tube and mixed for 3 to 5 seconds on the Vortex-Genie® 2 vortex mixer

5. CTAB/β-mercaptoethanol solution containing pulverized leaf material was heated to 65°C on the Thermolyne® 17600 Dri-Bath for 30 minutes

6. The mixture was cooled to room temperature; 600 µl of chloroform-isoamyl alcohol 24:1 is added to the eppendorf tube and gently agitated for 30 minutes on a The Belly Dancer® laboratory shaker

7. Lysate was subjected to centrifugation in an Eppendorf 5417C centrifuge for 10 minutes at 12,500g

8. Upper aqueous layer was transferred to a new 2ml eppendorf tube, to which 700 µl of pre chilled isopropanol is added prior to incubation at -20°C for 1 hour

9. Sample was centrifuged in an Eppendorf 5417C centrifuge for a further 5 minutes at 26,500g prior to residual liquid being discarded

10. 500 µl of 70% ethanol chilled at 3°C was added to the eppendorf tube and centrifugation is repeated for a further 5 minutes at 26,500g

11. Residual 70% ethanol was decanted and samples were dried for 15 minutes at 55°C on the Thermolyne® 17600 Dri-Bath

12. Dried DNA pellets were re-suspended with 40 µl ddH_20 and left at room temperature overnight or subjected to a further 1 hour at 55°C on the Thermolyne® 17600 Dri-Bath to resuspend the DNA.

13. Samples were stored at -20°C.

Gel Electrophoresis of Extract. Gel electrophoresis was conducted to check the presence of genomic DNA in a horizontal mini-gel system. A 2% agarose gel solution was cast, a sodium borate buffer was employed and electrophoresis was conducted at 175v for 35 minutes. DNA extract samples with Blue/Orange 6x Loading Dye (Promega, Madison, Wisconsin) and run simultaneously with a Benchtop 1KB DNA Ladder (Promega, Madison, Wisconsin). Gels were stained with GelRed Nucleic Acid Stain 10,000x (Biotium, Hayward, California) by gently agitating a solution 15 µl GelRed Nucleic Acid Stain 10,000x , 5ml Sodium Borate buffer and 55ml distilled water. Visualization of gel electrophoresis products were conducted using the BioDoc-It® 220 Imaging system (UPV LLC, Upland, Calif.).

Quantification of Extract. DNA extracts were also quantified using a Nanodrop 1000 Spectrophotometer V3.7 system (Thermo Fisher Scientific, Waltham, Mass.), 1 µl of each extract was utilized from each sample.

Polymerase Chain Reaction (PCR). A PCR procedure was conducted in a PCR Sprint Thermocycler (Thermo Fisher Scientific, Waltham, Mass.). High Fidelity PCR EcoDry™ premixed tubes (Clontech Laboratories, Mountain View, Calif.) were employed to which 0.5 µl of DNA template, 2.0 µl of each primer at a concentration of 10 mM and 20.5 µl of ddH_20 were added. Primers employed are listed in table 8. The cycling conditions for the amplification of both *MatK* and *rbcL* were adopted from Plant Working Group (2009) and are detailed in Table 9.

Table 9. Primers employed in PCR amplification of the *MatK* and *rbcL* plastid gene regions, as suggested by the Consortium for the Barcode of Life's Plant Working Group (Hollingsworth, 2009).

Primer Name	Gene Region	Direction	Sequence (5' to 3')
3F_Kim f	*MatK*	Forward	CGTACAGTACTTTTGTGTTTACGAG
1R_Kim r	*MatK*	Reverse	ACCCAGTCCATCTGGAAATCTTGGTTC
rbcL*a*_F	*rbcL*	Forward	GTAAAATCAAGTCCACCRCG
rbcL*a*_R	*rbcL*	Reverse	ATGTCACCACAAACAGAGACTAAAGC

Table 10. Cycling conditions employed in the PCR reaction for amplification of the *MatK* and *rbcL* plastid gene regions as suggested by the Consortium for the Barcode of Life's Plant Working Group (Hollingsworth, 2009).

MatK			
	Time	**Temperature**	**Cycles**
Initial Denaturation	1 minute	94^0C	1
Denaturation	30 seconds	94^0C	35
Annealing	20 seconds	52^0C	
Extension	50 seconds	72^0C	
Final Extension	5 minutes	72^0C	1
rbcL			
	Time	**Temperature**	**Cycles**
Initial Denaturation	4 minutes	95^0C	1
Denaturation	30 seconds	94^0C	35
Annealing	55 seconds	55^0C	
Extension	1 minute	72^0C	
Final Extension	10 minutes	72^0C	1

DNA Purification. Purification of DNA samples was conducted employing the Nanosep® 30K, Red Centrifugal Device with Omega™ Membrane (Pall Corporation, Port Washington, N.Y.). DNA samples were applied to the Omega™ membrane with 200 µl of ddH_20. Centrifugation was conducted in an Eppendorf 5417C centrifuge at 4,700g for 20 minutes. 40 µl of ddH_20 was then transferred directly onto the Omega™ Membrane and mixed, subsequently the 40 µl of ddH_20 containing the residual DNA was removed and transferred into a new 2ml eppendorf tube. DNA samples were stored at -20°C.

DNA Sequencing. Sequencing was conducted at Eurofins MWG Operon (Huntsville, Ala.). Samples containing 10 µl of DNA template at 20-50ng/µl and 10 µl of each of the required primers at 2µM were shipped overnight. Sequencing in a forward and reverse direction was

conducted on a ABI 3730 XL DNA Sequencer (Applied Biosystems, Foster City, Calif.), resultant electropherogram and sequence files were sent within a 36 hour time period after sample submission.

DNA Sequence Editing. Editing of DNA sequences was conducted in BioEdit 7.1.3.0 (Hall, 1999) and Geneious Pro™ 5.6 (Kearse, *et al*. 2012). BioEdit 7.1.3.0 was first employed while verifying and correcting ambiguous base calls by consulting the respective electropherogram. Alignment and composition of consensus sequences between the forward and reverse sequences were conducted using Geneious Pro™ 5.6. Consensus sequences were aligned employing the GENEIOUS algorithm and sequences were trimmed and augmented to form a combined *MatK* and *rbcL* sequence of 1236bp as per CBOL data standards (Hanner, 2009).

Bayesian Analysis. Model testing was first performed via jModelTest 0.1.1 (Posado, 2008) and the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The models deemed most appropriate for the data was TPM1uf+G for AIC and F81 for BIC. Analysis employing both models was conducted on a Mac Pro "Quad Core" 3.1 using the MrBayes 3.2. program (Huelsenbeck & Ronquist, 2001). Four Markov Monte Carlo chains were employed, three warm and one cool with default parameters. Generations were set at 5,000,000 repetitions with sampling every 1000 replicates. Burn-in was conducted on 500,000 trees (10%). Default settings were utilized for the analysis with the following expectations listed in Table 10. The outgroup was selected as *C. cusiskii.*

Table 11. Settings altered for parameters and priors from default for implementation of Bayesian Analysis via MrBayes 3.1 for phylogenetic analysis.

Parameter	Default Setting	Altered Setting	Description
Lset	nst = 1	nst = 6	modification of the sustitution type from 1 (all rates equal) to 6 (general-time reversability)
Lset	rates = equal	rates = gamma	modification of rate variation from equal (no rate variation) to gamma (rate variation)
PRset	shapepr = uniform (0,0)	shapepr = fixed (0)	modification of the gamma shape parameter from uniform (between 0 and 1) to fixed (0.05)
PRset	pinvarpr = uniform (0,0)	pinvarpr = fixed (0)	modification of the prior for proportion of invariable site from uniform (between 0 and 1) to fixed (0.1)

Uncorrected "P" Distance Matrix. Analysis was conducted employing PAUP* 4.0. Genetic distances were calculated using the uncorrected P algorithm by setting DSet Distances to P and executing the *showdist* command.

Maximum Parsimony Analysis. Analysis was conducted employing a Power Mac G4 and the PAUP 4.0* program (Swofford, 2001). All characters employed in the analysis were unordered and unweighted. A heuristic search was conducted with a Stepwise Addition, the default Tree-Bisectional-Reconnection (TBR) was employed for synthesis of the initial tree. Bootstrap replicates were set to 1000 with one tree held at each Stepwise Addition. MulTrees was activated yet the deepest descendant option was not utilized. Outgroups were designated as *A. sisalana*, *C. cusickii* plus *Y. gigantea* and the 50% majority rule was employed in consensus tree generation.

Nucleotide Network Analysis. Nucleotide Network Analysis is a technique, used as an alternative to phylogenetic inference, capable of mapping nucleotide polymorphisms. The method utilizes a distance matrix to calculate divergence between haplotypes. Employment of the technique for discrimination between potential DNA barcodes is favorable as opposed to phylogenetic inference due to the capacity of algorithms employed to discriminate between closely related DNA sequences. Analysis was conducted using HapStar 0.7 (Teacher & Griffiths, 2011). The Distance Estimation analysis was employed employing p-distance substitution model including transitions and transversions. Rates were set to uniform and pattern to homogeneous. All three codon positions were selected for analysis.

Results

Bayesian Analysis - TPM1uf+G. The generated cladogram demonstrated only two clades (Figure 51). Both the outgroups of *C. cusickii* and *Y. gigantea* were positioned external to both clades, however the third outgroup, *A. sisalana*, was unable to be distinguished from 13 *Manfreda* specimens. The basal clade included all *Manfreda* specimens and was supported by a moderate posterior probability (PP) of 0.79. Within the clade, however, relationships between most specimens of *Manfreda* was unresolved and no intrageneric inference as to relationships could be surmised. Within the basal clade, a secondary clade, strongly supported by a PP of 1.00, was present. The secondary clade contained six *Manfreda* specimens, yet again no interspecifc relationships were observed.

Bayesian Analysis - F81. Both models, TPM1uf+G and F81, exhibited the exact same topology. The only difference between the two models is that F81 inferred greater support for the main clade with a PP of 98.00 (Figure 52).

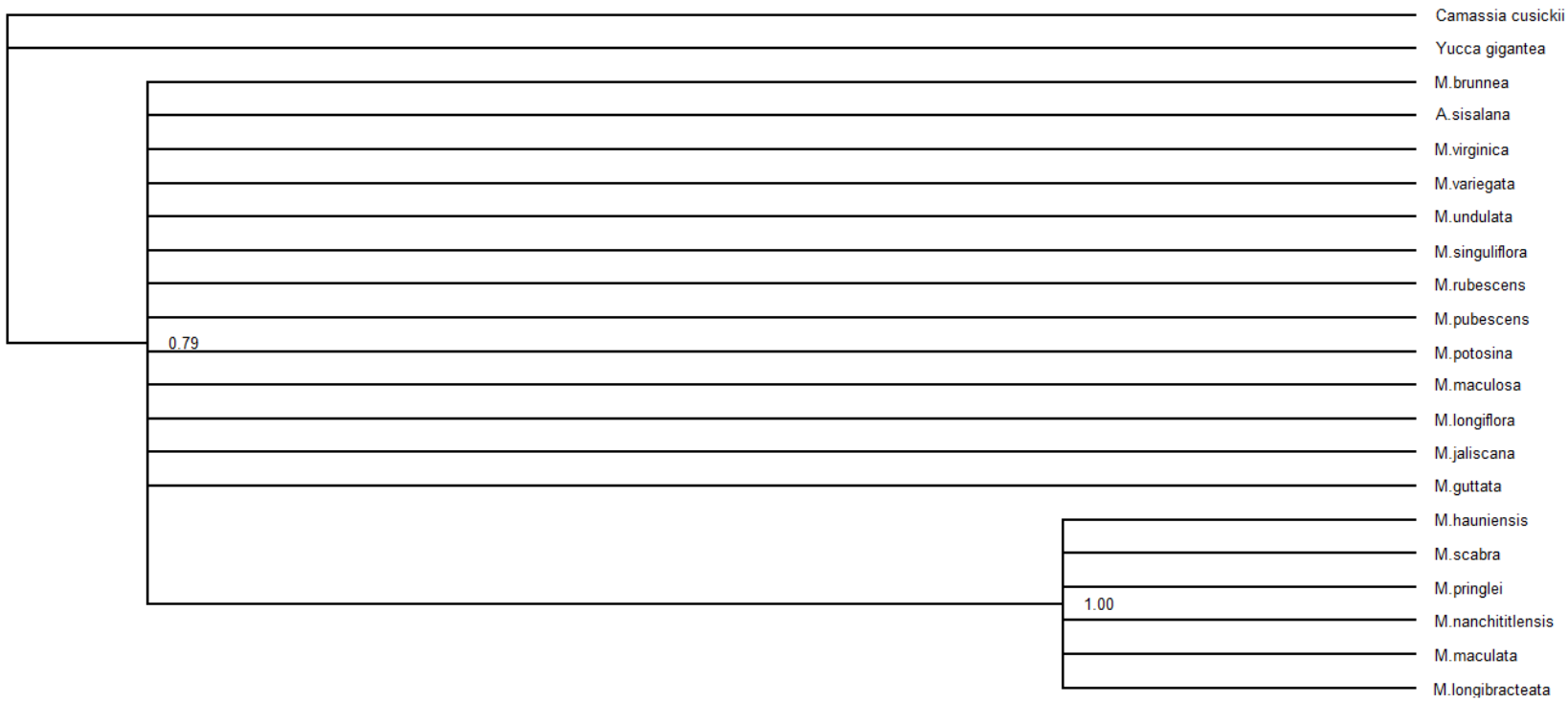


Figure 50: Cladogram generated employing TPM1uf+1 model of nucleotide evolution for Bayesian Analysis of the *MatK* and *rbcL* gene regions.

