

**ON THE ORIGIN OF HOPS:  
GENETIC VARIABILITY, PHYLOGENETIC RELATIONSHIPS,  
AND ECOLOGICAL PLASTICITY OF *HUMULUS* (CANNABACEAE)**

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF  
THE UNIVERSITY OF HAWAII AT MĀNOA IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

BOTANY

MAY 2014

By

Jeffrey R. Boutain

DISSERTATION COMMITTEE:

Will C. McClatchey, Chairperson

Mark D. Merlin

Sterling C. Keeley

Clifford W. Morden

Stacy Jørgensen

Copyright © 2014  
by  
Jeffrey R. Boutain

This dissertation is dedicated to my family tree.

## ACKNOWLEDGEMENTS

There are a number of individuals to whom I am indebted in many customs. First and foremost, I thank my committee members for their contribution, patience, persistence, and motivation that helped me complete this dissertation. Specifically, thank you Dr. Will McClatchey for the opportunity to study in a botany program with you as my advisor and especially the encouragement to surf plant genomes. Also with great gratitude, thank you Dr. Sterling Keeley for the opportunity to work on much of this dissertation in your molecular phylogenetics and systematics lab. In addition, thank you Dr. Mark Merlin for numerous brainstorming sessions as well as your guidance and expert perspective on the Cannabaceae. Also, thank you Dr. Cliff Morden for the opportunity to work in your lab where the beginnings of this molecular research took place. Thank you Dr. Jianchu Xu for welcoming me into your lab group at the Kunming Institute of Botany, Chinese Academy of Sciences (CAS) and the opportunity to study the Yunnan hop.

In many ways, major contributions towards the completion of this dissertation have come from my family, and I thank you for your unconditional encouragement, love, and support. Additionally, thank you to my hānai/adopted family in Hawaii because I could not have completed this dissertation without the countless little and big things you have done for me to succeed.

Thank you collaborators, counterparts, and friends in both the United States of America and the People's Republic of China that were instrumental in making this dissertation possible. My sincere gratitude, appreciation, and aloha to Dr. Tim Gallaher and his family for all things botany and not botany, but most importantly thank you for your friendship and moral support. Thank you Seana Walsh for your support, encouragement, and assistance. Also, thank you Dr. Alison Sherwood, Dr. Art Whistler, Dr. Ernest Small, Dr. Michael Allen, Dr. Mark Nickum, Dr. Dake Zhao, Dr. Pei-Luen Lu, Dr. Jennifer Bufford, Dr. Jason Cantley, Dr. Marian Chau, Dan Amato, David Visi, Dave Reedy, Katie Kamelamela, Matt Lurie, and Rob Clarke. Thank you lab, office, and teaching colleagues and fellow graduate students in the Botany and Biology Departments at the University of Hawai'i at Mānoa (UH). Thank you local field guides, plant

collectors, and herbarium curators for your assistance in the field, with specimen preservation, and especially with destructive sampling for DNA analysis. Thank you UH staff and faculty who assisted me in many ways while I was in graduate school.

This dissertation was made possible through support from the Botany, Biology, and Tropical Plant and Soil Sciences Departments at UH; Botanical Research Institute of Texas; Kunming Institute of Botany, CAS; World Agroforestry Centre, China Program; Centre for Mountain Ecosystem Studies ICRAF-China; Institute of Botany, CAS; Agriculture and Agri-Food Canada; Department of Botany, Smithsonian Institute, National Museum of Natural History; United States Department of Agriculture; Michigan State University Extension in Leelanau County, Michigan; Michigan Hop Alliance; and New Mission Organics.

Funding for this dissertation came from the following sources: Research and Travel grants from Graduate Student Organization at UH, Anne S. Chatham Fellowship in Medicinal Botany from the Garden Club of America, National Science Foundation - East Asia and Pacific Summer Institutes for U.S. Graduate Students (EAPSI), National Science Foundation - Ecology and Evolution of Infectious Disease Program through Brown University, Beatrice Krauss Fellowship Fund in Botany from UH, and the Botanical Research Institute of Texas.

## ABSTRACT

Introgression with the domesticated hop plant (*Humulus lupulus* L., Cannabaceae) makes some morphological indentifications difficult for wild or heirloom hop plants found on wild or fallow lands. As a result, this dissertation takes an approach using DNA barcodes to investigate the proposed polytypic model for *Humulus*, which is mostly based on morphology. Results from two standard DNA barcode studies with fresh and herbarium materials show the genetic variability and phylogenetic relationships. In addition, putatively wild as well as unknown hop plants were sorted into three species (*H. lupulus*, *H. scandens*, and *H. yunnanensis*), while *H. lupulus* was further split into Western to Central Eurasia, East Asia, and the New World clades. Low resolution was found for the putative varieties of *H. lupulus* from East Asia and the New World. Furthermore, results from two chloroplast genome (plastome) barcode studies show the conserved relationships and unique evolutionary history within the Cannabaceae (*s.s.*). The phylogenomic analyses presented here suggest the Cannabaceae (*s.s.*) is much more ancient than previously proposed. With a mid-Cretaceous origin based on high bootstrap and posterior probability support on a polyphyletic tree with basal East Asian taxa, a Laurasian migration hypothesis is probable for *Humulus*. Compared to single or several DNA regions used to barcode plants, the plastome as a single DNA barcode supports the unity of the *H. lupulus* complex in a polytypic model. Further duplicate sampling on the phyloplastome tree is required to test the varietal relationships of *H. lupulus* from East Asian and the New World. More broadly, a large scale phylogenomic study on the Cannabaceae (*s.l.*)/Celtidaceae remains as a high priority for future research.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	iv-v
ABSTRACT.....	vi
LIST OF TABLES .....	x-xi
LIST OF FIGURES .....	xii-xiii
PREFACE .....	xiv
CHAPTER 1: Introduction .....	1
Welcome to the Family .....	1
Genetic Branches on the Family Tree.....	6
The Wild and Cultivated Hop Branch on the Family Tree .....	8
Hop Taxonomy .....	14
Fossil and Phylogenetic Roots of the Hop Branch on the Family Tree.....	19
Hypothesizing Phylogenic Roots of the Hop Branch on the Family Tree.....	25
Main Research Hypotheses.....	26
CHAPTER 2: Rediscovering Wild Michigan Hops ( <i>Humulus lupulus</i> L.) .....	29
Abstract .....	29
Keywords .....	29
Introduction.....	30
Objectives .....	32
Hypothesis.....	32
Materials and Methods.....	34
Results.....	46
Discussion .....	59
Conclusion .....	67
Acknowledgements.....	68
CHAPTER 3: DNA Support for the Endemic Yunnan Hop ( <i>Humulus yunnanensis</i> Hu).69	
Abstract .....	69
Keywords .....	69
Introduction.....	70

Objectives .....	72
Hypothesis.....	72
Materials and Methods.....	75
Results .....	90
Discussion .....	103
Conclusion .....	115
Acknowledgements .....	116
CHAPTER 4: Draft Chloroplast Genome of a Wild American Hop ( <i>Humulus lupulus</i> var. <i>neomexicanus</i> A. Nelson & Cockerell).....	117
Abstract .....	117
Keywords .....	117
Introduction.....	118
Objectives .....	119
Hypothesis.....	119
Materials and Methods.....	120
Results.....	127
Discussion .....	136
Conclusion .....	137
Acknowledgements .....	138
CHAPTER 5: The Origin of <i>Humulus</i> : A Phylogenomic Surfing Approach.....	139
Abstract .....	139
Keywords .....	139
Introduction.....	140
Objectives .....	143
Hypothesis.....	143
Materials and Methods.....	144
Results.....	152
Discussion .....	161
Conclusion .....	164
Acknowledgements .....	165

CHAPTER 6: Outcomes, Applications, and Directions for Cannabaceae Research.....	166
Introduction.....	166
The Cannabaceae Conundrum .....	168
Main Research Hypotheses Revisited.....	171
Unique Contributions to the Advancement of Cannabaceae Science.....	172
Overall Conclusion .....	173
REFERENCES: .....	174

## LIST OF TABLES

<b>Table 1.1.</b> Numbers of putative and recognized species in the Cannabaceae .....	2
<b>Table 1.2.</b> Comparison of DNA regions in three Cannabaceae outgroup species. ....	7
<b>Table 1.3.</b> Taxonomy of <i>Humulus</i> according to Tropicos.org. ....	17
<b>Table 1.4.</b> Taxonomy of <i>Humulus</i> according to The Plant List .....	18
<b>Table 2.1.</b> Plant specimens sampled for DNA in Michigan hops study.....	35
<b>Table 2.2.</b> DNA regions and primer pairs for PCR in Michigan hops study .....	42
<b>Table 2.3.</b> DNA sequence variations in Michigan hops study .....	46
<b>Table 2.4.</b> Evolutionary divergence between groups in Michigan hops study .....	47
<b>Table 2.5.</b> Haplotypes estimates in Michigan hops study .....	48
<b>Table 2.6.</b> Plant samples sorted by lineage/clade in Michigan hops study .....	63
<b>Table 2.7.</b> Summary distinguishing the varieties of <i>Humulus lupulus</i> .....	65
<b>Table 2.8.</b> Evolutionary divergence between New World and Old World groups of <i>Humulus lupulus</i> in Michigan hops study.....	66
<b>Table 3.1.</b> Plant specimens sampled for DNA in China hops study .....	76
<b>Table 3.2.</b> DNA regions and primer pairs for PCR in China hops study .....	84
<b>Table 3.3.</b> DNA sequence variations in China hops study.....	90
<b>Table 3.4.</b> Evolutionary divergence between groups in China hops study. ....	91
<b>Table 3.5.</b> Haplotypes estimates in China hops study.....	92
<b>Table 3.6.</b> Plant Samples sorted by lineage/clade in China hops study. ....	109
<b>Table 3.7.</b> Summary distinguishing the species and varieties of <i>Humulus</i> .....	112
<b>Table 3.8.</b> Evolutionary divergence between East Asia, New World, and Old World groups of <i>Humulus lupulus</i> in China hops study. ....	113
<b>Table 4.1.</b> Parameters for <i>de novo</i> assembler to optimize word size. ....	123
<b>Table 4.2.</b> Parameters for SEREAL in reference assembler. ....	125
<b>Table 4.3.</b> Taxa and Genbank accessions for plastome sequences. ....	126
<b>Table 4.4.</b> Supplementary QC report of over-represented sequences .....	128
<b>Table 4.5.</b> Summary of optimized word score of 14 for <i>de novo</i> assembler.....	130
<b>Table 4.6.</b> Formulas to convert the number of base pairs to picograms .....	130

<b>Table 4.7.</b> Summary of SEREAL output from reference assembler .....	131
<b>Table 4.8.</b> Summary of SEREAL consensus used in reference assembler. ....	132
<b>Table 5.1.</b> Plant specimens sampled for total genomic DNA sequencing .....	146
<b>Table 5.2.</b> Summary of reference assembler for 8-plexed sequencing run .....	154
<b>Table 5.3.</b> DNA sequence variations in plastome study. ....	156
<b>Table 5.4.</b> Evolutionary divergence between taxa in plastome study. ....	157
<b>Table 5.5.</b> Plastome matrix comparisons for <i>Humulus</i> spp. and <i>Cannabis</i> .....	157

## LIST OF FIGURES

<b>Fig. 1.1.</b> Female inflorescence of hop ( <i>Humulus</i> ) and hemp ( <i>Cannabis</i> ) plants .....	3
<b>Fig. 1.2.</b> Evolutionary history of the Cannabaceae ( <i>s.s.</i> ).....	4
<b>Fig. 1.3.</b> Comparisons on the Urticlean rosids family tree .....	5
<b>Fig. 1.4.</b> Gladular peltate trichomes on the bracts of hops. ....	11
<b>Fig. 1.5.</b> Experimental low and high hop trellis systems in Hawaii .....	12
<b>Fig. 1.6.</b> High-performance liquid chromatography analysis of cultivar ‘Cascade’ .....	13
<b>Fig. 1.7.</b> Macrofossils of <i>Humulus</i> from the Eocene of Colorado .....	22
<b>Fig. 1.8.</b> Photomicrograph of <i>Humulus lupulus</i> pollen.....	23
<b>Fig. 1.9.</b> Maps of relative continental positions from early Jurassic to present era.....	24
<b>Fig. 2.1.</b> Photo of putatively wild and cultivated hops in Michigan.....	33
<b>Fig. 2.2.</b> Maximum Likelihood tree of ITS2 region in Michigan hops study .....	51
<b>Fig. 2.3.</b> Bayesian inference phylogeny of ITS2 region in Michigan hops study .....	52
<b>Fig. 2.4.</b> Maximum Likelihood tree of <i>petL-psbE</i> region in Michigan hops study .....	54
<b>Fig. 2.5.</b> Bayesian inference phylogeny of <i>petL-psbE</i> region in Michigan hops study .....	55
<b>Fig. 2.6.</b> Maximum Likelihood tree of combined data set in Michigan hops study .....	57
<b>Fig. 2.7.</b> Bayesian inference phylogeny of combined data set in Michigan hops study .....	58
<b>Fig. 2.8.</b> Simplified phylogenetic tree of combined data set in Michigan hops study .....	62
<b>Fig. 3.1.</b> Three recognized species of <i>Humulus</i> .....	73
<b>Fig. 3.2.</b> Collection of <i>H. lupulus</i> var. <i>cordifolius</i> from Yunnan Province of China .....	74
<b>Fig. 3.3.</b> Maximum Likelihood tree of ITS2 region in China hops study .....	95
<b>Fig. 3.4.</b> Bayesian inference phylogeny of ITS2 region in China hops study .....	96
<b>Fig. 3.5.</b> Maximum Likelihood tree of <i>petL-psbE</i> region in China hops study .....	98
<b>Fig. 3.6.</b> Bayesian inference phylogeny of <i>petL-psbE</i> region in China hops study.....	99
<b>Fig. 3.7.</b> Maximum Likelihood tree of combined data set in China hops study.....	101
<b>Fig. 3.8.</b> Bayesian inference phylogeny of combined data set in China hops study .....	102
<b>Fig. 3.9.</b> Simplified phylogenetic tree of combined data set in China hops study .....	114
<b>Fig. 4.1.</b> PGM run summary using a 316 chip.....	127
<b>Fig. 4.2.</b> Optimized word size using the default settings in <i>de novo</i> assembler .....	129

<b>Fig. 4.3.</b> Map drawings of the draft chloroplast genomes of <i>Humulus</i> and <i>Cannabis</i> ....	134
<b>Fig. 4.4.</b> progressiveMauve alignment of conserved and rearranged plastome regions..	135
<b>Fig. 5.1.</b> Total genomic DNA interpreted as the entire volume of water in a wave .....	141
<b>Fig. 5.2.</b> PGM summary for an 8-plexed barcode sequencing run using a 318 chip.....	152
<b>Fig. 5.3.</b> Maximum Likelihood plastome phylogeny of <i>Humulus</i> .....	159
<b>Fig. 5.4.</b> Bayesian inference plastome phylogeny of <i>Humulus</i> .....	160

## PREFACE

*“...“Domestic” implies that these species have come in or been brought under civilization’s roof, which is true enough; yet the house-y metaphor encourages us to think that by doing so they have, like us, somehow left nature, as if nature were something that only happens outside (xxiii-xxvi)...” “...Domesticated species don’t command our respect the way their wild cousins often do. Evolution may reward interdependence, but our thinking selves continue to prize self-reliance. The wolf is somehow more impressive to us than the dog (xvi)...” Michael Pollan (2001).*

# CHAPTER 1

## Introduction

### Welcome to the Family

The Cannabaceae (*s.s.*) family tree traditionally includes the two sister genera, *Humulus* (hop) and *Cannabis* (hemp) (Fig. 1.1). More broadly, the Cannabaceae (*s.l.*) family tree also includes the Celtidaceae, which are sister to the Ulmaceae, Moraceae, and Urticaceae as members of the Urticalean rosids. These stem relationships on the Urticalean rosids family tree have high support (Fig. 1.2) (Angiosperm Phylogeny Group 2009, Sytsma et al. 2002, Yang et al. 2013). Observed variations within the Urticalean rosids family tree can be seen in Figure 1.3. Initial divergence of the family group was probably in very warm temperate regions or tropical regions. As a result of tectonic plate movements, whether the early diverging lineages of the Cannabaceae arose in the New or Old world is debatable (Donoghue 2008, Sytsma et al. 2002), but they probably originated in Laurasia (Johnson 2002, MacGinitie 1953 and 1969, Manchester 2001, Tiffney 1986, specifically Weber 2003).

More narrowly, the placement of *Humulus* and *Cannabis* on the Cannabaceae (*s.l.*) family tree is under scrutiny (Sattarian 2006, Sytsma et al. 2002, Yang et al. 2013). Likewise with the genera relationships, problems resolving species level relations abounds. Also with a Central Asian origin for the genus *Cannabis* (Hillig 2005), the origin of *Humulus* should to be reexamined based on the clear findings that the Cannabaceae (*s.s.*) family is within the weakly supported Celtidaceae/Cannabaceae (*s.l.*) family (Figs. 1.2 and 1.3). For example, in the genus *Cannabis*, putative species separations are becoming clarified with the draft genome and transcriptome by van Bakel et al. (2011). The differences between two hemp cultivars ('Finola' and 'USO31') and two marijuana strains (Purple Kush and Chemdawg) suggests additional analysis of diverse germplasm is warranted to investigate the evolutionary history and the molecular impact of domestication and breeding on *Cannabis*. "Outstanding areas that might be addressed by further genomic investigation include whether the genus is composed of one

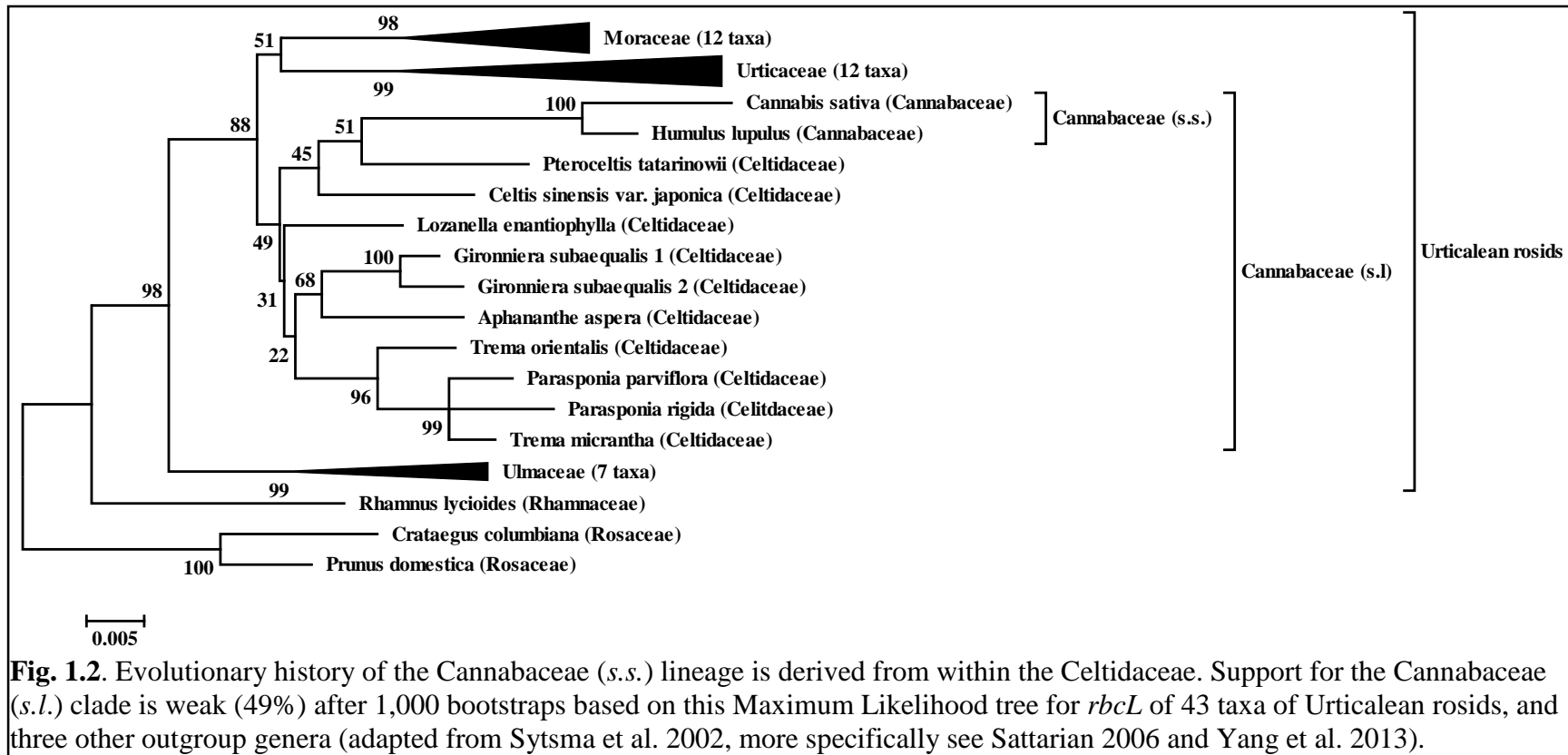
or several species, the existence of ‘sativa’ and ‘indica’ gene pools, the relative contributions that wild ancestors have made to modern hemp and marijuana germplasm, and the process by which cannabis was first domesticated by humans (van Bakel et al. 2011).” Very similarly, if species delimitation within *Humulus* and the broader relationships within the Celtidaceae/Cannabaceae (*s.l.*) are to be resolved (Table 1.1), new characters and new methods of analysis need to be investigated (i.e., DNA barcodes, genomics, high-throughput sequencing) (Boutain et al. 2010a and 2010b, Hillig 2005, Hillig and Mahlberg 2004, Kavalier et al. 2009, Yang et al. 2013).

**Table 1.1.** Numbers of putative and recognized species in the Cannabaceae (*s.s.* and *s.l.*) (adapted from Clarke and Merlin 2013, The Plant List 2013, and Yang et al. 2013).

<b>Genera</b>	<b>Putative Family</b>	<b>Putative species</b>	<b>Accepted species</b>
<i>Aphananthe</i>	( <i>s.l.</i> )	-	5
<i>Cannabis</i>	( <i>s.s.</i> )	1 (cf. <i>C. ruderalis</i> )	2
<i>Celtis</i>	( <i>s.l.</i> )	36	73
<i>Chaetachme</i>	( <i>s.l.</i> )	-	1
<i>Gironniera</i>	( <i>s.l.</i> )	-	6
<i>Humulus</i>	( <i>s.s.</i> )	1 (cf. <i>H. yunnanensis</i> )	2
<i>Lozanella</i>	( <i>s.l.</i> )	-	2
<i>Parasponia</i>	( <i>s.l.</i> )	5	5
<i>Pteroceltis</i>	( <i>s.l.</i> )	-	1
<i>Trema</i>	( <i>s.l.</i> )	30	12
Total ( <i>n</i> =10)		73	109



**Fig. 1.1.** A-Female inflorescence of the hop plant (*Humulus*) commonly called hops or cones. B-Female inflorescence of the hemp plant (*Cannabis*) that some refer to as buds.



**Fig. 1.2.** Evolutionary history of the Cannabaceae (s.s.) lineage is derived from within the Celtidaceae. Support for the Cannabaceae (s.l.) clade is weak (49%) after 1,000 bootstraps based on this Maximum Likelihood tree for *rbcL* of 43 taxa of Urticalean rosids, and three other outgroup genera (adapted from Sytsma et al. 2002, more specifically see Sattarian 2006 and Yang et al. 2013).

		+straight embryo hypanthium -curved embryo no hypanthium	+2-3 pore pollen 2° leaf vein not terminal -4-6 pore pollen 2° leaf vein teeth terminal	<u>Laticifers</u> +throughout *bark only -absent	<u>Fruit</u> +achene *drupe -samara	Base chromosome number
	Ficeae	-	+	+	+	x=13 & 14
	Castilleae					
	Dorst./Mor.					
	Moreae					
	Moreae	-	+	★	+	x=13 & 14
	Artocarpeae					
	Urtic./Lecanth.					
	Lecantheae					
	Poikilospermum					
	Cecropiaceae	-	+	+	+	x=8 & 10
	Boehmeriaceae					
	Parietarieae					
	Cannabaceae (s.s.)					
	Humulus	-	+	-	★/-	x=10, 13 & 15
	Cannabis					
	Pteroceltis					
	Celtis					
	Trema					
	Parasponia					
	Gironniera					
	Chaetachme					
	Lozanella	+	-	-	-	x=14
	Aphananthe					
	Ulmus					
	Zelkova					
	Hemiptelea					
	Holoptelea					
	Ampelocera					

**Fig. 1.3.** Observed comparisons on the Urticalean rosids family tree, which consists of the Ulmaceae, Celtidaceae/Cannabaceae (*s.l.*), Cannabaceae (*s.s.*), Urticaceae, and Moraceae (adapted from Judd et al. 1994, Sytsma et al. 2002, and Yang et al. 2013).

### **Genetic Branches on the Family Tree**

Molecular markers (DNA barcodes) have been used to address the stem relationships of the families within the Urticalean rosids clade (Angiosperm Phylogeny Group 2009, Sytsma et al. 2002, Yang et al. 2013), but none have yet been able to provide the level of resolution necessary to establish precise evolutionary relationships within the group. In two studies that investigated the broader Plantae family tree, Shaw et al. (2005 and 2007) generated >670 kb of sequence data from three closely related species in each of 10 seed plant lineages in an attempt to compare the systematic utility of 34 noncoding chloroplast genome (cpDNA) regions. From these two studies, no single region was found to be the best across all taxonomic lineages. Instead, top choices should be screened before committing to an all-out ‘Sanger’ sequencing effort in order to determine which of these regions are the most suitable in a given lineage (Shaw et al. 2007). Thus, the crown relationships (i.e., monophyly of genera) within the Cannabaceae family tree has strong phylogenetic support, but the relationships among them were largely unresolved (see Yang et al. 2013). Including all recognized genera, the phylogenetic analysis of Yang et al. (2013) indicated that a *Humulus-Cannabis* clade is sister to a low supported group containing both a *Celtis* clade and a *Pteroceltis-Chaetachme-Trema-Parasponia* clade. These relationships of genera and relations of species remain for further study (Angiosperm Phylogeny Group 2009, Sytsma et al. 2002, Yang et al. 2013).

In an example of DNA barcodes for species level relationships in a phylogenetic framework with *Trema*, Yesson et al. (2004) used the nuclear ribosomal DNA (rDNA) internal transcribed spacer (ITS) region and found considerable doubt on the current use of species names in most regions. In comparison, high similarity was observed for the ITS region of *Humulus* and *Cannabis* (Mukherjee et al. 2008, Murakami 2000, Pillay and Kenny 1996 and 2006). The commonly studied gene regions, such as *rbcL*, *trnL*, *ndhF*, and *matK*, for *Humulus* and *Cannabis* have provided low resolution for molecular phylogenetics (Murakami 2000, Pillay and Kenny 1994, Song et al. 2001), which is why spacer regions should be used too. A molecular phylogeny of wild *Humulus lupulus* L. and *Humulus* sp. as the outgroup was conducted with nuclear DNA as well as 12 non-

coding cpDNA regions (Murakami et al. 2006a and 2006b). Murakami et al. (2006a and 2006b) found high support for a European lineage and moderate support for an Asian-North American lineage; however, the research had ambiguous origins of hop plants sampled from China and did not include the third putative species, *H. yunnanensis*. Although, this molecular phylogeny study did find high genetic diversity in North American hops considered to have migrated from Asia. Further research using the systematic framework of Shaw et al. (2007) to find a high resolution cpDNA region most suitable to address intra- and interspecific taxonomic problems and to test the evolutionary relationships among taxa with current phylogenetic methods is a priority for *Humulus*, even for the Cannabaceae (Table 1.2). An alternative approach with the entire chloroplast genome (plastome), which acts as a single non-recombining maternal locus, may be required as the DNA barcode.

**Table 1.2.** Primer regions, number of potential informative characters (PICs), and rank out of 34 for the region with the most PICs found by Shaw et al. (2007) in three Cannabaceae outgroup species (i.e., *Prunus*, Rosaceae, Rosales, Eurosoid I).

<b>Region</b>	<b>Total PICs</b>	<b>Rank</b>	<b>Region</b>	<b>Total PICs</b>	<b>Rank</b>
ITS	-	-	<i>rpl32-trnL</i>	62	1
<i>trnL-trnF</i>	14	25	<i>psbD-trnT</i>	50	2
<i>ndhF-rpl32</i>	36	6	<i>petL-psbE</i>	41	3

### **The Wild and Cultivated Hop Branch on the Family Tree**

Both genera on the Cannabaceae (*s.s.*) family tree, *Cannabis* (hemp) and *Humulus* (hop), have long histories of use by humans. *Cannabis* has been documented for millennia as an important economic plant for textile, food, oil, medicine, and as a recreational/religious drug (Clarke and Merlin 2013, Jiang et al. 2006, Merlin 2003, Russo 2008, Schultes et al. 1974). Also for an extensive time, *Humulus* has been recognized as a traditional medicine to alleviate migraines, inflammation, insomnia, bladder problems, urogynecological infections, symptoms of menopause, and to treat other central nervous system and skin problems (Bennett 2007, Zanolli and Zavatti 2008).

The hop plant is indigenous to the Northern Hemisphere between 35° and 55° N latitude and has been introduced below the equator between 35° and 55° S latitude. Annual and perennial plants occur that have a dioecious (but some monocious), anemophilous, and dextrorse-twining habit. Plants are herbaceous to woody vines that climb with trichomes along a twisting stem instead of tendrils typically found on vines. Most hop plants reach over seven meters in length as they climb up trees, shrubs, or other supportive structures. Wild hop plants flourish in well drained terraces, river and stream bottoms, thickets, hedgerows, roadsides, and disturbed sites (Hampton et al. 2001, Small 1978).

The most familiar role of the hop plant has been as a flavor for fermented ale and lager beverages that are infused with the female flowers (hops) (Rösch 2008). Hops are also used economically and ethnobotanically for traditional medicine, pharmaceuticals, salads, making bread, garden ornaments, fiber, and fodder. Hops, leafy catkins or imbricated heads, are cone-like (strobili) and produce spherical, resinous glands of lupulin having a pleasant odor and bitter taste (Fig. 1.4). These glands of lupulin are responsible for providing aromatics and flavors to fermented beverages, while additionally acting as a preservative by inhibiting gram-positive bacterial growth (De Keukeleire and Heyerick 2005).

Probably due to the bacteriostatic properties, German brewers used wild hops to flavor their ales and lagers for hundreds of years before worldwide cultivation began (Carter et al. 1990; also see Rösch 2008 for a discussion on the remains of an ancient

wooden bottle from the 6<sup>th</sup> century that contained a single pollen grain of *Humulus* suggesting the use of hops to flavor a beverage). Except for Pliny the Elder of Rome (23 C.E.-79 C.E.), the first written record of hops in Europe dates to 768 C.E. (DeLyser and Kasper 1994, Wilson 1975). Furthermore and long before the hop plant was brought into cultivation (see Behre 1999 for first record of cultivation in Bavaria in 859 C.E.), wild hops were collected from sites known as *Humlonarias*, which most likely is a place-name noted for wild hops or places to gather hops, not a direct reference to hop gardens (DeLyser and Kasper 1994; also see Wilson (1975:645) for land ‘sometimes called Humanton’ as well as ‘reference to hoplands’). *Humlonarias* are where wild hop plants grew naturally and flourished "in marshy or wet hollows in fen carr and moist alder-oak woods" (Wilson 1975). Evidently, this ecotone was commonly mistaken in the literature for cultivated hop gardens.

In the Kalevala, an ancient oral folklore and mythology, a passage from Rune XX states:

*“Hop-vine was the son of Remu,  
Small the seed in earth was planted,  
Cultivated in the loose soil,  
Scattered like the evil serpents  
On the brink of Kalew-waters,  
On the Osmo-fields and borders.  
There the young plant grew and flourished,  
There arose the climbing hop-vine,  
Clinging to the rocks and alders.”* (Crawford 1888).

As described in this passage and in many other works of literature, the hop plant is a cultural keystone species requiring conservation for world heritage (Garibaldi and Turner 2004).

From the Iron Age onwards, a reduction of natural wet environments along with intense deforestation and vegetation succession after the plague provided new habitats for hop plants, like forest margins and around clearings within the forest (DeLyser and Kasper 1994, Yeloff and van Geel 2007). On the other hand, Hampton et al. (2001) found

the resilience of native hop populations in well-drained terraces of river basins in Midwestern North America a problem, and some study sites sustained complete loss or serious damage. Therefore, protection of sensitive riparian habitats occupied by truly wild hop plants is warranted as changes in climate impacts the availability, yield, and quality of hop resources (Hampton et al. 2001, Kučera and Krofta 2009, Mozny et al. 2009). Until sufficient wild hop germplasm from locations lacking in the current repositories are secured and deposited for further genetic investigations and use, the riparian habitat of wild *Humulus* remains a priority for conservation (Hampton et al. 2001, Hummer 2003 and 2005, Hummer et al. 2002 and 2003, Wu et al. 2003). “Not only may expeditions provide for the improvement of crop plants in areas already under cultivation, but their findings may also enable agriculture to be carried to new areas for which there is currently no variety or race suitable for cultivation.” (Baker 1978) (see Figs. 1.5 and 1.6 for Homegrown Hops in Hawaii).

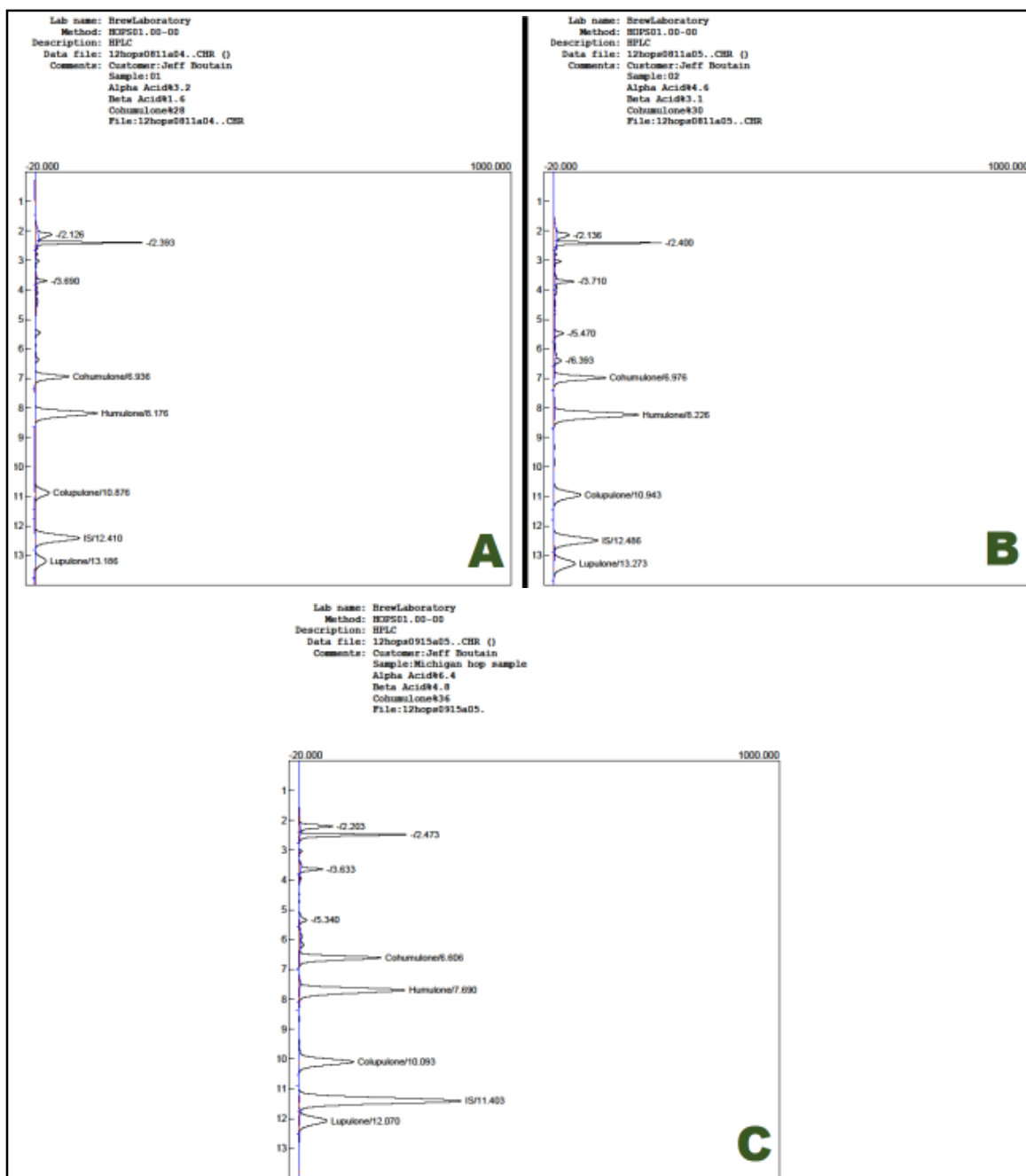
Today, most hop crops are monocultures of selected varieties for optimal growth and sustainable production in a particular locations on high or low trellis systems. Many hop cultivars are selections of the European native hop, *Humulus lupulus* L. var. *lupulus* (Small 1978), a perennial predominantly found from western Europe to central and southwest Asia. Typically, hop plants are asexually propagated from their rhizomes to continue fostering their desired aroma and flavor characteristics, although these chemical traits vary among individual plants of the same variety from year to year and among plants of different varieties (Fig. 1.6; Mozny et al. 2009). In addition, the chemical qualities in hops have a multipotent, bioactive, and antibiotic/antiviral effect (Behr and Vogel 2009, Chadwick et al. 2006, De Keukeleire and Heyerick 2005, Flythe 2009, Shen and Sofos 2008, Srinivasan et al. 2004, Suzuki et al. 2008, Van Cleemput et al. 2009, Wang et al. 2004, Yamaguchi et al. 2009, Zanolli and Zavatti 2008). More importantly, hops, as with ales and lagers in moderation, draw attention to the benefits they could bring to consumer health (Foster et al. 2009, Lamy et al. 2008, Magalhães et al. 2008, Stevens and Page 2004). New hop varieties are incorporated in local gardens, commercial farms, and international breeding programs after showing positive agronomic traits, such as disease, pest, and drought resistance, even an aesthetic value by growing a single plant.



**Fig. 1.4.** Glandular trichomes found in the bracts of hops contain lupulin. A- Triploid cultivar ‘Columbia’ grown in Michigan among diploid cultivars. Male flowers are in the center left. B-Yellow glandular peltate trichomes on the bracts of cultivar ‘Centennial.’



**Fig. 1.5.** A-High trellis approximately 7 meters in Kaimuki, Oahu. B-Low trellis approximately 3 meters in Manoa Valley, Oahu.



**Fig. 1.6.** High-performance liquid chromatography (HPLC) analysis of cultivar ‘Cascade.’ Hops were harvested in the 2012 season after the plants were grown for three consecutive years in same location (i.e., Oahu and Michigan). A-Sample 01 is the chemical profile of one individual plant grown on Oahu. B-Sample 02 is the chemical profile of the combined harvest from 10 plants grown on Oahu. The alpha acid values are 3.2% and 4.6%, and beta acid values are 1.6% and 3.1% for the individual plant and combined harvest respectively. C-Michigan sample of ‘Cascade’ hops with 6.4% alpha acid and 4.8% beta acid. Phytochemical traits differs due to *terrior*.