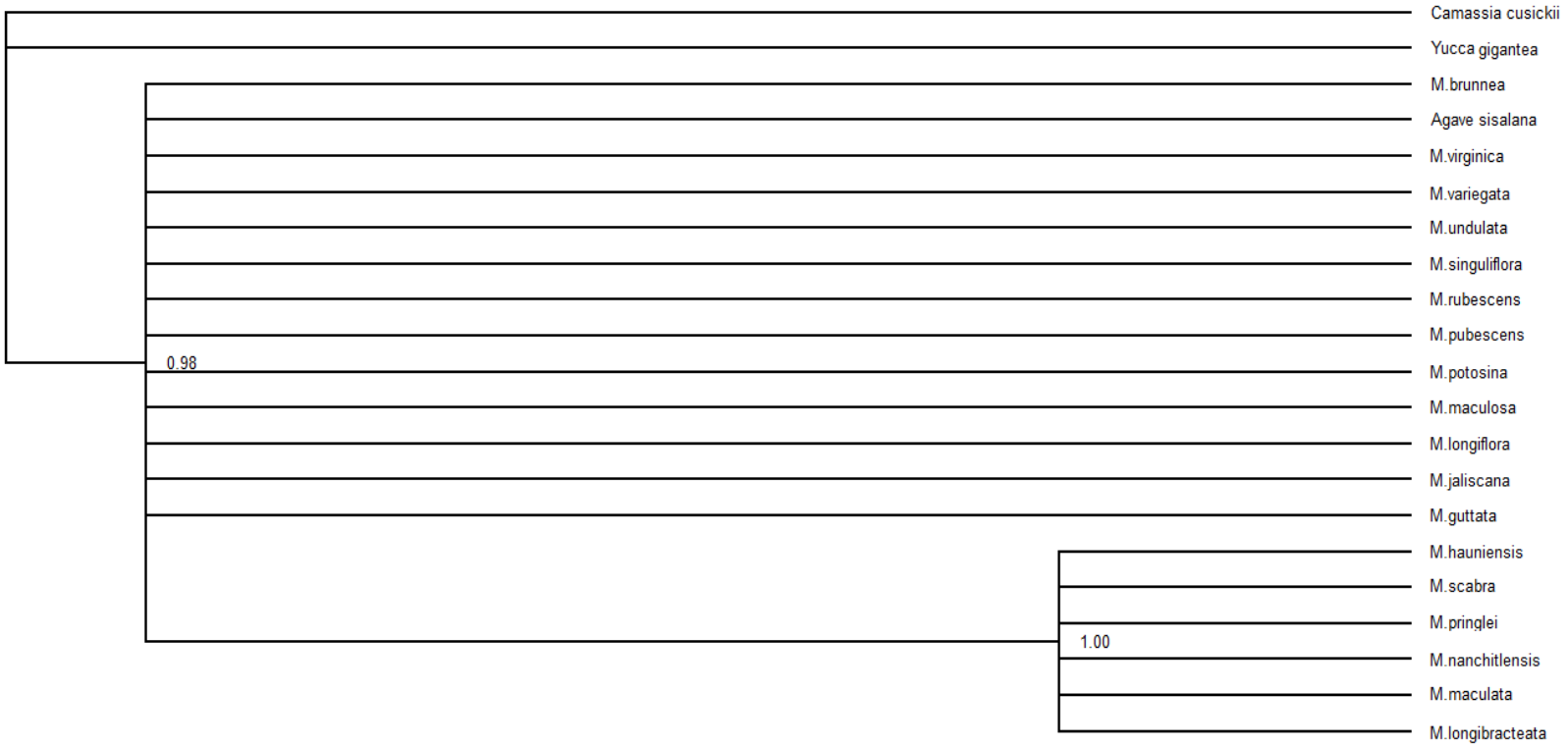


Figure 51: Cladogram generated employing F81 model of nucleotide evolution for Bayesian Analysis of the *MatK* and *rbcL* gene regions.

Uncorrected "P" Distance Matrix. Survey of the Uncorrected "P" Distance Matrix revealed that the greatest variation was between *M. longibracteata* Verh-Will. and *C. cusickii*. Interestingly, one of three designated outgroups, *A. sisalana*, possessed its greatest distance from *M. longibracteata* with a genetic distance of 0.00324 and its closest affinity to 11 species of *Manfreda* with a genetic distance of 0.00081. The greatest variation between two constituent species of *Manfreda* was between *M. longibracteata* and *M. brunnea* S. Watson. with a genetic distance of 0.004. The lowest genetic distance observed was 0.00 which occurred between 56 pairwise relationships (Appendix 4).

Maximum Parsimony. The resultant cladogram from the Maximum Parsimony Analysis is rooted by *A. sisalana* which was positioned outside the basal clade (Figure 53). All specimens of *Manfreda* as well as outgroups *C. cusickii* and *Y. gigantea* were included but in a large poorly supported basal clade. Two further clades were present within the basal clade: the first consisted of *C. cusickii* and *Y. gigantea* supported by a bootstrap value of 70 and the second contained six species of *Manfreda*. The clade consisting of six *Manfreda* species is supported by a bootstrap value of 85 and was identical to the internal clades of both Bayesian Analyses, no interspecific relationship can be surmised however. The remaining 12 constituent species of *Manfreda* contained in the basal clade were unresolved.

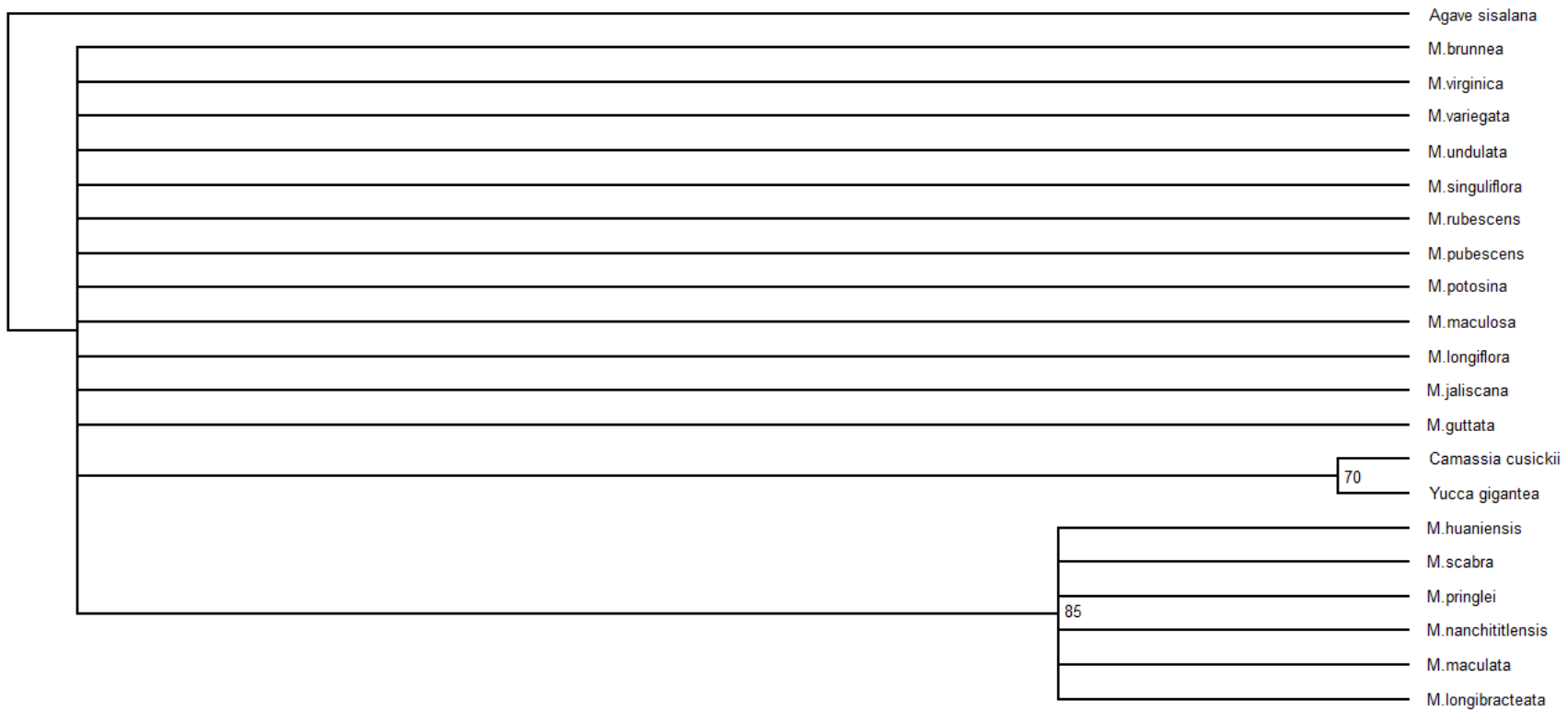


Figure 52: Cladogram generated (50% Majority Rule) employing Maximum Parsimony for phylogenetic analysis of the *MatK* and *rbcL* gene regions.

Nucleotide Network Analysis. The nucleotide network generated via HapStar 0.7 demonstrated an ability to differentiate all species of *Manfreda* and utilized outgroups exhibiting sufficient variation for identification by way of DNA barcoding (Figure 4). Furthermore, the analysis identified two basal species, *M. guttata* (Jacobi & C.D.Bouché) Rose. and *M. hauniensis*, which contained perceived plesiomorphic characters, from which autapomorphies and speciation occurred. Diversification was limited however with only a singular terminal node associated with each species, with the exception of *Y. gigantea.*

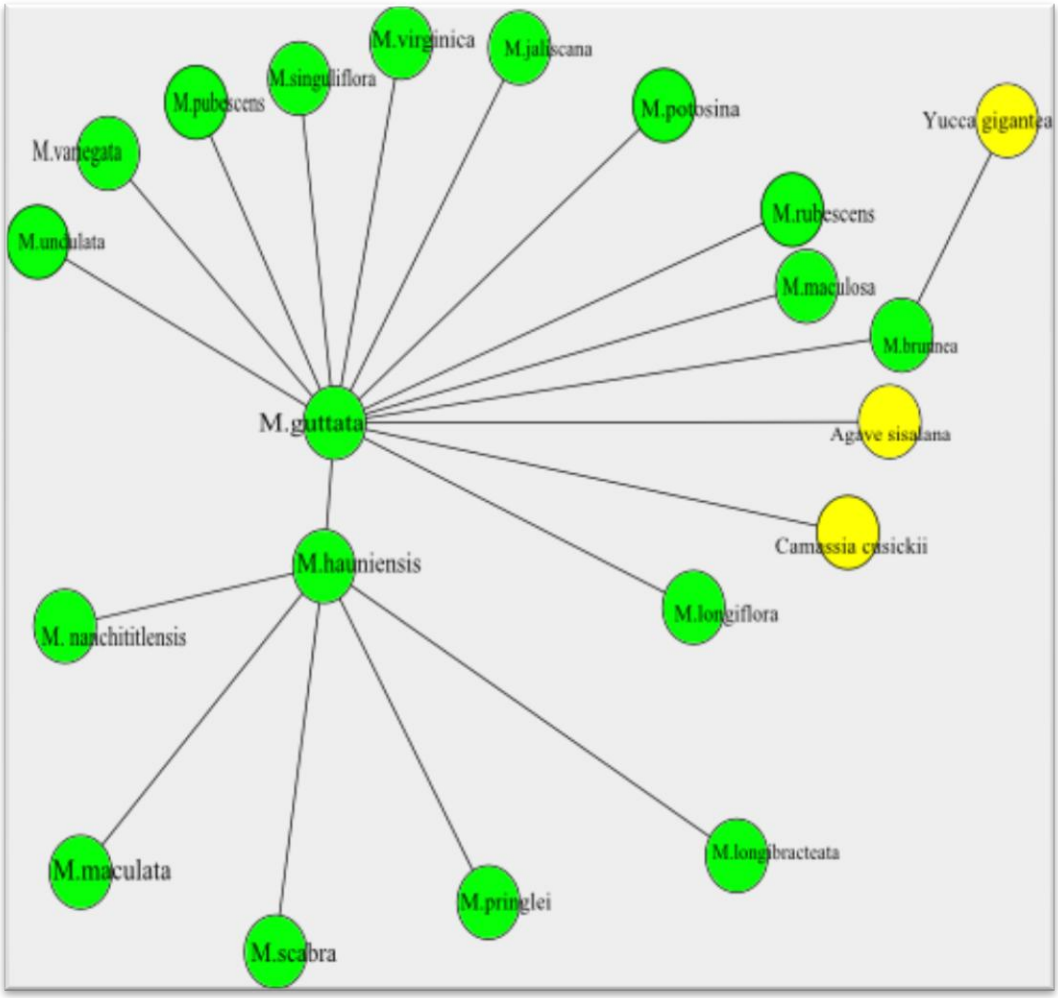


Figure 53. Nucleotide network exhibiting nucleotide divergence between specimens of *Manfreda* and selected outgroups based on *MatK* and *rbcL* gene sequences.

Discussion

Phylogenetic Studies. The phylogenetic analyses conducted exhibited minimal speciation and support for divergence. As such, the topology of both the Bayesian Analysis and Maximum Parsimony were nearly, identical with the majority of the taxa unresolved. The prowess of aforementioned phylogenetic techniques, however, are dependent on quality sequences containing an array of parsimony informative traits and autapomorphies for each species. Sequence diversification within the data set was low, of the 1236 characters: 1217 were constant, 14 were variable yet parsimony uninformative and a mere five were parsimony informative.

The position of the outgroups in both phylogenetic analyses was of interest. *C. cusickii* and *Y. gigantea* were selected as outgroups due to their evolutionary distance from *Manfreda* as two of the most distant relatives within subfamily Agavoideae. *Agave sisalana* was selected due to close affinity between *Agave* and *Manfreda*. It was surmised, based on contemporary taxonomic systems, that all three possessed sufficient diversification to function well as outgroups to *Manfreda.* In Bayesian Analysis via both the TPM1uf+G and F81 models, *A. sisalana* was positioned within a clade alongside species of *Manfreda*.

The relationships within each basal clades containing *A. sisalana* was unresolved, yet the polyphyly of *Manfreda* was observed previously. Bogler and Simpson (1996), based on Internal Transcribed Spacer (ITS) sequences subjected to Maximum Parsimony analysis, found specific taxa to be polyphyletic to nested specimens of *M. scabra* and *M. virginica*. Similarly, Good-Avila *et al*. (2006) employing Maximum Likelihood analysis to *trnL* and *trnL-trnF* sequences found *M. hauniensis, M. nanchititlensis* and *M. potosina* to also be polyphyletic interspersed with specimens of *Agave, Polianthes* and *Prochnyanthes* S. Watson. Molecular systematic studies to-date therefore imply, in contradiction to morphological studies, that *Manfreda* is not a

monophyletic genus and thus not a functional taxonomic group (APG, 2009). The alternative would be to include *Manfreda* in *Agave* sensu lato as many taxonomists have previously favored (Linnaeus, 1753; Berger, 1915), however more comprehensive phylogenetic evidence would be required.

The phylogenetic analysis, by means of both Bayesian Analysis and Maximum Parsimony, alluded to an internal clade consisting of six *Manfreda* species. The clade was defined by a transversion from thymine to guanine at the 712bp position in the barcode sequence or 162bp of *MatK*. It is unlikely however that the single nucleotide polymorphism could be correlated to taxonomic inference based on existing systematic treatments of *Manfreda*. Morphologically, the six members possessed substantial variation, geographical as well as phenological ranges showed minimal overlap and no documented hybridization between the six taxa has been recorded (table 11).

Table 12. Outline of distrubutions, phenology and hybridization of the six taxa allocated to an clade independant of all other taxa of *Manfreda* when subjected to Maximum Parsimony and Bayesian Analysis employing the *MatK* and *rbcL* gene regions.

Name	Distrubution	Phenology	Documented Hybridization
Manfreda hauniensis	Mexico State, Guerrero, Morelos, Oaxaca	October	None
Manfreda longibracteata	Jalisco, Michoacan	July - September	None
Manfreda maculata	Mexico State, Guerrero, Oaxaca	July - September	None
Manfreda nanchititlensis	Mexico State	October	None
Manfreda pringlei	Federal District, Mexico State, Hidalgo, Jalisco, Michoacan, Morelos, Oaxaca, Puebla	July - Novemeber	None
Manfreda scabra	Aguascalientes, Chiapas, Federal District, Mexico State, Durango, Guerrero, Guanajuato, Hidalgo, Jalisco, Michoacan, Morelos, Nayarit, Oaxaca, Puebla, Queretaro, San Luis Potosi, Veracruz, Zacatecas	July - September	with *Manfreda virginica* and *M.maculosa* (Verhoek-Williams, 1975)

The performance of the CBOL Plant DNA barcode consisting of the *MatK* and *rbcL* plastid regions was inadequate for a meaningful inferred phylogeny of the specific composition of *Manfreda*. Of the five parsimony informative characters, all were gleaned from the *MatK* gene region and no informative variation was present in the *rbcL.* The lesser extent of variation within the *rbcL* is well documented (Plant Working Group, 2009), however the uniformity of the region for inference of phylogeny in the genus *Manfreda* is impeding advances in taxonomic understanding. It is therefore deemed that the CBOL Plant DNA barcode possesses insufficient utility for taxonomic inference and that alternative gene regions will need to be sought to advance molecular systematics in the genus *Manfreda.*

Nucleotide Network Study. For the nucleotide network analysis both parsimony informative and uninformative characters are valuable, thus 19 variable characters were employed in the analysis. The nucleotide network was able to differentiate all taxa included in the study, meeting the single requirement in DNA barcoding. Although differentiation between taxa was limited to singular polymorphisms, sufficient variation was present for successful utilization of the DNA barcoding region.

With preliminary success in distinguishing between species of *Manfreda* via CBOL plant DNA barcodes via nucleotide networks, further, more comprehensive studies, would be of value. By increasing the number of taxa and individuals utilized in further research data could be collected as to the potential and limitations of the techniques for identification. In horticulture, the identification of specific taxa will aid breeding with robust identification of potential parent species. The CBOL Plant DNA Barcode would be of limited utility in the identification of F1 or F2 hybrids however, due to the uniparental inheritance of the plastid gene regions employed. Therefore, a horticultural specific DNA barcode would be required for the identification of

cultivated taxa yet, both the CBOL Plant DNA Barcode and a horticulture specific barcode would be of value.

Glossary of Terms

50% Majority Rule. A procedure in consensus tree construction were only speciation events documented in greater than 50% of trees generated are retained.

Autapomorphies. A derived character unique to a terminal group or taxon

Cytonuclear Disequilibrium. The presence of cytoplasmic and nuclear DNA in hybrids from alternative parents.

Genetic Distance. The distance calculated from the number of modified characters (Nucleotide or numeric) in distance matrices.

Heuristic Search. An abbreviated search technique for the most parsimonious trees utilizing a series of branch rearrangements.

Homoplasy. A resemblant character state not derived from a common ancestor.

Parsimony Informative Character. A nucleotide position with a minimum of two alternate states, each of which is represented by at least two taxa understudy.

Parsimony Uninformative Character. A nucleotide position with a minimum of two alternate states, yet represented by a single taxon.

Phenotypic Plasticity. The potential of a single genotype to exhibit alternative phenotypes due to environmental factors.

Plesiomorphic. A character state which is primitive/ancestral.

Polyphyly. An artificial phylogenetic group which does not share a common ancestor

Reticulation. The recombination of divergent species via hybridization

Transition. Point mutation which alter a purine to another purine or pyramidine to another pryamindine.

Transversion. Point mutation resulting in an alteration from a purine to pryamidine or a pryamidine to a purine.

Tree-Bisectional-Reconnection (TBR). A basic branch swapping technique which employs 'pruning' of tree sections and reattachment to survey for greater parsimony.

Literature Cited

Angiosperm Phylogeny Group. 2009. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. Bot J. Linnean Soc. 161: 105-121.

Berger, A. 1915. Die Agaven. (1st ed). Gustav-Fischer, Jena, Germany.

Bogler, D. J. and B. B. Simpson. 1995. A chloroplast DNA study of the Agavaceae. Systematic Bot. 20(2): 191-205.

Bogler, D. J. and B. B. Simpson. 1996. Phylogeny of Agavaceae based on its rDNA sequence variation. Amer. J. Bot. 83(9): 1225-1235.

Burgess, K. S., A.J. Fazekas, P.R. Kesanakurti, S.W. Graham, B.C. Husband, S.G. Newmaster, D.G. Percy, M. Hajibabaei, and S.C. Barrett. 2011. Discriminating plant species in a local temperate flora using the *rbcL* - *MatK* DNA barcode. Methods Ecol. Evolution 2: 233-240.

Castillo-Cruz, C. 2009. Sistemático del genero *Manfreda*. Natl Autonomous Univ. Mexico., Mexico City, Mexico, PhD Diss.

CBOL Executive Committee. 2009. CBOL approves *matK* and *rbcL* as the barcode regions for land plants. Consortium for the Barcode of Life, Washington, D.C.

Chupov, V. S. and N.G. Kutiavina. 1981. The comparative immunoelectrophoretic investigation of seed proteins of Liliaceae. Bot. Zhurnal Leningrad 63: 473-493.

De Vere, N., T.C. Rich, C.R. Ford, S.A. Trinder, C. Long, C.W. Moore, D. Satterthwaite, H. Davies, J. Allaingullaume, S. Ronca, T. Tatarinova, H. Garbett, K. Walker, and M.J. Wilkinson. 2012. DNA barcoding the native flowering plants and conifers of Wales. Public Library of Sci. 7(6): 1-12.

Good-Avila, S. V., V. Souza, B. S. Gaut, and L.E. Eguiarte. 2006. Timing and rate of speciation in *Agave* (Agavaceae). Proc. Natl. Acad. Sci. U.S. 103(24): 9124-9129.

Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acid Symp. Ser. 41: 95-98.

Hollingsworth, P.M. 2009. Interim notes on regions discussed in Plant Working Group (2009). Consortium for the Barcode of Life, Washington, D.C. 25 July 2012. <http://www.barcoding.si.edu/pdf/informationonbarcodeloci.pdf>.

Huelsenbeck, J. P. and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics Applications Notes 17(8): 754-755.

Kearse, M., R. Moir, A. Wilson, S. Stone-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes and A. Drummond. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics Applications Notes 28(12): 1647–1649.

Keb-Llanes, M., G. Gonzalez, B. Chi-Manzanero, and D. Infante. 2002. A rapid and simple method for small-scale DNA extraction in Agavaceae and other tropical plants. Plant Mol. Biol. Rptr. 20: 299a-299e.

Khew, G. S. and T.F. Chia. 2011. Parentage determination of *Vanda* 'Miss Joaquim' (Orchidaceae) through two chloroplast genes *rbcL* and *MatK*. AoB Plants, plr018.

Levin, A. 2009. A barcode for botany. The Botanics 39: 4-7.

Linnaeus, C. 1753. Species Plantarum: exhibentes plantas rite cognitas, ad genera relatas, cum differentiis specificis, nominibus trivialibus, synonymis selectis, locis natalibus, secundum systema sexuale digestas. 1st ed. Tomus, Stockholm, Sweden.

Lou, S. K., K.L. Wong, M. Li, S.K. Tsui, and P.C. Shaw. 2010. An integrated web medicinal materials DNA database: MMDBD (medicinal materials DNA barcode database). BMC Genomics 11(402): 1-8.

McKelvey, S. D., and K. Sax. 1933. Taxonomic and cytological relationships of *Yucca* and *Agave*. J. Arnold Arboretum 15: 76-80.

Mortiz, C., and C. Cicero. 2004. DNA barcoding: Promise and pitfalls. Public Library of Sci. 10(2): 1529-1531.

Newmaster, S. G., A.J. Fazekas, and S. Ragupathy. 2006. DNA barcoding in land plants: Evaluation of the *rbcL* in multigene tiered approach. Can. J. Bot. 84: 335-341.