## **Hop Taxonomy**

Many names were ascribed to the common hop plant, while also debated are the number of true species. For example, the ancient Latin name for the hop plant was *Lupus salictarius*—willow wolf (Gledhill 2002). However, Linnaeus (1753(2):1028) gave the hop plant the name, *Humulus*, referring to humus and the plant's soil-hugging habit if not supported (Marafioti 1970), and *lupulus*, referring to the wolf that likes to climb on top of others. This plant was the European hop, *H. lupulus* ( $2n=20$ ), with a perennial rootstock and annual climbing vines that die back each winter season. After 1753, the following names were used in opposition to that of Linnaeus: *Lupulus humulus* Mill. (1768), *Cannabis Lumulus* (L.) Scopoli (1772:263), *Lupulus scandens* Lam. (1778(2):217), *L. communis* Gaertn. (1788:358), *L. amarus* Gilib. (1792(2):451-452), *H. volubilis* Salisb. (1796:176), *H. vulgaris* Gilib. (1798(1):373), and *H. aculeatus* Nutt. (1848b:182).

Later on during the mid-1800s, Siebold and Zuccarini (1846:213) described an annual, Asian hop, *H. japonicas* ( $2n=17$  in males;  $2n=16$  in females). Nuttall (1848a:23 and 1848b:181-182) observed differences between European and American hop populations and named the latter, *H. americanus*. Miquel (1866:133) described *H. cordifolius* separating it based on the characters lobed versus unlobed leaves. However, Gray (1867) retained only the single species name, *H. lupulus*, for observed northern United States variations. Maximowicz (1879:489) proposed *H. lupulus* var. *cordifolius* based on a form with the upper and middle leaves entire, but the lower leaves divided. De Candolle (1883) later included three species in the genus, *H. lupulus*, *H. cordifolius*, and *H. japonicas*, and he did not question the validity of *H. cordifolius*. Roemer (1892) named *H. japonicas* var. *variegates*. Nelson and Cockerell (1903:45) proposed the variety *H. lupulus* var. *neomexicanus*, and Rydberg (1917:208) raised the rank of var. *neomexicanus* to species. Meanwhile, Zapalowicz (1908:94) named var. *brachystachyus*. Then, Merrill (1935:138) renamed *H. japonicas* to *H. scandens* after Loureiro's (1790(2):617) description of *Antidesma scandens*. Moldenke (1935:169) named *H. scandens* var. *variegates*. In continental Asia, Hu (1936) first named the putative perennial and endemic *H. yunnanensis* ( $2n$ =unknown). Davis (1957) observed unusually pubescent leaves of plants from the American Midwest but conserved *H. lupulus* to be

applied to the wild perennial hop plants from Europe and North America, while still recognizing *H. japonicus*.

In the most widely accepted taxonomy, Small (1978, 1980, 1981) kept the names *H. japonicas* Siebold & Zuccarini and *H. yunnanensis* Hu, then divided the species *H. lupulus* into five varieties:

- 1) *H. lupulus* var. *cordifolius* (Miguel) Maximowicz confined to eastern Asia and mainly Japan;
- 2) *H. lupulus* var. *neomexicanus* Nelson and Cockerell of the western American Cordillera;
- 3) *H. lupulus* var. *pubescens* E. Small of the Midwestern United States;
- 4) *H. lupulus* var. *lupuloides* E. Small of central and eastern North America; and
- 5) *H. lupulus* var. *lupulus* introduced from Europe to North America for brewing, as an ornamental, and has been grown elsewhere around the globe.

Shortly after Small's widely accepted nomenclature, *Humulus lupulus* ssp. *americanus* (Nutt.) Löve and Löve (1982) was named based on *Humulus americanus* (syn: *H. lupulus* var. *lupuloides*) and chromosome counts ( $2n=20$ ) of a specimen from Manitoba, Canada. An important note to be made is the chromosome counts are  $2n=20$  for the described varieties of *H. lupulus* by Small (1978) and the subspecies *americanus* by Löve and Löve (1982), with some rare exceptions that are triploid ( $2n=30$ ) or tetraploid ( $2n=40$ ).

Recent name changes by Grudzinskaja (1988) created *Humulopsis scandens* (Lour.) Grudz. after Loureiro's (1790) *A. scandens* description. Fu (1992) named a new variety *Humulus lupulus* var. *fengxianensis* that was putatively included in the Flora of the Republic Popularis Sinica (FRPS, Flora of China, Chinese edition) (Siushih 1998:220-224). However, a legitimate record of the publication for the variety *fengxianensis* has yet to be found and therefore invalid (nom. nud.). In 1994, the first volume of the revised English edition of the Flora of China (FOC) was published (Ma and Clemants 2006). In the FOC, Wu et al. (2003, Bartholomew personal communication) recognized three species: *H. lupulus*, *H. scandens*, and *H. yunnensis*, although no chromosome count or other information on genetic diversity, phylogenetic, or ecological relationship was given for the latter taxon. Wu et al. (2003) also commented

that, “A number of different varieties is usually recognized (Small, Syst. Bot. 3: 37-76. 1978), with the populations in China represented by at least by both *H. lupulus* var. *lupulus*, a predominantly European to C and SW Asian variety, and var. *cordifolius* (Miquel) Maximowicz, a predominantly Japanese variety. In addition, the populations in S Gansu and N Sichuan may possibly be another, distinct variety. Further study is needed to sort out the pattern of varietal occurrences in China, which is complicated by the introduction and escape of the cultivated var. *lupulus* for commercial production of hops for beer.” Not only is further research on the pattern of varietal occurrences of *H. lupulus* in China warranted, the collection and preservation of living germplasm of the putative endemic, *H. yunnanensis*, would be a conservation priority if it is found to be a truly distinct taxonomic entity. Tables 1.3 and 1.4 summarize the accepted taxonomy for *Humulus* according to Tropicos.org (2013) and The Plant List (2013), respectively.

**Table 1.3.** Taxonomy for *Humulus* species according Tropicos.org with accepted names and authorities.

Accepted name	Synonym (S) or Basionym (B)
<i>Humulus americanus</i> Nutt. 1848	
<i>Humulus cordifolius</i> Miq. 1866	(B) <i>Humulus lupulus</i> var. <i>cordifolius</i> Maxim. 1879
<i>Humulus lupulus</i> L. 1753 var. <i>lupulus</i>	
<i>Humulus lupulus</i> var. <i>fengxianensis</i> J.Q. Fu 1992	
<i>Humulus lupulus</i> var. <i>lupuloides</i> E. Small 1978	
<i>Humulus lupulus</i> var. <i>pubescens</i> E. Small 1978	
<i>Humulus neomexicanus</i> (A. Nelson & Cockerell) Rydb. 1917	(B) <i>Humulus lupulus</i> var. <i>neomexicanus</i> A. Nelson & Cockerell 1903
<i>Humulus scandens</i> (Lour.) Merr. 1935	(S) <i>Antidesma scandens</i> Lour. 1790; (S) <i>Humulopsis scandens</i> (Lour.) Grudz. 1988; (S) <i>Humulus japonicus</i> Siebold & Zucc. 1846
<i>Humulus scandens</i> var. <i>variegatus</i> (Siebold & Zucc.) Moldenke 1935	(B) <i>Humulus japonicus</i> Siebold & Zucc. 1846
<i>Humulus japonicus</i> var. <i>variegatus</i> F. Roem. 1892	
<i>Humulus yunnanensis</i> Hu 1936	

**Table 1.4.** Taxonomy for *Humulus* species according The Plant List, Version 1.1 released September 2013. Accepted names are in bold font. Dates for the accession are 2012-03-23, except for *H. lupulus* var. *lupulus* on 2012-04-18.

<b>Name</b>	<b>Status</b>	<b>Confidence</b>	<b>Source</b>
<i>Humulus aculeatus</i> Nutt.	Synonym	2/3	WCSP (in review)
<i>Humulus americanus</i> Nutt.	Synonym	2/3	WCSP (in review)
<i>Humulus cordifolius</i> Miq.	Synonym	2/3	WCSP (in review)
<i>Humulus japonicus</i> Siebold & Zucc.	Synonym	2/3	WCSP (in review)
<i>Humulus japonicus</i> var. <i>variegatus</i> F.Roem.	Synonym	1/3	WCSP (in review)
<b><i>Humulus lupulus</i> L.</b>	Accepted	2/3	WCSP (in review)
<i>Humulus lupulus</i> subsp. <i>americanus</i> (Nutt.) Á.Löve & D.Löve	Synonym	1/3	WCSP (in review)
<i>Humulus lupulus</i> var. <i>cordifolius</i> (Miq.) Maxim. ex Franch. & Sav.	Synonym	1/3	WCSP (in review)
<b><i>Humulus lupulus</i> var. <i>lupuloides</i> E.Small</b>	Accepted	1/3	WCSP (in review)
<i>Humulus lupulus</i> var. <i>lupulus</i>	Synonym	1/3	TRO
<b><i>Humulus lupulus</i> var. <i>neomexicanus</i> A.Nelson &amp; Cockerell</b>	Accepted	1/3	WCSP (in review)
<i>Humulus neomexicanus</i> (A.Nelson & Cockerell) Rydb.	Synonym	2/3	WCSP (in review)
<b><i>Humulus scandens</i> (Lour.) Merr.</b>	Accepted	2/3	WCSP (in review)
<i>Humulus scandens</i> var. <i>variegatus</i> (Siebold & Zucc.) Moldenke	Synonym	1/3	WCSP (in review)
<i>Humulus volubilis</i> Salisb. [Illegitimate]	Synonym	2/3	WCSP (in review)
<i>Humulus vulgaris</i> Gilib.	Synonym	2/3	WCSP (in review)
<b><i>Humulus yunnanensis</i> Hu</b>	Accepted	2/3	WCSP (in review)

### **Fossil and Phylogenic Roots of the Hop Branch on the Family Tree**

The Russian plant geneticists Vavilov (1992) theorized the historical geographic location with the most number of species/varieties of cultivated plants within a given group (e.g., barely, corn, and wheat) is likely to be recognized as the center of origin for that group of cultivated plants. Accordingly then, China is hypothesized as the likely place of origin for *Humulus* because three species of hop plants are recognized in the FOC (Neve 1991; Murakami et al. 2006a; Small 1978; Wu et al. 2003). In addition, molecular dating methods suggest the time of divergence between *H. lupulus* and *H. scandens* was approximately 6.38 mya based on noncoding cpDNA regions. Moreover, wild *H. lupulus* populations in Europe and others in China, Japan, and North America diverged approximately 1.05-1.27 million years ago, and the differentiation within the latter occurred approximately 0.46-0.69 million years ago (Murakami et al. 2006a). However, Murakami et al. (2006a and 2006b) has ambiguous origins of hop plants sampled from China and does not include *H. yunnanensis* in the analyses, although this research did find high genetic diversity in North American hops considered to have migrated from Asia. Alternative hypotheses for the geographic origin of *Humulus* besides China include: 1) Russia by Linneaus and Vavilov and 2) North America (see Chapter 5 in this dissertation).

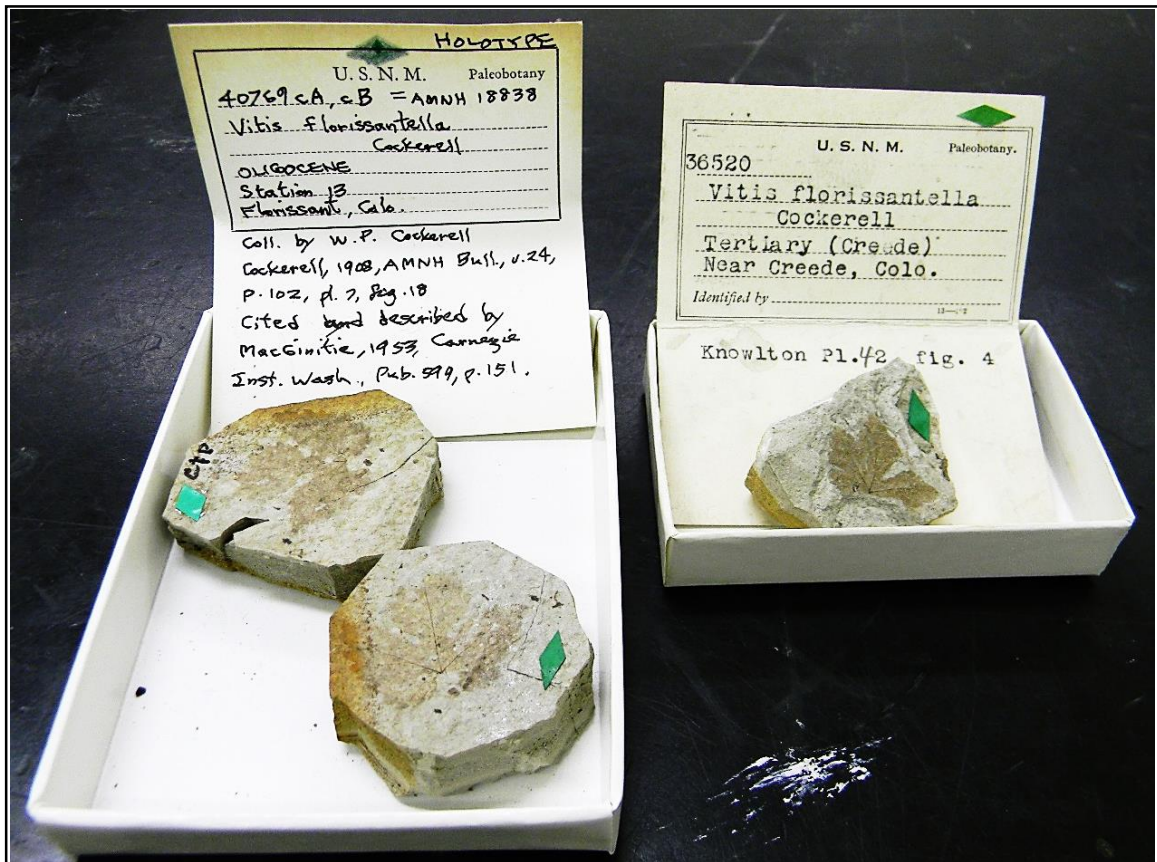
Support for the North American hypothesis is drawn from the macrofossil record for *Humulus* that dates at the K/T event in North Dakota (70.6 - 65.5 mya; Johnson 2002) and the Eocene in Colorado (37.2 - 33.9 mya; Fig. 1.7; MacGinitie 1953). Other macrofossils of *Humulus* were found from the Miocene in the Russian Federation (11.6 - 5.3 mya; Dorofeev 1963), and the Pliocene in Germany (3.6 - 2.6 mya; Mai and Walther 1988) (for further information see Paleobiology Database 2013; also see Collinson 1989 and Weber 2003). In regards to microfossils, *Humulus* pollen can be confused with *Cannabis* pollen (Fig. 1.8; Clarke and Merlin 2013), but sufficient evidence of *Humulus* pollen from Middle to East Asia dates from approximately 20 mya through the present (Jiang and Ding 2008, Jiang et al. 2011, Kou et al. 2006, Xu et al. 2008, Xu et al. 2012). As depicted from the macro- and microfossil record, *Humulus* or the now extinct members of the Cannabaceae (s.s.) could have migrated into Eastern Asia from natural

flood plains and riparian zones along the uplifting mountain ranges in western North America that began before the Jurassic Period (Fig. 1.9). Once in Eurasia, the early ancestors of the Cannabaceae (*s.s.*) may have radiated into the extant genera, *Cannabis* ( $2n=20$ ) and *Humulus*, with a loss of chromosomes in the latter associated with a shift to annual habit. Further support for this hypothesis is seen in overlapping geographic distributions of the recognized species in the sister taxa, *Cannabis* and *Humulus*, which follows:

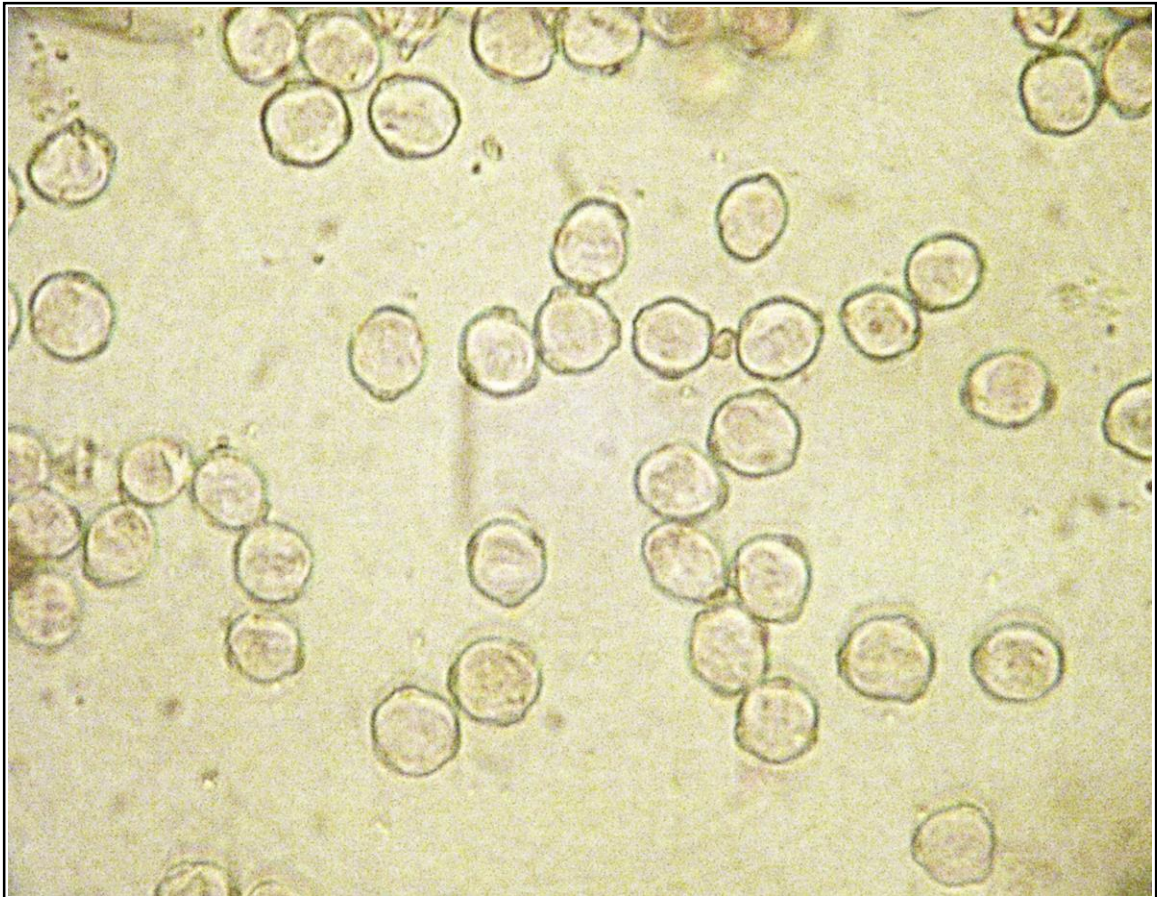
- 1) *H. lupulus* and *C. sativa* are originally distributed across the northern temperate zone;
- 2) *H. yunnanensis* and *C. indica* are originally localized near the Himalayan mountains; and
- 3) *H. scandens* and *C. ruderalis* are cosmopolitan weeds arising *in-situ*.

Due to the similar morphological traits exhibited in the Cannabaceae (*s.s.*) and the conundrum of their phylogenetic placement within the Urticalean rosids (Figs. 1.2 and 1.3; specifically see Judd et al. 1994), the origin of extant *Cannabis* was not determined until genetic evidence indicated Central Asia along the steppe plains (Clarke and Merlin 2013, Hillig 2005). For the broader evolution of the Cannabaceae (*s.s.*), I hypothesize that grasslands of Eurasia promoted the erect annual habit seen in *Cannabis*, and the twining perennial habit of *Humulus* was favored along meandering rivers in the mountains of Western North America soon after the K/T extinction event (Fig. 1.9). Whether an erect annual habit or twining perennial habit evolved first is debatable (Clarke and Merlin 2013). Nonetheless, this major extinction event that is commonly known to have wiped out the dinosaurs subsequently allowed for the rise of angiosperms and mammals, particularly fruits and primates (Ni et al. 2013, Perelman et al. 2011, Williams et al. 2010, also see early treeshrews (Scandentia), colugos (Dermoptera), and Plesiadapiformes). Also, this extinction event opened new niches to allow the speciation of many members in the rosid clade, which contains more than one-fourth of all angiosperms (~70,000) and includes most of the extant temperate and tropical forests lineages (Wang et al. 2009). The rapid radiation of the rosid clade corresponds with the rapid rise of angiosperm dominated forests approximately 108-83 mya, as well as the rapid rise of those clades

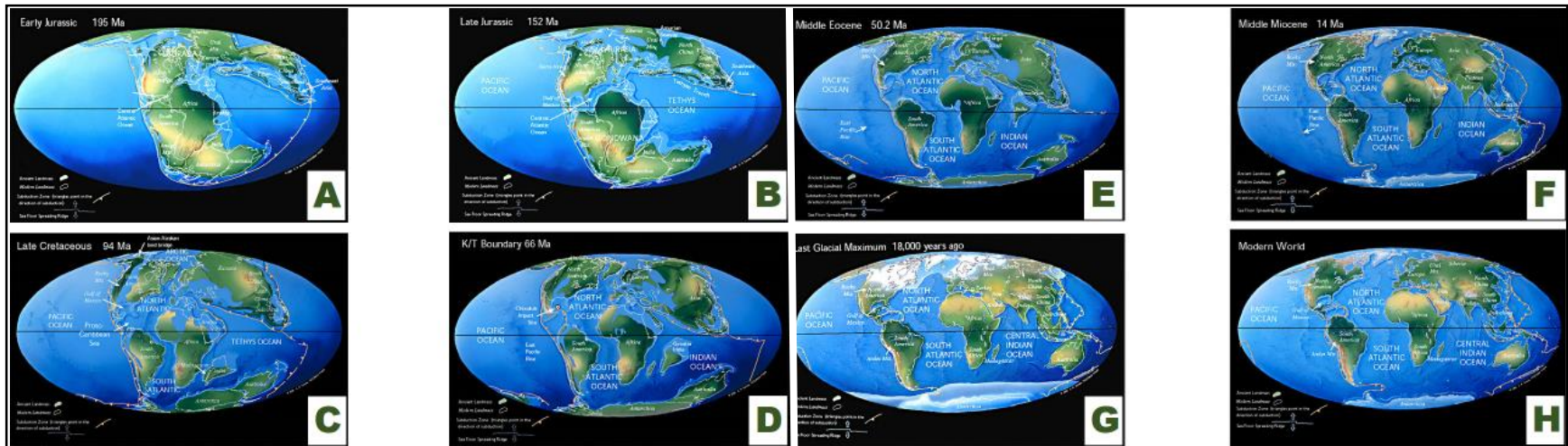
affiliated with habitation in these forests (i.e., amphibians, ants, placental mammals, and ferns) (Wang et al. 2009). Even with the lack of older fossils from Eurasia, Europe has many subfossils and is the place where *Humulus* was harvested from the wild before domestication approximately 1,200 years ago for brewing (Behre 1999, also see Heiser 1990 for an outline of the major hypothesis for the “invention” of agriculture).



**Fig. 1.7.** Macrofossils of *Humulus* from the Eocene of Colorado (37.2 - 33.9 mya; MacGinitie 1953 and 1969, Manchester 2001).



**Fig. 1.8.** *Humulus lupulus* pollen from a plant germinated on Oahu in 2007. The plant is still alive in 2014 and is in flower; however, the plant did not produced flowers from 2011-2013 (see Clarke and Merlin (2013), Fleming and Clarke (1998), and Whittington and Gordon (1987) for pollen identification of *Humulus* and *Cannabis*).



**Fig. 1.9.** Relative continental positions of the Early Jurassic (A), Late Jurassic (B), Late Cretaceous (C), K/T Boundary (D), Middle Eocene (E), Middle Miocene (F), Last Glacial Maximum (G), and the present day (H). The Cannabaceae (*s.s*) may have arose at the time of the K/T event in the New World close to the Rocky Mountains (D and E). Weber (2003) states, “Whatever the varying opinions may be at the present time concerning generic limits and alleged monophyletic or polyphyletic groups, it is most useful phytogeographically to discuss distribution patterns with reference to ‘small’ genera, or what may be considered by the dominant group of taxonomists as subgenera. Often, subgenera are morphologically quite distinct and have had different biohistories. Subgenera often are the small groups exhibiting great disjunctions and most likely represent monophyletic lines.” Specifically, Weber (2003) classifies *Humulus* within the Moraceae and concludes, “*H. lupulus* is oroboreal. A Tertiary relictual distribution pattern as evidenced by the fossil record of the species in the Florissant beds of Colorado (MacGinitie, 1953) and its present localization mostly along the eastern edge of the front Range. MacGinitie’s plate is misidentified as a species of *Vitis*.” Maps are from Scotese (2002).

### **Hypothesizing Phylogenic Roots of the Hop Branch on the Family Tree**

To support the genetic branches and roots on the Cannabaceae family tree, Maximum Likelihood (ML) and Bayesian inference (BI) test the hypothesized evolutionary relationships (Holder and Lewis 2003; Lemey et al. 2009; Hall 2011). Both ML and BI use a prior model of sequence evolution; a hypothesis that describes the relative probability of various genetic events (i.e., chance of a transition relative to the chance of a transversion for each nucleotide) (Holder and Lewis 2003). ML corrects for multiple mutational events at the same site, judges how well a hypothesis predicts the observed data by assessing confidence with bootstrapping the original data (i.e., random resampling), and results in a phylogenetic tree with the highest probability of producing the observed sequences shown with percentages that support the observed groups (for an example see Fig. 1.2). BI focuses on the posterior probability of hypotheses, which is proportional to the product of the prior probability and the likelihood; thus, BI has a strong connection to the ML method but may be faster than bootstrapping. However, the prior distributions for the parameters in BI must be specific, and it can be difficult to determine whether the Markov Chain Monte Carlo (MCMC) approximation has run for an enough number of generations (Holder and Lewis 2003). Depending on the length and number of sequences tested, the major obstacle of ML and BI is the burden of having a computer run the algorithms for hours, days, or sometimes weeks to month. Accurately reconstructing the hypothesized relationships between aligned homologues sequences that have been separated for a long time or are evolving rapidly requires a comparison of both ML and BI phylogenies.

## **Main Research Hypotheses**

The main hypotheses in this dissertation include:

Chapter 2) H<sub>2.0</sub> = Due to human actions, New World varieties of *Humulus lupulus* are not readily distinguishable from the cultivated *H. lupulus* var. *lupulus* (Davis 1957). H<sub>2.1</sub> = The New World varieties of *Humulus* are clearly distinguishable using forensic, DNA barcode regions.

[see Rediscovering Wild Michigan Hops (*Humulus lupulus* L.)].

Chapter 3) H<sub>3.0</sub> = The Yunnan hop is not a separate species but a variety of the common hop, *H. lupulus*. H<sub>3.1</sub> = The Yunnan hop is a distinct species (Small 1978).

[see DNA Support for the Endemic Yunnan Hop (*Humulus yunnanensis* Hu)].

Chapter 4) H<sub>4.0</sub> = Due to the highly conserved nature of the plastome, little genomic difference will be observed between the Cannabaceae (*s.s.*), other Urticalean rosids, and eudicots. H<sub>4.1</sub> = The entire plastome as a DNA barcode is ideal for genomic studies within the highly derived Cannabaceae (*s.s.*) compared to plastomes from the other Urticalean rosids and eudicots.

[see Draft Chloroplast Genome of a Wild American Hop (*Humulus lupulus* var. *neomexicanus* A. Nelson & Cockerell)].

Chapter 5) H<sub>5.0</sub> = Phylogenomic methods support the genus *Humulus* originated in the Old World (i.e., China) and migrated to the New World (Murakami et al. 2006a, Neve 1991). H<sub>5.1</sub> = After the K/T extinction event around 65 million years ago, open niches along riparian areas in the New World mountains (i.e., Colorado Rockies) allowed the perennial binning habit of *Humulus* to thrive in forest ecotones and subsequently migrate to the Old World, where the genus underwent adaptive radiation into an two additional annual species that are putatively not interfertile within the *Humulus lupulus* complex.

[see The Origin of *Humulus*: A Phylogenomic Surfing Approach].

Species concepts adopted by various authors eventually requires further revision of new and old taxa by extensive sampling of herbarium specimens for genetic material to clarify pertinent information that has become lost in taxonomic translation. Furthermore, wild hop populations around the world could provide novel traits worth millions (if not billions) of dollars annually to the brewing and pharmaceutical industries, if their distribution, taxonomic, and phylogenetic relationships are understood. New genetic variations could also lead to new flavors in ales and lagers, affecting the economies of many: the individual hops grower, local and national hops distributors, brewers and beverage distributors, hotel and restaurant owners, patrons, and food producers whose products are consumed with flavored beverages (Haglund 2013, Karabín et al. 2013). Although, this dissertation research addresses the myriad applications and benefits of hops used in the brewing and distillation sciences, a more narrow focus is on *Humulus* botanical research.

The broader perspectives of this dissertation research are in evolutionary biology, crop conservation, and phylogenetics, as well as advancing the theory of the origin of the hop plant by using the newest available DNA sequencing technology and phylogenomic approaches. A major purpose of this dissertation is to elucidate the taxonomic and evolutionary relationship of *Humulus* in the Cannabaceae (s.s.), as well as assist research efforts on other economic plants in the Urticalean rosids. *Humulus* relationships were studied through field and laboratory observations, the study of herbaria material (including significant historical specimens), and DNA sequence analyses.

This dissertation concludes with outcomes, applications, and directions for Cannabaceae research, emphasizing the importance of crop wild relatives, herbaria specimens for genomic resources, and conserving biodiversity with changes in climate. Lastly, unconventional perspectives for genome education, methods outlined to conduct and replicate research, and the knowledge disseminated to the public (e.g., surfing a genome) is anticipated to advance both the scientific theory and the equipment used for DNA sequencing and phylogenomic approaches. The future generations of plant biologists and botanists will have exceptional hardware and software literally at their fingertips, which is now just beginning to be accessible on remote servers and as

miniaturized devices that fit into a pocket. Whether the direction is for medicinal plants or for personalize medicine, an –omics perspective for *Humulus* as a model organism is presented.

## CHAPTER 2

### Rediscovering Wild Michigan Hops (*Humulus lupulus* L.)

#### **Abstract**

Michigan farmers recently reintroduced the cultivated hop plant (*Humulus lupulus* L., Cannabaceae) aimed at market security for the state's growing microbrewery industry. These farmers, as well as local brewers, are interested in developing new beverage flavors and new cultivars using native germplasm. Wild hop plants can be found along riparian zones and disturbed sites, while putatively wild plants can also be remnant heirloom cultivars now growing on fallow lands. Hundreds of years of introgression between cultivated and wild hop plants has caused problems with the morphological identification of putatively wild plants. An alternative approach using genetic tests can confirm a plant's origin. A collection of DNA from cultivars, putatively wild plants, and herbarium specimens was tested using short DNA sequences (barcodes) and phylogenetic methods to determine if the putatively wild hop plants are native Michigan germplasm or escaped European cultivars. Results show little genetic distance within the cultivated and wild plants sampled; however, two clades formed within *H. lupulus*: 1) Old World pedigrees and 2) New World pedigrees. Accordingly, Michigan does contain wild hop plants, and future research must include a comprehensive survey and collecting expedition throughout the state's two peninsulas for native North American hop plants.

#### **Keywords**

hops, cpDNA, *petL-psbE*, rDNA, ITS2, DNA barcodes, phylogenetics, herbarium specimen

## **Introduction**

### ***Background of the Hop Plant and Significance of Michigan Hop Cultivation***

Hop plants (*Humulus* L., Cannabaceae) are clockwise-twining, herbaceous vines with male and female flowers found on separate plants (Fig. 2.1; Small 1978, Hampton et al. 2001). The female flowers are commonly called hops. Indigenous to the Northern Hemisphere in both North America and Eurasia, wild hops are found along streams, river bottoms, forest margins, and disturbed sites. Cultivated hops are also widely distributed to suitable climates around the world as an important crop for the brewing industry.

Traditionally, Europeans and their colonies used wild hops to flavor beverages before the plant was brought into cultivation (Behre 1999, Rösch 2008). Also, wild hops were used in food and medicine preparations prior to flavoring beverages (Moerman 1998, 2009, 2010, Zanoli and Zavatti 2008). For the purpose of brewing, the cultivated hop plant (*H. lupulus* var. *lupulus*) was introduced to the United States from England in 1629 (Burgess 1964, Carter et al. 1990, Neve 1991, Small 1978). Today, the successful cultivation of hops is the basis of the global multi-million/billion dollar brewing industry (Prescott-Allen and Prescott-Allen 1986). Cultivated hops are produced on a limited scale in the Upper Midwest for local markets (Carter et al. 1990, Sirrine 2010, Tomlan 1992, Turner et al. 2011). To encourage market security for small scale microbreweries, the cultivation of the hop plant has expanded to other suitable regions, such as Michigan (Sirrine 2010). This expansion may lead to the development of new flavors in ales and lagers (Takoi et al. 2009).

Due to crop security and sustainability issues that address national safety and security, farmers in Michigan cannot keep growing the same asexually propagated plants from the National Clean Plant Network in Washington. Local brewers and farmers are also interested in developing new hops varieties using novel Michigan germplasm (Boutain personal observation). For example, the hop cultivar ‘Cascade’ grown in Michigan is, in essence, a new horticultural variety because it tastes different due to the *terroir* (i.e., unique geography, geology, and climate of a location interacting with a plant’s genetics – “a sense of place”) (also see Fig. 1.6 in this dissertation). Furthermore, the wild hop plants found in North America (*H. lupulus* var. *lupuloides*, *H. lupulus* var.

*pubescens*, and *H. lupulus* var. *neomexicanus*) have undergone introgression with cultivated hop plants for hundreds of years, causing a combination of overlapping morphological traits in the known botanical varieties. As a result, the morphological identification of putatively wild Michigan hop plants and heirloom cultivars that have escaped to fallow lands is problematic. Therefore, the genetic confirmation of the origins of putative wild Michigan hops can lead to plant selection and contribute to breeding new varieties of hops. To date, there has not been an extensive collecting expedition for hop germplasm in the State of Michigan, so hop plants with known origins (i.e., cultivars) can be compared to those that are unknown and putatively wild.

### ***DNA barcoding***

A DNA barcoding approach is a method to quickly identify a species by extracting standardized short sequences of DNA typically between 400-800 base pairs long. Typical plant DNA exons (e.g., chloroplast *rbcL* and *matK*, and nuclear 18S) are not always helpful in barcoding because of lack of variation (Kress and Erickson 2012); therefore, more variable spacer regions (e.g., introns; *trnH-psbA* or *petL-psbE*) are included. Specific ITS2 primers for forensic barcoding of *Cannabis* L. are easily amplified for *Humulus* (Gigliano 1998, Gigliano et al. 1997, Murakami 2000). Plastid spacer regions with a larger number of potentially informative characters must be selected because of phylogenetic affinities within the Cannabaceae (Shaw et al. 2005 and 2007, Sytsma et al. 2002, Yang et al. 2013).

### ***Herbarium Specimens***

DNA from herbarium specimens is widely used for phylogenetic studies, while Lavoie (2013) points out the use of molecular analyses to investigate herbarium specimens is still relatively unexplored from ancient DNA, biogeographical and environmental points of view. In regards to historical herbarium samples, if a plant collector's sampling is biased to the most colorful flowers on non-abrasive plants (Schmidt-Lebuhn et al. 2013, Wolf et al. 2011), then studies of morphological differences alone, not coupled with DNA studies, can skew the taxonomical interpretation of species. Schmidt-Lebuhn et al. (2013) in their study of the Asteraceae found green and brown inflorescences were under collected, along with spiny plants collected only about half as

often as expected. Consequently, potential collecting bias may exist for *Humulus*, which has green flowers on a plant with barbed-trichomes that climbs high into the forest canopy. Such bias potentially leads to samples ‘lost in taxonomic translation,’ and simple methods of molecular biology on herbarium specimens can solve such conundrums. Small (1978) documented geographical and morphological units in an extensive study of *Humulus* herbarium specimens; therefore, DNA studies on both *Humulus* and *Cannabis* from herbarium material could be successful to support the operational taxonomic units.

### **Objectives**

This project had three main objectives:

- 1) Collection of cultivated and putatively wild hop plants in Michigan,
- 2) Destructive sampling of herbarium specimens for DNA analysis, and
- 3) Testing short DNA sequences to determine if putatively wild hop plants are native germplasm and not escaped European cultivars of *Humulus lupulus* var. *lupulus*.

### **Hypothesis**

H<sub>2.0</sub> = Due to human actions, New World varieties of *Humulus lupulus* are not readily distinguishable from the cultivated *H. lupulus* var. *lupulus* (Davis 1957).

H<sub>2.1</sub> = The New World varieties of *Humulus* are clearly distinguishable using forensic, DNA barcode regions.



**Fig. 2.1.** Photo A shows the female flowers, hops, of the triploid cultivar Columbia grown on a low trellis in Southeast Michigan. Photo B shows a putatively wild hop plant growing in Northwest Michigan.

## **Materials and Methods**

Fresh plant material (leaves/flowers) collected in the field, cultivated hop yards, purchased online, or samples sent by collaborators were preserved in silica desiccant for later DNA extraction. Pressed voucher specimens were also prepared for collection and deposited in HAW. Additionally, herbarium specimens were destructively sampled to produce DNA samples (Table 2.1). Putative wild hop samples comprised 19 collections from Michigan that included one originally from the wild in Idaho, as well as one wild collection from Canada.