Oldfield, S. 1997. Cactus and succulent plants: status survey and conservation action plan. International Union for Conservation of Nature Media, Cambridge, England.

Plant Working Group. 2009. A DNA barcode for land plants. Proc. Natl. Acad. Sci. U.S. 106(31): 12794–12797.

Rose, J.N. 1905. Contributions for the United States National Herbarium, volume VIII: Mexican, Central American and Porto Rican plants. Government Press Office, Washington, D.C.

Roy, S., A. Tyagi, V. Shukla, A. Kumar, U.M. Singh, L.B. Chaudhary, B. Datt, S.K. Bag, P.K. Singh, N.R. Nair, T. Husan, and R. Tuli. 2010. Universal plant DNA barcode loci may not work in complex groups: A case study with Indian *Berberis* species. Public Library of Sci. 5(10): 1-14.

Sāto, D. 1935. Analysis of the karyotype in *Yucca*, *Agave* and related genera with special reference to the phylogenetic significance. Jpn. J. Genet. 11: 272-278.

Tapia-Tussel, R., A. Quijano-Raymayo, R. Rojas-Herrera, A. Larque-Savverda, and D. Perez-Brito. 2005. A fast, simple, and reliable high-yielding method for DNA extraction from different plant species. Mol. Biotechnology 31: 137-139.

Teacher, A. G. and D.J. Griffiths. 2011. Hapstar: automated haplotype network layout and visualization. Mol. Ecol. Resources 11: 151-153.

Verhoek-Williams, S.E. 1975. A study of the tribe Poliantheae (including *Manfreda*) and revision of *Manfreda* and *Prochnyanthes* (Agavaveae). Cornell Univ., Ithaca., PhD Diss.

Whittaker, T. W. 1934. Chromosome constitution in certain monocotyledons. J Arnold Arboretum 15: 135-143.

Chapter 3

DEVELOPMENT OF A DNA BARCODE FOR IDENTIFICATION OF INTERSPECIFIC HYBRIDS BETWEEN *AGAVE* L., *MANFREDA* SALISB. AND *POLIANTHES* L. (ASPARAGACEAE) BASED ON NUCLEAR RIBOSOMAL DNA INTERNAL AND EXTERNAL TRANSCRIBED SPACERS

Introduction

Breeding efforts led by Dr Jon T. Lindstrom at the University of Arkansas, Fayetteville aim to develop ornamental perennial plant hybrids suitable to the climatic conditions of Arkansas. One of the plant groups included in the project is allied taxa of subfamily Agavoideae, *Agave* L., *Manfreda* Salisb. and *Polianthes* L. Intergeneric crosses between *Manfreda* and *Agave* as well as *Manfreda* and *Polianthes* have been conducted to aid the development of intergeneric hybrids with increased cold-hardiness and improved aesthetics for cut-flower production. The project began in 2003 with the acquisition of *Manfreda virginica* (L.) Salisb. ex. Rose and *Polianthes tuberosa* L. from Yucca Do Nursery in Giddings, Tex. To date, 40 intergeneric and three trigeneric crosses have been achieved (Lindstrom, 2006).

The first description of a cross between constituent taxa of tribe Poliantheae was achieved between *P. geminiflora* (Lex.) Rose and *Prochnyanthes mexicana* (Zucc.) Rose by Anonymous (1899). Worsely (1911) documented hybridization between species of *Polianthes*, yet no documentation of breeding efforts concerning *Manfreda* are available from the early 20th century. A bimodal chromosomal complement of high uniformity was reported between genera of the former Agavaceae by McKelvey & Sax (1933), Sāto (1935) and Granick (1944) alluding

to the possibility of further potential for intergeneric hybridization within Agavoideae (Verhoek-Williams, 1975).

On the basis of the aforementioned evidence and reports of *Manfreda* × *Polianthes* hybrids being produced for cut flowers in California, Verhoek-Williams (1975) undertook the first documented breeding program to include species of *Manfreda* and *Polianthes*. Breeding trials were completed in partial fulfilment of her dissertation research, of which 130 crosses were completed and 33 viable progeny were produced (Verhoek-Williams, 1975).

Although infrequently utilized in horticulture to date, *Manfreda* and associated intergeneric hybrids have been discussed favorably for employment in landscapes of the American South in a number of horticultural texts. Suitability to low input landscapes, shade tolerance and winter hardiness of *M. virginica* or progeny have been cited as appealing characteristics to gardeners (Irish & Irish, 2000; Howard, 2001; Hannon, 2002).

DNA Barcoding. Identification of plant materials can be challenging, particularly with newly developed cultivars and selections. Correct identification of cultivated plant materials is critical in avoidance of mislabelling, incorrect attribution of royalties and evasion of commission payments. At present, the majority of identifications are based on personal knowledge of derived morphological characters, which is highly inconsistent between employees of plant sales centers (Pryer *et al*., 2010). To safeguard the rights of plant breeders to commission payments under Plant Variety Protection Act (1970) regulations and prevention of illegal trade, an accurate technique for identification has been sought (Goodall, 2006; Sass *et al*., 2011). The identification method must be robust enough to be viable evidence in a court of law, unlike traditional morphological techniques and require minimal taxonomic training. DNA barcoding, in light of

the successful utilization of the technique in ecological studies, has been championed for employment in horticulture (Goodall, 2006).

DNA barcoding in horticulture has been employed in a limited number of studies as the technique is still in its infancy (Pryer *et al*. 2010; Njuguna and Bassil, 2011). A major challenge is the identification of plant materials of recent hybridization due to the biparental recombination, undetectable by uniparental plastid DNA barcoding regions used conventionally in DNA barcoding. Initial studies into the creation of DNA barcodes appropriate for hybridized plant materials is currently being conducted (Burgess, 2007; Njuguna and Bassil, 2011), yet minimal data has been generated as to appropriate gene regions.

The aim of the study was to investigate the feasibility of employing two nuclear ribosomal DNA spacer regions, the Internal Transcribed Spacer (ITS) and the External Transcribed Spacer (ETS), for unique identification of intergeneric taxa of the genera *Agave*, *Manfreda* and *Polianthes*. The main objective of the study is to generate a barcode capable of identifying cultivated plant materials produced by the University of Arkansas' Agavoideae breeding program. Such a technique would allow for the correct identification of cultivars at each stage of ontogenesis and aid commercialization, once appropriate plant variety protective rights are acquired. A secondary objective was the generation of a DNA barcode library of all cultivated materials produced by the Agavoideae breeding program at the University of Arkansas.

Materials and Method

Sample Collection. Plant samples were sourced from the Arkansas Agricultural Research and Extension Center in Fayetteville, Ark. The collection, held by Dr. J T Lindstrom consists of *Agave*, *Polianthes* and *Manfreda* species and hybrids from which samples of mapo 01-04-07 (*M. virginica* × *P. tuberosa*), mapo 05-04-02 (*M. maculosa* (Hook) Rose × *P. geminiflora*) and maag 01-07-13 (*M. maculosa* × *A. polianthiflora* Gentry) were obtained. A single sample of *Rudbeckia hirta* L. 'Prairie Sun' was acquired from the Horticulture Display Gardens at the University of Arkansas in Fayetteville, Ark. and employed as a control. The selection of *R. hirta* 'Prairie Sun' was due to its close affinity to *Calycadenia* DC. in tribe Heliantheae of Asteraceae, in which successful amplification of the ETS gene region was achieved employing the 18S-IGS/26S-IGS primer set by Baldwin and Markos (1998). Leaf tissue samples for the three intergeneric hybrids and the control were dissected to 5 mm × 5 mm leaf segments, stored in tea bags submerged in Ziploc® bags containing silica gel and stored at room temperature.

DNA Extraction. The following protocol was adapted from Keb-Llanes *et al*. (2002), Tapia-Tussel *et al*. (2005) and personal communication from Dr. Gerardo Salazar, National Autonomous University of Mexico, Mexico and optimized for the attributes of *Manfreda* leaf tissue.

1. Hexadecyltrimethylammonium bromide (CTAB) solution (Appendix 3) was warmed to 65°C on a Thermolyne® 17600 Dri-Bath.

2. 5 mm × 5 mm leaf segments were submerged in liquid nitrogen and pulverized using a mortar and pestle.

3. 2µl of β-mercaptoethanol was added to 500 µl of pre-warmed CTAB solution and mixed on a Vortex-Genie® 2 vortex mixer.

4. Pulverized leaf material was added to the CTAB/β-mercaptoethanol solution in an 2ml eppendorf tube and mixed for 3 to 5 seconds on the Vortex-Genie® 2 vortex mixer.

5. CTAB/β-mercaptoethanol solution containing pulverized leaf material was heated to 65°C on the Thermolyne® 17600 Dri-Bath for 30 minutes.

6. The mixture was cooled to room temperature; 600 µl of chloroform-isoamyl alcohol 24:1 was added to the eppendorf tube and gently agitated for 30 minutes on a The Belly Dancer® laboratory shaker.

7. Lysate was subjected to centrifugation in an Eppendorf 5417C centrifuge for 10 minutes at 12,500g.

8. Upper aqueous layer was transferred to a new 2ml eppendorf tube, to which 700 µl of pre chilled isopropanol is added prior to incubation at -20°C for 1 hour.

9. Sample was centrifuged in an Eppendorf 5417C centrifuge for a further 5 minutes at 26,500g prior to residual liquid being discarded.

10. 500 µl of 70% ethanol chilled at 3°C was added to the eppendorf tube and centrifugation was repeated for a further 5 minutes at 26,500g.

11. Residual 70% ethanol was decanted and samples were dried for 15 minutes at 55°C on the Thermolyne® 17600 Dri-Bath.

12. Dried DNA pellets were re-suspended with 40 µl ddH_20 and left at room temperature overnight or subjected to a further 1 hour at 55°C on the Thermolyne® 17600 Dri-Bath.

13. Samples were stored at -20°C.

Gel Electrophoresis of Extract. Gel electrophoresis was conducted in a horizontal mini-gel system to check the presence of genomic DNA (Figure 55). A 2% agarose gel solution was cast, a sodium borate buffer (Appendix 3) was employed and electrophoresis was conducted at 175v for 35 minutes. DNA extract samples with Blue/Orange 6x Loading Dye (Promega, Madison, Wis.) were run simultaneously to a Benchtop 1KB DNA Ladder (Promega, Madison, Wis.) in Lane 1. The agarose gel was stained with GelRed Nucleic Acid Stain 10,000x (Biotium, Hayward, Calif.) by gently agitating a solution of 15 µl GelRed Nucleic Acid Stain 10,000x, 5ml sodium borate buffer and 55ml of distilled water. Visualization of gel electrophoresis products were conducted using the BioDoc-It® 220 Imaging system (UPV LLC, Upland, Calif.).

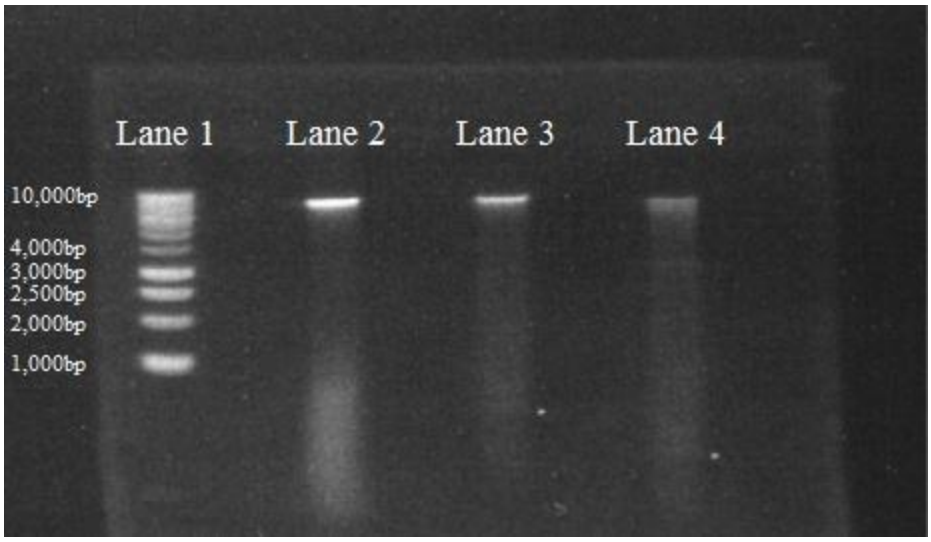


Figure 54. Gel electrophoresis of genomic DNA extract for the three intergeneric cultivars: 1kb DNA ladder (Lane 1), mapo 01-04-07 (lane 2), mapo 05-04-02 (Lane 3) and maag 01-07-13 (Lane 4).

Quantification of Extract. DNA extracts were also quantified using a Nanodrop 1000 Spectrophotometer V3.7 system (Thermo Fisher Scientific, Waltham, Mass.), 1 µl of each extract was utilized from each sample.

ITS - Polymerase Chain Reaction (PCR). The PCR procedure was conducted in a PCR Sprint Thermocycler (Thermo Fisher Scientific, Waltham, Mass.). High Fidelity PCR EcoDry™ premixed tubes (Clontech Laboratories, Mountain View, Calif.) were employed to which 0.5 µl of DNA template, 2.0 µl of each primer at a concentration of 10 mM and 20.5 µl of ddH_20 were added. Primers employed are listed in Table 12. The cycling conditions for the amplification of ITS was adopted from Bogler and Simpson (1996) (Table 13).

Table 13. Primers employed for amplification of the ITS gene region via PCR for Asparagaceae intergeneric hybrids; mapo 01-04-07, mapo 05-04-02 and maag 01-07-13.

Primer Name	Gene Region	Direction	Sequence (5' to 3')
ITS5 (White *et al.*, 1990)	ITS	Forward	GGAAGTAAAAGTCGTAACAAGG
AB102 (Douzery *et al.*, 1999)	ITS	Reverse	TAGAATTCCCCGGTTCGCTCGCCGTTAC

Table 14. PCR amplification parameters for amplification of the ITS gene region for Asparagaceae intergeneric hybrids; mapo 01-04-07, mapo 05-04-02 and maag 01-07-13.

Internal Transcribed Spacer PCR Amplification			
	Time	**Temperature**	**Cycles**
Initial denaturation	1 minute	95°C	1
Denaturation	30 seconds	95°C	28
Annealing	30 seconds	68°C	
Extension	30 seconds	68°C	
Final extension	1 minute	68°C	1

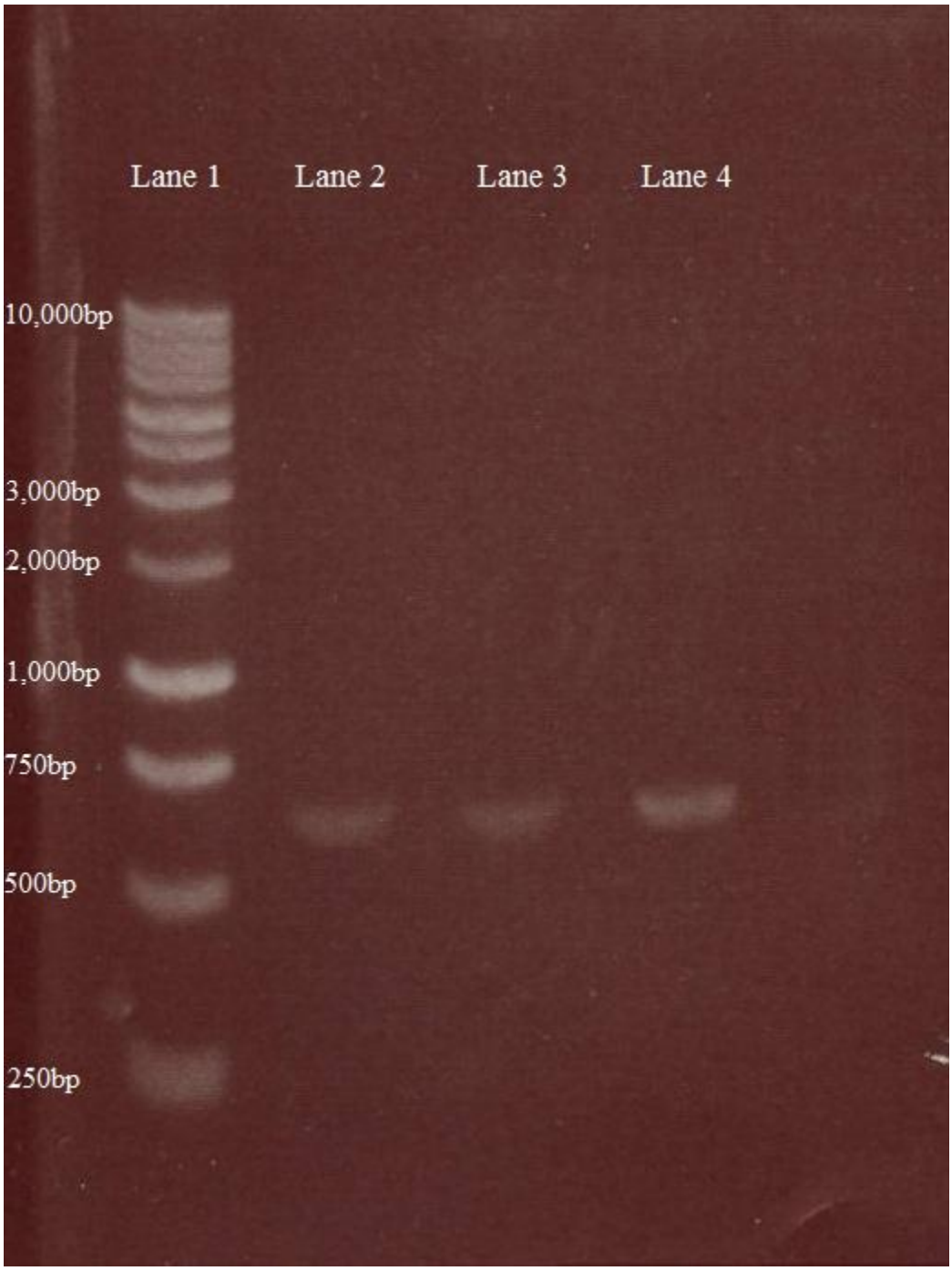


Figure 55. Gel electrophoresis of ITS amplicons for intergeneric cultivars: 1kb DNA ladder (Lane 1), mapo 01-04-07 (Lane 2), mapo 05-04-02 (Lane 3) and maag 01-07-13 (lane 4).

ITS - DNA Sequencing. Sequencing was conducted at Eurofins MWG Operon (Huntsville, Ala.). Samples containing 10 µl of DNA template at 20-50ng/µl and 10 µl of each of the required primers at 2µM were shipped overnight. Sequencing in a forward and reverse direction was conducted on a ABI 3730 XL DNA Sequencer (Applied Biosystems, Foster City, Calif.), resultant electropherograms and sequence files were sent within a 36 hour time period after sample submission.

ETS - Polymerase Chain Reaction (PCR). To design primers for the ETS region, amplification of the encompassing Intergenic Spacer (IGS) region must first be conducted and sequenced due to the lack of known conserved sites flanking the ETS region.

Three alternative regimes consisting of three different variables were tested for amplification of an IGS amplicon. The variables tested were primers (Table 14), PCR cycle number and PCR annealing temperature, each of the intergeneric hybrid samples and the control were subjected to each regime. With the exception of stated alterations to the parameters of the PCR reaction, cycling conditions employed are detailed in Table 15 and were adopted from Baldwin & Markos (1998). Details of the amplification trial are detailed in Table 16.

Due to the potential range in size of IGS amplicons from 3 kb to 6 kb an Advantage® Genomic LA Polymerase Kit was employed (Hershkovitz *et al.*, 1999). All reagents were thawed on ice, excluding the Advantage® Genomic LA Polymerase which was stored at -20°C until use. Components were compiled in 0.5 ml eppendorf tubes as outlined in Appendix 5 with the appropriate primer listed in table 14. Samples were briefly spun in an Eppendorf 5417C and loaded into the PCR Sprint Thermocycler.

Table 15. List of primers employed in the ETS Amplification Trial for nine parameters employing Asparagaceae intergeneric hybrids; mapo 01-04-07, mapo 05-04-02 and maag 01-07-13 and *Rudbeckia hirta* 'Prairie Sun'.

Primer name	Gene region	Direction	Sequence (5' to 3')
18S-IGS (Baldwin and Markos, 1998)	ETS	Forward	GAGACAAGCATATGACTACTGG CAGGATCAACCAG
26S-IGS (Baldwin and Markos, 1998)	ETS	Reverse	GGATTGTTCACCCACCAATAGG GAACGTGAGCTG
CSA1 (Hsieh *et al*., 2004)	ETS	Forward	AGGTTAGTTTTACCCTACT
CSA1-R (Hsieh *et al*., 2004)	ETS	Reverse	GCAGGATCAACCAGGTAGCA
5SRNA (Vilgalys, 2001)	ETS	Forward	ATCAGACGGGATGCGGT
5SRNAR (Vilgalys, 2001)	ETS	Reverse	ACQGCATCCCGTCTGAT

Table 16. Standard PCR amplification parameters for amplification of the ETS gene region of *Manfreda* cultivars; mapo 01-04-07, mapo 05-04-02 and maag 01-07-13 and *Rudbeckia hirta* 'Prairie Sun'.

External Transcribed Spacer PCR Amplification			
	Time	**Temperature**	**Cycles**
Initial Denaturation	1 minute	94°C	1
Denaturation	30 seconds	94°C	35
Annealing	3 minutes	68°C	
Extension	3 minutes	68°C	
Final Extension	10 minutes	68°C	1

Table 17. Table of the three different parameters for each of the three experiments included in the IGS amplification trial

Experiment 1: Primers												
Specimen	M1	M2	M3	R	M1	M2	M3	R	M1	M2	M3	R
Ladder Position	1	2	3	4	5	6	7	8	9	10	11	12
Primers	18S-IGS/26S-IGS				CSA1/CSA1-R				5SRNA/5SRNA-R			
Experiment 2: Cycle Number												
Specimen	M1	M2	M3	R	M1	M2	M3	R	M1	M2	M3	R
Ladder Position	14	15	16	17	18	19	20	21	22	23	24	25
Cycle No	25				30				40			
Experiment 3: Annealing Temperature												
Specimen	M1	M2	M3	R	M1	M2	M3	R	M1	M2	M3	R
Ladder Position	27	28	29	30	31	32	33	34	35	36	37	38
Temperature	53^{0}C				58^{0}C				63^{0}C			

Results - ITS

Electropherograms were received from sequencing on an ABI 3730 XL DNA Sequencer at Eurofins MWG Operon (Huntsville, Alabama). Sequencing was conducted in both forward and reverse directions with forward sequencing employing the ITS5 primer and reverse sequencing employing the AB102 primer. The vast majority of base calls were ambiguous with multiple peaks present of similar size. Alignment via the ClustalW algorithm employed by BioEdit 7.1.3.0 (Hall, 1999) or the GENEIOUS algorithm by Geneious Pro™ 5.6 (Drummond *et al*., 2006) could not find sufficient motifs to align any of the sequences obtained.

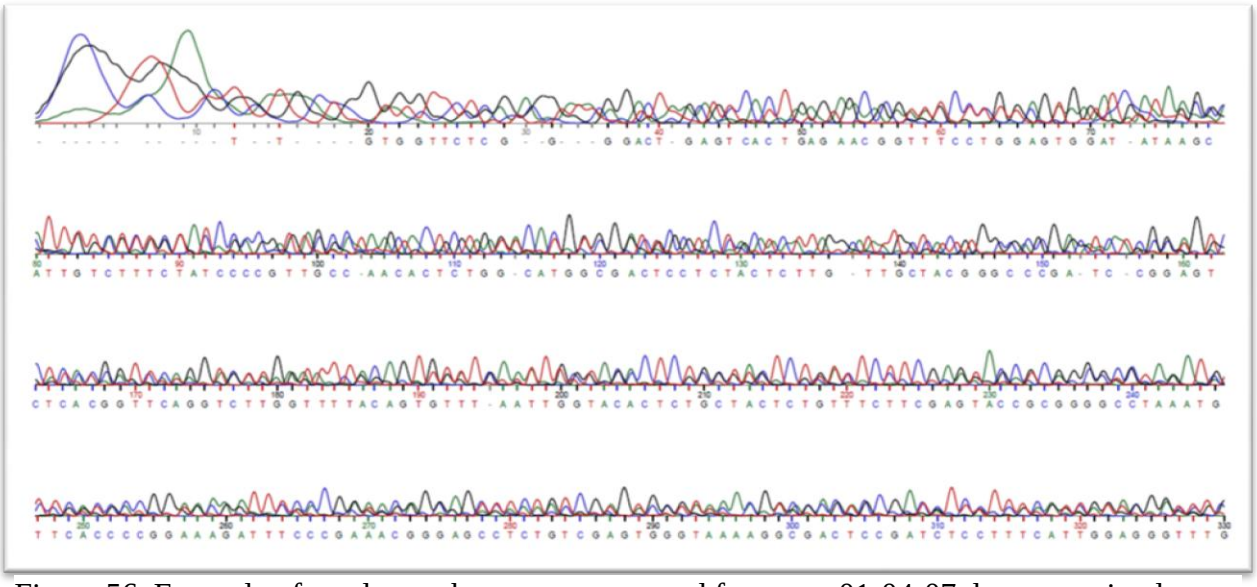


Figure 56. Example of an electropherogram generated for mapo 01-04-07 demonstrating low signal strength and intra-individual polymorphisms between ITS copies.