<b>Table 2.1.</b> Plant specimens sampled for DNA. Samples were taken from leaves unless noted.				
<b>Sample (H=Humulus)</b>	<b>DNA#</b>	<b>Gender</b>	<b>Locality</b>	<b>Collection ; Date ; Herbarium ; Note</b>
Cannabis_sp	152	No flowers	Lawai Valley, Koloa District, Kauai, Hawaii, U.S.A	D.H. Lorence #7204 and G. Nace ; 14 May 1992 ; PTBG#015426 and PTBG#021494
H_lupulus_Hallertau	3	Female	Hops pellets purchased from Adventures in Homebrewing, Taylor, Michigan, U.S.A.	JRBoutain #303
H_lupulus_StMary	12	Female	St. Mary's School, Lake Leelanau, Michigan, U.S.A.	JRBoutain #304
H_lupulus	13	Female	West of Zarrentin, Götting. Schleswig-Holstein, Germany	L. Holm-Nielsen and K. Larsen #285 ; 25.8.1976 ; BRIT/AAU
H_lupulus_cf_lupuloides	14	Female	Mogollon Mountains, New Mexico, U.S.A.	R.D. Worthington #7604 ; 22 Aug. 1981 ; BRIT
H_lupulus_lupuloides	15	Female	Rockingham County, North Carolina, U.S.A.	A.E. Radford #18614 ; 28 Sept. 1956 : BRIT/VDB #13063
H_lupulus_neomexicanus	16	Female	Gila National Forest, Mogollon, Catron County, New Mexico, U.S.A.	D. Demaree #48609 ; 7-25-1963 ; BRIT/SMU
H_lupulus_cf_pubescens	17	Female	Benton County, Arkansas, U.S.A	E.B. Smith #3856 ; 23 Sept. 1984 ; BRIT/VDB
H_lupulus_Nugget	21	Female	Sutton's Bay, Michigan, U.S.A.	JRBoutain #305

<b>Table 2.1.</b> continued.				
H_lupulus_Cascade	22	Female	Sutton's Bay, Michigan, U.S.A.	JRBoutain#306 ; leaf extracted
H_lupulus_Cascade	23	Female	Sutton's Bay, Michigan, U.S.A.	JRBoutain#306 ; flower extracted
H_lupulus_Waldy	24	cf. Female? ; No flowers when collected	Sutton's Bay, Michigan, U.S.A.	Michigan State Agricultural Extension Unit ; Waldy is a putative wild hop from Idaho that is approximately 80 years old. No herbarium specimen was collected.
H_lupulus_Bling	25	cf. Female? ; No flowers when collected	Sutton's Bay, Michigan, U.S.A.	Michigan State Agricultural Extension Unit. Bling is a putative Michigan wild hop. No herbarium specimen collected.
H_lupulus_UnknownX	26	Female	Michigan; N 44°57.598, W 85°29.648	JRBoutain#307
H_lupulus_UnknownY	27	Female	Michigan; N 44°57.598, W 85°29.648	JRBoutain#308
H_lupulus_UnknownZ	28	Female	Michigan; N 44°57.598, W 85°29.648	JRBoutain#309
H_lupulus_Claire	29	Female	Riverview, Michigan, U.S.A.	JRBoutain#310
H_lupulus_TCBohemia	30	Female	Traverse City, Michigan, U.S.A	JRBoutain#311
H_lupulus_Perle	32	Female	Michigan; N 44°53'4", W 85°40'45"	JRBoutain#312
H_lupulus_Williamette	33	Female	Michigan; N 44°53'4", W 85°40'45"	JRBoutain#313

<b>Table 2.1.</b> continued.				
H_lupulus_Goldings	34	Female	Michigan; N 44°53'4", W 85°40'45"	JRBoutain#314
H_lupulus_Tettanger	35	Female	Michigan; N 44°53'4", W 85°40'45"	JRBoutain#315
H_lupulus_Centennial	36	Female	Michigan; N 44°53'4", W 85°40'45"	JRBoutain#316
H_lupulus_Cascade	37	Female	Michigan; N 44°53'4", W 85°40'45"	JRBoutain#317
H_lupulus_BrewersGold	38	Female	Michigan; N 44°53'4", W 85°40'45"	JRBoutain#318
H_lupulus_Recycler	30	cf. Female? ; No flowers when collected	Michigan; N 44°53'4", W 85°40'45"	JRBoutain#319
H_lupulus_Galena	40	Female	Michigan; N 45°4'9", W 85°36'0"	JRBoutain#320
H_lupulus_Chinook	41	Female	Michigan; N 45°4'9", W 85°36'0"	JRBoutain#321
H_lupulus_lupuloides	42	Female	Southern Ottawa, Canada	JRBoutain#322; Sent by E. Small from clone of original wild population on bank of Black Rapids Creek at Merivale Road.
H_lupulus_lupulus	43	Female	Eastern Ottawa, Canada	JRBoutain#323; Sent by E. Small from clone of old brewery cultivar growing by roadside (now ruderal). Unknown cultivar is originating and persisting from nearby mid- 20 <sup>th</sup> century brewery hop plantation.

**Table 2.1.** continued.

H_lupulus_SorachiAce	44	Female	Whole leaf hops purchased from Freshops, Oregon, U.S.A	JRBoutain#324
H_lupulus_Saaz	45	Female	Whole leaf hops purchased from Freshops, Oregon, U.S.A	JRBoutain#325
H_lupulus_MtHood	46	Female	Whole leaf hops purchased from Freshops, Oregon, U.S.A	JRBoutain#326
H_lupulus	47	Male	Osage County, Kansas, U.S.A.	G. Tucker #6185 ; 6 Aug. 1967 ; BRIT/SMU
H_scandens	48	Female on sheet sampled	Peoria County, Illinois, U.S.A.	S.R. Hill #28569 with T. Kompare and P. Tessene ; 28 Aug. 1996 ; BRIT
H_scandens	50	Female	Richardson County, Nebraska, U.S.A.	R.B. Kaul #7493 and S. Rolfsmeier ; 28 Sept. 1996 ; BRIT
H_scandens	51	No flowers	Pendleton County, Kentucky, U.S.A.	M. Whitson 2007-0013 3/5 with L. Trauth and A. Tullis ; 28 July 2007 ; BRIT
H_scandens	52	Female	Clarion, Wright County, Iowa, U.S.A.	J. Ross and A.P. Bowman ; 16 Oct. 1944 ; BRIT/SMU
H_lupulus_Unknown1	54	No flowers	Michigan, U.S.A.	JRBoutain#327
H_lupulus_Unknown2	55	Female	Michigan, U.S.A.	JRBoutain#328
H_lupulus_Unknown3	56	Female	Michigan, U.S.A.	JRBoutain#329
H_lupulus_Unknown4	57	Female	Michigan, U.S.A.	JRBoutain#330
H_lupulus_Unknown5	58	Female	Michigan, U.S.A.	JRBoutain#331

---

**Table 2.1.** continued.

H_lupulus_Unknown6	59	Female	Michigan, U.S.A.	JRBoutain#332
H_lupulus_Unknown7	60	Female	Michigan, U.S.A.	JRBoutain#333
H_lupulus_Unknown8	61	Female	Michigan, U.S.A.	JRBoutain#334
H_lupulus_Unknown9	62	Female	Michigan, U.S.A.	JRBoutain#335
H_lupulus_Unknown10	63	Female	Michigan, U.S.A.	JRBoutain#336
H_scandens	67	Female	Dutchess County, New York, U.S.A.	S.R. Hill #28479 ; 13 Aug. 1996 ; BRIT
H_scandens	76	No flowers	Calhoun County, Alabama (33°38'56" N, 85°49'55" W), U.S.A.	D.D. Spaulding #11,497 and E. "Skeeter" Cole, Jr. ; 31 July 2002 ; BRIT/VDB
H_lupulus_pubescens	81	Female	North of East Peoria, Tazewell County, Illinois, U.S.A.	V.H. Chase #14819 ; 2 Sept. 1959 ; BRIT/VDB #42841
H_lupulus_pubescens	82	Female	Plattsmouth, Cass County, Nebraska, U.S.A.	D. Demaree #54175 ; 8-3-1966, BRIT/VDB #80336
H_lupulus_pubescens	83	Female	Plattsmouth, Cass County, Nebraska, U.S.A.	D. Demaree #54175 ; 8-30-1966 ; BRIT/VDB #59781
H_lupulus_pubescens	84	Female	Cass County, Missouri, U.S.A	N.C. Henderson #67-1565 ; 29 Aug. 1967 ; BRIT/VDB #51530
H_scandens	85	Male	Calhoun County, Alabama, (33°38'12" N, 85°49'58" W), U.S.A.	D.D. Spaulding #12,051 ; 18 Sept. 2003 ; BRIT/VDB

---

---

**Table 2.1.** continued.

H_scandens	89	Male	South Mountain Reservation, Millburn, Essex Co. County, New Jersey, U.S.A.	K. Barringer #10646 ; 19 Aug. 2003 ; BRIT/VDB
H_lupulus	90	Male	Williamette Park, Corvallis, Benton County, Oregon, U.S.A.	R.R. Halse #5436 ; 30 July 1998 ; BRIT/VDB
H_scandens	95	Female on sheet sampled	Jackson Park, Chicago, Illinois, U.S.A.	A.P. Anderson #2024 ; Aug. 1912 ; BRIT/SMU
H_lupulus	96	Female	Grobbendovk arboretum, Prov d'Anvers, Belgium	J.E. De Langhe ; Sept. 1964 ; BRIT/SMU
H_lupulus	97	Male	4 km north of Gnissau village and west of Ahrensbök municipality, Ostholstein in Schleswig-Holstein, Germany	L. Holm-Nielsen, I. Nielsen, S.P. Pinnerup #223 ; 9.8.1974 ; BRIT/SMU/AAU
H_scandens	104	Male	Arnold Arboretum, Harvard University, U.S.A.	G.P. DeWolf and P. Bruns #2179 ; 21 Sept. 1967 ; BRIT/SMU
H_scandens	121	Male	Brunswick, Chariton County, Missouri, U.S.A.	H.N. Moldenke #23160 ; 16 Oct. 1966 ; BRIT/SMU

---

---

<b>Table 2.1.</b> continued.				
H_lupulus_Saaz	153	Female	Whole leaf hops purchased from Adventures in Homebrewing, Taylor, Michigan, U.S.A.	JRBoutain#337

---

Approximately 0.03-0.05g of silica dried leaf, flower, or herbaria material was used from each sample ( $n=63$ ) for total genomic DNA extraction with a Qiagen DNeasy Plant Mini Kit (Qiagen, Valencia, California, USA) following the manufacture protocol with the only modification of 1.5 hours incubation in a 65°C water bath. Nuclear ribosomal DNA (rDNA) (ITS2; Murakami 2000, 2001, Murakami et al. 2003) and chloroplast DNA (cpDNA) (*petL-psbE*; Shaw et al. 2007) regions were amplified via a standard polymerase chain reaction (PCR) procedure modified from Keeley et al. (2007), Murakami (2000 and 2001), and Shaw et al. (2007) (Table 2.2).

**Table 2.2.** DNA regions and primer pairs used for PCR amplification and sequencing. The ITS2 primers are from Murakami (2000), and the *petL-psbE* primers are from Shaw et al. (2007).

Region	Primer Name	Sequence (5' to 3')
ITS2	ITS2F	TTGCAGAATCCCGTGAACCATCG
	ITS2R	CCAAACAACCCGACTCGTAGACAGC
<i>petL-psbE</i>	<i>petL</i>	AGTAGAAAACCGAAATAACTAGTTA
	<i>psbE</i>	TATCGAATACTGGTAATAATATCAGC

PCR amplification reactions were carried out in 25 µL of a PCR mixture consisting of 14 µL sterile water, 2.0 µL of 2.5 mM dNTP Mix (BIOLINE), 2.5 µL of 10 x NH<sub>4</sub> Reaction Buffer (BIOLINE), 1.25 µL of 50 mM MgCl<sub>2</sub> (BIOLINE), 0.5 µL of 10mg/mL Bovine Serum Albumin (Sigma), 1.0 µL of 10 µM concentration forward primer, 1.0 µL of 10 µM concentration reverse primer, 0.25 µL of 1u/µL BIOLASE Red DNA Polymerase (BIOLINE), and 2.5 µL of extracted template DNA. The DNA template was not standardized to a particular concentration after DNA extraction nor prior to PCR. The amount of template DNA was adjusted, if necessary, to generate sufficient PCR products for DNA sequencing. Positive and negative controls were used for every set of PCR reactions to ensure no contamination bias.

PCR amplification was carried out on a C1000 Thermal Cycler (BIO-RAD). Thermal cycler settings for ITS2 were an initial preheating at 94°C for 2 minutes, followed by 34 cycles of 1 minute denaturation at 94°C, 1 minute annealing at 65°C, and primer extension for 2 minutes at 72°C, ending with an additional 7 minute extension at 72°C for the completion of unfinished DNA strands before held at 4°C (Murakami 2000 and 2001). The total ITS2 program run time was 3 hours and 34 minutes. For *petL-psbE*,

the thermal cycler program (*rp116*) was template denaturation at 80°C for 5 minutes, followed by 30 cycles of denaturation at 95°C for 1 minute, primer annealing at 50°C for 1 minute, followed by a ramp of 0.3°C /s to 65°C, and primer extension at 65°C for 4 minutes; ending with a final extension step of 5 minutes at 65°C before held at 4°C (Shaw et al. 2005 and 2007). The total *petL-psbE* program run time was 5 hours and 55 minutes. A single product for each 25 µL PCR was determined using 1% agarose gel electrophoresis and a 100 bp DNA ladder (HyperLadder IV by BIONLINE). PCR products were stored at 4°C until purified prior to sequencing with an ExoSAP (USB Products, Affymetrix, Inc., Cleveland, Ohio, USA) purification reaction mixture of 1.6 µL of 10 un/µL Exonuclease I, 3.2 µL of Shrimp Alkaline Phosphatase 10X Reaction Buffer, and 8.0 µL of PCR DNA template. This method of purification without loss of PCR products (no filtration, precipitation, or washing needed) is especially important for DNA extracted from herbarium specimens, which are sometimes only weakly amplified and barely yield sufficient PCR product for sequencing (Keeley et al. 2007). The ExoSAP purification reaction was run on a PTC-100 thermal cycler (MJ Research) incubated at 37°C for 15 minutes to degrade any remaining primers and nucleotides and then 80°C for 15 minutes to inactivate the ExoSAP mixture. Sequencing reactions of 3.0 µL of purified PCR DNA template, 3.2 µL of 1 µM concentration forward primer (or reverse primer), plus 2.0 µL sterile water were sent to the Advanced Studies in Genomics, Proteomic and Bioinformatics (ASGPB) at the University of Hawai‘i at Mānoa for runs using Applied Biosystems BigDye terminator chemistry on ABI 3730XL capillary-based DNA sequencer. The resulting sequences from both strands of each PCR product were examined, compared, corrected, and edited into contigs using Sequencher® 4.10.1 (Ann Arbor, Michigan, USA). Multiple consensus sequences for each DNA region were aligned into a nucleotide data matrix using the program MUSCLE (Edgar 2004) implemented in the MEGA5.2 software package (Tamura et al. 2011). Default settings in MEGA5.2, DnaSP v5.10.01 (Rozas 2009), and CLC Genomics Workbench v7 (CLC bio Inc, Aarhus, Denmark) were used for comparisons of nucleotide sequences, haplotype reconstructions, and phylogenies.

Phylogenies for each separate rDNA and cpDNA matrix and for the combined nucleotide matrix were constructed using MEGA5.2 and BEAST 2.0.2. For maximum likelihood (ML) phylogenies using MEGA5.2, each nucleotide matrix was used to find the optimal model of sequence evolution for ML by generating an automatic neighbor-joining tree, using all sites, and applying a branch swap filter of very strong. The resulting model chosen for ML analysis had the lowest BIC score (Bayesian Information Criterion), which is considered to describe the substitution pattern the best. Generation of the ML phylogenies was conducted using the maximum number of threads available on the computer with the following parameters: 1) use all sites, 2) the models of nucleotide evolution were Tamura 3-parameter for *petL-psbE* and Hasegawa-Kishino-Yano (HKY) for ITS2 and the combined matrix, 3) nearest-neighbor-interchange; 4) an initial neighbor-joining tree made automatically, 5) applying a branch swap filter of very strong, and 6) 1000 bootstrap replicates.

Generation of a Bayesian inference (BI) phylogeny using BEAST 2.0.2 (Bouckaert et al. 2013) was done by exporting each aligned nucleotide matrix from MEGA5.2 as a Nexus (PAUP 4.0) with the following: 1) all sites were displayed; 2) the datatype was changed from nucleotide to DNA; and 3) the file was saved with the extension .nex. Next, the .nex file alignment is imported to BEAUti (included in the BEAST 2 software package) where the parameters set for the Site Model tab included a substitution rate of 1.0, gamma category count of 4, shape 1.0 with the estimate box checked, proportion invariant 0.1 with the estimate box checked, and the add-on substitution model of Reversible-jump Based substitution (v.1.0.3) (RB). The RB allows the MCMC chain to switch between nucleotide substitution models to search for the best fit model for the data set. For the Clock Model tab, strict clock was selected. The default parameters in the Priors tab were left at estimate for the Yule Model tree. The MCMC Chain parameters were: 1) length was set at 100,000,000 cycles, 2) stored every -1, 3) Pre-Burnin 0, and 4) logged every 1000 cycles to generate 100,001 trees. After the priors were set, the file was saved in BEAUti with the standard .xml extension. After opening the .xml in BEAST 2 and selecting the automatic thread pool size, a .log file was generated by BEAST 2 after the run came to completion. The .log file output from

BEAST 2 was viewed with Tracer v1.5 (available at URL: <http://tree.bio.ed.ac.uk/software/tracer/>), which depicts the outcome of parameters set in BEAUti with each tree sampled along the MCMC chain for likelihood and posterior effective sample size (ESS) (e.g., need ESS to be all black color font or >200). Using the TreeAnnotator (included in the BEAST 2 software package) to combine only the last 11,001 trees sampled after a burnin of 89,000 trees on the MCMC chain for final likelihood and posterior probability estimates, the .trees file was saved with the extension .tree. The annotated .tree file output from TreeAnnotator was viewed using FigTree v1.4.0 (available at URL: <http://tree.bio.ed.ac.uk/software/figtree/>). Topologies for the ITS2, *petL-psbE*, and combined data set for both ML and BI phylogenies were compared, and if similar, the bootstrap values and posterior probabilities for each main node were added to a final consensus tree.

## **Results**

Extraction, amplification, sequencing, and alignment were unproblematic for field collected, purchased samples, and herbarium material for the rDNA and cpDNA regions studied in *H. lupulus*, *H. scandens* (Lour.) Merr., and *Cannabis*. The final lengths of the aligned ITS2 and *petL-psbE* nucleotide matrices were 473bp and 1,078bp, respectively, for a combined total length of 1,551bp. Table 2.3 summarizes the number of conserved, variable, parsimony-informative, and singleton sites for ITS2, *petL-psbE*, and the combined nucleotide matrices. The estimates of average evolutionary divergence over all sequence pairs for the number of base differences per sequence were 6.911 nucleotide differences for ITS2 (p-distance=0.015), 7.206 nucleotide differences for *petL-psbE* (p-distance=0.007), and 14.197 for the combined nucleotide matrices (p-distance=0.010). Between groups mean distance estimates ranged from 0.044 to 0.099 for ITS2, 0.018 to 0.029 for *petL-psbE*, and 0.026 to 0.051 for the combined data set (Table 2.4). As for ITS2 alone, a total of 13-14 gaps were observed between the in-group of *Humulus* spp. and the out-group of *Cannabis*. In addition, 52-60 differences were observed between the in-group and out-group. Furthermore, 0-1 gaps and 0-29 differences were found within the *Humulus* spp. sampled. For *petL-psbE* alone, a total of 40-43 gaps and 69-71 differences were found between *Humulus* spp. and *Cannabis*. The number of gaps and differences within the *Humulus* spp. sampled were 0-18 and 0-37, respectively. The number of haplotypes (h) discovered for ITS2, *petL-psbE*, and the combined data set corresponded to h=10 (Hd: 0.7191), h=4 (Hd=0.6457), and h=14 (Hd=0.8218), accordingly (Table 2.5).

**Table 2.3.** Number of conserved, variable, parsimony-informative, and singleton sites for aligned DNA sequences conducted in MEGA5.2.

<b>Region</b>	<b># conserved</b>	<b># variable</b>	<b># parsimony-informative</b>	<b># singletons</b>
ITS2	413/473	60/473	30/473	30/473
<i>petL-psbE</i>	1013/1078	39/1078	20/1078	19/1078
combined	1426/1551	99/1551	50/1551	49/1551

**Table 2.4.** Estimates of evolutionary divergence over sequence pairs between groups conducted in MEGA5.2

<b>Species 1</b>	<b>Species 2</b>	p-distance		
		<b>ITS2</b>	<b><i>petL-psbE</i></b>	<b>combined</b>
<i>H. lupulus</i>	<i>Cannabis</i> sp.	0.084	0.027	0.045
<i>H. lupulus</i>	<i>H. scandens</i>	0.044	0.018	0.026
<i>Cannabis</i> sp.	<i>H. scandens</i>	0.099	0.029	0.051

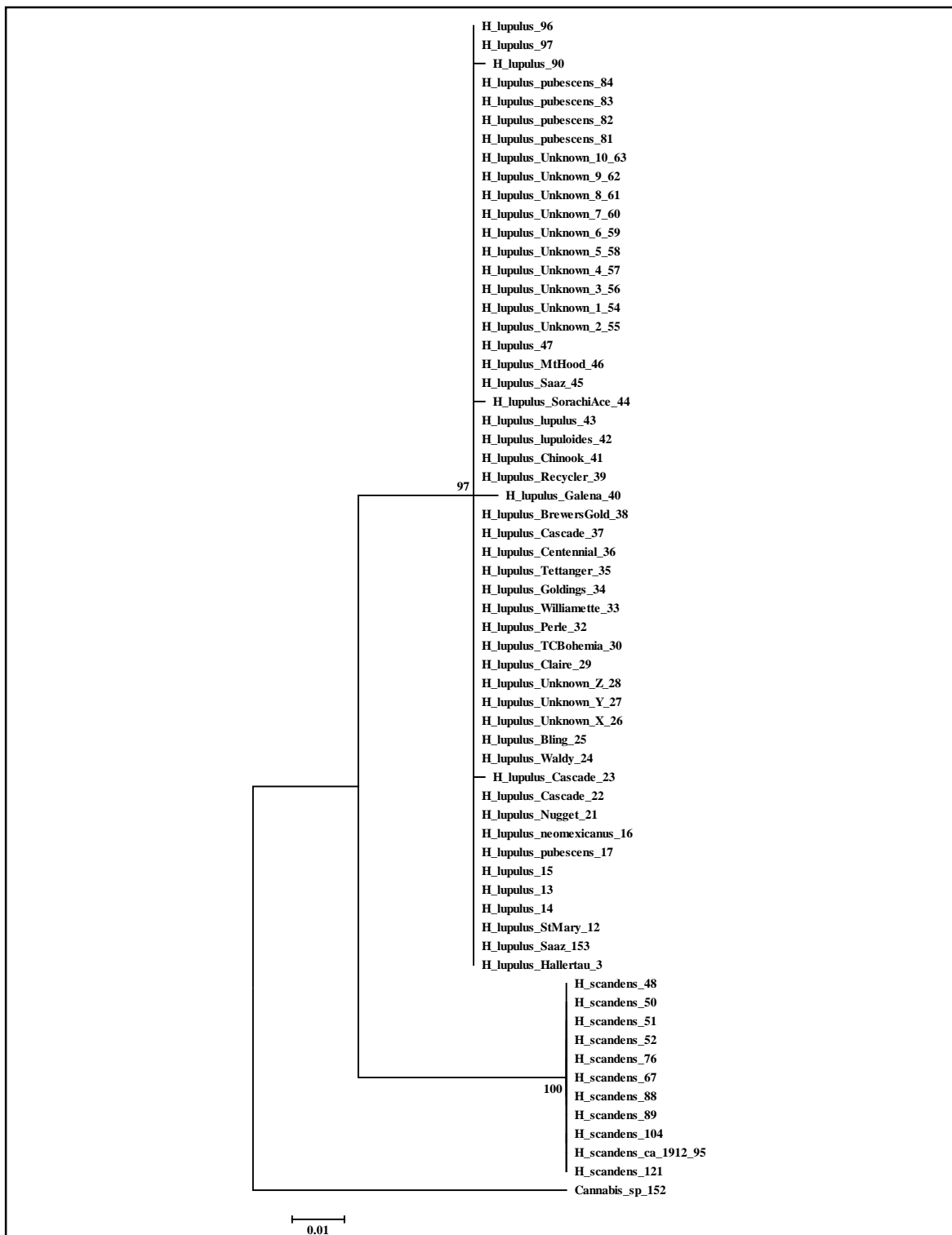
**Table 2.5.** Haplotype estimates for ITS2, *petL-psbE*, and the combined data set conducted in DnaSP v5.10.01. Prior to the haplotype reconstructions for ITS2 and combined data sets, the nucleotide data matrix was reconstructed using PHASE (Stephens et al. 2001, Stephens and Donnelly 2003) to incorporate the IUPAC nucleotide ambiguity codes that represent heterozygous sites. If the extractions numbers for the ITS2 and combined data sets are followed by \_1 or \_2, they signify two possible genotypes for the diploid sequences, otherwise the two sequences generated by PHASE are identical in that haplotype.

	region		
	ITS2	<i>petL-psbE</i>	combined
Number of haplotypes	10	4	14
Haplotype diversity	0.7191	0.6457	0.8218
Summary of haplotype distribution by extraction #	<b>Hap_1 (n=62):</b> 3, 12, 13, 21_1, 22_1, 24_2, 25, 26, 27, 28, 29_1, 30_2, 32, 33, 34, 35, 36, 37_1, 38_1, 39_2, 41, 43, 45, 46, 54, 55_2, 56, 57, 58, 59, 60, 61_2, 62_2, 63_2, 96, 97, 153 <b>Hap_2 (n=2):</b> 152 <b>Hap_3 (n=16):</b> 14, 15, 16, 17, 42, 47, 82_1, 83_1, 84 <b>Hap_4 (n=9):</b> 21_2, 22_2, 23_2, 37_2, 38_2, 44_2, 81_2, 82_2, 83_2 <b>Hap_5 (n=3):</b> 23_1, 29_2, 44_1 <b>Hap_6 (n=11):</b> 24_1, 30_1, 39_1, 40, 55_1, 61_1, 62_1, 63_1, 81_1, 90_2 <b>Hap_7 (n=13):</b> 48, 51, 52, 67_2, 76, 89, 95	<b>Hap_1 (n=29):</b> 3, 12, 13, 22, 23, 24, 25, 26, 27, 28, 30, 32, 33, 34, 35, 37, 39, 43, 45, 46, 54, 55, 61, 62, 63, 90, 96, 97, 153 <b>Hap_2 (n=1):</b> 152 <b>Hap_3 (n=22):</b> 14, 15, 16, 17, 21, 29, 36, 38, 40, 41, 42, 44, 47, 56, 57, 58, 59, 60, 81, 82, 83, 84 <b>Hap_4 (n=11):</b> 48, 50, 51, 52, 67, 76, 88, 89, 95, 104, 11	<b>Hap_1 (n=45):</b> 3, 12, 13, 22_1, 24_2, 25, 26, 27, 28, 30_2, 32, 33, 34, 35, 37_1, 39_2, 43, 45, 46, 54, 55_2, 61_2, 62_2, 63_2, 96, 97, 153 <b>Hap_2 (n=2):</b> 152 <b>Hap_3 (n=16):</b> 14, 15, 16, 17, 42, 47, 82_1, 83_1, 84 <b>Hap_4 (n=17):</b> 21_1, 29_1, 36, 38_1, 41, 56, 57, 58, 59, 60 <b>Hap_5 (n=6):</b> 21_2, 38_2, 44_2, 81_2, 82_2, 83_2 <b>Hap_6 (n=3):</b> 22_2, 23_2, 37_2 <b>Hap_7 (n=1):</b> 23_1 <b>Hap_8 (n=8):</b> 24_1, 30_1, 39_1, 55_1, 61_1, 62_1, 63_1, 90_2 <b>Hap_9 (n=2):</b> 29_2, 44_1 <b>Hap_10 (n=3):</b> 40, 81_1

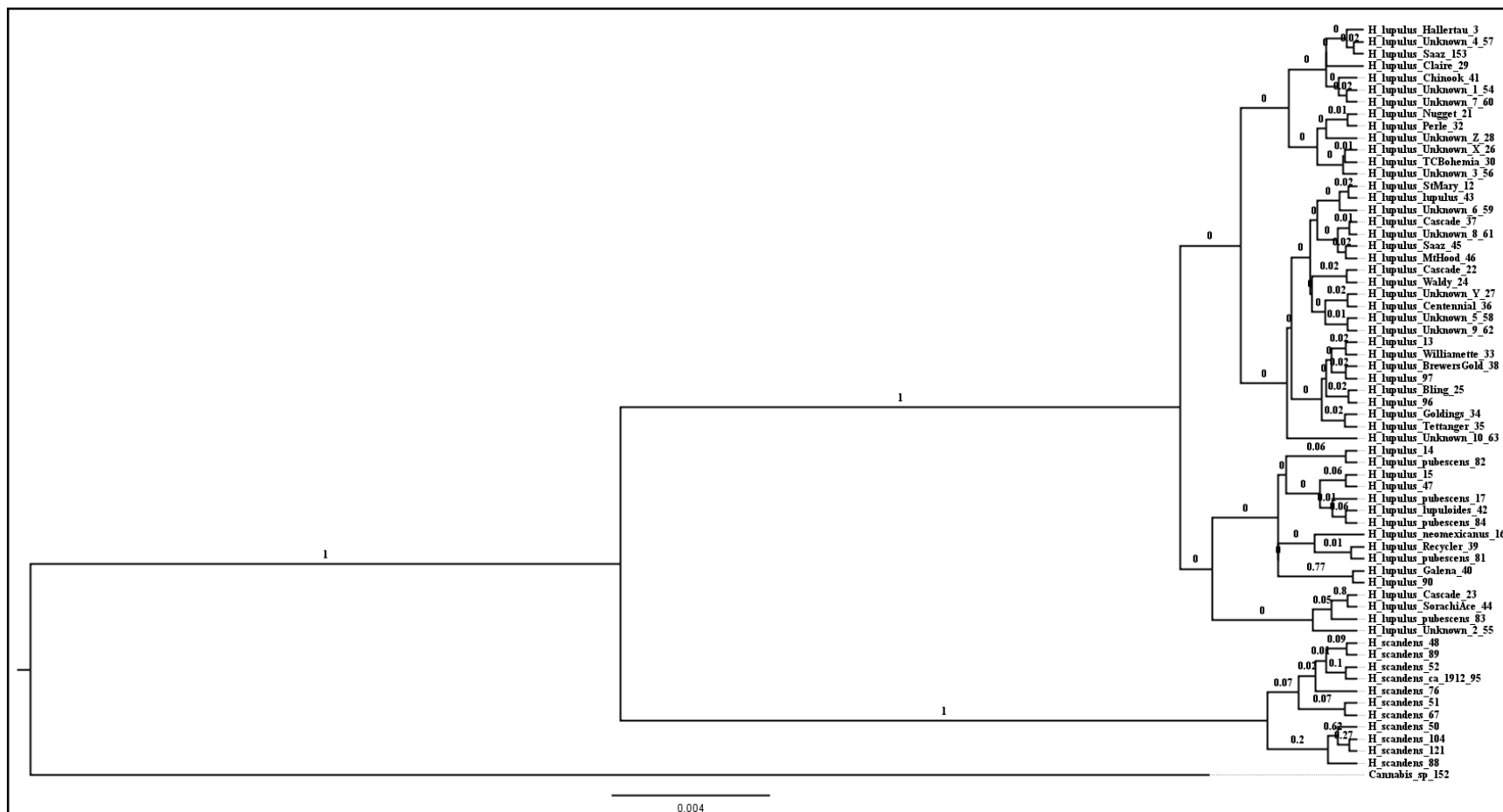
**Table 2.5.** continued.

	region		
	ITS2	<i>petL-psbE</i>	combined
Summary of haplotype distribution by extraction #	<b>Hap_8 (<i>n</i>=8):</b> 50, 67_1, 88_2, 104, 121 <b>Hap_9 (<i>n</i>=1):</b> 88_1 <b>Hap_10 (<i>n</i>=1):</b> 90_1		<b>Hap_11 (<i>n</i>=13):</b> 48, 51, 52, 67_2, 76, 89, 95 <b>Hap_12 (<i>n</i>=8):</b> 50, 67_1, 88_2, 104, 121 <b>Hap_13 (<i>n</i>=1):</b> 88_1 <b>Hap_14 (<i>n</i>=1):</b> 90_1

ML and BI phylogenies for ITS2 show two clades of *Humulus* with *Cannabis* as the outgroup (Figs. 2.2 and 2.3). The ML ITS2 tree has high support at nodes with polytomies for both the *H. lupulus* (97%) and *H. scandens* (100%) clades. Similarly, the BI ITS2 tree shows high support at the main nodes for both the *H. lupulus* and *H. scandens* clades (1.0). Furthermore for BI ITS2 tree, low support was found for the interior nodes of the *H. lupulus* clades (0-0.06), with the exception of one interior node at 0.77 supporting the samples H\_lupulus\_Galena\_40 and H\_lupulus\_90. Low support for the internal nodes of the *H. scandens* clades is similar to internal nodes of the *H. lupulus* clade. The only exception is an internal branch that includes four *H. scandens* samples at 0.2 posterior probability supported with two additional internal branches with 0.62 and 0.27 posterior probabilities.

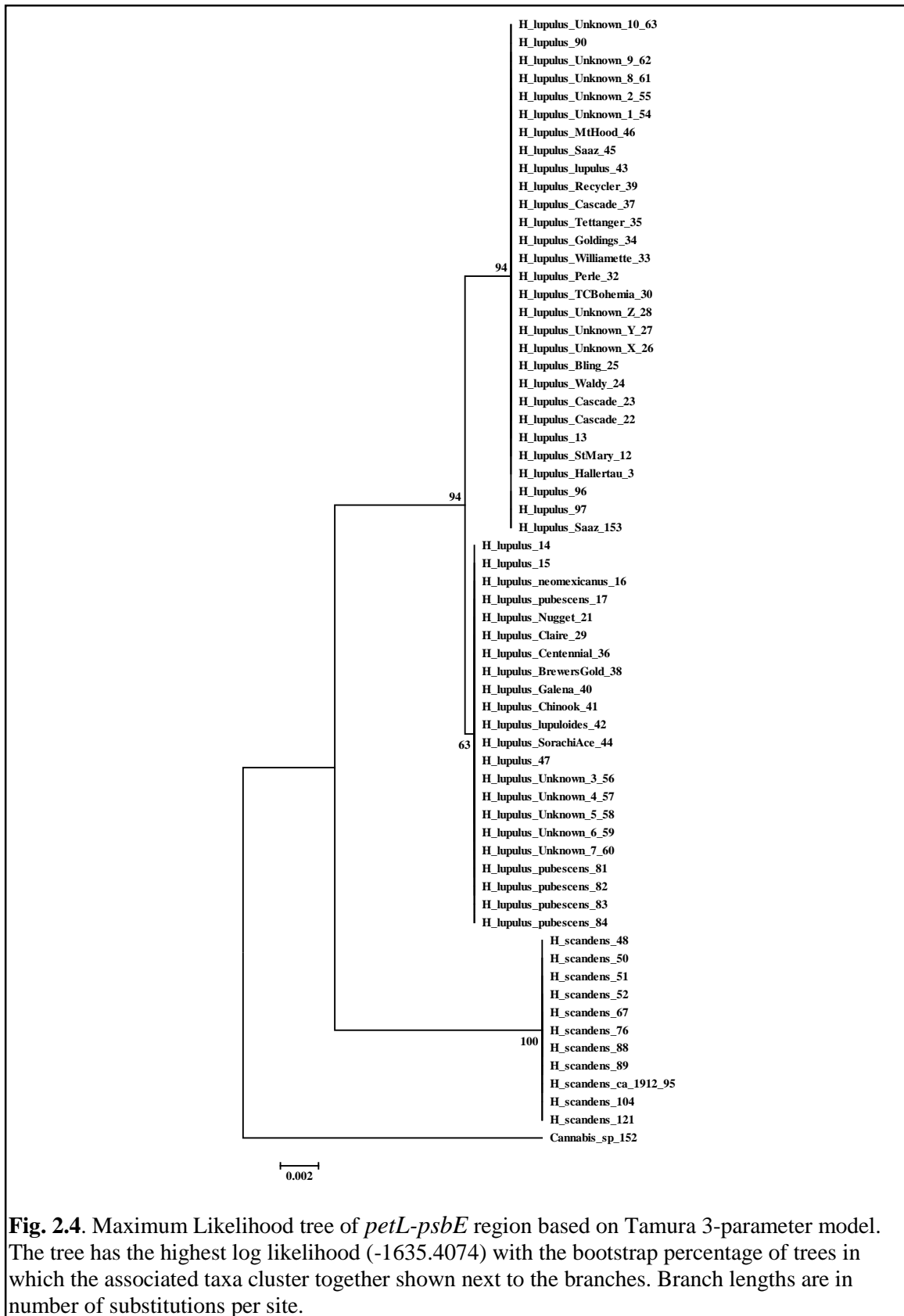


**Fig. 2.2.** Maximum Likelihood tree of ITS2 region based on Hasegawa-Kishino-Yano model. The tree has the highest log likelihood (-1040.7436) with the bootstrap percentage of trees in which the associated taxa cluster together shown next to the branches. Branch lengths are in number of substitutions per site.

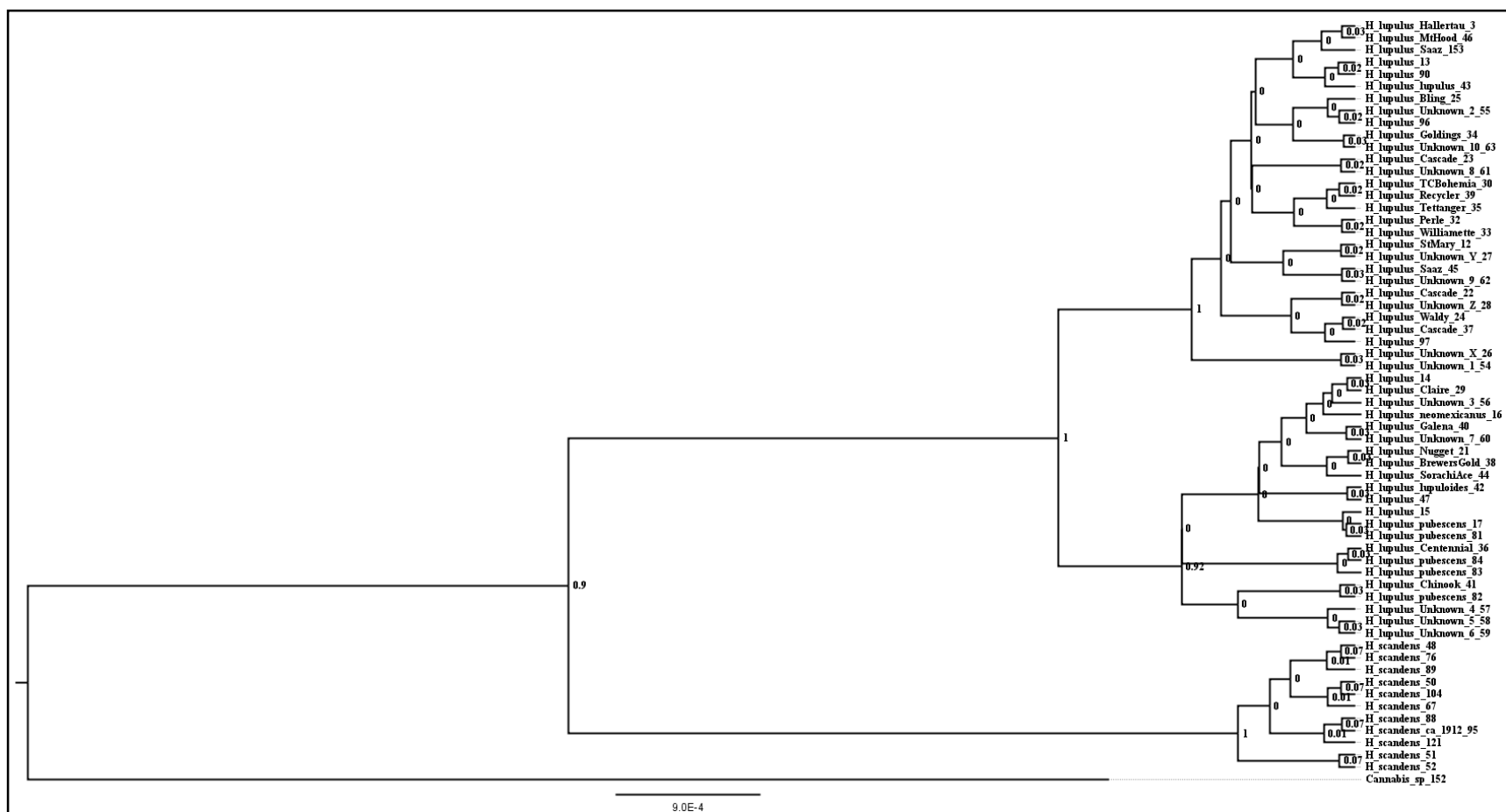


**Fig. 2.3.** Bayesian inference phylogeny of ITS2 depicted as a maximum clade credibility tree based on Reversible-jump Based substitution model. Posterior probabilities in which the associated taxa cluster together is shown next to the branches. Branch lengths are in number of substitutions per site. The highest log clade credibility is -320.13369418797964.