Results - ETS

The IGS amplification trial was largely unsuccessful. Of the 36 modified parameters only a single successful amplication was achieved in a control specimen alluding to a lack of utility of all three primers in both the integeneric hybrids and the control, *R. hirta* 'Prairie Sun' (Figure 58).

Primers. The 18S-IGS/26S-IGS (Lanes 1-4) and 5SRNA/5SRNAR (Lanes 9-12) primer sets resulted in no applification of any PCR product in either the intergeneric hybrids or *R. hirta* 'Prairie Sun'. The CSA1/CSA1-R primer set similarily resulted in no amplification in any of the intergeneric hybrids (Lanes 5-7) but an amplicon approximated to be 5,000 to 10,000bp in size was exhibited for the control specimen of the CSA1/CSA1-R primer set (Lane 8). Large smears were observable in all specimens with the exception of Lane 8, it is surmised that cycle number may be too high. The lack of successful amplification suggests that none of the three primer sets tested are a viable option for the amplification of the IGS region in intergeneric hybrids between *Agave, Manfreda* or *Polianthes.*

Cycle Number. Utilizing the 18S-IGS/26S-IGS primer set, cycling conditions did not aid amplification with no amplicon present in any sample. Employment of 25 cycles for PCR of the DNA samples in Lanes 14 -17 demonstrated minimal smearing, yet exhibited potential primer dimer at the apex of the gel. Both 30 and 40 cycles exhibited substantial smearing with the reason surmised to be excessive cycles.

Annealing Temperature. Modification of annealing temperatures did not influence amplification of PCR products employing the 18S-IGS/26S-IGS primer set. Samples of mapo 01-04-07 and mapo 05-04-02 at 53ºC resulted in a product of approximately 2,000bp, yet due to

the size of the amplicon and the presence of excess smearing prior to the amplicon, it can only be surmised that the product is non-specific amplification. The control at 53º and maag 01-07-13 at 58ºC exhibited terminal primer dimer and all other samples consisted of large smears with no PCR product.

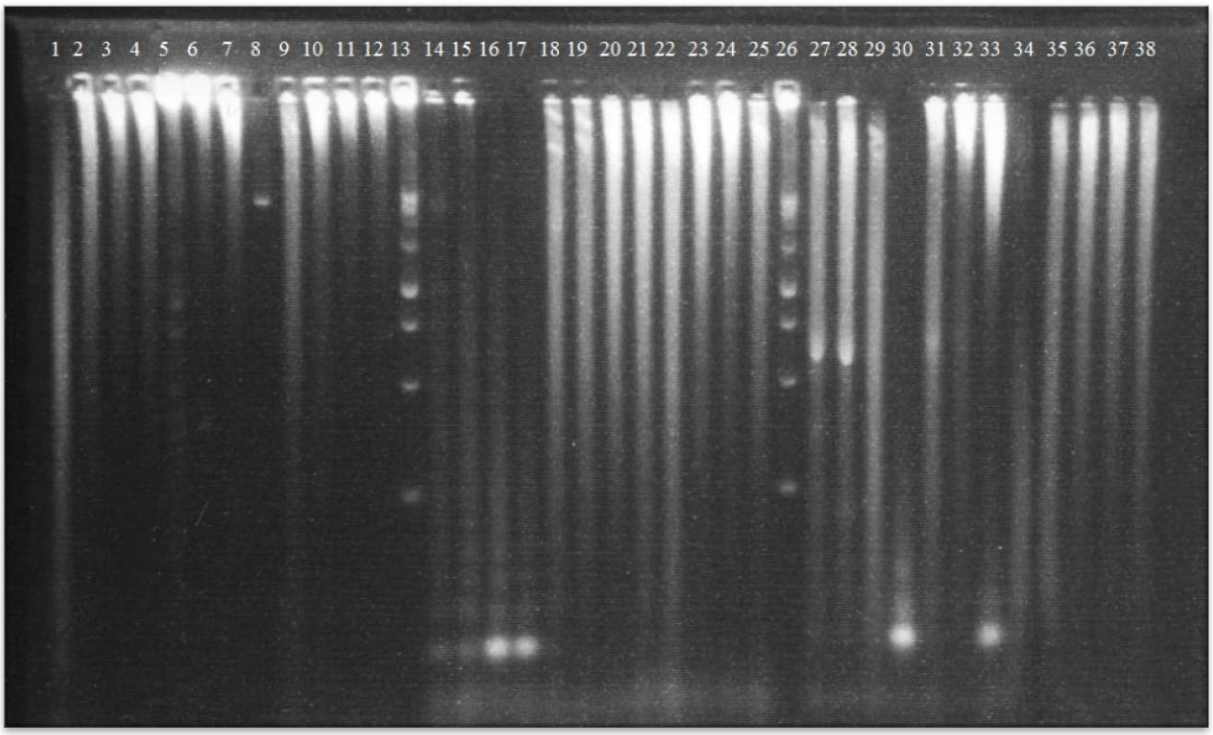


Figure 57. PCR amplification trial of nine alternative parameters conducted for three intergeneric hybrids of *Agave*, *Manfreda* and *Polianthes*: mapo 01-04-07, mapo 05-04-02 and maag 01-07-13 and *Rudbeckia hirta* 'Prairie Sun'. The trial included the following parameters; (1) mapo 01-04-07 with primers 18S-IGS/26S-IGS (2) mapo 05-04-02 with primers 18S-IGS/26S-IGS (3) maag 01-07-13 with primers 18S-IGS/26S-IGS (4) *Rudbeckia hirta* 'Prairie Sun' with primers 18S-IGS/26S-IGS (5) mapo 01-04-07 with primers CSA1/CSA1-R (6) mapo 05-04-02 with primers CSA1/CSA1-R (7) maag 01-07-13 with primers CSA1/CSA1-R (8) *Rudbeckia hirta* 'Prairie Sun' with primers CSA1/CSA1-R (9) mapo 01-04-07 with primers 5SRNA/5SRNA-R (10) mapo 05-04-02 with primers 5SRNA/5SRNA-R (11) maag 01-07-13 with primers 5SRNA/5SRNA-R (12) *Rudbeckia hirta* 'Prairie Sun' with primers 5SRNA/5SRNA-R (13) mapo 01-04-07 with 25 PCR cycles (14) mapo 05-04-02 with 25 PCR cycles (15) maag 01-07-13 with 25 PCR cycles (16) *Rudbeckia hirta* 'Prairie Sun' with 25 PCR cycles (17) mapo 01-04-07 with 30 PCR cycles (18) mapo 05-04-02 with 30 PCR cycles (19) maag 01-07-13 with 30 PCR cycles (20) *Rudbeckia hirta* 'Prairie Sun' with 30 PCR cycles (21) mapo 01-04-07 with 40 PCR cycles (22) mapo 05-04-02 with 40 PCR cycles (23) maag 01-07-13 with 40 PCR cycles (24) *Rudbeckia hirta* 'Prairie Sun' with 40 PCR cycles (25) mapo 01-04-07 with an annealing temperature of 53ºC (26) mapo 05-04-02 with an annealing temperature of 53ºC (27) maag 01-07-13 with an annealing temperature of 53ºC (28) *Rudbeckia hirta* 'Prairie Sun' an annealing temperature of 53ºC (29) mapo 01-04-07 with an annealing temperature of 58ºC (30) mapo 05-04-02 with an annealing temperature of 58ºC (31) maag 01-07-13 with an annealing temperature of 58ºC (32) *Rudbeckia hirta* 'Prairie Sun' with an annealing temperature of 58ºC (33) mapo 01-04-07 with an annealing temperature of 63ºC (22) mapo 05-04-02 with an annealing temperature of 63ºC (23) maag 01-07-13 with an annealing temperature of 63ºC (24) *Rudbeckia hirta* 'Prairie Sun' with an annealing temperature of 63ºC.

Discussion

The performance of primers designed to amplify the IGS region of nuclear ribosomal DNA was poor with amplification only observed in a single control specimen and no amplification of specimens representing intergeneric hybrids between *Agave, Manfreda* and *Polianthes*. The aforementioned primers were designed to be located within the 18S region, that is highly conserved. The lack of universality however and the limited utility of 18S-IGS/26S-IGS primer set to small taxonomic groups are indicative of a region subjected to only moderate evolutionary constraints. The position of such primers in close proximity to the contiguous Non Transcribed Spacer region, of known hypervarability, could have influenced the rate of evolution (Hershkovitz *et al*., 1999).

The availability of alternative primers for the amplification of the IGS region and subsequent development of taxon-specific ETS primers is limited. The seminal publications for the amplification of the ETS region for use in molecular ecological studies were the studies of Baldwin & Markos (1998) of *Calycadenia* DC. and Markos & Baldwin (2001) of *Lessingia* Cham., both of Asteraceae and tribes Heliantheae and Astereae, respectively. Both studies employed the 18S-IGS/26S-IGS primer set, that has also been employed in subsequent studies, yet demonstrated limited utility in intergeneric hybrids between *Agave*, *Manfreda* or *Polianthes* (Andreasen & Baldwin, 2001; Becerra, 2003; Kelch & Baldwin, 2003). A limited number of other novel primers have been published for plant species, including but not exclusive to *Cannabis* L. and *Calyptridium* Nutt., nevertheless no successful utilization of an IGS primer has been documented in the Asparagales order (Hseih, 2004; Guilliams, 2009).

The theoretical potential of ETS regions for DNA barcoding of cultivated taxa has yet to be tested, due primarily to the lack of universal primers and challenges associated with

development of universal primers. The challenging amplification of the ETS region in a range of plant groups hinders any potential utilization for barcoding efforts, unless specific to a particular taxon and a research capacity to invest substantial resources into the development of primers is available. Therefore, the likelihood of a DNA barcode for the identification of intergeneric hybrids between *Agave, Manfreda* and *Polianthes* employing the ETS gene region is low.

Despite the successful amplification of the ITS region, sequencing detected intra-individual polymorphisms between multiple copies of the ITS region in all of the intergeneric hybrids. Occurrence of such could be due to inefficient sequencing or non-concerted evolution of the multiple ITS copies. Although generally thought of as rare in plants, non-concerted evolution of the ITS has been frequently documented (Mayol and Rosselò, 2001; Andreasen and Baldwin, 2003; Ruggiero and Procaccini, 2004).

A study of *Mammillaria* Haw. by Harpke and Petersen (2006) also encountered such a phenomenon. The study noted the detection of multiple bands of 550-600bp and 700bp in size and multiple copies of the ITS detected within each band visible in the electropherogram. The study concluded the presence of deletions in the ITS2 region was responsible for the varied length akin to Hartmann *et al*. (2001).

A range of possible scenarios could have resulted in the multiple ITS copies detected in hybridized taxa of Asparagaceae including slow concerted evolution of parental lineages, presence of pseudogenes and hybridization. Although hybridization seems the most plausible cause of multiple ITS copies due to recent breeding, multiple copies of the ITS region were also detected in specific taxa (not documented).

The potential for utilization of the ITS region in a DNA barcode for cultivated taxa of Asparagaceae is negligible, due to the presence of multiple ITS copies. Although primers performed efficiently, the need for further cloning of ITS copies via ligation into a plasmid vector and selection of the appropriate ITS copy via conserved motifs would be labor intensive. Therefore neither the ETS nor ITS nuclear ribosomal DNA regions merit further investigation into the feasibility of their employment in a DNA barcode for identification of intergeneric hybrids between *Agave*, *Manfreda* and *Polianthes*.

Further study is needed to investigate a wider selection of biparental gene regions, screening for potential primers, ease of amplification and performance in detection of taxa of hybrid origin in the family Asparagaceae.

Literature Cited

Andreasen, K. and B.G. Baldwin. 2001. Unequal evolutionary rates between annual and perennial lineages of Checker Mallows (*Sidalcea*, Malvaceae): Evidence from 18s–26s rDNA internal and external transcribed spacers. Mol. Biol. and Evolution 18: 936-944.