ML and BI phylogenies for *petL-psbE* show three clades of *Humulus* with *Cannabis* as the outgroup (Figs. 2.4 and 2.5). High bootstrap and posterior probability support was found for the *H. scandens* clade. (100% and 1.0). Additionally, high bootstrap and posterior probability support was found for the *H. lupulus* clade (94% and 1.0), and this clade has two further groups: one with high bootstrap and posterior probability support (94% and 1.0) and another with slightly lower bootstrap and posterior probability support (63% and 0.92). The internal nodes of both ML and BI *petL-psbE* trees are low (0% and 0-0.07).

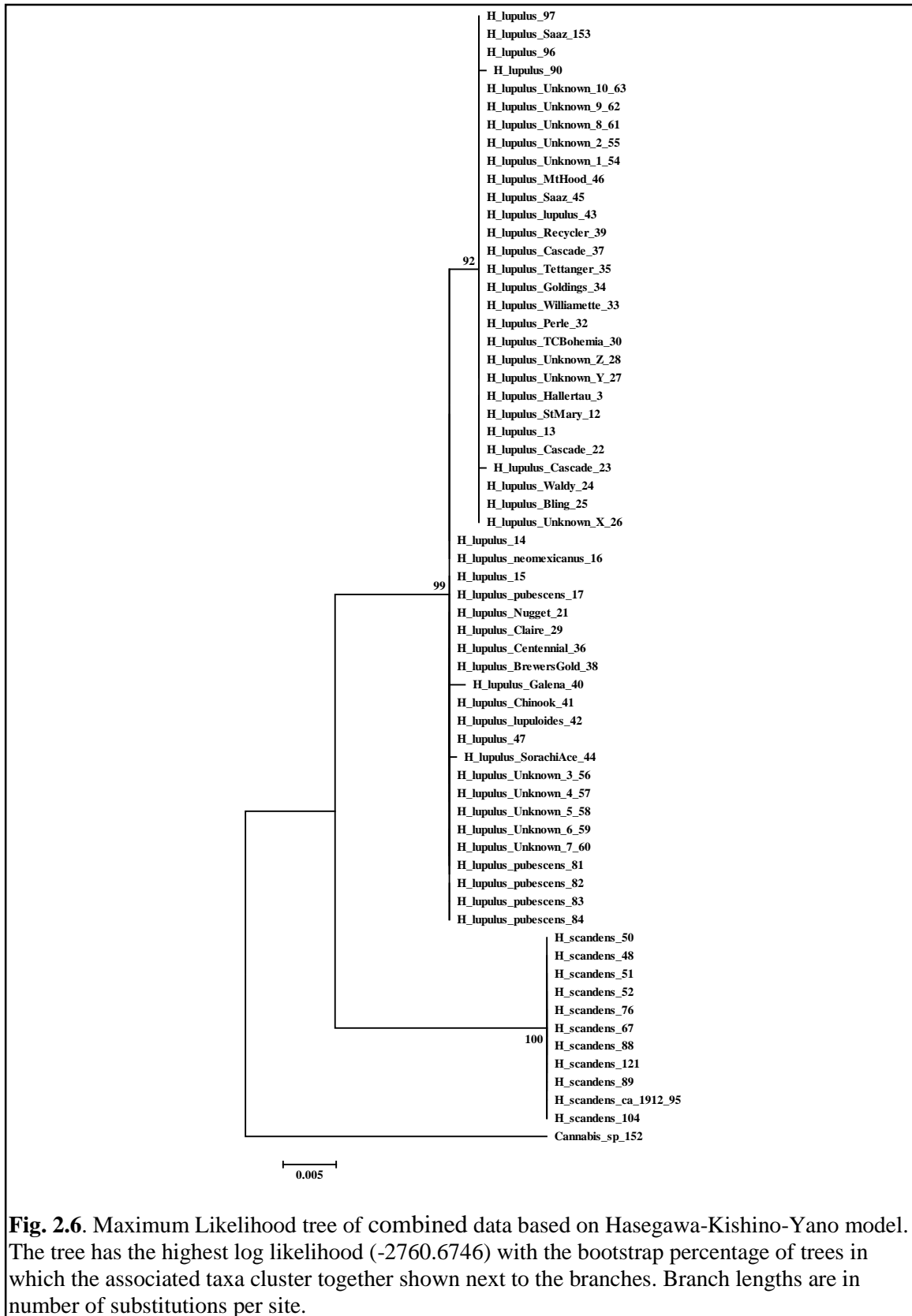


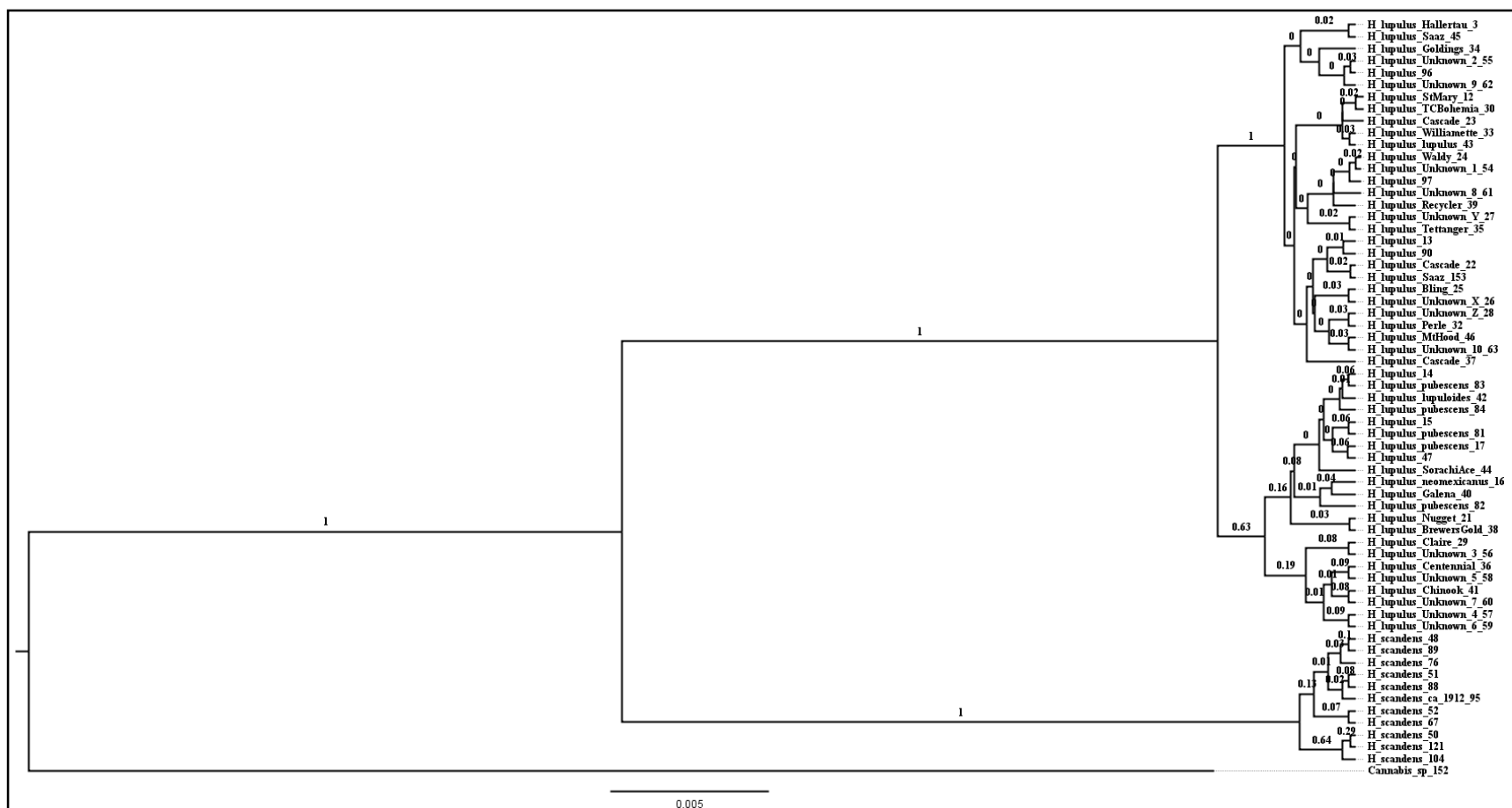
**Fig. 2.4.** Maximum Likelihood tree of *petL-psbE* region based on Tamura 3-parameter model. The tree has the highest log likelihood (-1635.4074) with the bootstrap percentage of trees in which the associated taxa cluster together shown next to the branches. Branch lengths are in number of substitutions per site.



**Fig. 2.5.** Bayesian inference phylogeny of *petL-psbE* depicted as a maximum clade credibility tree based on Reversible-jump Based substitution model. Posterior probabilities in which the associated taxa cluster together is shown next to the branches. Branch lengths are in number of substitutions per site. The highest log clade credibility is -323.9328661964444.

ML and BI phylogenies for the combined data set has a similar topology to the *petL-psbE* trees, showing three separate clades of *Humulus* (Figs. 2.6 and 2.7). High bootstrap and posterior probability support were found for the *H. scandens* clade. (100% and 1.0). Additionally, high bootstrap and posterior probability support were found for the *H. lupulus* clade (94% and 1.0), but this clade is separated into two further groups: one with high bootstrap and posterior probability support (94% and 1.0) and another with lower bootstrap and posterior probability support (63% and 0.92).





**Fig. 2.7.** Bayesian inference phylogeny of combined data depicted as a maximum clade credibility tree based on Reversible-jump Based substitution model. Posterior probabilities in which the associated taxa cluster together is shown next to the branches. Branch lengths are in number of substitutions per site. The highest log clade credibility is -276.3908123592795.

## **Discussion**

The ML and BI phylogenies for the ITS2 region clearly depict two separate species, *H. lupulus* and *H. scandens*, with high bootstrap percentages and posterior probabilities (Fig. 2.8). Alternatively, the *petL-psbE* and combined data sets show a topology supporting *H. scandens* as a sister clade to two *H. lupulus* clades. These two clades of the common hop segregate the samples used in this study into primarily New World and Old World pedigrees. Relying on the evolutionary histories presented here in the phylogenies, the putative wild hop samples collected in Michigan were sorted into their respective lineages/clades (Table 2.6). The Old World group contains samples of the highly cultivated *H. lupulus* var. *lupulus* and those samples with similar maternal pedigrees. On the other hand, the New World clade contains *H. lupulus* var. *lupuloides*, *H. lupulus* var. *neomexicanus*, and *H. lupulus* var. *pubescens*, as well as the common cultivars with significant degrees of New World hop ancestry (e.g., hybrids). For example, both ‘Nugget’ and ‘Galena’ have ‘Brewer’s Gold’ in their pedigrees (Townsend and Henning 2009). Specifically, Salmon (1934) developed ‘Brewer’s Gold’ via open pollination of the native American female BB1 collected in 1916 near Morden, Manitoba, Canada (Burgess 1964). The BB1 genotype most probably is a representative of the variety *H. lupulus* var. *lupuloides* (Hampton et al. 2002 and 2003, Townsend and Henning 2009), and this maternal line is represented in the chloroplast region *petL-psbE* and combined data phylogenies.

The ITS2, *petL-psbE*, and the combined data mirrors results from previous investigators for a split between New World and Old World hop pedigrees (Cerenak et al. 2009, McAdam 2013, Murakami et al. 2006a and 2006b, Patzak et al. 2010a and 2010b, Peredo et al. 2010, Stajner et al. 2008, Townsend and Henning 2009). The *petL-psbE* region highlights the New World common hop germplasm, specifically *H. lupulus* var. *lupuloides* compared to the Old World germplasm of *H. lupulus* var. *lupulus*. Much of the wild germplasm collected in this study in the State of Michigan resembles the morphology of *H. lupulus* var. *lupuloides* (Boutain personal observation). Thus, Michigan does have truly wild hop germplasm (putatively *H. lupulus* var. *lupuloides*) as

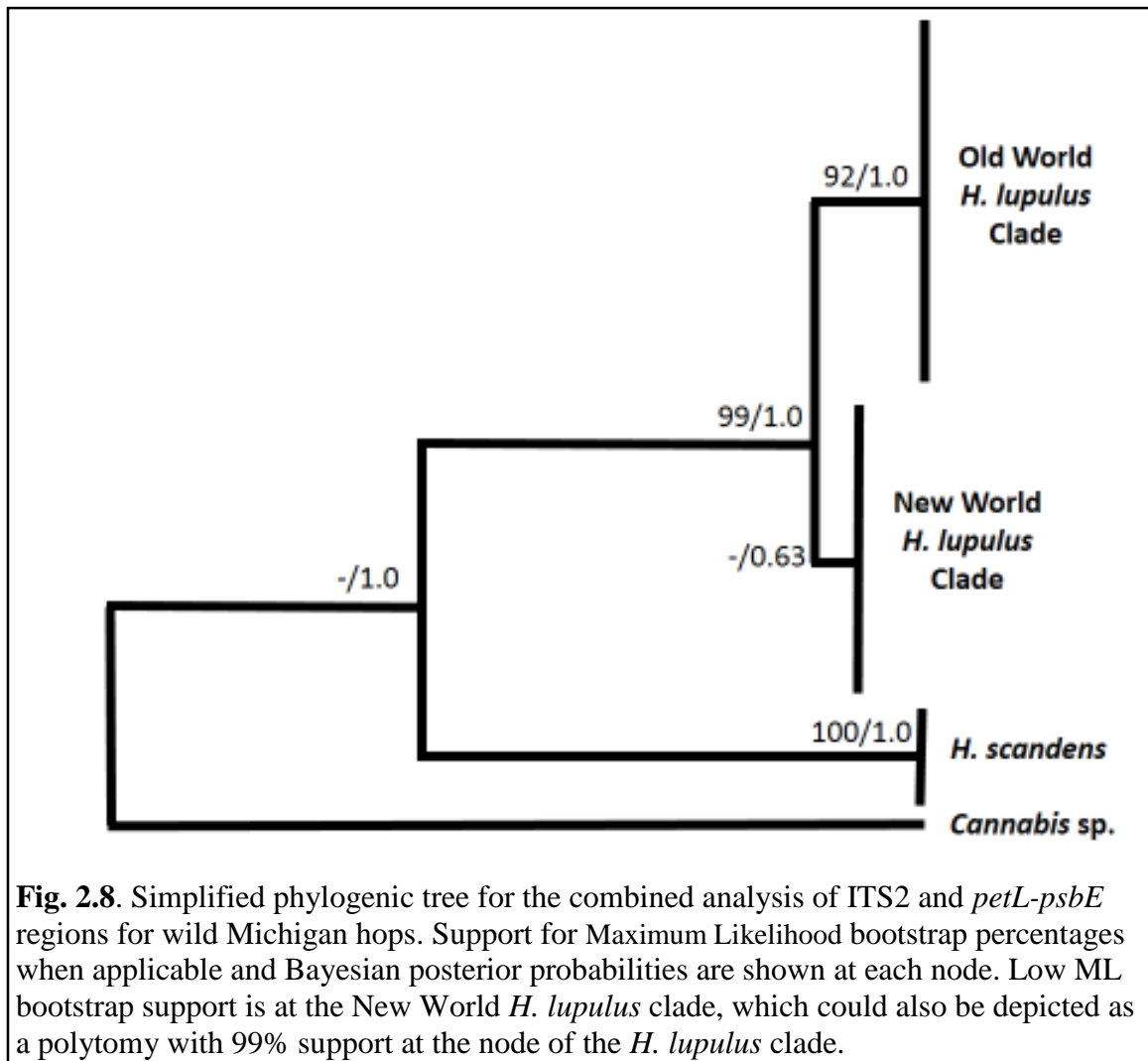
well as escaped and possibly heirloom cultivars (e.g., Traverse City Bohemian Hop and the St. Mary's Hop collected in Leelanau County, Northwest Michigan).

When the Traverse City Bohemian Hop specimen was collected, the owners of the land indicated that this plant was introduced in the 1870s from Bohemia when their family immigrated to the U.S.A. The St. Mary's Hop specimen was collected along the fence line next to the new playground at St. Mary's elementary school in Leelanau County. This plant dominates the vegetation of the nearby forested region on the other side of the fence and is most likely a remnant cultivar now gone fallow (see Fig. 2.1). Also, the new cultivar Claire, collected in Wayne County, Southeast Michigan was found to belong to the New World clade. Whether this plant is a result of a hybridization event between early European cultivars of *H. lupulus* var. *lupulus* and a native Michigan hop (cf. *H. lupulus* var. *lupuloides*) remains for further research. Areas for additional research include genomic, agronomic, and chemical testing the putatively wild Michigan hops sampled in this study to improve local hop varieties by selecting for noteworthy traits. After all, the native American female BB1 genotype most probably is a representative of the variety *H. lupulus* var. *lupuloides* that has contributed significantly to hop plant breeding programs.

Moreover, the *petL-psbE* region distinguished the common hop based on the maternal origins, while the ITS2 region proved useful for species recognition between *H. lupulus* and *H. scandens*. Although the morphological differences between *H. lupulus* and *H. scandens* can easily distinguish the two species, due to phenotypic plasticity within the *H. lupulus* complex, their morphological traits can be blurred upon a continuum as a result of introgressive hybridization (Table 2.7). Sometimes but not often, *H. lupulus* and *H. scandens* are misidentified and mislabeled on herbarium specimens, especially if the specimens are male or were collected in an early juvenile period without flowers. Using the ITS2 region helped distinguish some of these cryptic herbarium specimens. The oldest amplified and successfully sequenced *Humulus* specimen was of material collected from Jackson Park, Chicago, Illinois, U.S.A in 1912. This historical DNA example from *Humulus* highlights the significance of herbaria for botanical knowledge, conservation,

and education (Ahrends et al. 2011, Case et al. 2007, Fleet et al. 2006, Prather et al. 2004, Pyke and Ehrlich 2010, Sundberg et al. 2011).

The phylogeny of hop specimens presented here indicates: 1) a single clade of *H. scandens* and 2) two groups of the common hop plant. These latter two groups could be classified according to Small (1978) as: 1) a clade of cultivars of *H. lupulus* var. *lupulus* with Old World/European genetics and 2) a clade of cultivars and plants with significant New World genetics represented by *H. lupulus* var. *lupuloides*, *H. lupulus* var. *neomexicanus*, and *H. lupulus* var. *pubescens*. An alternative taxonomy conserves those samples in the *H. lupulus* var. *lupulus* clade to *H. lupulus* and lumps the clade with the New World genetics as *H. americanus* Nutt. Also, a third taxonomy could be made based on a phylogenetic and DNA sequence species concept, which conserves the samples restricted to the *H. lupulus* var. *lupulus* clade and lumps the clade containing those samples with the New World genetics (e.g., *H. lupulus* var. *americanus* J. Boutain, var. nov.). Still based on the non-coding regions analyzed for the samples in this study, little genetic distance was found between the Old World and New World clades of *H. lupulus* (p-distance=0.002; Table 2.8), so these two clades could be lumped to a single species given they are interfertile. In another approach with AFLPs, Reeves and Richards (2011) examined five different species criteria for wild North American *H. lupulus*. They found support to recognize vars. *neomexicanus* and *pubescens* as species; however, Reeves and Richards (2011) withheld a species recommendation for var. *lupuloides* until further sampling of genetic variation is complete or a stable biological process can be identified to explain its observed genetic divergence. In general, the use of DNA for plant species determinations must include reviews of the organism's taxonomy and herbarium collections (Hajibabaei et al. 2007, Padial et al. 2010, Tautz et al. 2003, Wandeler et al. 2007).



**Table 2.6.** List of samples with DNA numbers sorted by lineage/clade.

<b>Sample (H=Humulus)</b>	<b>DNA#</b>	<b>Clade in phylogeny</b>
H_lupulus_cf_lupuloides	14	New World
H_lupulus_lupuloides	15	New World
H_lupulus_var_neomexicanus	16	New World
H_lupulus_cf_pubescens	17	New World
H_lupulus_Nugget	21	New World
H_lupulus_Claire	29	New World
H_lupulus_Centennial	36	New World
H_lupulus_BrewersGold	38	New World
H_lupulus_Galena	40	New World
H_lupulus_Chinook	41	New World
H_lupulus_lupuloides	42	New World
H_lupulus_SorachiAce	44	New World
H_lupulus	47	New World
H_lupulus_Unknown3	56	New World
H_lupulus_Unknown4	57	New World
H_lupulus_Unknown5	58	New World
H_lupulus_Unknown6	59	New World
H_lupulus_Unknown7	60	New World
H_lupulus_pubescens	81	New World
H_lupulus_pubescens	82	New World
H_lupulus_pubescens	83	New World
H_lupulus_pubescens	84	New World
H_lupulus_Hallertau	3	Old World
H_lupulus_StMary	12	Old World
H_lupulus	13	Old World
H_lupulus_Cascade	22	Old World
H_lupulus_Cascade	23	Old World
H_lupulus_Waldy	24	Old World
H_lupulus_Bling	25	Old World
H_lupulus_UnknownX	26	Old World
H_lupulus_UnknownY	27	Old World
H_lupulus_UnknownZ	28	Old World
H_lupulus_TCBohemia	30	Old World
H_lupulus_Perle	32	Old World
H_lupulus_Williamette	33	Old World
H_lupulus_Goldings	34	Old World

<b>Table 2.6.</b> continued.		
H_lupulus_Tettanger	35	Old World
H_lupulus_Cascade	37	Old World
H_lupulus_Recycler	30	Old World
H_lupulus_lupulus	43	Old World
H_lupulus_Saaz	45	Old World
H_lupulus_MtHood	46	Old World
H_lupulus_Unknown1	54	Old World
H_lupulus_Unknown2	55	Old World
H_lupulus_Unknown8	61	Old World
H_lupulus_Unknown9	62	Old World
H_lupulus_Unknown10	63	Old World
H_lupulus	90	Old World
H_lupulus	96	Old World
H_lupulus	97	Old World
H_lupulus_Saaz	153	Old World
H_scandens	48	scandens
H_scandens	50	scandens
H_scandens	51	scandens
H_scandens	52	scandens
H_scandens	67	scandens
H_scandens	76	scandens
H_scandens	85	scandens
H_scandens	89	scandens
H_scandens	95	scandens
H_scandens	104	scandens
H_scandens	121	scandens
Cannabis_sp	152	Outgroup

**Table 2.7.** Distribution, DNA sequence clade, and morphological trait of floral leaf midrib hairs distinguishing *Humulus lupulus* var. *lupulus* from *H. lupulus* var. *lupuloides*, *H. lupulus* var. *neomexicanus*, and *H. lupulus* var. *pubescens*. Due to introgression, the morphological characters can be skewed (data adapted from Smith et al. (2006) and Small (1978)).

	taxa			
	var. <i>lupulus</i>	var. <i>neomexicanus</i>	var. <i>lupuloides</i>	var. <i>pubescens</i>
Distribution	Eurasia but naturalized in eastern North America	<b>western North America</b>	<b>central North America</b>	<b>south central North America</b>
ITS2 clade	<i>H. lupulus</i>	<i>H. lupulus</i>	<i>H. lupulus</i>	<i>H. lupulus</i>
<i>petL-psbE</i> clade	Old World	<b>New World</b>	<b>New World</b>	<b>New World</b>
Floral leaf midrib hairs	<20/cm	<b>&gt;20/cm</b>	<b>&lt;100/cm</b>	<b>&gt;100/cm</b>

**Table 2.8.** Estimates of evolutionary divergence over sequence pairs between New World and Old World groups of *Humulus lupulus* conducted in MEGA5.2

<b>Species 1</b>	<b>Species 2</b>	p-distance
		<b>combined data set</b>
Old World <i>H. lupulus</i> clade	<i>Cannabis</i> sp.	0.045
Old World <i>H. lupulus</i> clade	New World <i>H. lupulus</i> clade	0.002
<i>Cannabis</i> sp.	New World <i>H. lupulus</i> clade	0.043
Old World <i>H. lupulus</i> clade	<i>H. scandens</i>	0.026
<i>Cannabis</i> sp.	<i>H. scandens</i>	0.051
New World <i>H. lupulus</i> clade	<i>H. scandens</i>	0.026

## **Conclusion**

This study illustrates that hop crop species and their wild congeners are phylogenetically discriminated. The main objectives were accomplished to support the hypothesis (H<sub>2.1</sub>). Hop breeding programs in Michigan can use simple genetic tests as a first step for the identification of putatively wild plants, which can lead to the development of new cultivars and germplasm resources for the world hop industry (see Steiger et al. 2002 for an example from *Coffea arabica* cultivars). In Michigan, wild hop germplasm may be better adapted to regional and local disease causing organisms and changes in climate. Additional genomic and chemical analyses on the wild Michigan hop plants and escaped cultivars can be done to determine traits of value to the microbreweries and pharmaceutical industries around the State of Michigan (Buck et al. 2009, Srećec et al. 2012, Steele and Pires 2011, Whittock et al. 2009). Future directions for Michigan hop research must include a comprehensive survey and collecting expedition throughout the state's two peninsulas with an emphasis on the sensitive riparian habitats and plant communities associated with native North American hop plants (Boutain and Gelderloos 2006).

### **Acknowledgements**

I thank Elizabeth, Michael, and Emma Patterson; Richard, Barbara, Matthew, and Claire Boutain; Hugh and Jean Zeits; Ernest Small, Atsushi Murakami, Rob Sirrine, Brian Tennis, Graham Kelly, Joanne Birch, Tim Gallaher, Jason Cantley, the Botanical Research Institute of Texas, the Michigan Hop Alliance, the Michigan State University Extension, Agriculture and Agri-Food Canada, the Botany Department at the University of Hawai‘i at Mānoa (UH), the Graduate Student Organization at UH, numerous office, lab, and herbarium colleagues and counterparts that made this study possible. Any errors are solely by the author.

## CHAPTER 3

### DNA Support for the Endemic Yunnan Hop (*Humulus yunnanensis* Hu)

#### **Abstract**

The Yunnan hop, *Humulus yunnanensis* Hu (Cannabaceae), is one of the recognized species in the genus, while China is hypothesized as the center of origin for *Humulus*. A three species concept adopted by morphological investigations and extensive molecular analyses for both *H. lupulus* and *H. scandens* does not include molecular data to support the putative taxon (*H. yunnanensis*). Therefore, an approach with short DNA sequences (barcodes) can support the hypothesis:  $H_0$  = The Yunnan hop is not a separate species but a variety of the common hop, *H. lupulus*. A collection of cultivars, putatively wild plants, and herbarium specimens was tested using DNA barcodes and phylogenetic methods to determine if the putatively wild hop plants are native germplasm or escaped European cultivars. Phylogenetic and evolutionary divergence results show the Yunnan hop is a distinct species. Further field work and genomic study of DNA from herbarium material is warranted to further sort out the observed patterns of speciation and varietal relationships in *Humulus* spp., which will likely require a taxonomic revision of the genus.

#### **Keywords**

hops, cpDNA, *petL-psbE*, rDNA, ITS2, DNA barcodes, phylogenetics, herbarium specimen, Xinjiang, China

## **Introduction**

The Yunnan hop, *Humulus yunnanensis* Hu (Cannabaceae), is one of the three recognized species in the genus (Fig. 3.1, Hu 1936, Small 1978, Wu et al. 2003). On the one hand, the authenticity of *H. yunnanensis* as a species has been questioned because of a close resemblance to the common hop, *H. lupulus* L. (Small 1978, Small personal communication). Also, few collections of *H. yunnanensis* exist in herbaria, and identifications appear to be incorrect in many cases (Boutain personal observation). This misidentification is a result of overlapping variation in morphological characters for leaves and flowers, as well as the limited knowledge and distribution of the putative Yunnan hop in the proposed region of origin for the genus (i.e., China). To date, only one molecular study included DNA sequences of the putative *H. yunnanensis* (Yang et al. 2013, Yang 005 (KUN), Yang personal communication), and this plant was later determined to be *H. lupulus* var. *cordifolius* Maxim. (syn: *Humulus cordifolius* Miq.) (Fig 3.2, Boutain personal observation). As a result, the phylogeny and genetic diversity of *Humulus* remains unknown. The purpose of this study is to clarify the taxonomy and phylogenetic relationship of *H. yunnanensis* within the genus using putatively wild *Humulus* samples, herbarium specimens, and DNA barcodes.

## ***Humulus in China***

Hop plants (*Humulus* spp.) are broadly distributed in China from the northeast to the southwest provinces (Wu et al. 2003). Three species (i.e., *H. lupulus*, *H. scandens* (Lour.) Merr., and *H. yunnanensis*) and two botanical varieties (i.e., *H. lupulus* var. *lupulus* and *H. lupulus* var. *cordifolius*) are generally recognized based on morphological characters (Small 1978, Wu et al. 2003). Since this is the only area of the world in which all three putative species are found, China is hypothesized to be the center of origin for the genus. In addition, an extensive history of use of hops, the female flowers of the hop plant, exists for a wide range of traditional Chinese medicinal preparations (Li and Luo 2003). Identifying the genetic origin and potentially greatest biodiversity of *Humulus* spp. is important for the conservation of wild hop plants for scientific and commercial uses, especially for brewing and medicine (Hampton et al. 2001, Hummer 2003 and 2005, Hummer et al. 2002 and 2003, Smith et al. 2006, Steele and Pires 2011).

### ***DNA barcoding herbarium specimens***

Plant DNA is easily extracted from small amounts of fresh, silica dried, or herbarium material; however, the DNA molecule degrades to approximately 500-100 base pairs the older the specimen ages (Almakarem et al. 2012, Fulton 2012, Herrmann and Hummel 1994, Hummel 2003, Rogers and Bendich 1985, 1994, Russo et al. 2008, Särkinen et al. 2012, Shapiro and Hofreiter 2012). For cryptic species, the identification of operational taxonomic units through DNA barcoding has been achieved by the retrieval of short sequences of DNA, typically between 400-800 base pairs from a standardized area of the genome (i.e., exons) (Hajibabaei et al. 2007, Li et al. 2007), although, more variable introns must be included for the identification of some plant species (CBOL Plant Working Group et al. 2009, China Plant BOL Group et al. 2011, Kress and Erickson 2012). For example, specific ITS2 primers for forensic barcoding of *Cannabis* L. easily discriminate *Humulus* (Gigliano 1998, Gigliano et al. 1997, Murakami 2000). Biparentally inherited regions of DNA, such as ribosomal DNA (rDNA), have shown incongruence relative to the chloroplast (Kim and Jansen 1998, Cronn et al. 2002). For this study, both the ITS2 nuclear and *petL-psbE* chloroplast regions were chosen on the basis of the greater reliability and better resolution of lower level variation as shown by CBOL Plant Working Group et al. (2009), China Plant BOL Group et al. (2011), Hollingsworth (2011), Shaw et al. (2005 and 2007), and specifically Chapter 2 in this dissertation.

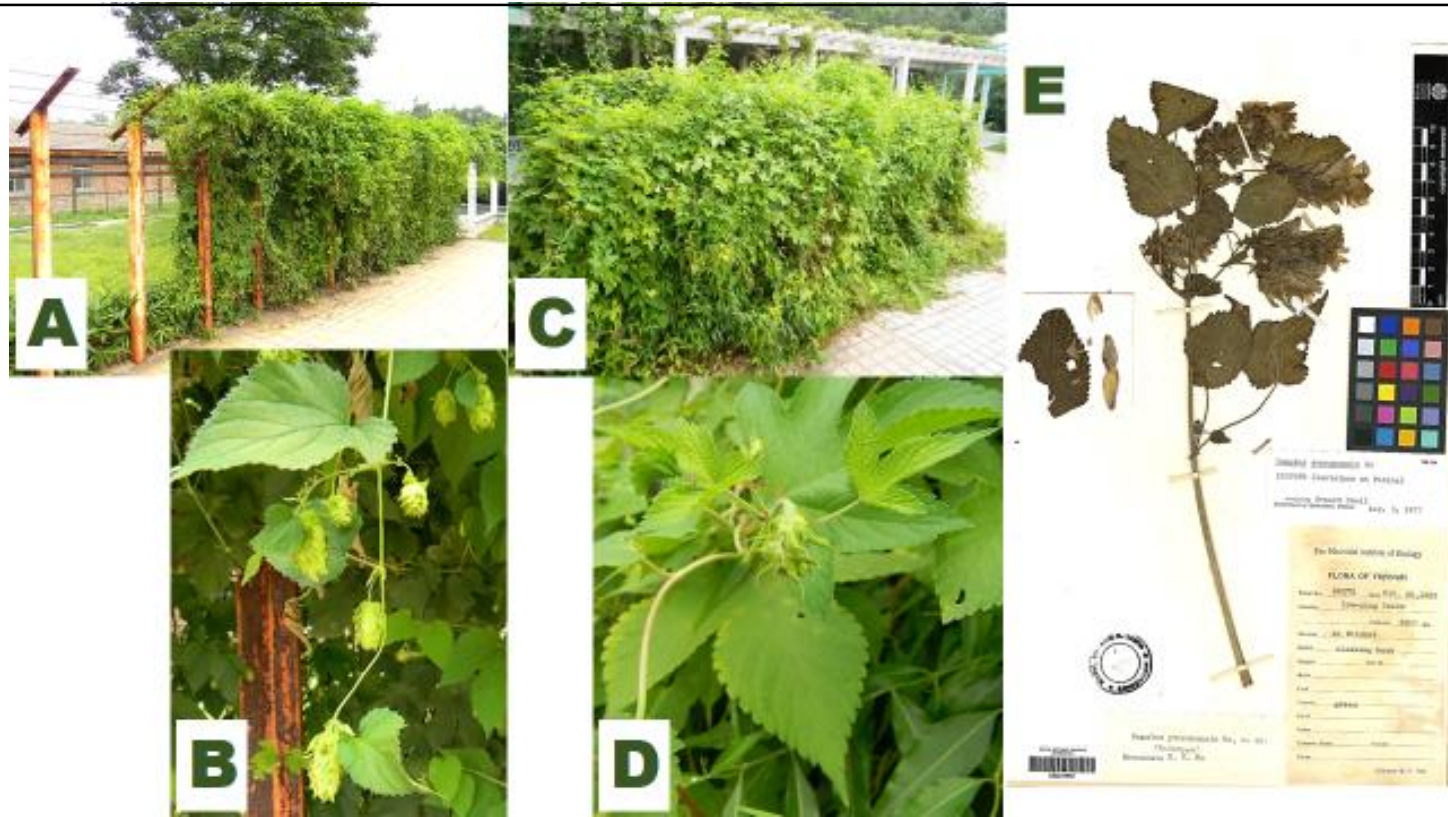
## **Objectives**

This project had four main objectives:

- 1) Collecting cultivated and putatively wild hop plants in China,
- 2) Sampling of herbarium specimens for historical DNA analysis,
- 3) Testing short DNA sequences to determine if putatively wild hop plants are native germplasm or escaped European cultivars of *Humulus lupulus* var. *lupulus*, and
- 4) Clarifying the existence of and potential phylogenetic relationship of the taxon, *H. yunnanensis*, within the Cannabaceae.

## **Hypothesis**

H<sub>3.0</sub> = The Yunnan hop is not a separate species but a variety of the common hop, *H. lupulus*. H<sub>3.1</sub> = The Yunnan hop is a distinct species (Small 1978).



**Fig. 3.1.** The three recognized species of *Humulus*. Photo A is *H. lupulus* growing up a 3 meter trellis. Photo B is a close up of the female flowers of the plants in Photo A. Photo C is the habitat of *H. scandens* climbing over a small hedge approximately 1.5 meters tall. Photo D is a close up of the female flowers of the plants in Photo C, which have smaller bracts than *H. lupulus*. Photo E is the Yunnan hop, *H. yunnanensis*, with larger bracts than *H. lupulus* > *H. scandens*. Photos A-D were taken in the botanical garden at the Institute of Botany, Chinese Academy of Sciences in Beijing. Photo E is a type specimen held at the Royal Botanic Garden Edinburgh Herbarium (E) (available online at the URL: <<http://data.rbge.org.uk/herb/E00275605>>).



**Fig. 3.2.** Collection of *H. lupulus* var. *cordifolius* Maxim. (syn: *Humulus cordifolius* Miq.) grown in the botanical garden at the Kunming Institute of Botany, CAS.

## **Materials and Methods**

Fresh plant material (leaves/flowers) collected in the field, cultivated hop yards, purchased online, or samples sent by collaborators were preserved in silica desiccant for later DNA extraction. Pressed voucher specimens were also prepared for collection and deposit in HAW or KUN. Additionally, herbarium specimens were destructively sampled to produce DNA samples (Table 3.1). Putative wild hop samples from China include collections of *H. lupulus* from Xinjiang and Yunnan, as well as *H. scandens* from Beijing and Zhejiang Province. A total of four *H. yunnanensis* specimens were included from destructive sampling permission from KUN.

**Table 3.1.** Plant specimens sampled for DNA. Samples were taken from leaves unless noted. This table includes the same samples as in Table 2.2 of this dissertation, as well as the newly added samples from China denoted with a \* symbol towards the end of the table.

Sample (H=Humulus)	DNA #	Gender	Locality	Collection ; Date ; Herbarium ; Note
Cannabis_sp	152	No flowers	Lawai Valley, Koloa District, Kauai, Hawaii, U.S.A.	D.H. Lorence #7204 and G. Nace ; 14 May 1992 ; PTBG#015426 and PTBG#021494
H_lupulus_Hallertau	3	Female	Hops pellets purchased from Adventures in Homebrewing, Taylor, Michigan, U.S.A.	JRBoutain #303
H_lupulus_StMary	12	Female	St. Mary's School, Lake Leelanau, Michigan, U.S.A.	JRBoutain #304
H_lupulus	13	Female	West of Zarrentin, Götting. Schleswig-Holstein, Germany	L. Holm-Nielsen and K. Larsen #285 ; 25.8.1976 ; BRIT/AAU
H_lupulus_cf_lupuloides	14	Female	Mogollon Mountains, New Mexico, U.S.A.	R.D. Worthington #7604 ; 22 Aug. 1981 ; BRIT
H_lupulus_lupuloides	15	Female	Rockingham County, North Carolina, U.S.A.	A.E. Radford #18614 ; 28 Sept. 1956 : BRIT/VDB #13063
H_lupulus_neomexicanus	16	Female	Gila National Forest, Mogollon, Catron County, New Mexico, U.S.A.	D. Demaree #48609 ; 7-25-1963 ; BRIT/SMU
H_lupulus_cf_pubescens	17	Female	Benton County, Arkansas, U.S.A.	E.B. Smith #3856 ; 23 Sept. 1984 ; BRIT/VDB

<b>Table 3.1.</b> continued.				
H_lupulus_Nugget	21	Female	Sutton's Bay, Michigan, U.S.A.	JRBoutain #305
H_lupulus_Cascade	22	Female	Sutton's Bay, Michigan, U.S.A.	JRBoutain#306 ; leaf extracted
H_lupulus_Cascade	23	Female	Sutton's Bay, Michigan, U.S.A.	JRBoutain#306 ; flower extracted
H_lupulus_Waldy	24	cf. Female? ; No flowers when collected	Sutton's Bay, Michigan, U.S.A.	Michigan State Agricultural Extension Unit ; Waldy is a putative wild hop from Idaho that is approximately 80 years old. No herbarium specimen was collected.
H_lupulus_Bling	25	cf. Female? ; No flowers when collected	Sutton's Bay, Michigan, U.S.A.	Michigan State Agricultural Extension Unit. Bling is a putative Michigan wild hop. No herbarium specimen collected.
H_lupulus_UnknownX	26	Female	Michigan; N 44°57.598, W 85°29.648	JRBoutain#307
H_lupulus_UnknownY	27	Female	Michigan; N 44°57.598, W 85°29.648	JRBoutain#308
H_lupulus_UnknownZ	28	Female	Michigan; N 44°57.598, W 85°29.648	JRBoutain#309
H_lupulus_Claire	29	Female	Riverview, Michigan	JRBoutain#310
H_lupulus_TCBohemia	30	Female	Traverse City, Michigan	JRBoutain#311
H_lupulus_Perle	32	Female	Michigan; N 44°53'4", W 85°40'45"	JRBoutain#312
H_lupulus_Williamette	33	Female	Michigan; N 44°53'4", W 85°40'45"	JRBoutain#313

---

**Table 3.1.** continued

H_lupulus_Goldings	34	Female	Michigan; N 44°53'4", W 85°40'45"	JRBoutain#314
H_lupulus_Tettanger	35	Female	Michigan; N 44°53'4", W 85°40'45"	JRBoutain#315
H_lupulus_Centennial	36	Female	Michigan; N 44°53'4", W 85°40'45"	JRBoutain#316
H_lupulus_Cascade	37	Female	Michigan; N 44°53'4", W 85°40'45"	JRBoutain#317
H_lupulus_BrewersGold	38	Female	Michigan; N 44°53'4", W 85°40'45"	JRBoutain#318
H_lupulus_Recycler	30	cf. Female? ; No flowers when collected	Michigan; N 44°53'4", W 85°40'45"	JRBoutain#319
H_lupulus_Galena	40	Female	Michigan; N 45°4'9", W 85°36'0"	JRBoutain#320
H_lupulus_Chinook	41	Female	Michigan; N 45°4'9", W 85°36'0"	JRBoutain#321
H_lupulus_lupuloides	42	Female	Southern Ottawa, Canada	JRBoutain#322; Sent by E. Small from clone of original wild population on bank of Black Rapids Creek at Merivale Road.
H_lupulus_lupulus	43	Female	Eastern Ottawa, Canada	JRBoutain#323; Sent by E. Small from clone of old brewery cultivar growing by roadside (now ruderal). Unknown cultivar is originating and persisting from nearby mid- 20 <sup>th</sup> century brewery hop plantation.

---

**Table 3.1.** continued.

H_lupulus_SorachiAce	44	Female	Whole leaf hops purchased from Freshops, Oregon, U.S.A.	JRBoutain#324
H_lupulus_Saaz	45	Female	Whole leaf hops purchased from Freshops, Oregon, U.S.A.	JRBoutain#325
H_lupulus_MtHood	46	Female	Whole leaf hops purchased from Freshops, Oregon, U.S.A.	JRBoutain#326
H_lupulus	47	Male	Osage County, Kansas, U.S.A.	G. Tucker #6185 ; 6 Aug. 1967 ; BRIT/SMU
H_scandens	48	Female on sheet sampled	Peoria County, Illinois, U.S.A.	S.R. Hill #28569 with T. Kompare and P. Tessene ; 28 Aug. 1996 ; BRIT
H_scandens	50	Female	Richardson County, Nebraska, U.S.A.	R.B. Kaul #7493 and S. Rolfsmeier ; 28 Sept. 1996 ; BRIT
H_scandens	51	No flowers	Pendleton County, Kentucky, U.S.A.	M. Whitson 2007-0013 3/5 with L. Trauth and A. Tullis ; 28 July 2007 ; BRIT
H_scandens	52	Female	Clarion, Wright County, Iowa, U.S.A.	J. Ross and A.P. Bowman ; 16 Oct. 1944 ; BRIT/SMU
H_lupulus_Unknown1	54	No flowers	Michigan, U.S.A.	JRBoutain#327
H_lupulus_Unknown2	55	Female	Michigan, U.S.A.	JRBoutain#328
H_lupulus_Unknown3	56	Female	Michigan, U.S.A.	JRBoutain#329
H_lupulus_Unknown4	57	Female	Michigan, U.S.A.	JRBoutain#330

---

**Table 3.1.** continued.

H_lupulus_Unknown5	58	Female	Michigan, U.S.A.	JRBoutain#331
H_lupulus_Unknown6	59	Female	Michigan, U.S.A.	JRBoutain#332
H_lupulus_Unknown7	60	Female	Michigan, U.S.A.	JRBoutain#333
H_lupulus_Unknown8	61	Female	Michigan, U.S.A.	JRBoutain#334
H_lupulus_Unknown9	62	Female	Michigan, U.S.A.	JRBoutain#335
H_lupulus_Unknown10	63	Female	Michigan, U.S.A.	JRBoutain#336
H_scandens	67	Female	Dutchess County, New York, U.S.A.	S.R. Hill #28479 ; 13 Aug. 1996 ; BRIT
H_scandens	76	No flowers	Calhoun County, Alabama, (33°38'56" N, 85°49'55" W), U.S.A.	D.D. Spaulding #11,497 and E. "Skeeter" Cole, Jr. ; 31 July 2002 ; BRIT/VDB
H_lupulus_pubescens	81	Female	North of East Peoria, Tazewell County, Illinois, U.S.A.	V.H. Chase #14819 ; 2 Sept. 1959 ; BRIT/VDB #42841
H_lupulus_pubescens	82	Female	Plattsmouth, Cass County, Nebraska, U.S.A.	D. Demaree #54175 ; 8-3-1966, BRIT/VDB #80336
H_lupulus_pubescens	83	Female	Plattsmouth, Cass County, Nebraska, U.S.A.	D. Demaree #54175 ; 8-30-1966 ; BRIT/VDB #59781
H_lupulus_pubescens	84	Female	Cass County, Missouri, U.S.A.	N.C. Henderson #67-1565 ; 29 Aug. 1967 ; BRIT/VDB #51530
H_scandens	85	Male	Calhoun County, Alabama (33°38'12" N, 85°49'58" W), U.S.A.	D.D. Spaulding #12,051 ; 18 Sept. 2003 ; BRIT/VDB

---

<b>Table 3.1.</b> continued.				
H_scandens	89	Male	South Mountain Reservation, Millburn, Essex Co. County, New Jersey, U.S.A.	K. Barringer #10646 ; 19 Aug. 2003 ; BRIT/VDB
H_lupulus	90	Male	Williamette Park, Corvallis, Benton County, Oregon, U.S.A.	R.R. Halse #5436 ; 30 July 1998 ; BRIT/VDB
H_scandens	95	Female on sheet sampled	Jackson Park, Chicago, Illinois, U.S.A.	A.P. Anderson #2024 ; Aug. 1912 ; BRIT/SMU
H_lupulus	96	Female	Grobbendovk arboretum, Prov d'Anvers, Belgium	J.E. De Langhe ; Sept. 1964 ; BRIT/SMU
H_lupulus	97	Male	4 km north of Gnissau village and west of Ahrensbök municipality, Ostholstein in Schleswig-Holstein, Germany	L. Holm-Nielsen, I. Nielsen, S.P. Pinnerup #223 ; 9.8.1974 ; BRIT/SMU/AAU
H_scandens	104	Male	Arnold Arboretum, Harvard University, U.S.A.	G.P. DeWolf and P. Bruns #2179 ; 21 Sept. 1967 ; BRIT/SMU
H_scandens	121	Male	Brunswick, Chariton County, Missouri, U.S.A.	H.N. Moldenke #23160 ; 16 Oct. 1966 ; BRIT/SMU

<b>Table 3.1.</b> continued.				
H_lupulus_Saaz	153	Female	Whole leaf hops purchased from Adventures in Homebrewing, Taylor, Michigan, U.S.A.	JRBoutain#337
*H_lupulus_cordifolius	11	Monecious? - Female specimen collected	Nursery at Botanical Garden of the Kunming Institute of Botany, CAS, Yunnan, China	JRBoutain#338, #JRB_KIB01, and #JRB_KIB02 ; 7 Sept. 2010 ; KUN#0935629 and KUN#0935630 ; 5 Oct. 2012; KUN#1014725
*H_lupulus_cordifolius	249	Monecious? – Male specimen collected	Nursery at Botanical Garden of the Kunming Institute of Botany, CAS, Yunnan, China	JRBoutain#338, #JRB_KIB01, and #JRB_KIB02 ; 7 Sept. 2010 ; KUN#0935629 and KUN#0935630 ; 5 Oct. 2012; KUN#1014725
*H_lupulus_Xinjiang	156	cf. Female? ; No flowers	Ahlemala, Xinyuan, Yi Li, Xinjiang Province, China	JRBoutain#340 ; 01 July 2011 ;
*H_scandens	216	Female	Jiangxi Province, China	Chuang#2673? ; 27 Oct. 1968? ; KUN#525227 ; Chung Shan Botanical Garden collection
*H_scandens	217	Male	Shicheng County, Jiangxi Province, China	#4800 ; 30 August 30 1969 ; KUN#525230 ; Lushan Botanical Herbarium
*Humulus_sp	234	Male	Fuyuan County, Yunnan Province, China	Red River plant investigation group 892462 ; 23 June 1989 ; KUN#604961

<b>Table 3.1.</b> continued.				
* <i>Humulus</i> _sp	235	Male	Fuyuan County, Yunnan Province, China	Red River plant investigation group 892462 ; 23 June 1989 ; KUN#604962
* <i>H_scandens</i> _Tsinghua	240	cf. Female? ; No flowers	Tsinghua University, Beijing, China	JRBoutain#341; 17 June 2011 ;
* <i>H_scandens</i>	242	Female	Linan County, Zhejiang Province, China	Jie Liu #10690 & Zeng-Yuan Wu ; 6 Oct. 2010 ;
* <i>H_scandens</i>	244	cf. Female?	Linan County, Zhejiang Province, China	Jie Liu #10806 & Zeng-Yuan Wu ; 8 Oct. 2010 ;
* <i>H_scandens</i>	246	cf. Female?	Linan County, Zhejiang Province, China	Jie Liu #10681 & Zeng-Yuan Wu ; 6 Oct. 2010 ;
* <i>H_yunnanensis</i>	171	Female	Xinjiang Province?, China	Chen 7612 ; 23 Oct. 1976 ; KUN #525339 ; collection possibly indigenous to Xinjiang Province?
* <i>H_yunnanensis</i>	176	Female	Kunming Botanical Garden, Yunnan Province, China	Tao Yu Zhi and Chiou Bing Yun 771302 ; 22 Nov. 1977 ; KUN #525382 ; incorrect label of <i>Humulus lupulus</i>
* <i>H_yunnanensis</i>	232	Female	Yunnan Province, China	Tian Daike 9764 ; 04 Dec. 1997 cf. 1987? ; KUN#96812
* <i>H_yunnanensis</i>	233	Female	Yunnan Province, China	Tian Daike 9764 ; 04 Dec. 1997 cf. 1987? ; KUN#96813

Approximately 0.03-1.0 g of dried leaf, flower, or herbaria material was used from each sample ( $n=78$ ) for total genomic DNA extractions with: 1) Qiagen DNeasy Plant Mini Kit at the University of Hawai‘i at Mānoa or 2) a modified CTAB procedure (Doyle 1991, Doyle and Dickson 1987, Doyle and Doyle 1987, 1990) at the Plant Germplasm and Genomics Center, Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences. Nuclear ribosomal DNA (rDNA) (ITS2; Murakami 2000, 2001, Murakami et al. 2003) and chloroplast DNA (cpDNA) (*petL-psbE*; Shaw et al. 2007) regions were amplified via a standard polymerase chain reaction (PCR) procedure modified from Keeley et al. (2007), Murakami (2000 and 2001), and Shaw et al. (2007) (Table 3.2).

**Table 3.2.** DNA regions and primer pairs used for PCR amplification and sequencing. The ITS2 primers are from Murakami (2000), and the *petL-psbE* primers are from Shaw et al. (2007).

Region	Primer Name	Sequence (5' to 3')
ITS2	ITS2F	TTGCAGAATCCCGTGAACCATCG
	ITS2R	CCAAACAACCCGACTCGTAGACAGC
<i>petL-psbE</i>	<i>petL</i>	AGTAGAAAACCGAAATAACTAGTTA
	<i>psbE</i>	TATCGAATACTGGTAATAATATCAGC

The modified CTAB protocol follows. In 2 mL microcentrifuge tubes, a pinch of PVP (polyvinylpyrrolidone) was added to each sample after maceration into a fine powder with liquid nitrogen. For a 10 minute incubation period on ice, 900  $\mu$ L of carbohydrate removal buffer (6.2% PVP, 2% D-glucose anhydrous, ddH<sub>2</sub>O) was added to the powdered sample and mixed frequently to keep the particles suspended. After incubation, the 2 mL tubes were centrifuged at 9,000 RPM for 8 minutes, then the liquid was discarded and the pellet retained. The buffer composed of 1 mL of 4X CTAB (hexadecyltrimethylammonium bromide) buffer containing 2  $\mu$ L DTT (dithiothreitol) preheated in a water bath to 65°C, then added to each 2 mL extraction tube with the pellet and heated for 1 hour at 65°C in a water bath, regularly flicking the tubes to keep the particles in suspension. Afterward, the samples were removed from the water bath and allowed to cool to room temperature, before adding 500  $\mu$ L of chloroform:isoamyl alcohol in a 24:1 ratio to each extraction tube. The tubes were gently mixed for 8 minutes by continuously inverting the tubes. Each extraction tube was centrifuged at 11,000 RPM

for 8 minutes, and the supernatant transferred to a clean 2.0 mL microcentrifuge tube. Again, an additional 500  $\mu$ L of chloroform:isoamyl alcohol in a 24:1 ratio was added to each extraction tube, gently mixed for 8 minutes by continuously inverting the tubes, and centrifuged at 11,000 RPM for 8 minutes. The supernatant was transferred to a clean 2.0 mL tube where the DNA was precipitated by adding 600  $\mu$ L 0°C isopropanol. The tubes were gently inverted to ensure proper mixing and then left at -20°C for 4-5 days to precipitate any degraded DNA fragments at compromising the potential overall DNA quality (Almakarem et al. 2012, Rogers and Bendich 1985, 1994). Once the extraction samples were removed from -20°C, the tubes were centrifuged at 10,000 RPM for 10 minutes. The supernatant was carefully removed to not disturb or dislodge the DNA pellet in the bottom of the tube. Approximately, 1 mL of 70% ethanol was added to the tube and agitated vigorously to release and wash the pellet on the bottom. The DNA and 70% ethanol in the tube was left for 30 minutes at room temperature, and then centrifuged at 10,000 RPM for 10 minutes. Afterward, the 70% ethanol was removed without dislodging the DNA pellet, and 1 mL of 100% ethanol was added to the tube to release the pellet from the bottom. The DNA was washed and left in 100% ethanol for 10 minutes at room temperature. After centrifugation at 13,000 RPM for 5 minutes and removing the supernatant, the tube was invert to allow any remaining ethanol to drain away, taking care not to dislodge the pellet. Each pellet was placed in a vacuum centrifuge at 44°C for 5 minutes or until dry. The dried DNA pellet was dissolved in 30  $\mu$ L of elution buffer (TE), mixed thoroughly at room temperature to dissolve the DNA completely, and stored until further use at -20°C.

Each extracted DNA template was assigned a position on a 96-well plate for initial, purification, and sequencing PCR reactions. For PCR amplification, each reaction was carried out in 19.4  $\mu$ L of a PCR mixture consisting of 9.2  $\mu$ L sterile water, 9.2  $\mu$ L of 2 X Taq PCR MasterMix (composed of 0.1 U Taq Polymerase/ $\mu$ L, 500  $\mu$ M dNTP each, 20 mM Tris-HCl (pH 8.3), 100 mM KCl and 3 mM MgCl<sub>2</sub>), 0.5  $\mu$ L forward primer, 0.5  $\mu$ L reverse primer, and 0.8  $\mu$ L of extracted DNA template. If necessary to generate sufficient PCR products for DNA sequencing, the PCR mixture was adjusted to the methods used in Chapter 2 of this dissertation. PCR amplification was carried out on a

Veriti 96 Well Thermal Cycler (Applied Biosystems). Thermal cycler settings for ITS2 (Murakami 2000, 2001) were an initial preheating at 94°C for 2 minutes; followed by 34 cycles of 1 minute denaturation at 94°C, 1 minute annealing at 50°C, and primer extension for 2 minutes at 72°C; and ending with an additional 4 minute extension at 72°C before held at 4°C. For *petL-psbE* (Shaw et al. 2005, 2007), the thermal cycler program was template denaturation at 80°C for 5 minutes; followed by 35 cycles of denaturation at 95°C for 1 minute, primer annealing at 50°C for 1 minute, followed by a ramp of 0.3°C /s to 65°C (delta 0.3°C/delta 0.01 second), and primer extension at 65°C for 4 minutes; and ending with a final extension step of 5 minutes at 65°C before held at 4°C.

A single product for each 25 µL PCR was determined using 1% agarose gel electrophoresis and a 100 bp DNA ladder (HyperLadder IV by BIOLINE). PCR products were stored at 4°C until purified prior to sequencing with 2 µL ExoSAP-IT (USB Products, Affymetrix, Inc., Cleveland, Ohio, USA) for 5 µL PCR product. This method of purification without loss of PCR products (no filtration, precipitation, or washing needed) is especially important for DNA extracted from herbarium specimens, which are sometimes only weakly amplified and barely yield sufficient PCR product for sequencing (Keeley et al. 2007). The ExoSAP-IT purification reaction mixture was ran on the Veriti thermal cycler with an incubation at 37°C for 15 minutes, 80°C for 15 minutes, then held at 4°C.

For sequencing amplification, each reaction was carried out in 6.03 µL of a sequencing PCR mixture consisting of 3.75 µL sterile water, 0.3 µL BigDye terminator, 0.5 µL primer, 0.03 µL BSA, 1.05 µL sequence buffer, and 0.4 µL of purified PCR template. Sequencing amplification was carried out on a Veriti thermal cycler set to an initial denaturation of 30 seconds at 94°C; followed by 32 cycles of denaturation at 96°C for 10 seconds, annealing at 50°C for 5 seconds, and sequence extension at 60°C for 4 minutes; and then held at 4°C.

Before sequencing using standard Sanger protocols on the ABI 3730 xl instrument, 20 µL of preparation buffer for sequencing was added to the sequencing amplified PCR products on the 96-well plate, left for 30-60 minutes on ice, then the plate

was centrifuged at 3,700 RPM for 10 minutes. Plates were drained then centrifuged at 300 RPM for 4 minutes. Plates were washed with 150  $\mu$ L ice cold 70% ethanol, then centrifuged at 3,700 RPM for 10 minutes. The plates were drained, centrifuged at 300 RPM for 4 minutes, then dried in the Veriti thermal cycler at 75-80°C for 2-10 minutes with the lid off to allow complete ethanol evaporation. A final addition of 10  $\mu$ L HighDye was incubated on the thermal cycler at 95°C for 2 minutes with lid closed. Immediately after two minutes, the plate was put on ice to quickly cool and stored at 4°C before added to lab queue on the ABI 3730 xl instrument at the Plant Germplasm and Genomics Center, Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences. The resulting sequences from both strands of each PCR product were examined, compared, corrected, and edited into contigs using Sequencher® 4.10.1 (Ann Arbor, Michigan, USA). Multiple consensus sequences for each DNA region were added to the nucleotide matrix generated in Chapter 2 of this dissertation and aligned into a nucleotide data matrix using the program MUSCLE (Edgar 2004) implemented in the MEGA5.2 software package (Tamura et al. 2011). Default settings in MEGA 5.2, DnaSP v5.10.01 (Rozas 2009), and CLC Genomics Workbench v7 (CLC bio Inc., Aarhus, Denmark) were used for comparisons of nucleotide sequences, haplotype reconstructions, and phylogenies.

Phylogenies for each separate rDNA and cpDNA matrix and for the combined nucleotide matrix were constructed using MEGA5.2 and BEAST 2. For maximum likelihood (ML) phylogenies using MEGA5.2, each nucleotide matrix was used to find the optimal model of sequence evolution for ML by generating an automatic neighbor-joining tree, using all sites, and applying a branch swap filter of very strong. The resulting model chosen for ML analysis had the lowest BIC score (Bayesian Information Criterion), which is considered to be the best for an accurate description of the substitution pattern. Generation of the ML phylogenies was conducted using the maximum number of threads available on the computer with the following parameters: 1) all sites, 2) the models of evolution were Tamura 3-parameter+I for ITS2, Tamura 3-parameter for *petL-psbE*, and Hasegawa-Kishino-Yano+G with five discrete gamma categories for the combined matrix, 3) nearest-neighbor-interchange, 4) an initial

neighbor-joining tree made automatically, 5) applying a branch swap filter of very strong, and 6) 1000 bootstrap replicates.

Generation of a Bayesian inference (BI) phylogeny using BEAST 2.0.2 (Bouckaert et al. 2013) was done by exporting each aligned nucleotide matrix from MEGA5.2 as a Nexus file (PAUP 4.0) with the following: 1) all sites were displayed, 2) the data type was changed from nucleotide to DNA, and 3) the file was saved with the extension .nex. Next, the .nex file alignment is imported to BEAUti (included in the BEAST 2 software package) where the parameters set for the Site Model tab included a substitution rate of 1.0, gamma category count of 4, shape 1.0 with the estimate box checked, proportion invariant 0.1 with the estimate box checked, and the add-on substitution model of Reversible-jump Based substitution (v.1.0.3) (RB). The RB allows the MCMC chain to switch between nucleotide substitution models to search for the best fit model for the data set. For the Clock Model tab, strict clock was selected. The default parameters in the Priors tab were left at estimate for the Yule Model tree. The MCMC Chain parameters were: 1) length was set at 100,000,000 cycles, 2) stored every -1, 3) Pre Burnin 0, and 4) logged every 1000 cycles to generate 100,001 trees. After the priors were set, the file was saved in BEAUti with the standard .xml extension. After opening the .xml in BEAST 2 and selecting the automatic thread pool size, a .log file was generated by BEAST 2 after the run came to completion. The .log file output from BEAST 2 was viewed with Tracer v1.5 (available at URL: <<http://tree.bio.ed.ac.uk/software/tracer/>>), which depicts the outcome of parameters set in BEAUti with each tree sampled along the MCMC chain for likelihood and posterior effective sample size (ESS) (e.g., need ESS to be in black font or >200). Using the TreeAnnotator (included in the BEAST 2 software package) to combine only the last 11,001 trees sampled after a burnin of 89,000 trees on the MCMC chain for final likelihood and posterior probability estimates, the .trees file was saved with the extension .tree. The annotated .tree file output from TreeAnnotator was viewed using FigTree v1.4.0 (available at URL: <<http://tree.bio.ed.ac.uk/software/figtree/>>). Topologies for the ITS2, *petL-psbE*, and combined data set for both ML and BI phylogenies were compared,

and if similar, the bootstrap values and posterior probabilities for each main node were added to a final consensus tree.

## **Results**

Extraction, amplification, sequencing, and alignment were unproblematic for field collected, purchased samples, and herbarium material for the rDNA and cpDNA regions studied in *H. lupulus*, *H. scandens*, *H. yunnanensis*, and *Cannabis*. The final lengths of the aligned ITS2 and *petL-psbE* nucleotide matrices were 473bp and 1,078bp, respectively, for a combined total length of 1,551bp. Table 3.3 summarizes the number of conserved, variable, parsimony-informative, and singleton sites for ITS2, *petL-psbE*, and the combined nucleotide matrices. The estimates of average evolutionary divergence over all sequence pairs for the number of base differences per sequence were 9.175 nucleotide differences for ITS2 (p-distance=0.020), 8.537 nucleotide differences for *petL-psbE* (p-distance=0.008), and 17.712 for the combined nucleotide matrices (p-distance=0.012). Between groups mean distance estimates ranged from 0.020 to 0.099 for ITS2, 0.006 to 0.029 for *petL-psbE*, and 0.010 to 0.050 for the combined data set (Table 3.4). As for ITS2 alone, a total of 12-14 gaps were observed between the in-group of *Humulus* spp. and the out-group of *Cannabis*. In addition, 51-60 differences were observed between the in-group and out-group. Furthermore, 0-2 gaps and 0-30 differences were found within the *Humulus* spp. sampled. For *petL-psbE* alone, a total of 38-51 gaps and 68-78 differences were found between *Humulus* spp. and *Cannabis*. The number of gaps and differences within the *Humulus* spp. sampled were 0-31 and 0-46, respectively. The number of haplotypes (h) discovered for ITS2, *petL-psbE*, and the combined data set corresponded to h=13 (Hd: 0.7695), h=9 (Hd=0.7416), and h=20 (Hd=0.8652), accordingly (Table 3.5).

**Table 3.3.** Number of conserved, variable, parsimony-informative, and singleton sites for aligned DNA sequences conducted in MEGA5.2.

<b>Region</b>	<b># conserved</b>	<b># variable</b>	<b># parsimony-informative</b>	<b># singletons</b>
ITS2	411/473	62/473	34/473	28/473
<i>petL-psbE</i>	1018/1078	41/1078	22/1078	19/1078
combined	1429/1551	103/1551	56/1551	47/1551

**Table 3.4:** Estimates of evolutionary divergence over sequence pairs between groups conducted in MEGA5.2

<b>Species 1</b>	<b>Species 2</b>	p-distance		
		<b>ITS2</b>	<b><i>petL-psbE</i></b>	<b>combined</b>
<i>H. lupulus</i>	<i>Cannabis</i> sp.	0.084	0.027	0.045
<i>H. lupulus</i>	<i>H. scandens</i>	0.044	0.017	0.025
<i>Cannabis</i> sp.	<i>H. scandens</i>	0.099	0.029	0.050
<i>H. lupulus</i>	<i>H. yunnanensis</i>	0.032	0.013	0.019
<i>Cannabis</i> sp.	<i>H. yunnanensis</i>	0.084	0.027	0.045
<i>H. scandens</i>	<i>H. yunnanensis</i>	0.020	0.006	0.010

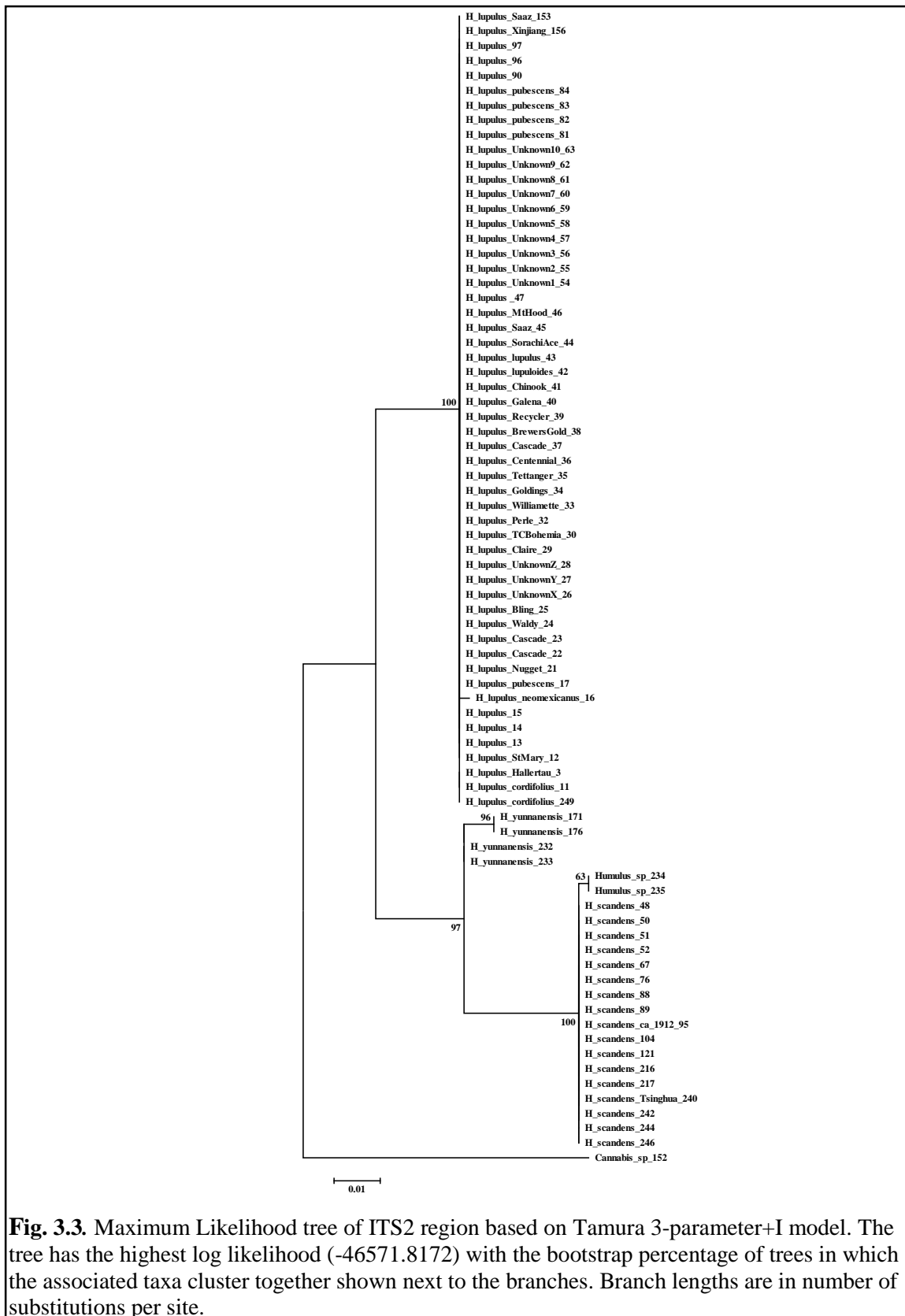
**Table 3.5.** Haplotype estimates for ITS2, *petL-psbE*, and the combined data set conducted in DnaSP v5.10.01. Prior to the haplotype reconstructions for ITS2 and combined data sets, the nucleotide data matrix was reconstructed using PHASE (Stephens et al. 2001, Stephens and Donnelly 2003) to incorporate the IUPAC nucleotide ambiguity codes that represent heterozygous sites. If the extractions numbers for the ITS2 and combined data sets are followed by \_1 or \_2, they signify two possible genotypes for the diploid sequences, otherwise the two sequences generated by PHASE are identical in that haplotype.

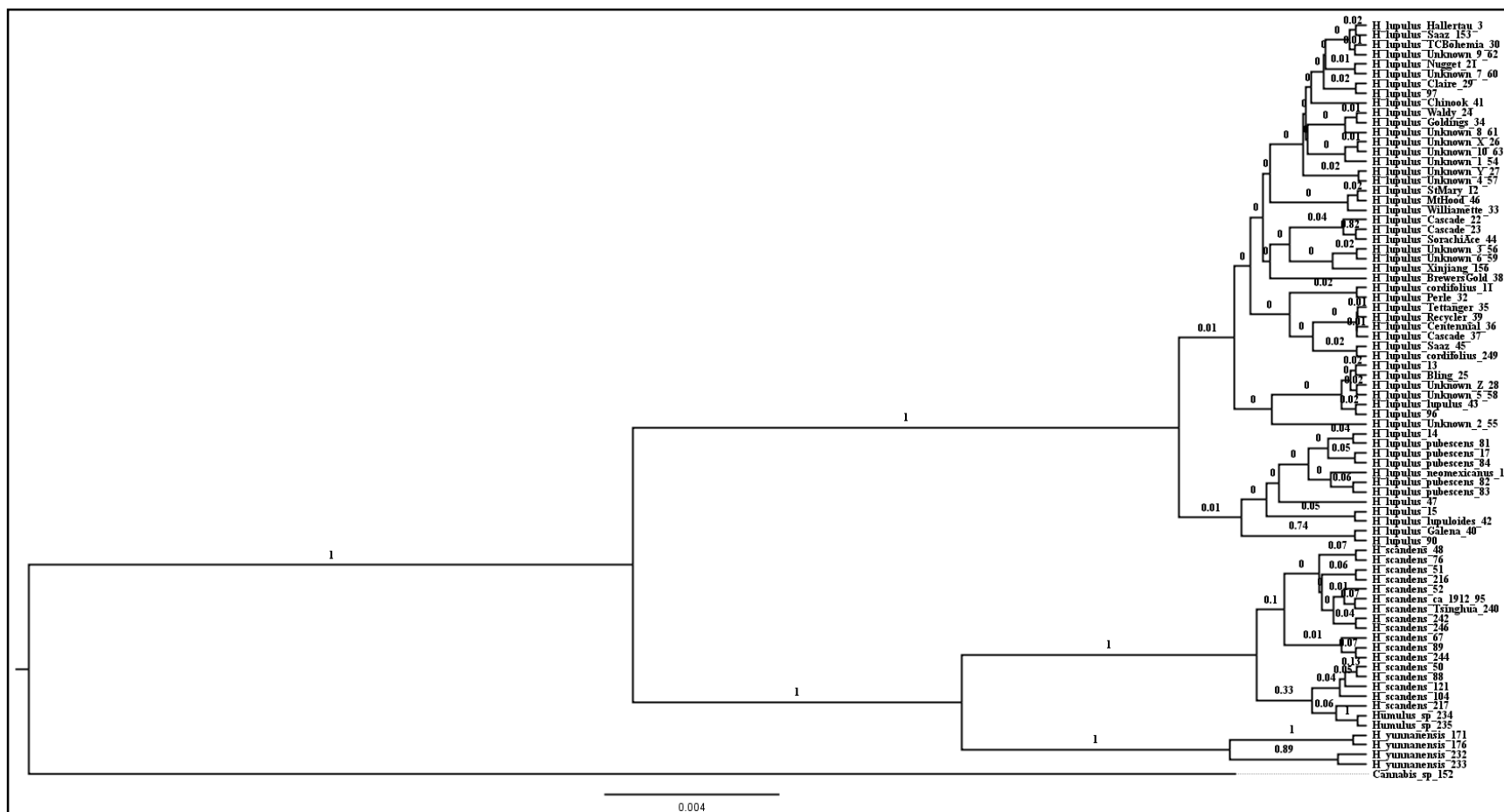
	region		
	ITS2	<i>petL-psbE</i>	combined
Number of haplotypes	13	9	20
Haplotype diversity	0.7695	0.7416	0.8652
Summary of haplotype distribution by extraction #	<b>Hap_1</b> ( <i>n</i> =68): 3, 11, 12, 13, 21_1, 22_1, 24_2, 25, 26, 27, 28, 29_1, 30_2, 32_2, 33_2, 34, 35, 36, 37_1, 38_1, 39_2, 41, 43, 45, 46, 54, 55_2, 56, 57, 58, 59, 60, 61_2, 62_2, 63_2, 96, 97, 153, 156, 249 <b>Hap_2</b> ( <i>n</i> =2): 152 <b>Hap_3</b> ( <i>n</i> =16): 14, 15, 16, 17, 42, 47, 82_1, 83_1, 84 <b>Hap_4</b> ( <i>n</i> =9): 21_2, 22_2, 23_2, 37_2, 38_2, 44_2, 81_2, 82_2, 83_2 <b>Hap_5</b> ( <i>n</i> =3): 23_1, 29_2, 44_1 <b>Hap_6</b> ( <i>n</i> =11): 24_1, 30_1, 39_1, 40, 55_1, 61_1, 62_1, 63_1, 81_1, 90_2	<b>Hap_1</b> ( <i>n</i> =30): 3, 12, 13, 22, 23, 24, 25, 26, 27, 28, 30, 32, 33, 34, 35, 37, 39, 43, 45, 46, 54, 55, 61, 62, 63, 90, 96, 97, 153, 156 <b>Hap_2</b> ( <i>n</i> =2): 11, 249 <b>Hap_3</b> ( <i>n</i> =1): 152 <b>Hap_4</b> ( <i>n</i> =22): 14, 15, 16, 17, 21, 29, 36, 38, 40, 41, 42, 44, 47, 56, 57, 58, 59, 60, 81, 82, 83, 84 <b>Hap_5</b> ( <i>n</i> =15): 48, 50, 51, 52, 67, 76, 88, 89, 95, 104, 121, 240, 242, 244, 246 <b>Hap_6</b> ( <i>n</i> =2): 216, 217 <b>Hap_7</b> ( <i>n</i> =2): 234, 235 <b>Hap_8</b> ( <i>n</i> =2): 171, 176 <b>Hap_9</b> ( <i>n</i> =2): 232, 233	<b>Hap_1</b> ( <i>n</i> =47): 3, 12, 13, 22_1, 24_2, 25, 26, 27, 28, 30, 32, 33, 34, 35, 37_1, 39_2, 43, 45, 46, 54, 55_2, 62, 63_2, 96, 97, 153, 156 <b>Hap_2</b> ( <i>n</i> =4): 11, 249 <b>Hap_3</b> ( <i>n</i> =2): 152 <b>Hap_4</b> ( <i>n</i> =16): 14, 15, 16, 17, 42, 47, 82_1, 83_1, 84 <b>Hap_5</b> ( <i>n</i> =17): 21_1, 29_1, 36-2, 38_1, 41, 56, 57, 58, 59, 60 <b>Hap_6</b> ( <i>n</i> =6): 21_2, 38_2, 44_2, 81_2, 82_2, 83_2 <b>Hap_7</b> ( <i>n</i> =3): 22_2, 23_2, 37 <b>Hap_8</b> ( <i>n</i> =1): 23_1 <b>Hap_9</b> ( <i>n</i> =8): 24_1, 30_1, 39_1, 55_1, 61_1, 62_1, 63_1, 90_2 <b>Hap_10</b> ( <i>n</i> =2): 29_2, 44_1 <b>Hap_11</b> ( <i>n</i> =3): 40, 81_1

**Table 3.5.** continued.

	<b>region</b>	
	<b>ITS2</b>	<b>combined</b>
Summary of haplotype distribution by extraction #	<b>Hap_7</b> ( <i>n</i> =21): 48, 51, 52, 67_2, 76, 89, 95, 216, 240, 242_2, 244, 246_2 <b>Hap_8</b> ( <i>n</i> =12): 50, 67_1, 88_2, 104, 121, 217, 242_1, 246_1 <b>Hap_9</b> ( <i>n</i> =1): 88_1 <b>Hap_10</b> ( <i>n</i> =1): 90_1 <b>Hap_11</b> ( <i>n</i> =4): 234, 235 <b>Hap_12</b> ( <i>n</i> =4): 171, 176 <b>Hap_13</b> ( <i>n</i> =4): 232, 233	<b>Hap_12</b> ( <i>n</i> =19): 48, 51, 52, 67_2, 76, 89, 95, 240, 242_2, 244, 246_2 <b>Hap_13</b> ( <i>n</i> =10): 50, 67_1, 88_2, 104, 121, 242_1, 246_1 <b>Hap_14</b> ( <i>n</i> =1): 88_1 <b>Hap_15</b> ( <i>n</i> =1): 90_1 <b>Hap_16</b> ( <i>n</i> =2): 216 <b>Hap_17</b> ( <i>n</i> =2): 217 <b>Hap_18</b> ( <i>n</i> =4): 234, 235 <b>Hap_19</b> ( <i>n</i> =4): 171, 176 <b>Hap_20</b> ( <i>n</i> =4): 232, 233

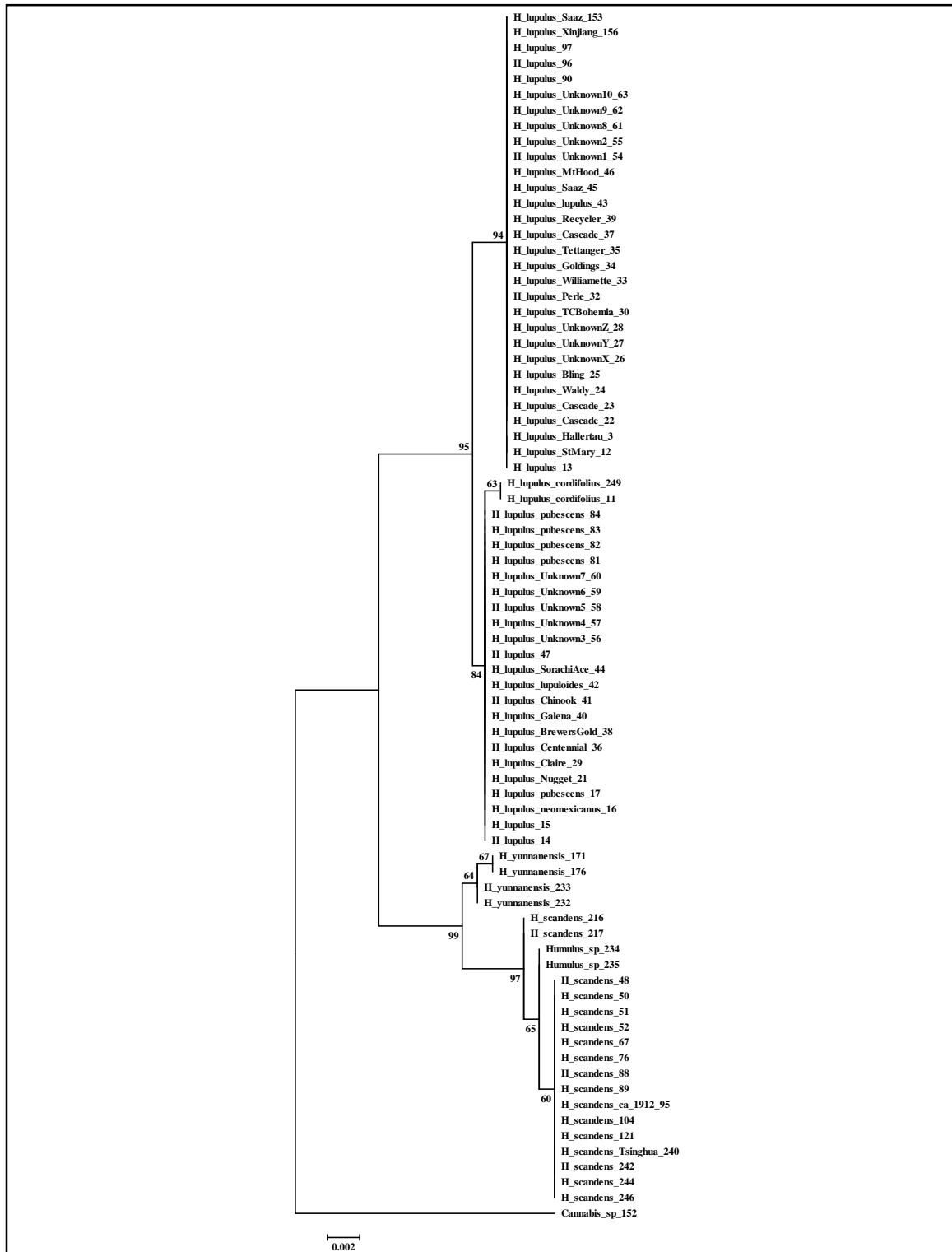
ML and BI phylogenies for ITS2 show three clades of *Humulus* with *Cannabis* as the outgroup (Figs. 3.3 and 3.4). The ML ITS2 tree has high support for a polytomy node for *H. lupulus* (100%). The *H. scandens* and *H. yunnanensis* clades are separated by a single node (97%). Both the *H. scandens* (63%) and *H. yunnanensis* (96%) clades have a single internal node. Similarly, the BI ITS2 tree shows high support at the main nodes for the *H. lupulus*, *H. scandens*, and *H. yunnanensis* clades (1.0). Furthermore for BI ITS2 tree, low support was found for the interior nodes of the *H. lupulus* clades (0-0.06), with the exceptions of the interior nodes at 0.74 supporting the samples H\_lupulus\_Galena\_40 and H\_lupulus\_90 and at 0.082 supporting the samples H\_lupulus\_Cascade\_23 and H\_lupulus\_SorachiAce\_44. Low support for the internal nodes of the *H. scandens* clades is similar to internal nodes of the *H. lupulus* clade (0-0.33). The only exception is an internal branch that includes two *H. scandens* samples at 1.0 posterior probability (i.e., Humulus\_sp\_234 and Humulus\_sp\_235). High support was found at the *H. yunnanensis* node (97% and 1.0). Two internal groups within the *H. yunnanensis* clade are highly supported: 1) the samples H\_yunnanensis\_232 and H\_yunnanensis\_233 at 97% and 0.89 and 2) the samples H\_yunnanensis\_171 and H\_yunnanensis\_176 at 96% and 1.0.