Baldwin, B. G. and S. Markos. 1998. Phylogenetic utility of the External Transcribed Spacer (ETS) of 18s–26s rDNA: Congruence of ETS and ITS trees of *Calycadenia* (Compositae). Mol. Phylogenetics Evolution 10(3): 449-463.

Becerra, J. X. 2002. Evolution of Mexican *Bursera* (Burseraceae) inferred from ITS, ETS, and 5S nuclear ribosomal DNA sequences. Mol. Phylogenetics Evolution 26: 300-309.

Bogler, D. J. and B.B. Simpson. 1996. Phylogeny of Agavaceae based on its rDNA sequence variation. Amer. J. Bot. 83(9): 1225-1235.

Burgess, K.S., A.J. Fazekas, P.R. Kesanakurti, D.M. Percy, M. Hajibabaei, S.W. Graham, B.C. Husbank, S.G. Newmaster, and S.C. Spencer. 2007. Applied DNA Barcoding in Plants. Mtg. Can. Plant Barcoding Group, Toronto. 2007.

Douzery, E.J., A.M. Pridgeon, P. Kores, H.P. Linder, H. Kurzweil and M.W. Chase. 1999. Molecular phylogenetics of *Diseae* (Orchidaceae): A contribution from nuclear ribosomal ITS sequences. Amer. J. Bot. 86: 887-899.

Guilliams, C.M. 2009. Phylogenetic reconstruction, character evolution, and conservation in the genus *Calyptridium* (Montiaceae). San Diego State Univ., San Diego, M.S. Thesis.

Hall, T. A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acid Symp. Ser. 41: 95-98.

Hannon, D. P. 2002. *Manfreda* and its allies. Cactus Succulent J. 74: 245-251.

Harpke, D. and A. Petersen. 2006. Non-concerted ITS evolution in *Mammillaria* (Cactaceae). Mol. Phylogenetics Evolution 41(3): 579-593.

Hartmann, S., J.D. Nason, and D. Bhattacharya. 2001. Extensive ribosomal DNA genic variation in the columnar cactus *Lophocereus*. J. Mol. Evolution 53: 124-134.

Hershkovitz, M. A., E.A. Zimmer, and W.J. Hahn. 1999. Ribosomal DNA sequences, p. 268-326. In: P. Hollingsworth, R. Bateman and R. Gornall (eds.), Molecular Systematics. Taylor and Francis, London, England.

Howard, T.M. 2001. Bulbs for warmer climates. University of Texas Press, Austin, TX.

Hseih, H.M., R.J. Hou, K.F. Chen, L.C. Tsai, S.W. Liu, K.L. Liu, A. Linacre, and J.C. Lee. 2004. Estimating the rDNA IGS Structure of *Cannabis sativa*. J. Forensic Sci. 49(3): 1-3.

Irish, M. and G. Irish. 2001. Arizona gardener's guide. Cool Spring Press, Brentwood, TN.

Kearse, M., R. Moir, A. Wilson, S. Stone-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, C. Duran, T. Thierer, B. Ashton, P. Meintjes and A. Drummond. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics Applications Notes 28(12): 1647–1649.

Keb-Llanes, M., G. Gonzalez, B. Chi-Manzanero, and D. Infante. 2002. A rapid and simple method for small-scale DNA extraction in Agavaceae and other tropical plants. Plant Mol. Biol. Rptr. 20: 299a-299e.

Kelch, D.G. and B.G. Baldwin. 2003. Phylogeny and ecological radiation of New World thistles (*Cirsium*, Cardueae-Compositae) based on ITS and ETS rDNA sequence data. Mol. Ecol. 12: 141-151.

Lindstrom, J.T. 2006. Intergeneric hybrids between *Polianthes* and *Manfreda*. Southern Nursery Assn. Res. Conf. 51: 599-601.

Markos, S. and B.G. Baldwin. 2001. Higher-level relationships and major lineages of *Lessingia* (Compositae, Astereae) based on nuclear rDNA Internal and External Transcribed Spacer (ITS and ETS) sequences. Systematic Bot. 26(1): 168-183.

Mayol, M. and J.A. Rosselo. 2001. Why nuclear ribosomalDNAspacers (ITS) tell different stories in Quercus. Mol. Phylogenetic Evolution 19: 167–176.

McKelvey, S. D. and K. Sax. 1933. Taxonomic and cytological relationships of *Yucca* and *Agave*. J. Arnold Arboretum 15: 76-80.

Njuguna, W. and N.V. Bassil. 2010. DNA barcoding: unsuccessful for species identification in *Fragaria* L. ACTA Hort. 918: 349 - 356.

Pyrer, K. M., E. Schuettpeltz, L. Huiet, A.L. Gusz, C.J. Rothfels, T. Avent, D. Swartz, and M.D. Windham. 2010. DNA barcoding exposes a case of mistaken identity. Mol. Ecol. Resources 10: 979-985.

Ruggiero, M.V. and G. Procaccini, G. 2004. The rDNA ITS region in the lessepsian marine angiosperm *Halophila stipulacea* (Forssk.) Aschers. (Hydrocharitaceae): intragenomic variability and putative pseudogenic sequences. J. Mol. Evolution 58: 115–121.

Sass, C., D.P. Little, D. D.W. Stevenson, and C.D. Specht. 2011. Vascular Plants and Paleobotany. 1st ed. Apple Academic Press, Point Pleasant, NJ.

Sāto, D. 1935. Analysis of the karyotype in *Yucca*, *Agave* and related genera with special reference to the phylogenetic significance. Jpn. J. of Genet. 11: 272-278.

Tapia-Tussel, R., A. Quijano-Raymayo, R. Rojas-Herrera, A. Larque-Savverda, and D. Perez-Brito. 2005. A fast, simple, and reliable high-yielding method for DNA extraction from different plant species. Mol. Biotechnology 31: 137-139.

Verhoek-Williams, S.E. 1975. A study of the tribe Poliantheae (including *Manfreda*) and revision of *Manfreda* and *Prochnyanthes* (Agavaveae). Cornell Univ., Ithaca, NY, PhD Diss.

Vilgalys, R. 2001. Conserved primer sequences for PCR amplification and sequencing from nuclear ribosomal RNA. Duke University, Durham, NC. 17 July 2012. <http://www.biology.duke.edu /fungi/mycolab/primers.htm>.

White, T. J., T. Bruns, S. Lee, and J. W. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. p. 315-322 In: Innis, M. A., D. H. Gelfand, J. J. Sninsky, and T. J. White (eds.) PCR protocols: A guide to methods and applications. Academic Press Inc, New York, NY.

CONCLUSIONS

The objective of Study 1 was to review existing specific delimitations by means of phylogenetic and Principal Component analyses to infer greater confidence or question existing interspecific classifications.

The lack of variation between qualitative morphological characters assessed via phylogenetic techniques resulted in both Maximum Parsimony and Bayesian analyses failing to elicit a greater understanding of any interspecific relationships within *Manfreda*. Principal Component Analysis employing quantitative characters alluded to strongly supported classifications for 20 species. Subsequent review of existing literature and herbarium data provided substantial evidence to support a further 27 specific classifications.

The study concluded that *M. maculata* and *M. pubescens* possessed only limited derived characters and previous independent specific classifications were overly reliant on anatomical proportions that proved to be highly variable, therefore a varietal system for *Manfreda maculata* was proposed consisting of *M. maculata* var. *maculata* and var. *pubescens*.

The objective of Study 2 was to test the utility of CBOL Plant DNA Barcodes for phylogenetic research and species identification.

Phylogenetic inference was limited to only five informative variables, neither Maximum Parsimony or Bayesian Analyses, employing two models of evolution, could determine well-supported monophyletic lineages for individual species. It was concluded that the CBOL Plant DNA barcode possessed limited utility for phylogenetic inference in the genus *Manfreda*.

Assessment of CBOL Plant DNA barcodes employing a nucleotide network utilized all 19 variable characters. All species of *Manfreda* were distinguished. Only a single node separated

each species except a *Yucca gigantea* Lem. outgroup suggesting limited polymorphisms, yet sufficient variation was observed to suggest that CBOL Plant DNA Barcodes are functional in the 19 taxa of *Manfreda* included in the study.

The objective of Study 3 was to investigate the potential utility of External Transcribed Spacer (ETS) and Internal Transcribed Spacer (ITS) nuclear ribosomal gene regions as DNA barcodes capable of detecting recent hybridization for employment in identification of intergeneric hybrids between *Agave*, *Manfreda* and *Polianthes*.

The study of the ETS gene region was hindered by the inability to amplify the encompassing Intergenic Spacer (IGS) region for subsequent development of primers. The lack of universal primers proved to be the greatest obstacle to study of the ETS region and utilization for DNA barcoding.

Amplification of the ITS gene region was not problematic with successful utilization of primers and PCR protocol. Subsequent sequencing however detected intra-individual polymorphisms of the multiple ITS copies appertained by the genome. Such a phenomenon is rare in plants but can only be overcome via labor intensive cloning operations. Therefore, the ITS gene region also proved to be inappropriate for inclusion in DNA barcodes for the identification of hybridized taxa in the between *Agave*, *Manfreda* and *Polianthes*.

APPENDIX 1

Character	Character description	Character type
Anther attachment	*Dorsifixed = 1 Basifixed = 2*	Categorical
Anther dehiscence type	*Longitudinal = 1 Other = 2*	Categorical
Anther length	*Length of anther in mm*	Quantitative
Anther type	*Dithecal = 1 Monothecal =2*	Categorical
Arial stem type	*Rosette = 1 Other = 2*	Categorical
Calyx fusion	*Synsepalous = 1 Aposepalous = 2*	Categorical
Calyx length	*Length from base of calyx to highest lobe (mm)*	Quantitative
Calyx symmetry	*Actinomorphic = 1 Zygomorphic = 2*	Categorical
Carpel number	*Number of carpel present*	Quantitative
Collector	*Collectors name*	Information
Filament length	*Length of filament in mm*	Quantitative
Flower arrangement	*Spicate, 1 flower per node = 1 Spicate, 2 flowers per node = 2 Paniculate = 3*	Categorical
Flower attachment	*Pedicillate = 1 Sessile = 2*	Categorical
Flower curvature	*Straight = 1 Recurved = 2*	Categorical
Flower length (-pedicel)	*Length from ped to calyx tip (mm)*	Quantitative
Flower symmetry	*Actinomorphic = 1 Zygomorphic = 2*	Categorical
Flower type	*Funnelform = 1 Cylindrical = 2*	Categorical
Flower width	*length from widest portion of calyx (mm)*	Quantitative
Full name	*Botanical name*	Information
Herbarium	*Institute holding specimen*	Information
Identifier	*Unique identifer for each specimen viewed*	Information

Inflorescence form	*Dense Raceme = 1 Semidense Raceme = 2 Open Raceme = 3*	Categorical
Inflorescence length	*Length of the entire peduncle (mm)*	Quantitative
Inflorescence type	*Scapose = 1 Other = 2*	Categorical
Leaf arrangement	*Rosette = 1 Other = 2*	Categorical
Leaf attachment	*Sessile = 1 Petiolate = 2*	Categorical
Leaf Blade apex	*Acute = 1 Spinose = 2*	Categorical
Leaf blade margin	*Entire = 1 Toothed = 2*	Categorical
Leaf blade shape	*Linear = 1 [Fat Middle] = 2*	Categorical
Leaf length (longest)	*Length of leaf from base to tip (mm)*	Quantitative
Leaf pubescence	*Pubescent = 1 Glabrous = 2*	Categorical
Leaf succulence	*Succulent = 1 Semi-Succulent = 2 Non-Succulent = 3*	Categorical
Leaf type	*Simple = 1 Compound = 2*	Categorical
Leaf width	*Width of leaf from widest points (mm)*	Quantitative
Native co-ordinates	*Estimated Coordinates for native locality*	Information
Native locality	*Location noted on herbarium specimen*	Information
Ovary length	*Length of Ovary in mm*	Categorical
Ovary position	*Inferior = 1 Superior = 2*	Categorical
Ovary shape	*Round = 1 Intermediate = 2 Elliptic = 3*	Categorical
Pedicel length	*Length of pedicel (mm)*	Quantitative
Perianth cycly	*Uniseriate = 1 Biseriate = 2*	Categorical
Perianth type	*Homochylamydeous = 1 Dichlamydeous = 2*	Categorical

Plant habit	*Description of habit in which specimen was located*	Information
Plant height	*From ground to leaf tip (mm)*	Quantitative
Root type	*Adventious = 1 Other = 2*	Categorical
Stamen cycly	*Uniseriate = 1 Biseriate = 2*	Categorical
Stamen fusion	*Apostemonous = 1 Other =2*	Categorical
Stamen insert	*1 series = 1 2 series = 2*	Categorical
Stamen merosity	*number of anthers present*	Quantitative
Stamen position	*Exserted = 1 Inserted = 2*	Categorical
Stamen type	*filamentous = 1 Laminar = 2*	Categorical
Stigma numbers	*Number of stigma present*	Quantitative
Stigma position	*Terminal = 1 Other = 2*	Categorical
Style length	*Length of style in mm*	Quantitative
Style number per Pistil	*3 locules and 3 lobes = 1 other = 2*	Categorical
Style position	*Exserted = 1 Inserted = 2*	Categorical
Underground stem	*Bulb = 1 Other = 2*	Categorical
Year	*Year collected*	Information

APPENDIX 2

Maximum Parsimony - Morphological Categorical Characters

Pairwise Distance Matrix

	M. brunnea	*M. chamelensis*	*M. elongata*	*M. fusca*	*M. guttata*	*M. hauniensis*
M. brunnea						
M. chamelensis	0.08108					
M. elongata	0.10811	0.05405				
M. fusca	0.11765	0.08824	0.08824			
M. guttata	0.10811	0.08108	0.08108	0.05882		
M. hauniensis	0.10811	0.10811	0.10811	0.11765	0.08108	
M. involuta	0.08108	0.05405	0.05405	0.11765	0.10811	0.08108
M . jaliscana	0.10811	0.05405	0.05405	0.11765	0.13514	0.05405
M. littoralis	0.11429	0.08571	0.08571	0.03125	0.00000	0.08571
M. longibracteata	0.08108	0.05405	0.05405	0.11765	0.10811	0.08108
M. longiflora	0.11111	0.13889	0.11111	0.12121	0.16667	0.22222
M. maculata	0.18919	0.16216	0.10811	0.17647	0.10811	0.13514
M. maculosa	0.08108	0.13514	0.16216	0.14706	0.18919	0.13514
M. nanchititlensis	0.16216	0.10811	0.05405	0.11765	0.13514	0.10811
M. parva	0.10811	0.05405	0.05405	0.05882	0.05405	0.13514
M. planifolia	0.10811	0.08108	0.08108	0.05882	0.00000	0.08108
M. potosina	0.16667	0.19444	0.22222	0.24242	0.22222	0.22222
M. pringlei	0.08108	0.05405	0.05405	0.02941	0.02703	0.10811
M. pubescens	0.13514	0.10811	0.10811	0.17647	0.10811	0.08108
M. revoluta	0.10811	0.08108	0.08108	0.14706	0.08108	0.05405
M. rubescens	0.08108	0.08108	0.02703	0.05882	0.05405	0.08108
M. scabra	0.05405	0.05405	0.11765	0.10811	0.08108	0.00000
M. sileri	0.02941	0.08824	0.08824	0.09677	0.11765	0.14706
M. singuliflora	0.16216	0.10811	0.16216	0.11765	0.16216	0.18919
M. umbrophila	0.10811	0.16216	0.13514	0.14706	0.10811	0.10811
M. variegata	0.05405	0.08108	0.10811	0.17647	0.16216	0.16216
M. virginica	0.10811	0.13514	0.13514	0.05882	0.10811	0.18919
Agave americana	0.16216	0.24324	0.27027	0.29412	0.21622	0.21622
Polianthes tuberosa	0.16216	0.18919	0.21622	0.23529	0.16216	0.16216

	M. involuta	*M. jaliscana*	*M. littoralis*	*M. longibracteata*	*M. longiflora*	*M. maculata*
M. jaliscana	0.05405					
M. littoralis	0.11429	0.14286				
M. longibracteata	0.00000	0.05405	0.11429			
M. longiflora	0.16667	0.16667	0.14286	0.16667		
M. maculata	0.10811	0.16216	0.11429	0.10811	0.16667	
M. maculosa	0.10811	0.10811	0.17143	0.10811	0.13889	0.21622
M. nanchititlensis	0.10811	0.05405	0.14286	0.10811	0.11111	0.10811
M. parva	0.05405	0.10811	0.05714	0.05405	0.13889	0.10811
M. planifolia	0.10811	0.13514	0.00000	0.10811	0.16667	0.10811
M. potosina	0.22222	0.22222	0.20000	0.22222	0.11111	0.22222
M. pringlei	0.08108	0.10811	0.02857	0.08108	0.13889	0.13514
M. pubescens	0.05405	0.10811	0.11429	0.05405	0.22222	0.05405
M. revoluta	0.02703	0.08108	0.08571	0.02703	0.19444	0.08108
M. rubescens	0.05405	0.08108	0.05714	0.05405	0.13889	0.10811
M. scabra	0.05405	0.11429	0.00000	0.16667	0.10811	0.10811
M. sileri	0.08824	0.11765	0.09375	0.08824	0.06061	0.17647
M. singuliflora	0.10811	0.16216	0.14286	0.10811	0.19444	0.21622
M. umbrophila	0.13514	0.13514	0.11429	0.13514	0.13889	0.13514
M. variegata	0.10811	0.10811	0.17143	0.10811	0.11111	0.21622
M. virginica	0.16216	0.18919	0.08571	0.16216	0.11111	0.21622
Agave americana	0.24324	0.27027	0.22857	0.24324	0.27778	0.24324
Polianthes tuberosa	0.18919	0.21622	0.17143	0.18919	0.27778	0.18919

	M. maculosa	*M. nanchititlensis*	*M. parva*	*M. planifolia*	*M. potosina*	*M. pringlei*
M. nanchititlensis	0.16216					
M. parva	0.13514	0.10811				
M. planifolia	0.18919	0.13514	0.05405			
M. potosina	0.19444	0.22222	0.25000	0.22222		
M. pringlei	0.16216	0.10811	0.02703	0.02703	0.25000	
M. pubescens	0.16216	0.16216	0.10811	0.10811	0.22222	0.13514
M. revoluta	0.13514	0.13514	0.08108	0.08108	0.19444	0.10811
M. rubescens	0.16216	0.08108	0.05405	0.05405	0.25000	0.02703
M. scabra	0.10811	0.05405	0.10811	0.22222	0.08108	0.05405
M. sileri	0.02941	0.14706	0.08824	0.11765	0.15152	0.08824
M. singuliflora	0.16216	0.21622	0.10811	0.16216	0.19444	0.13514
M. umbrophila	0.08108	0.13514	0.10811	0.10811	0.19444	0.13514
M. variegata	0.10811	0.16216	0.13514	0.16216	0.16667	0.13514
M. virginica	0.16216	0.18919	0.10811	0.10811	0.16667	0.08108
Agave americana	0.24324	0.32432	0.27027	0.21622	0.22222	0.24324
Polianthes tuberosa	0.24324	0.27027	0.21622	0.16216	0.27778	0.18919

	M. pubescens	*M. revoluta*	*M. rubescens*	*M. scabra*	*M. singuliflora*	*M. sileri*
M. revoluta	0.02703					
M. rubescens	0.10811	0.08108				
M. scabra	0.02703	0.05405	0.05405			
M. sileri	0.14706	0.11765	0.08824	0.08824		
M. singuliflora	0.16216	0.13514	0.16216	0.10811	0.14706	
M. umbrophila	0.13514	0.10811	0.13514	0.13514	0.08824	0.21622
M. variegata	0.16216	0.13514	0.13514	0.10811	0.05882	0.18919
M. virginica	0.21622	0.18919	0.10811	0.16216	0.08824	0.10811
Agave americana	0.18919	0.21622	0.24324	0.24324	0.20588	0.27027
Polianthes tuberosa	0.13514	0.16216	0.18919	0.18919	0.20588	0.21622

	M. umbrophila	*M. variegata*	*M. virginica*	*Agave americana*	*Polianthes tuberosa*
M. variegata	0.13514				
M. virginica	0.16216	0.16216			
Agave americana	0.21622	0.16216	0.21622		
Polianthes tuberosa	0.21622	0.21622	0.27027	0.18919	

APPENDIX 3

DNA Extraction and Electrophoresis Solutions

CTAB buffer. 100 m*M* Tris-HCl (pH 7.5), 700 mM NaC1, 50 mM EDTA (pH 8.0), 1% hexadecyltrimethylammonium bromide (CTAB) (w/v), 1% Polyvinylpyrrolidone (PVP) (w/v).

Sodium borate buffer. 10 mM sodium hydroxide (Adjusted to 8.5 pH with boric Acid).

APPENDIX 4

Maximum Parsimony - *MatK* and *rbcL* gene regions

Uncorrected ("p") Distance Matrix

	M. brunnea	*M. guttata*	*M. hauniensis*	*M. jaliscana*	*M. longibracteata*	*M. longiflora*
M. brunnea						
M. guttata	0.00162					
M. hauniensis	0.00324	0.00162				
M. jaliscana	0.00162	0	0.00162			
M. longibracteata	0.00405	0.00243	0	0.00243		
M. longiflora	0.00245	0.00081	0.00244	0.00081	0.00081	
M. maculata	0.00324	0.00162	0	0.00162	0.00082	0.00244
M. maculosa	0.00162	0	0.00162	0	0.00243	0.00081
M. nanchititlensis	0.00324	0.00162	0	0.00162	0.00081	0.00243
M. potosina	0.00162	0	0.00162	0	0.00243	0.00081
M. pringlei	0.00324	0.00162	0	0.00162	0.00082	0.00243
M. pubescens	0.00163	0	0.00162	0	0.00243	0.00081
M. rubescens	0.00162	0	0.00162	0	0.00244	0.00081
M. scabra	0.00325	0.00163	0	0.00163	0.00081	0.00244
M. singuliflora	0.00162	0	0.00162	0	0.00243	0.00081
M .undulata	0.00162	0	0.00162	0	0.00243	0.00081
M. variegata	0.00162	0	0.00162	0	0.00243	0.00081
M. virginica	0.00162	0	0.00162	0	0.00243	0.00081
Agave sisalana	0.00243	0.00081	0.00243	0.00081	0.00324	0.00162
Camassia cusickii	0.01052	0.00892	0.01052	0.00891	0.01134	0.00976
Yucca gigantea	0.00405	0.00405	0.00567	0.00405	0.00648	0.00488

	M. maculata	*M. maculosa*	*M. nanchititlensis*	*M. potosina*	*M. pringlei*	*M. pubescens*
M. maculosa	0.00162					
M. nanchititlensis	0	0.00162				
M. potosina	0.00162	0	0.00162			
M. pringlei	0	0.00162	0	0.00162		
M. pubescens	0.00162	0	0.00162	0	0.00162	
M. rubescens	0.00162	0	0.00162	0	0.00162	0
M. scabra	0	0.00163	0	0.00163	0	0.00163
M. singuliflora	0.00162	0	0.00162	0	0.00162	0
M. undulata	0.00162	0	0.00162	0	0.00162	0
M. variegata	0.00162	0	0.00162	0	0.00162	0
M. virginica	0.00162	0	0.00162	0	0.00162	0
Agave sisalana	0.00243	0.00081	0.00243	0.00081	0.00243	0.00081
Camassia cusickii	0.01054	0.00891	0.01052	0.00891	0.01053	0.00893
Yucca gigantea	0.00568	0.00405	0.00566	0.00405	0.00567	0.00405

	M. rubescens	*M. scabra*	*M. singuliflora*	*M. undulata*	*M. variegata*	*M. virginica*	*Agave sisalana*	*Camassia cusickii*
M. scabra	0.00163							
M. singuliflora	0	0.00163						
M. undulata	0	0.00163	0					
M. variegata	0	0.00163	0	0				
M. virginica	0	0.00163	0	0	0			
Agave sisalana	0.00081	0.00244	0.00081	0.00081	0.00081	0.00081		
Camassia cusickii	0.0089	0.01061	0.0089	0.00891	0.00891	0.0089	0.00971	
Yucca gigantea	0.00405	0.00571	0.00405	0.00405	0.00405	0.00405	0.00485	0.0089

APPENDIX 5

Components of PCR Reactions

ITS PCR Reaction	
Reagent	**Volume/Final Concentration per Reaction**
Sterile deionized H20	20.5 µl
DNA template	0.5 µl
Primer 1	2.0 µl
Primer 2	2.0 µl
EcoDry™ PCR Premix	Pre-packaged with eppendorf tube

ETS PCR Reaction	
Reagent	**Volume/Final Concentration per Reaction**
Sterile deionized H20	16.0 µl
10x Advantage Genomic LA Buffer	2.5 µl
dNTP Mixture	1.0 µl
Primer 1	1.5 µl
Primer 2	1.5 µl
Advantage Genomic LA Polymerase	2.5 µl