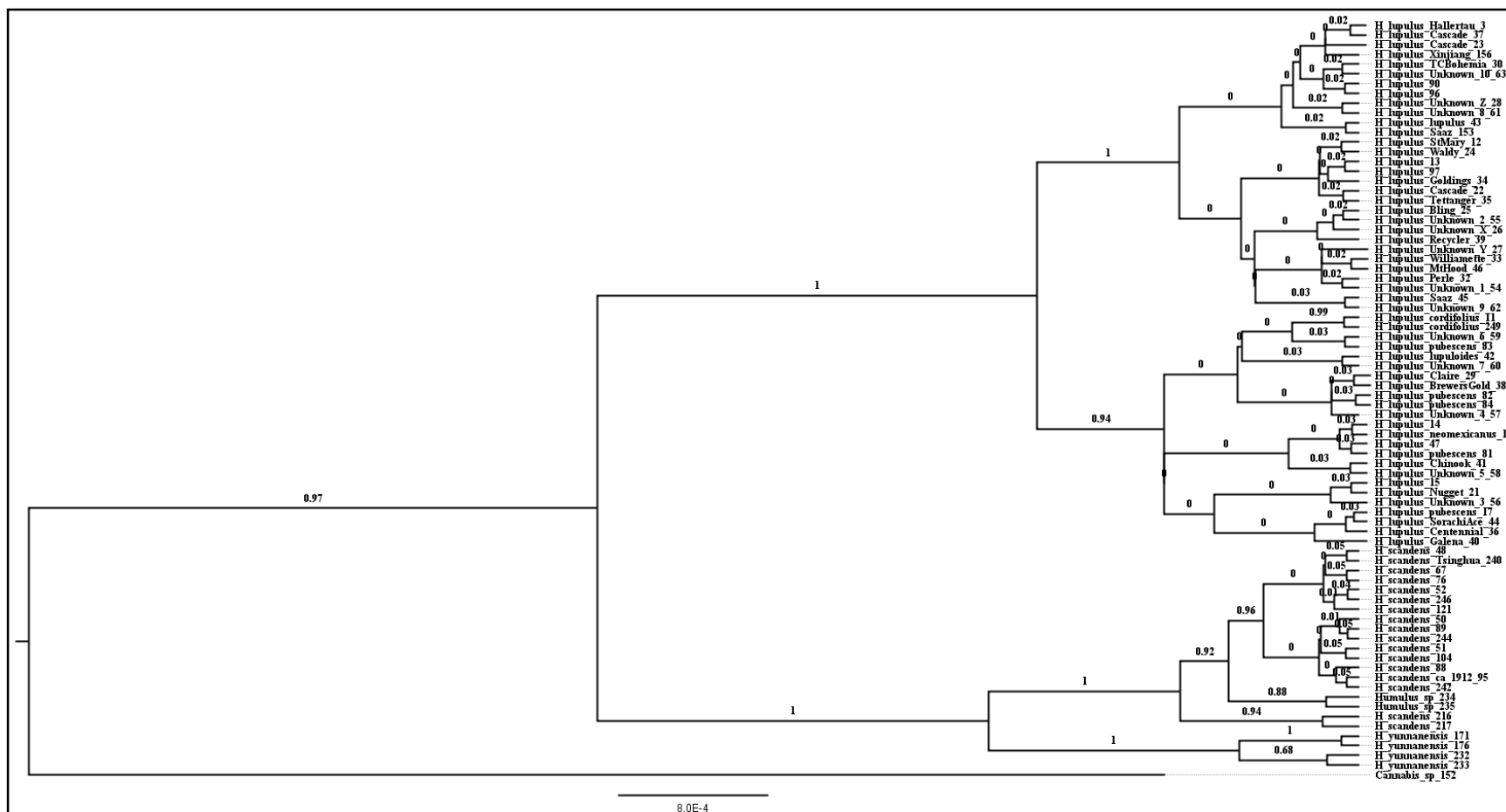


**Fig. 3.4.** Bayesian inference phylogeny of ITS2 depicted as a maximum clade credibility tree based on Reversible-jump Based substitution model. Posterior probabilities in which the associated taxa cluster together is shown next to the branches. Branch lengths are in number of substitutions per site. The highest log clade credibility is -380.295443163647.

ML and BI phylogenies for *petL-psbE* show three main clades of *Humulus* with *Cannabis* as the outgroup (Figs. 3.5 and 3.6). High bootstrap and posterior probability support was found for the *H. lupulus* clade. (95% and 1.0). A single node separates the *H. scandens* and *H. yunnanensis* clades (99% and 1.0). Within the *H. lupulus* clade, three groups formed in the ML tree with 94%, 84%, and 63% support at the nodes. For the BI tree of the *H. lupulus* clade, two main groups formed with 1.0 and 0.94 posterior probability support. Only one highly derived internal node at 0.99 separates the samples H\_lupulus\_cordifolius\_11 and H\_lupulus\_cordifolius\_249, while the other internal nodes have low support (0-0.03). Additionally, high bootstrap and posterior probability support was found for the *H. scandens* clade (97% and 1.0). The *H. scandens* clade has three further groups: one with high bootstrap and posterior probability support (97% and 0.94) and another two with slightly lower bootstrap but high posterior probability support (65%:0.88 and 60%:0.96). The internal nodes of both ML and BI trees for the *H. scandens* clade have low support (0% and 0-0.05). Lastly, medium bootstrap and high posterior probability support was found for the *H. yunnanensis* clade (64% and 1.0). The *H. yunnanensis* clade has two internal groups: 1) the samples H\_yunnanensis\_232 and H\_yunnanensis\_233 at 64 % and 0.68 and 2) the samples H\_yunnanensis\_171 and H\_yunnanensis\_176 at 67% and 1.0.

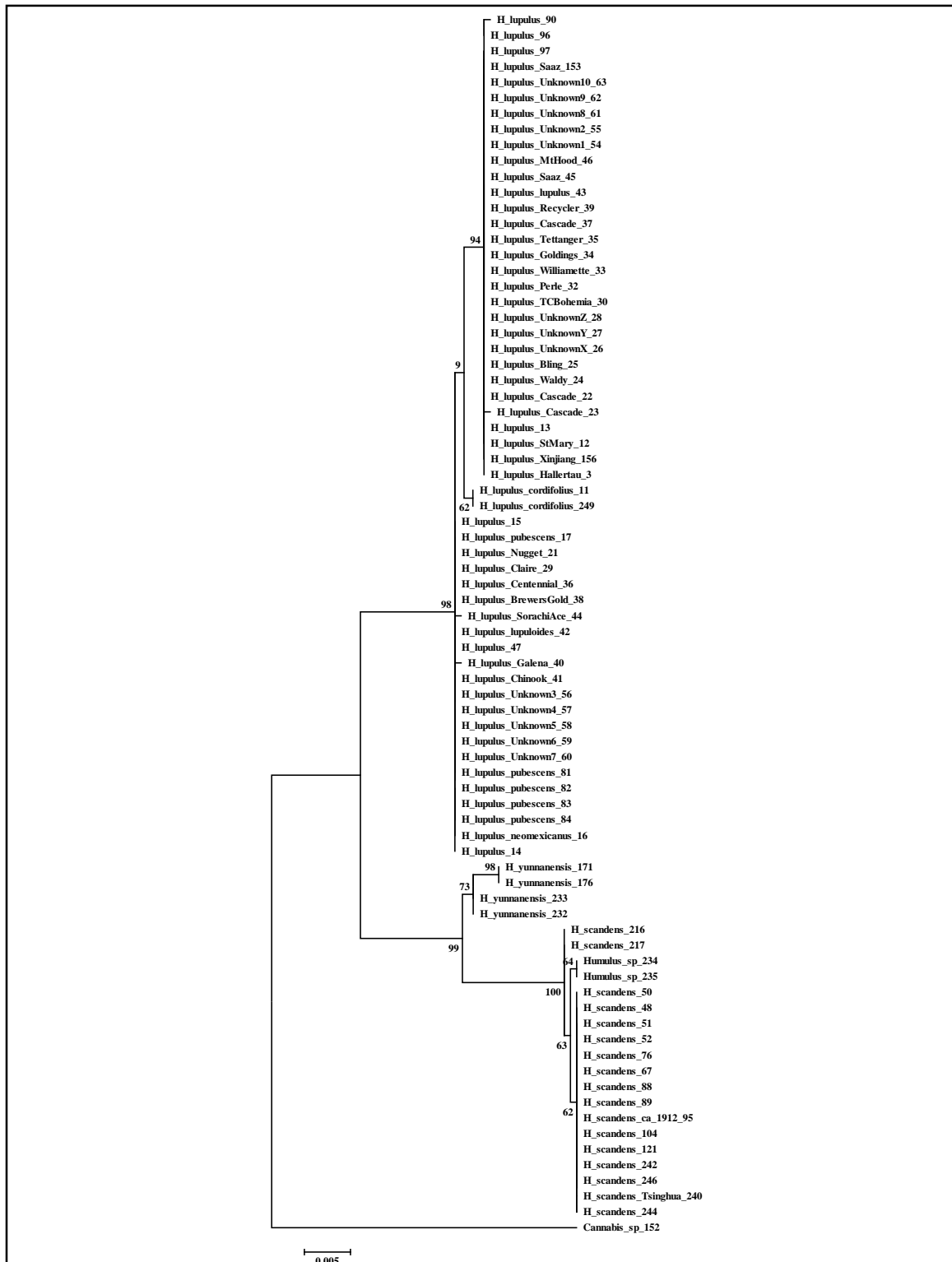


**Fig. 3.5.** Maximum Likelihood tree of *petL-psbE* region based on Tamura 3-parameter model. The tree has the highest log likelihood (-1685.4822) with the bootstrap percentage of trees in which the associated taxa cluster together shown next to the branches. Branch lengths are in number of substitutions per site.

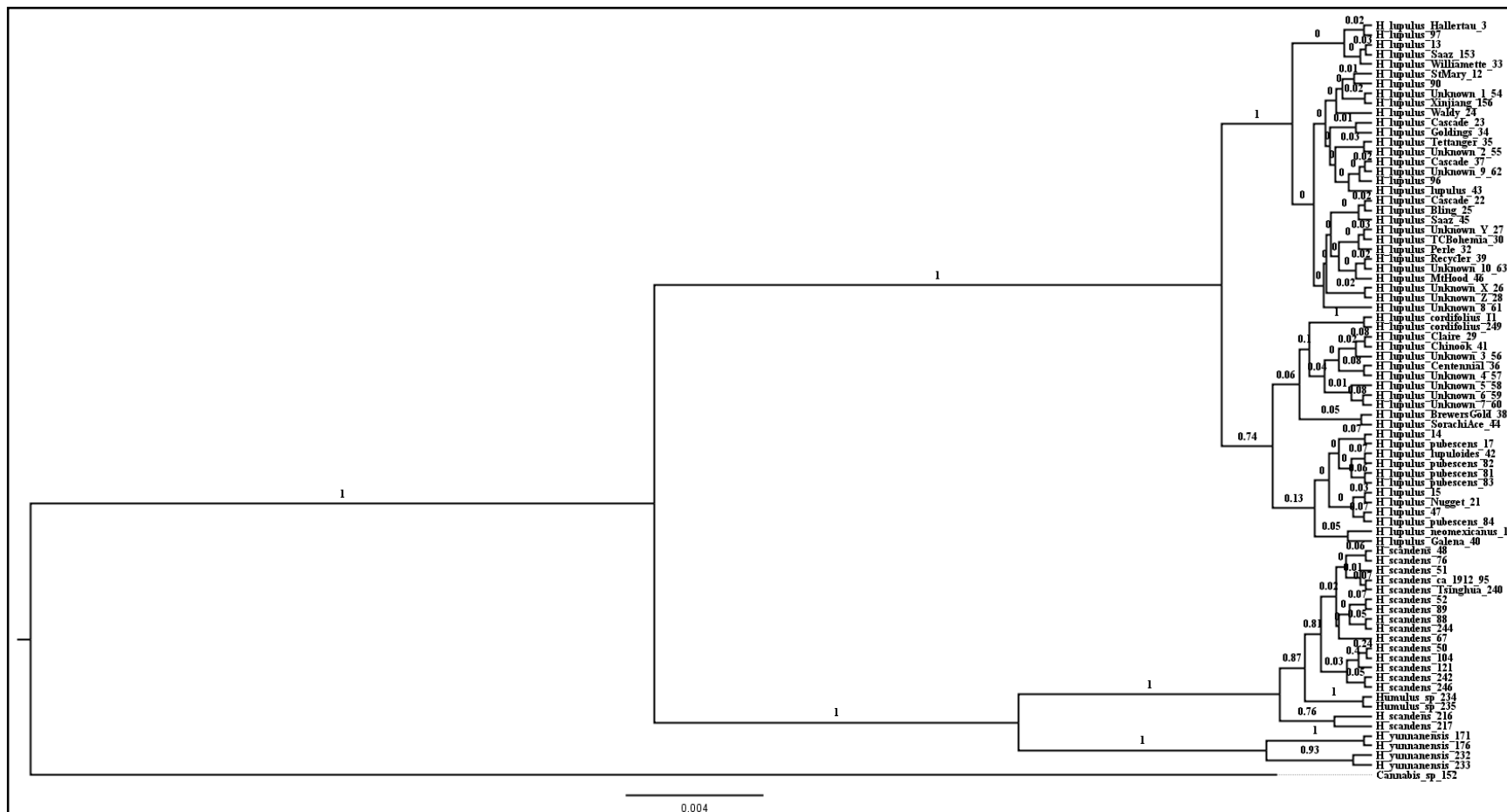


**Fig. 3.6.** Bayesian inference phylogeny of *petL-psbE* depicted as a maximum clade credibility tree based on Reversible-jump Based substitution model. Posterior probabilities in which the associated taxa cluster together is shown next to the branches. Branch lengths are in number of substitutions per site. The highest log clade credibility is -370.46162071889165.

ML and BI phylogenies for the combined data set has a similar topology to the *petL-psbE* trees, showing three separate clades of *Humulus* (Figs. 3.7 and 3.8). High bootstrap and posterior probability support was found for the *H. lupulus* (98% and 1.0), *H. scandens* (100% and 1.0), and *H. yunnanensis* clades (73% and 1.0). A single node separates the *H. scandens* clade from the *H. yunnanensis* clade (99% and 1.0). Within the *H. lupulus* clade, three groups formed in the ML tree with 98%, 62%, and 94% support at the nodes. For the BI tree of the *H. lupulus* clade, two main groups formed with 1.0 and 0.74 posterior probability support. Only one highly derived internal node at 1.0 separates the samples H\_lupulus\_cordifolius\_11 and H\_lupulus\_cordifolius\_249, while the other internal nodes have low support (0-0.13). Additionally, high bootstrap and posterior probability support was found for the *H. scandens* clade (100% and 1.0). The *H. scandens* clade has three further groups: one with high bootstrap and posterior probability support (100% and 0.76) and another two with slightly lower bootstrap but high posterior probability support (64%:1.0 and 62%:0.81). The internal nodes of both ML and BI trees for the *H. scandens* clade have low support (0% and 0-0.24). Lastly, high bootstrap and high posterior probability support was found for the *H. yunnanensis* clade (73% and 1.0). The *H. yunnanensis* clade has two internal groups: 1) the samples H\_yunnanensis\_232 and H\_yunnanensis\_233 at 73% and 0.93 and 2) the samples H\_yunnanensis\_171 and H\_yunnanensis\_176 at 98% and 1.0.



**Fig. 3.7.** Maximum Likelihood tree of combined data based on Hasegawa-Kishino-Yano model. The tree has the highest log likelihood (-2880.8018) with the bootstrap percentage of trees in which the associated taxa cluster together shown next to the branches. Branch lengths are in number of substitutions per site



**Fig. 3.8.** Bayesian inference phylogeny of combined data depicted as a maximum clade credibility tree based on Reversible-jump Based substitution model. Posterior probabilities in which the associated taxa cluster together is shown next to the branches. Branch lengths are in number of substitutions per site. The highest log clade credibility is -310.50202882398133.

## **Discussion**

The ML and BI phylogenies for the ITS2 region clearly depict three separate species, *H. lupulus*, *H. scandens*, and *H. yunnanensis*, with high bootstrap percentages and posterior probabilities. Alternatively, the *petL-psbE* region and combined data sets show a topology supporting the *H. scandens* and *H. yunnanensis* clades that are sister to three *H. lupulus* clades. The three clades of the common hop segregate the samples tested in this study into East Asia, New World, and Western/Central Eurasia pedigrees. Relying on the evolutionary histories presented here in the phylogenies, the putative wild hop samples collected in China and elsewhere were sorted into their respective lineages/clades (Table 3.6).

The Western/Central Eurasia group contains samples from Europe, the highly cultivated common hop (*H. lupulus* var. *lupulus*), and those samples with similar maternal pedigrees, including a collection from western Xinjiang (H\_lupulus\_Xinjiang). The New World clade contains *H. lupulus* var. *lupuloides*, *H. lupulus* var. *neomexicanus*, and *H. lupulus* var. *pubescens*, as well as the common cultivars with significant degrees of New World hop ancestry (e.g., hybrids). For example, both ‘Nugget’ and ‘Galena’ have ‘Brewer’s Gold’ in their pedigrees (Townsend and Henning 2009). Specifically, Salmon (1934) developed ‘Brewer’s Gold’ via open pollination of the native American female BB1 collected in 1916 near Morden, Manitoba, Canada (Burgess 1964). The BB1 genotype most probably is a representative of the variety *H. lupulus* var. *lupuloides* (Hampton et al. 2002 and 2003, Townsend and Henning 2009), and this maternal line is represented in the chloroplast region *petL-psbE* and combined data phylogenies. As for the East Asia clade with the samples H\_lupulus\_cordifolius\_11 and H\_lupulus\_cordifolius\_249, this group exhibits greater maternal lineage support to the New World clade than the Western/Central Eurasia clade (Figs. 3.4, 3.5, 3.6, and 3.7). Furthermore, the phylogenetic relationships and genetic diversity analyses presented here for the common hop plants from Western/Central Eurasia (*H. lupulus* var. *lupulus*), East Asia (*H. lupulus* var. *cordifolius*), and New World (*H. lupulus* var. *lupuloides*, *H. lupulus* var. *neomexicanus*, and *H. lupulus* var. *pubescens*) are similar to the results of Murakami et al. (2006a and 2006b) and other investigators (Cerenak et al. 2009, McAdam 2013,

Murakami et al. 2006a and 2006b, Patzak et al. 2010a and 2010b, Peredo et al. 2010, Stajner et al. 2008, Townsend and Henning 2009).

The *petL-psbE* region highlights the East Asia and New World common hop germplasm, specifically *H. lupulus* var. *cordifolius* and *H. lupulus* var. *lupuloides* compared to the Western/Central Eurasia germplasm of *H. lupulus* var. *lupulus*. The morphology of the sample from Xinjiang (H\_lupulus\_Xinjiang) mostly resembles the commonly cultivated *H. lupulus* var. *lupulus*. Although this plant was collected in Xinjiang in early July and did not have flowers, a single DNA barcode from the chloroplast (*petL-psbE*) clearly differentiated this sample, compared to a single DNA barcode from the nucleus (ITS2). Likewise, the morphology of the hop sample from the Nursery at Botanical Garden of the Kunming Institute of Botany, CAS in Yunnan mostly resembles the hop plants found in East Asia (*H. lupulus* var. *cordifolius*) (H\_lupulus\_cordifolius\_11 and H\_lupulus\_cordifolius\_249); however, because the plant may have been monecious (i.e., exhibited both male and female flowers; possibly  $2n=3x=27+XXY$ ) or a hybrid, the placement of the taxa *H. lupulus* var. *cordifolius* from Yunnan remains for further study.

*Humulus scandens* as a distinct species has high support from the ITS2, *petL-psbE*, and combined datasets. Analyses of the *petL-psbE* and combined regions for *H. scandens* highlight the possibility of three separate evolutionary histories. For example, the chromosome numbers of female and male *H. scandens* plants are  $2n=14+XX$  and  $2n=14+XY_1Y_2$ , respectively, while the chromosome numbers for the botanical varieties of *H. lupulus* are  $2n=18+(XX \text{ or } XY)$  (Grabowska-Joachimik et al. 2006). Therefore, the phylogenetic results presented here for *H. scandens* suggests three putative separate evolutionary histories, which are hypothesized to correlated with the reduction in chromosome numbers, the selection of X-chromosomes and two separate Y-chromosomes, and the shift to an annual from a perennial habit. A similar sex chromosome system as *H. scandens* is found in *Rumex acetosa* (Grabowska-Joachimik et al. 2006). Studies on *Tradescantia*, slipper orchids, and *Zamia* also show atypical chromosome numbers, probably as a result of stressful influences (Jones 1998). Furthermore, the nucleotide differences in the highly homologous intergenic spacer of

18S-26S rDNA (IGS) for the common hop samples from Western/Central Eurasia, East Asia, and New World as well as *H. scandens* indicate the slight variations in length are due to subrepeats (Murakami 2001). To the author's knowledge, no karyology studies have been conducted on *H. yunnanensis*, but due to the sister clade of *H. scandens*, *H. yunnanensis* is hypothesized to have a similar chromosomal relationship. Further genetic testing, such as sequencing the whole rDNA cistron and chloroplast genome, is required to validate the observed phylogenetic pattern of three separate evolutionary histories for *H. scandens*.

In regards to the Yunnan hop as a distinct species, high support for H<sub>3.1</sub> based on ML and BI phylogenies for the ITS2, *petL-psbE*, and combined datasets suggests the putative taxon is a separate species and not a variety of the common hop. If both *H. lupulus* and *H. yunnanensis* are narrowly adapted to habitats with specialized genotypes, then these taxa should have closer phylogenetic affinities and lower evolutionary divergence (Table 3.4, Fig. 3.7). Compared to *H. lupulus* with a specialized genotype for riparian habitats, the phylogenetic results presented here suggest *H. yunnanensis* has a general purpose genotype adapted to be flexible to colonize across a range of habitats, like *H. scandens*. Further support for this hypothesis comes from three duplicate herbarium specimens of *H. yunnanensis* that are putatively collected from the wild in Xinjiang (KUN #s 525337, 525338, and 525339). Therefore, the possible conclusions are:

- 1) the Yunnan hop is not endemic to the Yunnan Province,
- 2) the Yunnan hop is indigenous across a much wider range in central Eurasia,
- 3) the three duplicate KUN samples are incorrectly labelled as original introductions from the wild of Xinjiang,
- 4) the three duplicate KUN samples are examples of plant relics along the Silk Road,
- 5) the three duplicate KUN samples represent an ancient or recent introduction of the Yunnan hop to Xinjiang,

- 6) the Yunnan hop was brought to Xinjiang for the expansion of the brewing and hop industry (Pavlovič et al. 2006), later identified not to be the common hop, but the plant(s) continued to persist along road and waste sites, and/or
- 7) the Yunnan hop is an ancient hybrid between *H. lupulus* and *H. scandens* that arose approximately when the Indian Plate rammed into the Eurasian Plate creating the Himalayan Mountains and the Tibetan Plateau.

Due to deteriorating handwritten labels on decade old samples as well as the interpretation of the information on herbarium labels varying according to different generations of botanists, collection information may become lost or complicated.

With the only known DNA collection of *H. yunnanensis* from herbarium specimens, historical DNA highlights the significance of herbaria for botanical knowledge, conservation, and education (Ahrends et al. 2011, Case et al. 2007, Flannery 2011, Fleet et al. 2006, Prather et al. 2004, Pyke and Ehrlich 2010, Sundberg et al. 2011). Sometimes *H. lupulus* and *H. yunnanensis* are misidentified and mislabeled on herbarium specimens, especially if the specimens are male or were collected in an early juvenile period without flowers. The ITS2 region can be used distinguish these cryptic herbarium specimens to distinct *Humulus* species. Moreover, the *petL-psbE* region segregated the Yunnan hop samples into two groups with high support: 1) the samples H\_yunnanensis\_232 and H\_yunnanensis\_233 and 2) the samples H\_yunnanensis\_171 and H\_yunnanensis\_176. Specifically, H\_yunnanensis\_232 and H\_yunnanensis\_233 were collected in the Yunnan Province, while H\_yunnanensis\_171 and H\_yunnanensis\_176 were collected from possibly the wild of Xinjiang Province and the Kunming Botanical Garden, respectively. Based on the phylogeny presented here, if the distribution of *H. yunnanensis* extends on entire northern range of the Himalayan Mountains from Xinjiang to Yunnan Provinces, then the diversification and radiation of *Humulus* species in China is potentially much more extensive than previously hypothesized. A similar distribution pattern on the northern range of the Himalayan Mountains for *H. yunnanensis* may be present and correlate well to the distribution and cryptic taxa discovered for *Taxus* L. (Taxaceae) (Liu et al. 2011 and 2013). Until further collecting of *H. yunnanensis* and other wild *Humulus* taxa is conducted across the entire

range of the Himalayan Mountains and the Tibetan Plateau, the cryptic species not collected and deposited in herbarium could be lost with recent changes toward a drier climate. Alternatively, sampling current herbarium specimens correctly determined to be *H. yunnanensis* would be a much quicker and inexpensive option than an all-out collecting expedition into extreme and remote mountains.

The phylogeny of hop specimens presented here indicates: 1) three *Humulus* species in separate clades and 2) internal groups within each of the three clades. These results are similar to the morphological species concept for *Humulus* (Table 3.7). For the common hop clade (*H. lupulus*), three internal groups could be classified according to Small (1978) as: 1) a clade of cultivars of *H. lupulus* var. *lupulus* with Western/Central Eurasia genetics, 2) a clade of cultivars and plants with significant New World genetics represented by *H. lupulus* var. *lupuloides*, *H. lupulus* var. *neomexicanus*, and *H. lupulus* var. *pubescens*, and 3) a clade of *H. lupulus* var. *cordifolius* that is closely related to the plants with New World genetics. An second taxonomy conserves the samples in the *H. lupulus* var. *lupulus* clade to *H. lupulus*, lumps the clade with the New World genetics as *H. americanus* Nutt., and raises the rank of *H. lupulus* var. *cordifolius* to the species *H. cordifolius* Miq. A third taxonomy could be made based on a phylogenetic and DNA sequence species concept, which conserves the samples restricted to the *H. lupulus* var. *lupulus* clade, lumps the samples with the New World genetics to *H. lupulus* var. *americanus* J. Boutain, var. nov., and conserves those sample with East Asia genetics to *H. lupulus* var. *cordifolius*. In comparison and based on a phylogenetic and DNA sequence species concept, a four taxonomy could be made conserving the samples restricted to the *H. lupulus* var. *lupulus* clade and lumping the clades containing those samples with the New World and East Asia genetics to *H. lupulus* var. *laurasiana* J. Boutain, var. nov. Support for *H. lupulus* var. *laurasiana* is based on the low evolutionary divergence between *Humulus* samples from the New World and East Asia clades (p-distance=0.001; Table 3.8). Lastly, a more simplified taxonomy could conserve *H. lupulus* to only the domesticated hop plants from Western/Central Eurasia and classify the hop plants from East Asia and the New World as *H. phytolaurasiana* J. Boutain, sp. nov. After all, *Humulus* is a Tertiary relic that had a possible original distributed across

Laurasia (Johnson 2002, MacGinitie 1953 and 1969, Manchester 2001, Tiffney 1986, specifically Weber 2003).

Overall, based on the non-coding regions analyzed for the *H. lupulus* samples in this study, little genetic distance was found between Western/Central Eurasia, East Asia and the New World clades (p-distance=0.001-0.003; Table 3.10). Therefore, these three *H. lupulus* clades could be lumped to a single species given they are interfertile. In another approach with AFLPs, Reeves and Richards (2011) examined five different species criteria for wild North American *H. lupulus*. They found support to recognize vars. *neomexicanus* and *pubescens* as species; however, Reeves and Richards (2011) withheld a species recommendation for var. *lupuloides* until further sampling of genetic variation is complete or a stable biological process can be identified to explain its observed genetic divergence. In general, the use of DNA for plant species determinations must include reviews of the organism's taxonomy and herbarium collections (Hajibabaei et al. 2007, Padial et al. 2010, Tautz et al. 2003).

For the *H. scandens* clade presented here, the entire range of the taxon is not included, so new species concepts cannot be completely validated. Therefore, *H. scandens* is conserved here to include the three internal groups of the samples studied. Further analyses on samples collected across the proposed natural range of the Japanese hop will clarify cryptic botanical varieties (e.g., peninsular Southeast Asia and Korea, as well as the islands of Japan and Taiwan). In comparison, additional analyses of *H. yunnanensis* across the putative natural range will verify the plant as truly endemic to Yunnan Province or more broadly distributed across the northern Himalayan Mountains and the Tibetan Plateau into Xinjiang Province. Thus, *H. yunnanensis* is a unique species, which possibly may have unknown botanical varieties sharing a closer evolutionary history with *H. scandens* than *H. lupulus* (p-distance=0.010; Table 3.4; Fig. 3.9).

**Table 3.6.** List of plant samples with DNA numbers sorted by lineage/clade. Samples from China are denoted with a \* symbol for comparison to Table 2.8. The Old World clade in Chapter 2 of this dissertation is the same as the Western/Central Eurasia clade with H\_lupulus\_Xinjiang.

<b>Sample (H=Humulus)</b>	<b>DNA#</b>	<b>Clade in phylogeny</b>
H_lupulus_Hallertau	3	Western/Central Eurasia
H_lupulus_StMary	12	Western/Central Eurasia
H_lupulus	13	Western/Central Eurasia
H_lupulus_Cascade	22	Western/Central Eurasia
H_lupulus_Cascade	23	Western/Central Eurasia
H_lupulus_Waldy	24	Western/Central Eurasia
H_lupulus_Bling	25	Western/Central Eurasia
H_lupulus_UnknownX	26	Western/Central Eurasia
H_lupulus_UnknownY	27	Western/Central Eurasia
H_lupulus_UnknownZ	28	Western/Central Eurasia
H_lupulus_TCBohemia	30	Western/Central Eurasia
H_lupulus_Perle	32	Western/Central Eurasia
H_lupulus_Williamette	33	Western/Central Eurasia
H_lupulus_Goldings	34	Western/Central Eurasia
H_lupulus_Tettanger	35	Western/Central Eurasia
H_lupulus_Cascade	37	Western/Central Eurasia
H_lupulus_Recycler	30	Western/Central Eurasia
H_lupulus_lupulus	43	Western/Central Eurasia
H_lupulus_Saaz	45	Western/Central Eurasia
H_lupulus_MtHood	46	Western/Central Eurasia
H_lupulus_Unknown1	54	Western/Central Eurasia
H_lupulus_Unknown2	55	Western/Central Eurasia
H_lupulus_Unknown8	61	Western/Central Eurasia
H_lupulus_Unknown9	62	Western/Central Eurasia
H_lupulus_Unknown10	63	Western/Central Eurasia
H_lupulus	90	Western/Central Eurasia
H_lupulus	96	Western/Central Eurasia
H_lupulus	97	Western/Central Eurasia
H_lupulus_Saaz	153	Western/Central Eurasia
*H_lupulus_Xinjiang	156	Western/Central Eurasia
*H_lupulus_cordifolius	11	East Asia
*H_lupulus_cordifolius	249	East Asia
H_lupulus_cf_lupuloides	14	New World
H_lupulus_lupuloides	15	New World
H_lupulus_neomexicanus	16	New World

**Table 3.6.** continued.

H_lupulus_cf_pubescens	17	New World
H_lupulus_Nugget	21	New World
H_lupulus_Claire	29	New World
H_lupulus_Centennial	36	New World
H_lupulus_BrewersGold	38	New World
H_lupulus_Galena	40	New World
H_lupulus_Chinook	41	New World
H_lupulus_lupuloides	42	New World
H_lupulus_SorachiAce	44	New World
H_lupulus	47	New World
H_lupulus_Unknown3	56	New World
H_lupulus_Unknown4	57	New World
H_lupulus_Unknown5	58	New World
H_lupulus_Unknown6	59	New World
H_lupulus_Unknown7	60	New World
H_lupulus_pubescens	81	New World
H_lupulus_pubescens	82	New World
H_lupulus_pubescens	83	New World
H_lupulus_pubescens	84	New World
Cannabis_sp	152	Outgroup
H_scandens	48	scandens
H_scandens	50	scandens
H_scandens	51	scandens
H_scandens	52	scandens
H_scandens	67	scandens
H_scandens	76	scandens
H_scandens	85	scandens
H_scandens	89	scandens
H_scandens	95	scandens
H_scandens	104	scandens
H_scandens	121	scandens
*H_scandens	216	scandens
*H_scandens	217	scandens
*Humulus_sp	234	scandens
*Humulus_sp	235	scandens
*H_scandens_Tsinghua	240	scandens
*H_scandens	242	scandens
*H_scandens	244	scandens

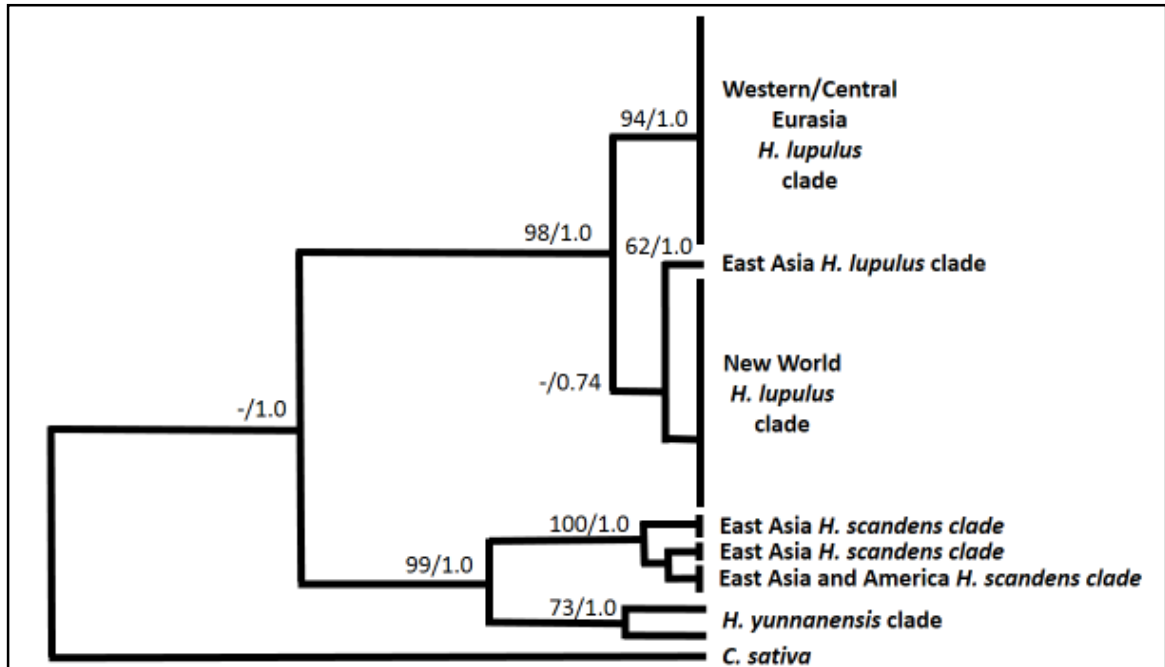
<b>Table 3.6.</b> continued.		
*H_scandens	246	scandens
*H_yunnanensis	171	yunnanensis
*H_yunnanensis	176	yunnanensis
*H_yunnanensis	232	yunnanensis
*H_yunnanensis	233	yunnanensis

**Table 3.7.** Distribution, DNA sequence clade, morphological, and chromosome traits distinguishing *Humulus lupulus* varieties, *H. scandens*, and *H. yunnanensis* (data adapted from Small (1978) and Wu et al. (2003)).

	taxa						
	var. <i>lupulus</i>	var. <i>cordifolius</i>	var. <i>neomexicanus</i>	var. <i>lupuloides</i>	var. <i>pubescens</i>	<i>scandens</i>	<i>yunnanensis</i>
Distribution	Eurasia but naturalized in eastern North America	East Asia	western North America	central North America	south central North America	eastern Asia but naturalized in eastern and central North America as well as Europe	Yunnan Province and putatively Xinjiang Province of China
ITS2 clade	<i>H. lupulus</i>	<i>H. lupulus</i>	<i>H. lupulus</i>	<i>H. lupulus</i>	<i>H. lupulus</i>	<i>H. scandens</i>	<i>H. yunnanensis</i>
<i>petL-psbE</i> clade	Western/Central Eurasia	East Asia	New World	New World	New World	<i>H. scandens</i>	<i>H. yunnanensis</i>
Flowers per bract	2	2	2	2	2	1	1
Infructescence (cm)	3-4	3-4	3-4	3-4	3-4	0.5-1.5(-2)	2-7
Female chromosomes	2n=18+ (XX)	2n=18+ (XX)	2n=18+ (XX)	2n=18+ (XX)	2n=18+ (XX)	2n=14+ (XX)	2n=unknown
Male chromosomes	2n=18+ (XY)	2n=18+ (XY)	2n=18+ (XY)	2n=18+ (XY)	2n=18+ (XY)	2n=14+ (XY <sub>1</sub> Y <sub>2</sub> )	2n=unknown

**Table 3.8:** Estimates of evolutionary divergence over sequence pairs between groups conducted in MEGA5.2

Taxon 1	Taxon 2	p-distance
		combined data set
Old World <i>H. lupulus</i> clade	East Asian <i>H. lupulus</i> clade	0.003
Old World <i>H. lupulus</i> clade	<i>Cannabis</i> sp.	0.046
East Asian <i>H. lupulus</i> clade	<i>Cannabis</i> sp.	0.044
Old World <i>H. lupulus</i> clade	New World <i>H. lupulus</i> clade	0.002
East Asian <i>H. lupulus</i> clade	New World <i>H. lupulus</i> clade	0.001
<i>Cannabis</i> sp.	New World <i>H. lupulus</i> clade	0.044
Old World <i>H. lupulus</i> clade	<i>H. scandens</i>	0.026
East Asian <i>H. lupulus</i> clade	<i>H. scandens</i>	0.026
<i>Cannabis</i> sp.	<i>H. scandens</i>	0.050
New World <i>H. lupulus</i> clade	<i>H. scandens</i>	0.025
Old World <i>H. lupulus</i> clade	<i>H. yunnanensis</i>	0.019
East Asian <i>H. lupulus</i> clade	<i>H. yunnanensis</i>	0.020
<i>Cannabis</i> sp	<i>H. yunnanensis</i>	0.045
New World <i>H. lupulus</i> clade	<i>H. yunnanensis</i>	0.019
<i>H. scandens</i>	<i>H. yunnanensis</i>	0.010



**Fig. 3.9.** Simplified phylogeny for the combined analysis of wild *Humulus* from China. Support for Maximum Likelihood (ML) bootstrap percentages when applicable and Bayesian Inference (BI) posterior probabilities are shown at the nodes. Low ML bootstrap support is at the New World *H. lupulus* clade, which could also be depicted as a polytomy with 98% and 1.0 support at the node of the *H. lupulus* clade. The clade sister to *H. lupulus* contains the two species, *H. scandens* and *H. yunnanensis*. The Yunnan hop is supported with high ML bootstrap percentages and BI posterior probabilities.

## **Conclusion**

Based on literature, herbarium records, and DNA barcodes, the three species of *Humulus* are confirmed and the four main objectives of this study were accomplished to reject the hypothesis (H<sub>3.0</sub>). The nuclear DNA barcode, ITS2, phylogenetically discriminates *H. lupulus*, *H. scandens*, and *H. yunnanensis*, while the chloroplast DNA barcode, *petL-psbE*, further differentiates the three *Humulus* species into putative subgroups. Hop breeding programs in China (Lou 2005, Pavlovič et al. 2006) and elsewhere can use simple genetic tests as a first step for the identification of putatively wild plants, which can lead to the development of new cultivars and germplasm resources for the world hop industry. Additional genomic and chemical analyses on the wild hop plants, escaped cultivars, and specifically *H. yunnanensis* must be done to determine traits of value to the brewing, medicinal, and pharmaceutical industries. Future directions for hop research must include a comprehensive survey and collecting expedition throughout the putative natural range of *H. yunnanensis* with an emphasis on the sensitive riparian habitats and plant communities associated with China's epicenter of biodiversity, the Three Parallel Rivers Region of Yunnan (Elvin 2004, Grumbine 2011, Marks 2012). As an alternative to an all-out collecting expedition, herbaria specimens have vast amounts of data ready for genomic analyses using Sanger and high-throughput DNA sequencing technologies (Atherton et al. 2010, Cronn et al. 2012, Egan et al. 2012, Flannery 2011, Glenn 2011, Neafsey and Haas 2011, Niedringhaus et al. 2011, Shokralla et al. 2012, Steele and Pires 2011, Steele et al. 2012, Quail et al. 2012, Zhang et al. 2011).

### **Acknowledgements**

I thank Professor Jianchu XU, Professor Dezhu LI, Professor Shengji PEI, Ende LIU, Junbo Yang, Qiaohong LI, Mingming WANG, Mei YANG, Yufang SU, Zhilin MU. Hongen JIANG, Jinghau WANG, Ram Poudel, Jie LIU, Zeng-Yuan WU, Dake ZHAO, Wen TANG, Yongtao ZHAO, Yana SHI, Xie YAN, Jun YAN, Genyi MENG, Pei-Luen LU, Edward Grumbine, Jacob Wickham, Ernest Small, Tom Ranker, Whitney Reyes, Ben Wadman, Tim Gallaher, Jason Cantley, the Kunming Institute of Botany, Chinese Academy of Sciences, the Institute of Botany, Chinese Academy of Sciences, the Ministry of Science and Technology of the People's Republic of China, Agriculture and Agri-Food Canada, the National Science Foundation - East Asia and Pacific Summer Institutes for U.S. Graduate Students (Award ID: 1108575), the Botanical Research Institute of Texas, the Botany Department at the University of Hawai'i at Mānoa (UH), the Graduate Student Organization at UH, the Anne S. Chatham Fellowship in Medicinal Botany from the Garden Club of America, and numerous office, lab, and herbarium colleagues and counterparts in China and the United States of America. A version of this paper was present at the Inaugural China-U.S. Young Scientist Forum in Beijing in August 2011 and at the 4<sup>th</sup> Biennial EcoHealth Conference in Kunming, Yunnan, China in October 2012. Any errors are solely by the author.

## CHAPTER 4

### Draft Chloroplast Genome of a Wild American Hop

(*Humulus lupulus* var. *neomexicanus* A. Nelson & Cockerell, Cannabaceae)

#### **Abstract**

With the rise in number of genomics projects driven by the affordability of high-throughput DNA sequencing, the *Cannabis* genome can now serve as a reference for *Humulus*. The main foci of this study were to generate a draft chloroplast genome (plastome) for a wild American hop plant (*Humulus lupulus* var. *neomexicanus*) and to visually compare the generated plastome to other plastomes in the Urticalean rosids and eudicots to establish the genetic relatedness of these taxa. High-throughput DNA sequencing of total genome DNA and Serial REmapping with ALignment (SEREAL) were used to capture chloroplast DNA sequences. Comparable to the size of the *Cannabis* plastome, the resulting draft hop plastome was ca. 151,716 bp with an average coverage of 11.31x and a maximum coverage of 38x in this study. The plastomes of *Humulus* and *Cannabis* share approximately 98% identity. The plastome, as a single DNA barcode, is a conserved locus that highlights a shared evolutionary history in the Cannabaceae and warrants further support for the sister taxa.

#### **Keywords**

hops, plastome, phylogenomics, DNA barcodes, Urticalean rosids, eudicot

## **Introduction**

In the Cannabaceae (*s.s.*), two genera, *Humulus* (Barth et al. 1994, Burgess 1964, Neve 1991, Small 1978, 1980, 1981, Zanolini and Zavatti 2008) and *Cannabis* (Clarke and Merlin 2013), are very closely related. The Cannabaceae share many similarities with the sister group, the Urticalean rosids (see Fig 1.3 in this dissertation). For many of the Urticalean rosids, phylogenetic comparisons involve several standard DNA barcodes. Although with advances in DNA sequencing technology reducing the total costs per project/genome, the entire chloroplast genome (plastome), a maternally inherited locus that behaves as a single non-recombining region, can be used for phylogenetic analysis as a single DNA barcode (Boutain et al. 2013). However, during the complicated computational task of *de novo* assembly of high-throughput sequencing reads in high-ploidy level plants, many reads of the plastome and mitochondrial genomes are bioinformatically removed as highly abundant/redundant ‘contamination’ (Matvienko et al. 2013, also see van Bakel et al. 2011).

Conversely, organelle DNA ‘contamination’ is useful for advancing phylogenomics, since whole plastomes have been shown to resolve low-level taxonomic (species specific) hypotheses (Cronn et al. 2008, Parks et al. 2009, Straub et al. 2011 and 2012, Stull et al. 2013, Whittall et al. 2010). Furthermore, these ‘contaminated’ data sets viewed as metagenomic samples of the whole environment within a cell are also processed for shallow-sequencing (low coverage) of the nuclear genome (Kunin et al. 2008, Straub et al. 2012). Shallow-sequencing nuclear genomes and multiplexing one flowcell lane has shown to be fruitful for obtaining nearly complete to entire plastomes at approximately 30-50x coverage (see Cronn et al. 2008, Parks et al. 2009, and Whittall et al. 2010 for *Pinus*; see Straub et al. 2011 and 2012 for *Asclepias*; see Stull et al. 2013 for plastome targeted enrichment). To the author’s knowledge, the utility of plastome capture from shallow-sequencing on the Ion Torrent’s Personal Genome Machine (PGM) has not been evaluated for plants. As the costs for high-throughput DNA sequencing decrease, ‘small’ genome projects exist at scalable levels for both the size in number of base pairs of the genome in question and the amount of available funds allocated to the project.

## **Objectives**

This project had four main objectives:

- 1) Shallow-sequencing the hop (*Humulus lupulus* L.) genome with reference mapping to the plastome of hemp (*Cannabis sativa* L.) to develop a draft hop plastome.
- 2) Test the PGM for the ease of capture of ‘contaminated’ reads from the chloroplast genome from an attempt at whole genome sequencing,
- 3) Compare reference mapping assemblies to members in the sister group of the Urticalean rosids, and
- 4) Visually compare the plastomes of the Cannabaceae, Urticalean rosids, and eudicots.

## **Hypothesis**

H<sub>4.0</sub> = Due to the highly conserved nature of the plastome, little genomic difference will be observed between the Cannabaceae (*s.s.*), other Urticalean rosids, and eudicots. H<sub>4.1</sub> = The entire plastome as a DNA barcode is ideal for genomic studies within the highly derived Cannabaceae (*s.s.*) compared to plastomes from the other Urticalean rosids and eudicots.

## **Materials and Methods**

### ***Plant material and DNA extraction***

Total genomic DNA from a wild American hop plant originally collected in Colorado (*H. lupulus* var. *neomexicanus* A. Neslon & Cockerell; USDA, HUM 1353.001, 2011-194, SH 8-24 [PI 635448]; [http://www.ars-grin.gov/cgi-](http://www.ars-grin.gov/cgi-bin/npgs/acc/search.pl?accid=%20PI+635448)

[bin/npgs/acc/search.pl?accid=%20PI+635448](http://www.ars-grin.gov/cgi-bin/npgs/acc/search.pl?accid=%20PI+635448)) was extracted from fresh leaves that were preserved in silica desiccant. A modified DNA extraction procedure used mini columns from Qiagen DNeasy® Plant Mini Kit (Qiagen, Valencia, California, USA) and reagents from Macherey-Nagel NucleoSpin® Plant II Mini Kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany). The total genomic DNA extraction protocol follows.

Approximately, 0.02-0.05g of a leaf was ground in a mortar and pestle with liquid nitrogen and sea sand. The macerated leaf material and sand were transferred to a 1.5 mL microcentrifuge tube, and 300 µl PL2 (Lysis Buffer) and 10 µl RNase A were added. The mixture was homogenized thoroughly with a vortex then incubated in a 65°C water bath for 90 minutes. After incubation, 75 µl PL3 (Precipitation Buffer) was added to the 1.5 mL tube, briefly mixed with a vortex, and placed on ice for 5 minutes. Afterwards, the tube was centrifuged for 5 minutes at full speed (16,000 x g). Leaving the pellet, the supernatant was transferred from 1.5 mL tube to a lilac QIAshredder Mini Spin Column then centrifuged for 2 minutes at 10,000 x g. The liquid flow through was transferred to a new 1.5 mL tube, 450 µl PC (Binding Buffer) was added, and then thoroughly mixed by pipetting up and down. The tube contents were transferred to a colorless DNeasy Mini Spin Column, and then centrifuged for 1 minute at 10,000 x g. Any remaining aliquots from the 1.5-mL tube were transferred to the column and centrifuged for 1 minute at 10,000 x g, discarding the flow through both times. Then, 400 µl PW1 (Wash Buffer) was added to the column and centrifuged for 1 minute at 10,000 x g. The flow through liquid was discarded, and 700 µl PW2 (Wash Buffer) was added to the column. After centrifuging for 1 minute at 10,000 x g and discarding the flow through, 200 µl PW2 (Wash Buffer) was added to the column and centrifuged for 2 minutes at 10,000 x g to dry the silica membrane in the column. The final flow through was discarded, and the dry column was transferred to a new 1.5 mL collection tube. To elute the DNA, 100 µl TE

Buffer (previously incubated at 65°C) was pipetted directly onto the silica membrane in the bottom of the column and incubated for 5 minutes at room temperature. After centrifuging the column for 1 minute at 10,000 x g, an additional 100 µl of TE Buffer was added directly to the silica membrane and centrifuged for an additional minute at 10,000 x g. A total of 200 µl of eluted DNA with approximately 125 ng/µl (25 µg) was stored until further use at -20°C.

### ***Ion Torrent Library Preparation and Sequencing***

NEBNext® Fast DNA Fragmentation and Library Prep Set for Ion Torrent (New England Biolabs® Inc., Ipswich, Massachusetts, USA; catalog # E6285L; version 3.2, 3/13) was followed according to the manufacturer's recommendations with 1µg of extracted DNA. Slight modification to the fragmentation and end repair of DNA protocol included incubation at ambient temperature (~22.5°C) for 10 minutes prior to the incubations in a water bath at 25°C for 20 minutes and 70°C for 10 minutes. After preparation and cleanup of adaptor-ligated DNA, 200 base-read library size (~330 bp median library size) selection was carried out on a FlashGel System, 2.2% 8+1 Two-Tier Well Format (Lonza, Rockland, Maine, USA; catalog #57022). PCR amplification of adaptor ligated DNA was carried out on a thermocycler with an initial denaturation at 98°C for 30 seconds; 6 cycles of 98°C denaturation for 10 seconds, 58°C annealing for 30 seconds, and 72°C elongation for 30 seconds; and a final extension cycle of 72°C for 5 minutes that was held at 4°C upon completion. Clean-up of the amplified library was done with AMPure XP beads and then quantified on an Applied Biosystems 7300 Real-Time PCR System according to Ion Library Quantitation Kit (Life Technologies catalog # 4468802). Based on quantitative PCR results, the library was diluted 1/4,000 before being prepared according to the manufacturer's protocol (publication part number 4478871Rev. A, revision date 4 June 2012) on the Ion OneTouch™ 200 template kit v2 (cat #4478316). Subsequent preparation of template-positive Ion OneTouch™ 200 Ion Sphere™ Particles (ISPs) was done on the Ion OneTouch™ ES instrument. Whole genomic DNA sequencing was performed on the Personal Genome Machine (Ion Torrent) with ISPs loaded on a 316 chip and using the Ion PGM™ Sequencing 200 Kit v2 (cat #4482006).

Initial filtering and trimming of raw reads for low quality and ambiguous bases was performed on the Torrent Suite™ Software 3.4.2 that runs on the Torrent Server and confirmed with *Core NGS Tools-Create Sequencing QC Report tool* on CLC Genomics Workbench (CLC GW).

***Optimized De Novo Assembly with CLC GW and Nuclear Genome Coverage***

After importing the SFF file generated from the Torrent Suite™ Software to the CLC GW, additional filtering was done by trimming reads at a 0.05 quality limit (minimum 13 Phred) and minimum length of 50 bp with the *Core NGS Tools-Trim Sequences tool*. Duplicate reads were removed with the *De Novo Sequencing-Remove Duplicate Reads tool* (Duplicate Reads Removal plugin version 1.0 beta 4). Assembly of the filtered nucleotide sequences into contigs was performed using the default settings in *De Novo Sequencing-De Novo Assembly tool* and also performed using the range of word sizes (12-64) to find the most optimal k-mer that gave the longest length of the total contigs assembled (Table 4.1; Haimine et al. 2011, Haznedaroglu et al. 2012). The Lander/Waterman equation was used for computing nuclear genome coverage, which is  $C = LN / G$ , where C stands for coverage, G is the haploid genome length, L is the read length, and N is the number of reads (Lander and Waterman 1988).

**Table 4.1.** Parameters set for CLC GW *De Novo* assembler to optimize word size.

Parameter	Option
Mapping mode	Map reads back to contigs (slow)
Update contigs	Yes
Automatic bubble size	Yes (Comment - Bubble size: 176)
Minimum contig length	200
Word size	Varied from 12-64
Perform scaffolding	Yes
Auto-detect paired distances	Yes
Mismatch cost	2
Deletion cost	3
Length fraction	0.5
Similarity fraction	0.8
Create list of un-mapped reads	Yes
Colospace alignment	No
Guidance only reads	No

### ***Reference Mapping and SEREAL Assembly of a Draft Chloroplast Genome***

With the plastome of *Cannabis sativa* L. as a reference (van Bakel et al. 2011) and the *Core NGS Tools-Map Reads to Reference tool*, a SERIAL REmapping with ALignment (SEREAL) method was performed, where the *Core NGS Tools-Extract Consensus Sequence tool* generates a continuous sequence filled with Ns for the zero coverage regions from four different reference mapping settings (Table 4.2). The four reference mappings with increasing similarity of reads and increasing length of reads were aligned with the *C. sativa* plastome using the MAFFT version 7.037 (2013/Apr/25) online server (<http://mafft.cbrc.jp/alignment/server/>), then manually edited by eye for a draft chloroplast genome of the wild American hop plant. The objective of the SEREAL method is to identify any regions of homopolymer-associated indel errors caused by the PGM sequencing platform that the *de novo* and reference mapping assemblies may have missed (Quail et al. 2012). Also, SEREAL compares regions of the draft *Humulus* plastome that putatively do not exist in the *Cannabis* plastome. Therefore, a final consensus sequence from the SEREAL method should introduce less errors for the future downstream analyses than using a single reference mapping to closely related species (i.e., *C. sativa*, *Morus indica* L., *Prunus persica* (L.) Batsch, *Castanea mollissima* Blume, *Theobroma cacao* L., *Arabidopsis thaliana* (L.) Heynh., and *Helianthus annuus* L.).

### ***Plastome Comparisons of the Cannabaceae, Urticalean rosids, and eudicots***

For reference mapping assemblies generated in CLC GW, the default parameters were used to map the *H. lupulus* var. *neomexicanus* reads to the plastomes of *C. sativa*, *M. indica*, *P. persica*, *C. mollissima*, *T. cacao*, *A. thaliana*, and *H. annuus*. Also, the program Mauve 2.3.1 (Darling et al. 2004 and 2010) was used to visualize conserved and rearranged regions in the draft plastome of *H. lupulus* var. *neomexicanus* when compared with *C. sativa*, *M. indica*, *P. persica*, *C. mollissima*, *T. cacao*, *A. thaliana*, and *H. annuus*. Plastome files for each species were downloaded from GenBank as fasta files, which were input individually into the Mauve 2.3.1 program (Table 4.3). In Mauve 2.3.1, the default parameters for a progressiveMauve alignment were used to generate a visualized genome alignment of the conserved (exons) and rearranged (inverted repeats) regions in the plastomes.

**Table 4.2.** Parameters set for SEREAL method using CLC GW *Core NGS Tools-Map Reads to Reference tool* with increasing similarity of reads and increasing length of reads. Subscripts a,b,c, and d are the individual settings for each of the four different reference mappings that were aligned with MAFFT version 7 and then manually edited by eye for the draft chloroplast genome of *Humulus lupulus* var. *neomexicanus*.

Parameter	Option
References	<i>Cannabis sativa</i> chloroplast, complete genome
Masking mode	No masking
Mismatch cost	2
Insertion cost	3
Deletion cost	3
Length fraction	0.5 <sub>a</sub> ; 0.5 <sub>b</sub> ; 0.8 <sub>c</sub> ; 0.8 <sub>d</sub>
Similarity fraction	0.7 <sub>a</sub> ; 0.8 <sub>b</sub> ; 0.8 <sub>c</sub> ; 0.99 <sub>d</sub>
Global alignment	No
Non-specific match handling	Map randomly
Output mode	Create reads track
Create report	Yes
Collect un-mapped reads	Yes
Reads mapped/% of reads (Similarity fraction)	9,987/1.29% (0.7 <sub>a</sub> ); 9,976/1.29% (0.8 <sub>b</sub> ); 9,381/1.21% (0.8 <sub>c</sub> ); 2,704/0.35% (0.99 <sub>d</sub> )
Positions of zero coverage/# regions of zero coverage (Similarity fraction)	972/21 (0.7 <sub>a</sub> ); 933/19 (0.8 <sub>b</sub> ); 1,516/24 (0.8 <sub>c</sub> ); 44,444/194 (0.99 <sub>d</sub> )

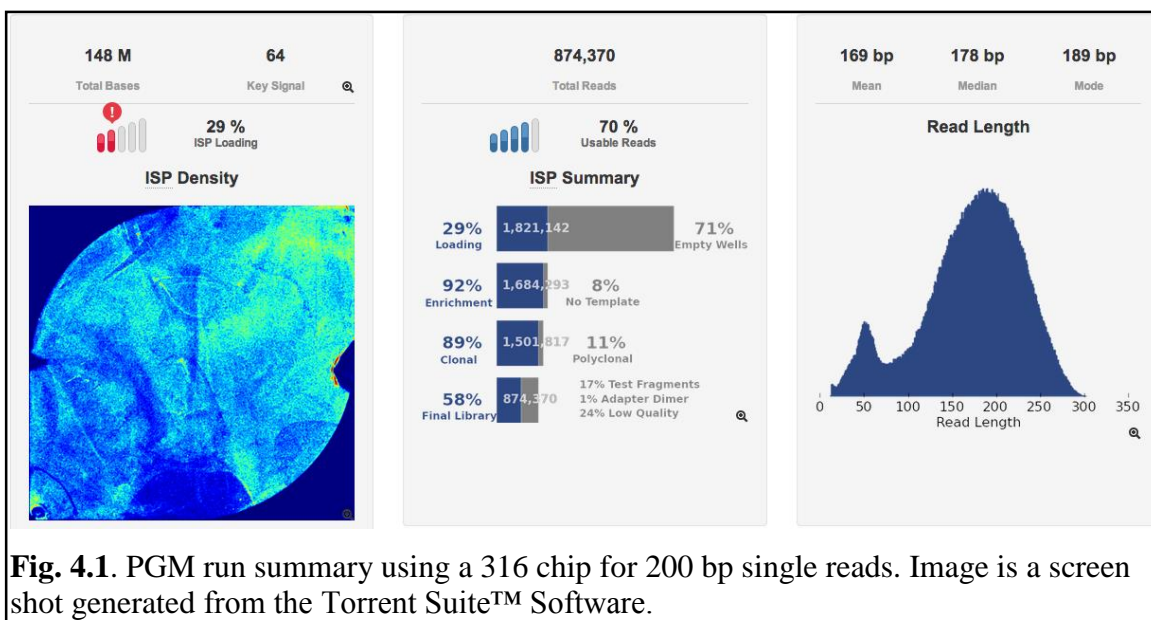
**Table 4.3.** List of taxa and GenBank accessions when applicable for the plastome sequences used in the progressiveMauve alignment and visualization of locally collinear blocks for conserved exons and rearranged inverted repeat regions.

<b>Taxa</b>	<b>GenBank Accession</b>
<i>Humulus lupulus</i> var. <i>neomexicanus</i>	TBA = Generated in this study
<i>Cannabis sativa</i> L.	N/A = PK_chloroplast available at the URL: < <a href="http://genome.ccbr.utoronto.ca/cgi-bin/hgTracks?hgsid=44159&amp;chromInfoPage=&gt;">http://genome.ccbr.utoronto.ca/cgi-bin/hgTracks?hgsid=44159&amp;chromInfoPage=&gt;</a>
<i>Morus indica</i> L.	NC_008359
<i>Prunus persica</i> (L.) Batsch	NC_014697
<i>Castanea mollissima</i> Blume	NC_014674
<i>Theobroma cacao</i> L.	NC_014676
<i>Arabidopsis thaliana</i> (L.) Heynh.	NC_000932
<i>Helianthus annuus</i> L.	NC_007977

## Results

### *Ion Torrent Sequencing*

One sequencing run on the PGM using a 316 chip for 200 bp single reads resulted in a total of 874,370 sequences of quality filtered reads with a mean read length of 169 bp, representing 148Mb of sequencing data from the Torrent Suite™ Software (Fig. 4.1).



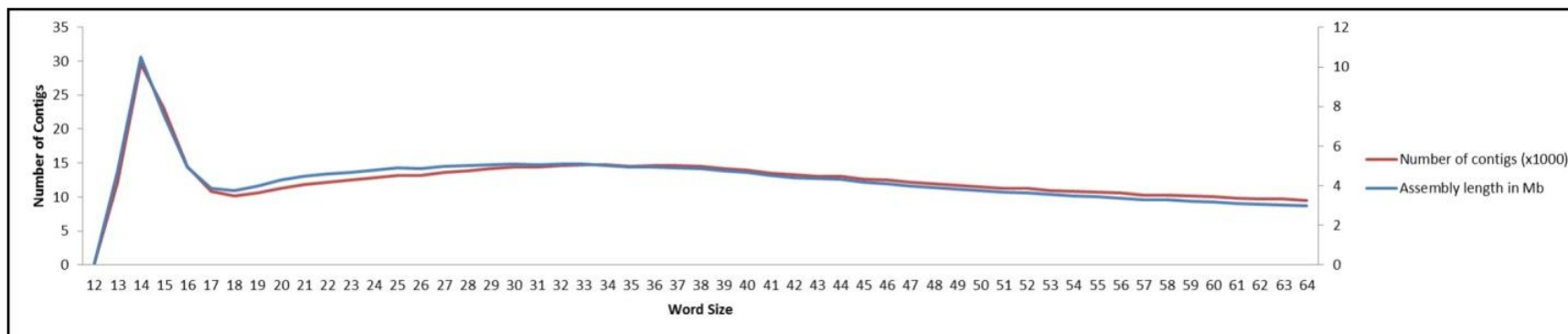
**Fig. 4.1.** PGM run summary using a 316 chip for 200 bp single reads. Image is a screen shot generated from the Torrent Suite™ Software.

## Optimized De Novo Assembly with CLC GW and Nuclear Genome Coverage

Of the remaining filtered sequence reads (i.e., 775,549 total reads of 94.69% remaining reads with 5.31% duplicate reads; Table 4.4), an optimized word score of 14 was found (Fig. 4.2). A minimum contig length of 200 bp resulted in a total of 29,675 contigs (i.e., 464,550 matched and 310,999 unmatched reads). The N75, N50, and N25 contig assemblies were 300, 382, and 475 bp, respectively. The total length of all contigs assembled was 10,518,355 bp (Table 4.5), representing approximately 0.0279x coverage of the *H. lupulus* var. *neomexicanus* nuclear genome [Table 4.6; ~2.2-2.5 Gb, Henning personal communication; ~3.0 pg = ~2.934 Gb (Grabowska-Joachim et al. 2006)] ( $C = LN / G$ , where C stands for coverage, G is the haploid genome length (2.934 Gb), L is the read length (176 bp), and N is the number of reads (464,550 matched)).

**Table 4.4.** CLC Genomics Workbench *Remove Duplicate Reads* tool supplementary QC report of over-represented sequences (sequence: the 5'-end of the sequence that has been found multiple times; abs: number of times this sequence has been observed).

sequence	abs
GTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGT	43
CTACTACTACTACTACTACTACTACTACTACTACTACTACTACT	20
GTTCAACTCGTTCGATATTTTTCCCCAAGAGATCTCATGGGTAAATGAAT	13
ATATAAGGCCAAAACTTGCGAAAATTGCCATCAAAAAGGGGAATTTTTTT	8
ATATTAGAGAAATCGAAGATTTCTGAAAGGGGTACCCCTTTGGTGAATTG	8
ATTTTCGGAGAAATTAGTGAAATCTCCAATTCACCAAAGGGGTACCCCTTT	8
GTTAAAAGAAAAGATTATTAAGATACTAGAATCCACAAAATGTAAGTTTA	8
ATACATTTTCGGAGAAATTAGTGAAATCTCCAATTCACCAAAGGGGTACCC	7
CCGAGTTCCCAAGATGTAAGTATGGGCTAGTCCGTAGGGTAAGCTGGTAA	7
GTCATATTAATCAAACCTTAGGTTAAAATTAATATTCTTAAACTATAGGT	7
ATGTTCAAATAAGTTAGTCCTAAGATTAGTCAGTGCACAATATTTACAC	6
GATTTTTTAACGCAAAGACCACCGCTGCCAACTCCATATCGTGAGTTGGAT	6
GTAAATTTATTTGTTAGATTTATGATAACTTATTGGAGCTTGAATTCAT	6
GTAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTAAGTAGTAGTAGTAGTAG	6
GTGGGAGTGTTAATCATAGATATGAACATCTATAGCTTCTGATGAAGAAG	6
GTTATTAATCTCAAGGTTATCTCTGAAAAATCCTTTTAGCATGAATTCTT	6
GTTCTTTGACTTGTTTCGTTACCAGCTTACCCTACGGACTAGCCCATACTT	6
GTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTTAGGGTTAGGG	6
TCTATTGGAGTTGAAACGGAGCCCTAAGTCAACATTTGAAAAAAATTCCC	6
ATAAATCTAACAAATAAATTTACTAACTTATTAATTCCTCGTGACTCCAC	5
ATAATATTGACTAGTGAGAGTAGAAGGATTTTAACCTTTGAACCACTTAAA	5
ATATATATATATATATATATATATTAATCACATACAAATAGATTAGAGATTA	5
ATATTAGAGAAATCGAAGATTTCTGAAAGGGGTACCCCTTTGGTGAATTT	5
ATCCATATTAGAGAAATCGAAGATTTCTGAAAGGGGTACCCCTTTGGTGA	5



**Fig. 4.2.** Optimization of word size using the default settings in *De Novo Sequencing-De Novo Assembly Tool*. The total length of 29,675 contigs assembled was 10,518,355 bp with a word size of 14.

**Table 4.5.** Summary of optimized word score of 14 from CLC GW *De Novo* assembly report.

Contig set	Contig count	Min contig length	Max contig length	Mean contig length	Standard deviation	Total contig length	% GC
N25 contigs	4,459	475	9,753	589.82	313.62	2,630,003	38.05
N50 contigs	10,679	382	9,753	492.51	219.71	5,259,504	38.18
N75 contigs	18,394	300	9,753	428.89	184.01	7,888,953	38.29
All contigs	29,675	42	9,753	354.45	174.79	10,518,355	38.46
Long contigs (>10,000bp)	0						
Short contigs (<200bp)	29,675	42	9,753	354.45	174.79	10,518,355	38.46

**Table 4.6.** Formulas for converting the number of base pairs to picograms of DNA, where 1 pg = 978 Mb (Doležel et al. 2003).

Output	Equation
Genome size (bp)	$(0.978 \times 10^9) \times \text{DNA content (pg)}$
DNA content (pg)	$\text{genome size (bp)} / (0.978 \times 10^9)$

### ***Reference Mapping with SEREAL Assembly of a Draft Chloroplast Genome***

Each of the four reference mappings covered 0.99 fraction of the complete chloroplast genome of *C. sativa*, with an exception of the strict parameters (i.e., 99% similarity and 80% length mapping) only covering 0.71 fraction of *C. sativa* (Table 4.7). Additionally, the strict parameters also had the greatest number of positions of zero coverage and the greatest number of regions of zero coverage with 44,444 bp and 194, respectively. The strict parameter consensus sequence included many highly conserved regions of the chloroplast (e.g., exons instead of introns). Also, the strict parameters helped to visualize the “backbone” of the draft chloroplast genome of the wild American hop plant when aligned manually by eye after alignment with the MAFFT online server. The SEREAL method produced a draft chloroplast genome of approximately 151,716 bp with 83 positions having zero coverage across three regions [(i.e., 90182-90210 (29bp), 99971-100011 (41bp), 151614-151626 (13bp)] when using the SEREAL consensus sequence as the reference for mapping to the original filtered reads (Table 4.8).

**Table 4.7.** Summary of CLC GW *Map Reads to Reference* tool report with subscripts a,b,c, and d showing results from each of the four different reference mappings parameters in Table 4.2.

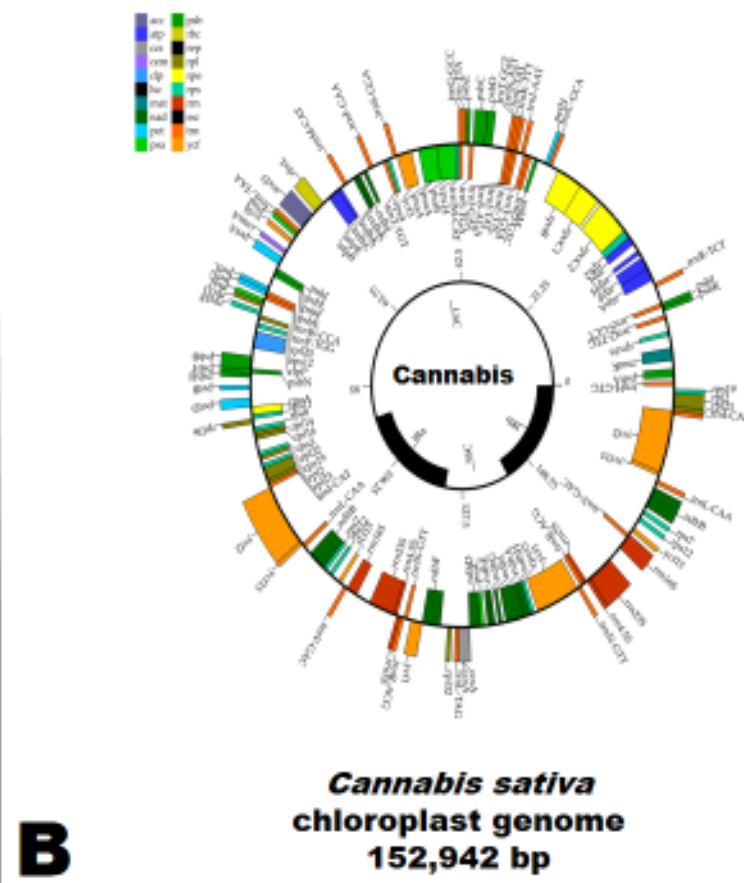
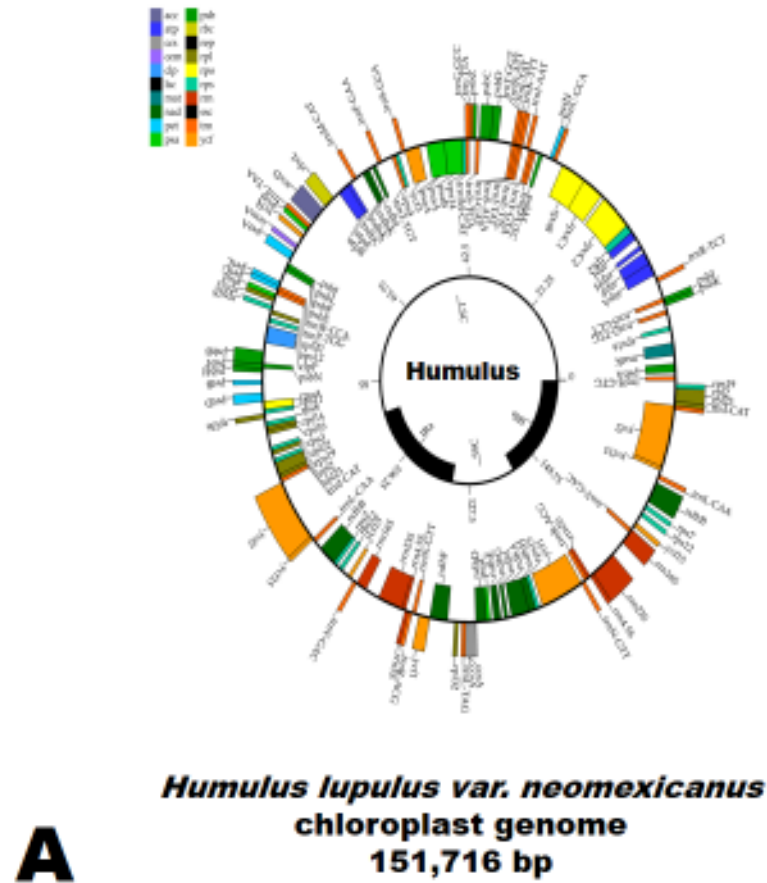
Parameter	SEREAL Output			
	Length 0.5 <sup>a</sup> Similarity 0.7 <sup>a</sup>	Length 0.5 <sup>b</sup> Similarity 0.8 <sup>b</sup>	Length 0.8 <sup>c</sup> Similarity 0.8 <sup>c</sup>	Length 0.8 <sup>d</sup> Similarity 0.99 <sup>d</sup>
Reference count	1	1	1	1
Type	Read mapping <sub>a</sub>	Read mapping <sub>b</sub>	Read mapping <sub>c</sub>	Read mapping <sub>d</sub>
Total reference length	152,942	152,942	152,942	152,942
GC contents in %	36.69	36.69	36.69	36.69
Total consensus length	152,641	152,662	152,687	152,967
Fraction of reference covered	0.99	0.99	0.99	0.71
Total read count	9,987	9,976	9,381	2,704
Mean read length	174.58	174.59	176.04	175.49
Total read length	1,743,504	1,741,696	1,651,413	474,523
Minimum coverage	0	0	0	0
Maximum coverage	38	38	37	24
Average coverage	11.11	11.10	10.71	3.09
Standard deviation coverage	4.81	4.78	4.81	3.13
Minimum excl. zero coverage regions	1	1	1	1
Average excl. zero coverage regions	11.18	11.17	10.82	4.36
Standard deviation excl. zero coverage regions	4.74	4.72	4.72	2.88
Zero coverage regions Count	21	19	24	194
Zero coverage regions Minimum length	1	1	1	1
Zero coverage regions Maximum length	315	315	315	1,934
Zero coverage regions Mean length	46.29	49.11	63.17	229.57
Zero coverage regions Standard deviation	66.73	70.08	68.78	294.57
Zero coverage regions Total length	972	933	1,516	44,444

**Table 4.8.** Summary of CLC GW *Map Reads to Reference tool* report with the SEREAL consensus sequence as the reference mapped to the original filtered reads of *Humulus lupulus* var. *neomexicanus*.

Parameter	Output
Reference count	1
Type	Read mapping
Total reference length	151,716
GC contents in %	36.88
Total consensus length	151,626
Fraction of reference covered	1
Total read count	10,023
Mean read length	174.5
Total read length	1,749,019
Minimum coverage	0
Maximum coverage	38
Average coverage	11.31
Standard deviation coverage	4.77
Minimum excl. zero coverage regions	1
Average excl. zero coverage regions	11.32
Standard deviation excl. zero coverage regions	4.77
Zero coverage regions Count	3
Zero coverage regions Minimum length	13
Zero coverage regions Maximum length	41
Zero coverage regions Mean length	27.67
Zero coverage regions Standard deviation	11.47
Zero coverage regions Total length	83

### ***Plastome Comparisons of the Cannabaceae, Urticalean rosids, and eudicots***

A draft hop plant chloroplast genome shares approximately 98% identity with *C. sativa* chloroplast genome (Fig 4.3). Furthermore, reference mapping to the complete chloroplast genomes of *M. indica*, *P. persica*, *C. mollissima*, *T. cacao*, *A. thaliana*, and *H. annuus* show 0.89, 0.89, 0.84, 0.83, 0.80, and 0.80 fraction of the reference covered, respectively. The results from the progressiveMauve alignment show the conserved and rearranged regions visualized as collinear block in the draft plastome of *H. lupulus* var. *neomexicanus* compared to the complete chloroplast genomes of *C. sativa*, *M. indica*, *P. persica*, *C. mollissima*, *T. cacao*, *A. thaliana*, and *H. annuus* (Fig. 4.4). The homology among these plastomes is high, particularly within *Humulus* and *Cannabis* (Jansen et al. 2011, Ravi et al. 2006, Sytsma et al. 2002; Yang et al. 2013). Results of the progressiveMauve alignment show the plastome of *Cannabis* to be the best fit reference for the *Humulus* plastome.



**Fig. 4.3.** Map drawings of the draft chloroplast genome of *Humulus lupulus* var. *neomexicanus* compared to the plastome of *Cannabis sativa*. *Humulus* (A) has approximately 1,226 base pairs less than *Cannabis* (B) that are roughly located in the intron regions between the transcribed genes.



**Fig. 4.4.** The progressiveMauve alignment showing the conserved and rearranged regions in the draft plastome of *H. lupulus* var. *neomexicanus* (A) compared to the complete chloroplast genomes of *C. sativa* (B), *M. indica* (C), *P. persica* (D), *C. mollissima* (E), *T. cacao* (F), *A. thaliana* (G), and *H. annuus* (H). Matching regions of the genome are partitioned into a minimum set of collinear blocks. Identically colored blocks indicate a matching region with a connecting line drawn to the collinear blocks in the comparison genomes. Blocks that are below the genome's center line are inverted in respect to the reference. The purple blocks in *Humulus*, *Cannabis*, and *Helianthus* are above the center line compared to the other genomes, and the yellow block in *Helianthus* shows the inverted region below the center line. Overall, the homology among these plastomes is high, mainly within *Humulus* and *Cannabis*.

## **Discussion**

The PGM, a compact and high-throughput benchtop sequencer, was found to be sufficient at capturing the ‘contaminated’ reads of the plastome. The shallow-sequencing results of this study were within the range of expected plastome reads as previously found for other plant taxa (Straub et al. 2011 and 2012). The SEREAL method is easy to perform when a closely related taxon’s genome is used as a reference. Also, SEREAL allows the biologist/user, not the computer, to perform a ‘correct’ alignment via manual visualization from serial reference mappings. With approximately 151,716 bp in the draft plastome of *H. lupulus* var. *neomexicanus*, the 83 positions with zero coverage can be considered non-significant contributions when considering downstream phylogenetic modeling at the entire plastome level. Further annotation of the wild hop plastome for gene order, conserved genes, and non-specific regions is a priority using specific programs developed for the genomes of organelles (e.g., DOGMA by Wyman et al. 2004; CpGAVAS by Liu et al. 2012a; Microbial Genome Finishing Module plugin by CLC GW). With a completely annotated hop plastome, strict modeling of individual genes can be conducted for an intensive understanding of *Humulus* spp. evolutionary history.

The conserved nature of the *Humulus* plastome as compared to that of *Cannabis* highlights the close evolutionary history of these two genera in the Cannabaceae (s.s.) and provides support for the circumscription of the Cannabaceae (s.l.) as now proposed by Yang et al. (2013). In this study, support for the closest sister taxon and therefore most appropriate outgroup for *Humulus* is *Cannabis*. Rational to use *Cannabis* as the correct outgroup for *Humulus* is the putative presence of an inverted sequence in the Cannabaceae (s.s.) plastomes (i.e., purple blocks in Fig. 4.4). In comparison, a large inversion was also found in the Asteraceae (i.e., *H. annus*) (Timme et al. 2007, Fig. 4.4). Other taxa have been suggested as closely related, but *M. indica*, *P. persica*, *C. mollissima*, *T. cacao*, *A. thaliana*, and *H. annus* are more distant than *Cannabis* (Fig. 4.4). Accordingly, the results presented here support the sequenced genome of *Cannabis* is the best reference for *Humulus* genome sequencing.

## **Conclusion**

Based on total genomic DNA sequencing and comparisons of plastomes, the four main objectives of this study were accomplished to support the hypothesis (H<sub>4.1</sub>). A DNA barcoding approach focusing on shallow-sequencing the nuclear genome to capture ‘contaminated’ reads from the chloroplast genome can easily be accomplished with the PGM. Also, the PGM is comparable to other high-throughput sequencing platforms (Cronn et al. 2008 and 2012, Dewey et al. 2012, Glenn 2011, Ku and Roukos 2013, Liu et al. 2012b, Loman et al. 2012, Neafsey and Haas 2011, Niedringhaus et al. 2011, Parks et al. 2009, Quail et al. 2012, Straub et al. 2011 and 2012, Stull et al. 2013, van Bakel et al. 2011, Whittall et al. 2010). To overcome the PGM’s homopolymer errors, the SEREAL method was developed. Of the Urticalean rosids samples used in this study, high homology was found between the *Humulus* and *Cannabis* plastomes. The phylogenetic placement for the intraspecific taxa of *Humulus* and *Cannabis* must be based on the plastome as a DNA barcode with the proper sister taxa in the Cannabaceae (*s.s.*) as the outgroup (Sytsma et al. 2002, Yang et al. 2013; also see Chen et al. 2012 for support of the Moraceae as the outgroup).

### **Acknowledgements**

I thank Michael Allen, David Visi, Richard Hale, Marnie Rout, Yan Zhang, Arthur Eisenberg, Vicki Funk, Keri McNew Barfield, Tiana Franklin Rehman, Alison Sherwood, Seana Walsh, Tim Gallaher, Jason Cantley, Marjorie and Thomas Day, the University of North Texas Health Science Center, the University of North Texas at Denton, the Botanical Research Institute of Texas, the United States Department of Agriculture National Clonal Germplasm Repository in Corvallis, Kim Hummer, John Henning, Nahla Bassil, Bruce Bartlett, the University of Hawai'i at Mānoa (UH), the Anne S. Chatham Fellowship in Medicinal Botany from the Garden Club of America, the Beatrice Krauss Fellowship from the Botany Department at UH, numerous office and lab colleagues and counterparts in Texas and Hawaii. Any errors are solely by the author.

## CHAPTER 5

### The Origin of *Humulus*: A Phylogenomic Surfing Approach

#### **Abstract**

A genome can now be generated quickly and inexpensively using semiconductor sequencers; as a result, this study focuses on a phylogenomic surfing approach for the entire chloroplast genome (plastome) as the minimum DNA barcode for plant species. Total genomic DNA was extracted from both fresh and herbarium material of *Humulus* spp. Shallow, high-throughput sequencing was conducted for an 8-plexed run that resulted in 4,528,200 sequence reads with a mean read length of 154 bp, representing 697 Mb of data. Reads from each indexed sample were assembled into draft plastomes with the reference *Cannabis sativa* L. Maximum Likelihood and Bayesian inference plastome phylogenies of three *Humulus* species., including the taxonomic varieties in the *Humulus lupulus* L. complex, show high bootstrap and posterior probabilities support as a polyphyletic tree with basal East Asian taxa and monophyly of the genus. Divergence estimates based on a single fossil calibration prior suggest the split between *Humulus* and *Cannabis* occurred approximately 87.28 million years ago (mya). The interspecific taxa of *Humulus* arose approximately 44.43 mya, while the intraspecific varieties of *H. lupulus* complex clade and the *H. scandens*/*H. yunnanensis* clade arose 26.1 and 15.95 mya, respectively. Compared to single or several DNA regions used to barcode plants, the phyloplastome model presented here supports the unity of the *H. lupulus* complex. Further duplicated sampling is require to test the varietal relationships.

#### **Keywords**

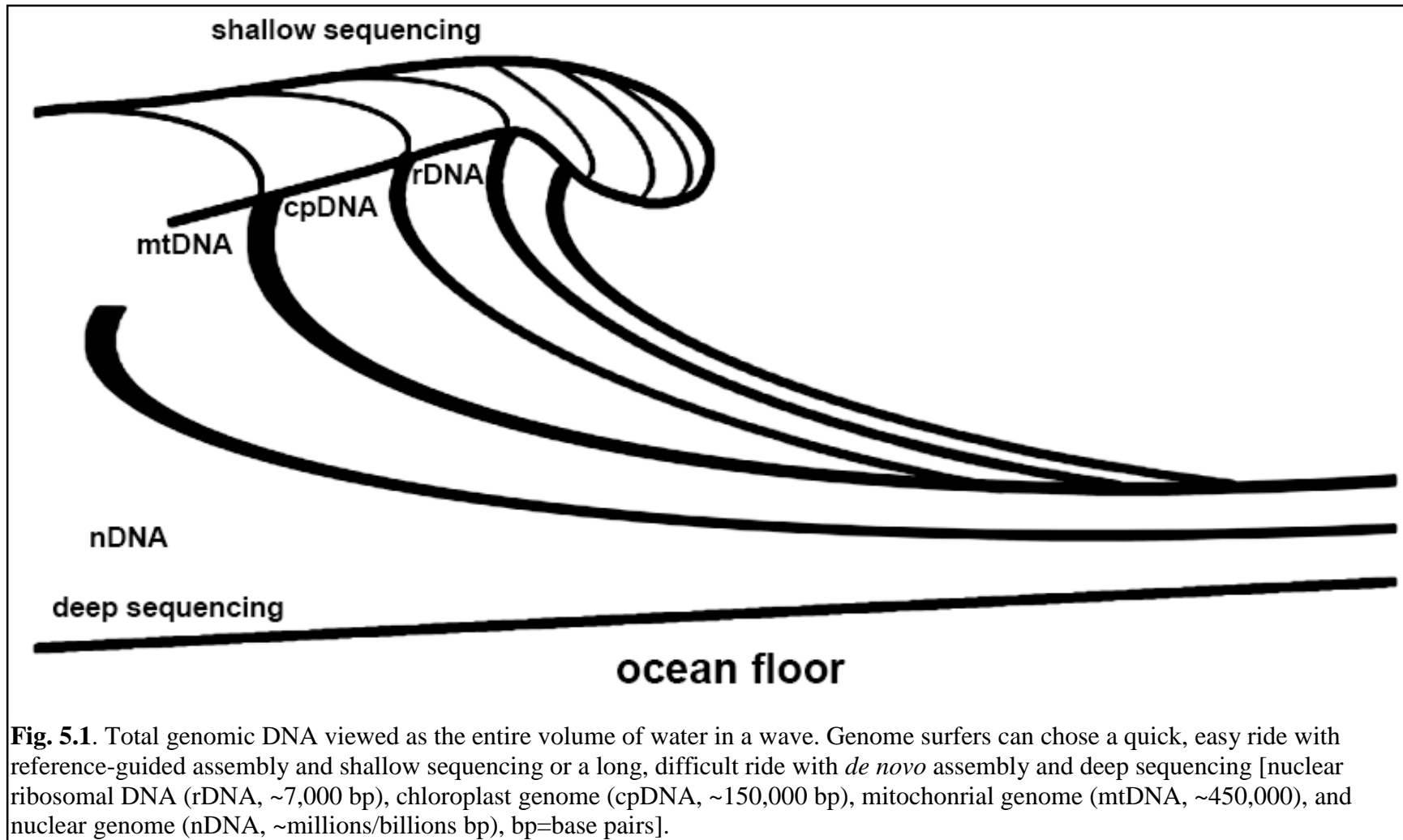
Common hop (*H. lupulus*), Japanese hop (*H. scandens*), Yunnan hop (*H. yunnanensis*), Cannabaceae, phylogenomics, fossil calibrated phylogeny, multiplexing, high-throughput sequencing, personal genome machine, herbarium specimen, DNA barcodes

## **Introduction**

### ***Surfing a Genome***

In Straub et al. (2012), the process of shallow-sequencing a genome is analogous to navigating the tip of the iceberg, where ribosomal DNA (rDNA) and chloroplast DNA (cpDNA) are represented in the 10% of ice that is visible above the water. Although this depiction is great, much of the world is not covered by arctic climates, and an alternative terminology for the process of shallow-sequencing is required to captivate those in tropical climates who are unfamiliar with icebergs. Therefore, a genome, a term originally coined by a botanist to describe the total DNA content of an organism (Lederberg and McCray 2001, Winkler 1920), can also be depicted as a wave (Fig 5.1).

For the genome surfer, the total genomic DNA (gDNA) of the study organism, in particular the hop plant (*Humulus*, Cannabaceae), is analogous to the entire volume of water in a wave. Four kinds of genetic data are found in the gDNA of plants and are visualized at different locations on the cresting wave. The total size in number of base pairs for each kind of data increases as the surfer travels further down the wave. Relatively small rDNA, cpDNA, and mitochondrial DNA are near the peak of the wave compared to the larger nuclear DNA at the base. Essentially, success at surfing either the shallow peak for gDNA components or deep in the entire wave depends on the type of surf board used: reference-guided or *de novo* assemblies. Surfing any organism's genome can be quick and easily done with reference-guided assembly or long and difficult with *de novo* assembly. Therefore in plants, a general and easy genome sequencing strategy is to sample the smaller sized rDNA and cpDNA from the top of the wave, just like dropping in from the top of a cresting wave is accomplished by a big wave surfer. The concept of surfing genomes is new and can be implemented in botany courses across K-12 and undergraduate education in the State of Hawaii, the United States of America, and abroad. Endorsement of this approach, the genome wave, is anticipated to stimulate the next generation of botanists, biologists, and non-scientists to understand and study the genomes of model and non-model organisms. As previously mentioned, the genome wave can also be conceptually interpreted as an iceberg (see Straub et al. 2012 to navigate the tip of the iceberg).



### ***Hops and DNA barcodes in the Next Generation***

Hops, the female flowers of the hop plant (*Humulus*), are used as a traditional medicine to alleviate migraines, inflammation, insomnia, bladder problems, uro-gynecological infections, symptoms of menopause, central nervous system irregularities, and skin problems (Boutain 2012a). The biological activities from hops also suggest application in cancer and osteoporosis prevention (Stevens and Page 2004). Today, hops are most commonly known as an important preservative and flavor in the production of ales and lagers (Boutain 2012b). For the brewing industry, the domesticated hop plant (*H. lupulus* var. *lupulus*) was introduced from England to the United States, where many new hops cultivars arose in the last two centuries (Boutain 2012c, Carter et al. 1990).

Since the introduction of the domesticated hop plant, varying degrees of introgression, either natural or human mediated with the native North American plants (*H. lupulus* var. *lupuloides*, var. *pubescens*, and var. *neomexicanus*), gave rise to the development of new cultivars, conservation of wild New World germplasm, and sustainable organic hops (Boutain 2012b and 2012c, Turner et al. 2011). Conversely, native North American hop plants can be difficult to identify with morphological characters alone because phenotypes are highly similar to the domesticated taxon as a result of hundreds of years of introgressive hybridization (also see Clarke and Merlin 2013 for a detailed discussion of how species identification problems apply to the closely related genus, *Cannabis*). Further confusion arises when heirloom cultivars escape into wild habitats, intermix with wild plants, and combine into new genotypes. As a result, DNA barcoding methods helped determine truly wild Native American hops (Boutain 2012a and 2012b), as well as understand the evolutionary relationships of the three hop species (*H. lupulus*, *H. scandens*, and *H. yunnanensis*) (Boutain and Xu 2012). Since the estimated nuclear genome of the hop plant exceeds two billion base pairs (Henning personal communication), this non-trivial sequencing project requires alternative approaches using high-throughput sequencers. The purpose of this study is to clarify the phylogenetic relationships of the three hop species and the described botanical varieties in the *H. lupulus* species complex using a whole chloroplast genome phylogeny coupled with fossil calibrated dating methods.

## **Objectives**

This project had four main objectives:

- 1) Surf the chloroplast genome (plastome) of the hop plant by shallow-sequencing the nuclear genome of the recognized taxonomic species and varieties of *Humulus* with reference mapping to the plastome of the hemp (*Cannabis* L.),
- 2) Construct Maximum Likelihood (ML) and Bayesian inference (BI) phylogenies of the draft hop plastomes to support or refute the recognized taxonomic species and varieties of *Humulus*,
- 3) Develop a high-throughput DNA sequencing protocol for capturing plastomes from herbarium specimens of *Humulus*, and
- 4) Test phylogeographic hypotheses for the origin of *Humulus* by using a fossil calibrated BI plastome phylogeny (BIPP).

## **Hypothesis**

H<sub>5.0</sub> = Phylogenomic methods support the genus *Humulus* originated in the Old World (i.e., China) and migrated to the New World (Murakami et al. 2006a, Neve 1991). H<sub>5.1</sub> = After the K/T extinction event around 65 million years ago, open niches along riparian areas in the New World mountains (i.e., Colorado Rockies) allowed the perennial binning habit of *Humulus* to thrive in forest ecotones and subsequently migrate to the Old World, where the genus underwent adaptive radiation into an two additional annual species that are putatively not interfertile within the *Humulus lupulus* complex.

## **Materials and Methods**

### ***DNA Extractions***

Total genomic DNA from the recognized taxonomic species and varieties of *Humulus* was extracted from fresh leaves that were preserved in silica desiccant or from herbarium material (Table 5.1). The manufacture protocol for the Qiagen DNeasy® Plant Mini Kit (Qiagen, Valencia, California, USA) or a modified DNA extraction procedure that used mini columns from Qiagen DNeasy® Plant Mini Kit and reagents from Macherey-Nagel NucleoSpin® Plant II Mini Kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany) was performed with an additional blank extraction to ensure minimal contamination to the rare herbarium specimen of *H. yunnanensis* (see Fulton 2012, Shapiro and Hofreiter 2012). The DNA extraction procedure follows. Approximately, 0.02-0.05g of leaf material was ground in a mortar and pestle with liquid nitrogen and sea sand before beginning both types of extractions. In the modified procedure, the macerated leaf material and sand were transferred to a 1.5 mL microcentrifuge tube, and 300 µl PL2 (Lysis Buffer) and 10 µl RNase A were added. The mixture was homogenized thoroughly with a vortex then incubated in a 65°C water bath for 90 minutes. After incubation, 75 µl PL3 (Precipitation Buffer) was added to the 1.5 mL tube, briefly mixed with a vortex, and placed on ice for 5 minutes. Afterwards, the tube was centrifuged for 5 minutes at full speed (16,000 x g). Leaving the pellet, the supernatant was transferred from 1.5 mL tube to a lilac QIAshredder Mini Spin Column then centrifuged for 2 minutes at 10,000 x g. The liquid flow through was transferred to new 1.5 mL tube, 450 µl PC (Binding Buffer) was added, and then thoroughly mixed by pipetting up and down. The tube contents were transferred to a colorless DNeasy Mini Spin Column, and then centrifuged for 1 minute at 10,000 x g. Any remaining aliquots from the 1.5 mL tube were transferred to the column and centrifuged for 1 minute at 10,000 x g, discarding the flow through both times. Then, 400 µl PW1 (Wash Buffer) was added to the column and centrifuged for 1 minute at 10,000 x g. The flow through liquid was discarded, and 700 µl PW2 (Wash Buffer) was added to the column. After centrifuging for 1 minute at 10,000 x g and discarding the flow through, 200 µl PW2 (Wash Buffer) was added to the column and centrifuged for 2 minutes at 10,000 x g to dry the silica membrane in the column. The final flow through

was discarded, and the dry column was transferred to a new 1.5 mL collection tube. To elute the DNA in both types of extractions, 100  $\mu$ l TE Buffer (previously incubated at 65°C) was pipetted directly onto the silica membrane in the bottom of the column and incubated for 5 minutes at room temperature. After centrifuging the column for 1 minute at 10,000 x g, an additional 100  $\mu$ l of TE Buffer was added directly to the silica membrane and centrifuged for an additional minute at 10,000 x g. A total of 200  $\mu$ l of eluted DNA was stored until further use at -20°C.

**Table 5.1.** Plant specimens sampled for total genomic DNA sequencing from the recognized taxonomic species and varieties of *Humulus*. The *H. yunnanensis* sample is from KUN.

<b>Sample (H=Humulus)</b>	<b>DNA # ; multiplex barcode</b>	<b>Gender</b>	<b>Locality</b>	<b>Collection ; Date ; Herbarium ; Note</b>
H_lupulus_var_lupulus	253 ; 1	Female	Wild from Kazakhstan	Received leaves in silica ; USDA - HUM 1025.007, 2011-194, SH 8-23 [PI 635262]
H_lupulus_var_lupulus	43 ; 2	Female	Eastern Ottawa, Canada	JRBoutain#323; Sent by E. Small from clone of old brewery cultivar growing by roadside (now ruderal). Unknown cultivar is originating and persisting from nearby mid-20 <sup>th</sup> century brewery hop plantation.
H_lupulus_var_cordifolius	427 ; 3	Moneicious ? – Female	Yunnan, China	JRBoutain#338 ; Oct. 2011 ; specimen with mature female cones; duplicates KUN#0935629, KUN#0935630, and KUN#1014725
H_lupulus_var_neomexicanus	254 ; 4	Female	Wild from Colorado, U.S.A.	Received leaves in silica ; USDA - HUM 1353.004; 2011-194, SH 8-24 [PI 635448]
H_lupulus_var_pubescens	255 ; 5	Female	Wild from Nebraska, U.S.A.	Received plant in-vitro ; USDA - HUM 489.002; HUM->HUMOH; 9-14-10, ALS, H.L.B., Nebraska [PI 5592132]
H_lupulus_var_lupuloides	42 ; 6	Female	Southern Ottawa, Canada	JRBoutain#322; Sent by E. Small from clone of original wild population on bank of Black Rapids Creek at Merivale Road.
H_scandens	282 ; 7	Male	Hubei Province, China	WUZY-2012195 ; 20 Aug. 2012

---

<b>Table 5.1.</b> continued.				
H_yunnanensis	434 ; 8	Female	Yunnan Province, China	Yuxidui 1151 ; 29 Oct. 1989 ; KUN#525329

---

### ***Ion Torrent Library Preparation and Sequencing***

The library preparation and sequencing methods have been slightly modified from Chapter 4 for: 1) the fragmentation, end repair, and ligation of DNA; 2) cleanup and the 200 base-read library (~330 bp median library size) selection; 3) addition of multiplex barcodes ( $n=8$ ); and 4) use of the 318 chip (Visi personal communication). NEBNext® Fast DNA Fragmentation and Library Prep Set for Ion Torrent (New England Biolabs® Inc., Ipswich, Massachusetts, USA; catalog # E6285L; version 3.2, 3/13) was followed according to the manufacturer's recommendations with 0.5-1 µg of extracted DNA. DNA fragmentation, end repair, and adaptor ligation were performed following manufacture protocols. PCR amplification of the adaptor ligated DNA was performed on a thermocycler with the following parameters: 40 µL of the 200 bp size selected adaptor ligated DNA, 4 µL of primers, 50 µl NEBNext High-Fidelity 2X PCR Master Mix, and 6 µL of sterile H<sub>2</sub>O to bring the reaction to a total volume of 100 µL. PCR cycling conditions were as follows: 1) initial denaturation at 98°C for 30 seconds; 2) denaturation at 98°C for 10 seconds, annealing of primers at 58°C for 30 seconds, and extension at 72°C for 30 seconds that was repeated for 10 cycles; and 3) a final extension at 72°C for 5 minutes, then held at 4°C. The forward barcoded ligated adaptors were from NEXTFlex DNA Barcodes for Ion Torrent (BIOO Scientific). Barcodes 1-8 were assigned to the recognized taxonomic species and variety of *Humulus* (Small 1978). After library preparation and cleanup of adaptor ligated DNA, 200 base-read library size (~330 bp median library size) selection was carried out using Agencourt Ampure XP Beads (Beckman Coulter, Inc.). To obtain ~200 bp fragments, a dual bead-based size selection was performed with two enrichments. 0.7X bead:DNA ratio of the fragmented DNA allowed the binding of large fragments and removing them from the solution. A subsequent 0.15X bead:DNA ratio size selected for the 200 bp fragments, which had a range of 310-370 that included the ligated Ion Torrent specific adaptors.

Final library clean-up was performed with AMPure XP Beads by adding 1X volume of beads (100 µl) to the sample, then followed the standard protocol. Approximately 20 µl of the final library was eluted in 0.1X TE and analyzed using the Agilent 2100 Bioanalyzer for quantification. The Agilent 2100 Bioanalyzer eliminated

the need for qPCR (as done in Chapter 4), and this capillary gel-electrophoresis unit allowed for the quantification of DNA as well as DNA fragment size, essential to high-throughput sequencing protocols. On the Bioanalyzer, 1 µl of each sample was added to separate wells, and after quantification, each sample was diluted to the appropriate amount (i.e., 26 pM). Samples of the recognized taxonomic species and varieties of *Humulus* ( $n=8$ ) were pooled in equimolar amounts and prepared according to the manufacturer's protocol (publication part number 4478871Rev. A, revision date 4 June 2012) on the Ion OneTouch™ 200 template kit v2 (cat #4478316). Subsequent preparation of template-positive Ion OneTouch™ 200 Ion Sphere™ Particles (ISPs) was done on the Ion OneTouch™ ES instrument. Whole genomic DNA sequencing was performed on the Personal Genome Machine (Ion Torrent) with ISPs loaded on a 318 chip and the Ion PGM™ Sequencing 200 Kit v2 (cat #4482006). Initial filtering and trimming of raw reads for low quality and ambiguous bases was performed on the Torrent Suite™ Software 3.4.2 that runs on the Torrent Server and confirmed with *Core NGS Tools-Create Sequencing QC Report tool* on CLC Genomics Workbench (CLC GW).

### ***Reference Mapping Assemblies with CLC GW and Plastome Matrix Analyses***

After importing the SFF files for each barcode generated from the Torrent Suite™ Software to the CLC GW, additional filtering was done by trimming reads at a 0.05 quality limit (minimum 13 Phred) and minimum length of 50 bp with the *Core NGS Tools-Trim Sequences tool*. Assembly of the filtered nucleotide sequences into draft plastomes was done using *Cannabis sativa* L. as a reference (van Bakel et al. 2011, PK\_chloroplast available at the URL: <<http://genome.ccbr.utoronto.ca/cgi-bin/hgTracks?hgsid=44159&chromInfoPage=>>) and the *Core NGS Tools-Map Reads to Reference tool*. With the default reference mapping parameters set to automatic word score (k-mer number), automatic bubble score, 80% similarity and 50% length mapping of the sequenced reads, the *Core NGS Tools-Extract Consensus Sequence tool* generated a continuous sequence filled with Ns for the zero coverage regions after default reference mapping to generate a draft plastome for each taxa. The draft *Humulus* plastomes were aligned with the *C. sativa* reference using the MAFFT version 7.037 (2013/Apr/25)

online server (<http://mafft.cbrc.jp/alignment/server/>). The resulting Cannabaceae (*s.s.*) plastome matrix was saved/exported as a fasta file for further analyses. Default settings in MEGA 5.2 (Tamura et al. 2011), DnaSP v5.10.01 (Rozas 2009), and CLC Genomics Workbench v7 (CLC bio Inc., Aarhus, Denmark) were used for comparisons of nucleotide sequences and haplotype reconstructions on the plastome matrix.

### ***Phylogenomic Analyses***

Phylogenies for the Cannabaceae (*s.s.*) plastome matrix were constructed with MEGA5.2 (Tamura et al. 2011) and BEAST 2 (Bouckaert et al. 2013). For Maximum Likelihood (ML) phylogenies, the optimal model of sequence evolution for the aligned plastome matrix was used done in MEGA5.2 by generating an automatic neighbor-joining tree, using all sites, and applying a branch swap filter of very strong. The resulting model chosen for ML analysis had the lowest BIC score (Bayesian Information Criterion), which is considered to describe the substitution pattern the best. Generation of the ML plastome phylogeny (MLPP) was conducted using the maximum number of threads available on the computer with the following parameters: 1) all sites, 2) the model of evolution was the General Time Reversal + Gamma (GTR+G) with five discrete gamma categories, 3) nearest-neighbor-interchange, 4) an initial neighbor-joining tree made automatically, 5) applying a branch swap filter of very strong, and 6) 1000 bootstrap replicates.

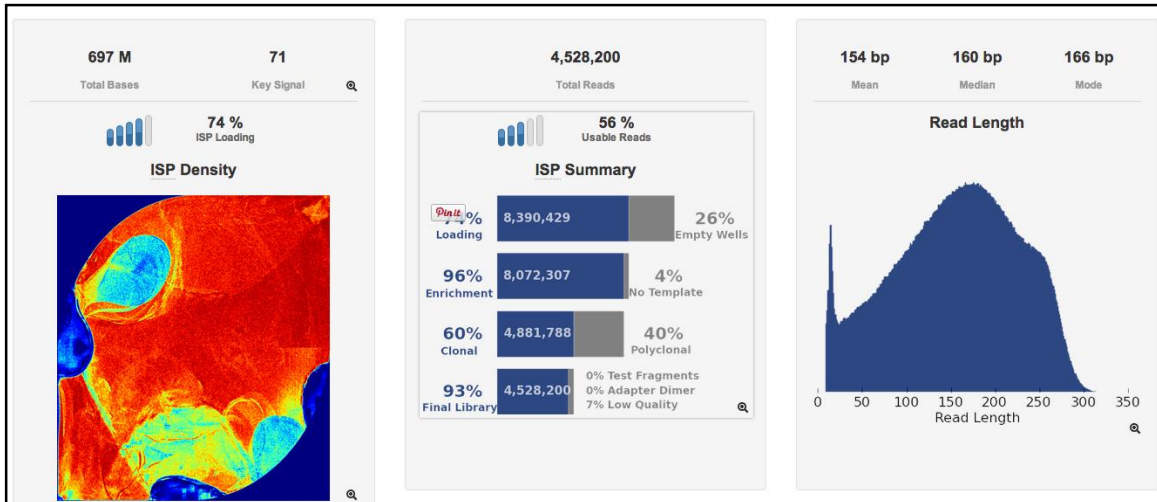
Generation of a Bayesian inference plastome phylogeny (BIPP) using BEAST 2 was done by exporting each aligned nucleotide matrix from MEGA5.2 as a Nexus file (PAUP 4.0) with the following: 1) all sites were displayed, 2) the data type was changed from nucleotide to DNA, and 3) the file was saved with the extension .nex. Next, the .nex file alignment is imported to BEAUti (included in the BEAST 2 software package) where the parameters set for the Site Model tab included a substitution rate of 1.0, gamma category count of 4, shape 1.0 with the estimate box checked, proportion invariant 0.1 with the estimate box checked, and the add-on substitution model of Reversible-jump Based substitution (v.1.0.3) (RB). The RB allows the MCMC chain to switch between nucleotide substitution models to search for the best fit model for the data set. For the

Clock Model tab, relaxed clock log normal was selected. The default parameters in the Priors tab were left at estimate for the Yule Model tree. A single fossil constraint was applied to the root of the monophyletic Cannabaceae (*s.s.*) clade with the uniform prior range of 66-150 million years. The lower fossil constraint at 66 million years correlates to the estimated age of leaf macrofossils for *Humulus* at the K/T event in North Dakota (70.6 - 65.5 mya; Johnson 2002). The upper constraint at 150 million years correlates to the estimated age of the dicot-monocot divergence (Bell et al. 2010, Chaw et al. 2004, Crepet et al. 2004). The MCMC Chain parameters were: 1) length was set at 100,000,000 cycles, 2) stored every -1, 3) Pre Burnin 0, and 4) logged every 1000 cycles to generate 100,001 trees. After the priors were set, the file was saved in BEAUti with the standard .xml extension. After opening the .xml in BEAST 2 and selecting the automatic thread pool size, a .log file was generated by BEAST 2 after the run came to completion. The .log file output from BEAST 2 was viewed with Tracer v1.5 (available at URL: <<http://tree.bio.ed.ac.uk/software/tracer/>>), which depicts the outcome of parameters set in BEAUti with each tree sampled along the MCMC chain for likelihood and posterior effective sample size (ESS) (e.g., need ESS to be all black color font or >200). Using the TreeAnnotator (included in the BEAST 2 software package) to combine only the last 11,001 trees sampled after a burnin of 89,000 trees on the MCMC chain for final likelihood and posterior probability estimates, the .trees file was saved with the extension .tree. The annotated .tree file output from TreeAnnotator was viewed using FigTree v1.4.0 (available at URL: <<http://tree.bio.ed.ac.uk/software/figtree/>>). Topologies for both MLPP and BIPP were compared, and if similar, the bootstrap values and posterior probabilities for each main node were added to a final consensus tree.

## Results

### *Ion Torrent Sequencing*

An 8-plexed barcode sequencing run on the Ion Torrent Personal Genome Machine (PGM) using a 318 chip for 200 base pair reads resulted in a total of 4,528,200 sequences of quality filtered reads with a mean read length of 154 bp, representing 697 Mb of sequencing data from the Torrent Suite™ Software (Fig 5.2).



**Fig. 5.2.** PGM summary for an 8-plexed barcode sequencing run on a 318 chip for 200 bp single reads. Image was generate from the Torrent Suite™ Software.

### *Reference Mapping Assemblies with CLC GW and Plastome Matrix Analyse*

DNA Extraction, library preparation, high-throughput sequencing, and plastome alignment were unproblematic for the recognized taxonomic species and varieties of *Humulus*, expect for the sample H\_lupulus\_var\_lupuloides that was excluded from the final analyses due to an insufficient number of total read count (Table 5.2). The length of the aligned nucleotide matrix for the seven remaining hop plastomes with *Cannabis* as the outgroup was 155,701 bp. Table 5.3 summarizes the number of conserved, variable, parsimony-informative, and singleton sites for plastome matrix. The estimates of average evolutionary divergence over all sequence pairs for the number of base differences per sequence was 909 nucleotide differences (p-distance=0.006). Between taxa, the number of base differences per sequence and the mean distance estimates ranged from 226-1,875 and 0.001-0.015, respectively (Table 5.4). A total of 1,460-2,409 gaps were observed

between the in-group of *Humulus* spp. and the out-group of *Cannabis*. In addition, 3,643-4,587 differences were observed between the in-group and out-group. Furthermore, 952-2,496 gaps and 1,195-3,661 differences were found within the *Humulus* spp. sampled (Table 5.5). The number of haplotypes (h) discovered for plastome matrix was h=8 (Hd: 1.0000).

**Table 5.2.** Summary of CLC GW *Map Reads to Reference tool* report for the recognized taxonomic species and varieties of *Humulus*. Sample H\_lupulus\_var\_lupuloides was excluded from the final plastome matrix due to a low total read count of 459.

Parameter	Taxa				
	pubescens	cordifolius	lupulus_253	lupulus_43	neomexicanus
Reference count	1	1	1	1	1
Type	Read mapping	Read mapping	Read mapping	Read mapping	Read mapping
Total reference length	152,942	152,942	152,942	152,942	152,942
GC contents in %	36.69	36.69	36.69	36.69	36.69
Total consensus length	152,937	152,814	152,842	152,846	152,863
Fraction of reference covered	0.99	0.99	0.99	0.99	0.98
Total read count	10,004	14,013	22,051	12,408	7,971
Mean read length	148.36	150.69	160.39	157.95	154.45
Total read length	1,484,206	2,111,689	3,536,731	1,959,840	1,231,160
Minimum coverage	0	0	0	0	0
Maximum coverage	28	39	70	43	32
Average coverage	9.11	13.44	22.60	12.47	7.82
Standard deviation coverage	3.89	5.35	9.55	5.55	3.81
Minimum excl. zero coverage regions	1	1	1	1	1
Average excl. zero coverage regions	9.21	13.53	22.81	12.57	7.95
Standard deviation excl. zero coverage regions	3.80	5.25	9.34	5.46	3.70
Zero coverage regions Count	33	28	32	29	76
Zero coverage regions Minimum length	1	1	1	1	1
Zero coverage regions Maximum length	509	118	416	391	296
Zero coverage regions Mean length	49.33	36.00	43.88	41.38	33.21
Zero coverage regions Standard deviation	92.62	33.16	71.74	70.61	36.69
Zero coverage regions Total length	1,628	1,008	1,404	1,200	2,524

**Table 5.2.** continued.

<b>Parameter</b>	<b>Taxa</b>		
	scandens	yunnanensis	lupuloides
Reference count	1	1	1
Type	Read mapping	Read mapping	Read mapping
Total reference length	152,942	152,942	152,942
GC contents in %	36.69	36.69	36.69
Total consensus length	153,031	152,851	153,082 (mostly Ns)
Fraction of reference covered	0.99	0.99	0.31
Total read count	15,979	11,313	459
Mean read length	159.25	160.70	146.10
Total read length	2,544,628	1,817,973	67,061
Minimum coverage	0	0	0
Maximum coverage	54	39	6
Average coverage	16.22	11.58	0.41
Standard deviation coverage	7.31	4.69	0.72
Minimum excl. zero coverage regions	1	1	1
Average excl. zero coverage regions	16.35	11.68	1.34
Standard deviation excl. zero coverage regions	7.01	4.58	0.68
Zero coverage regions Count	37	30	280
Zero coverage regions Minimum length	1	1	3
Zero coverage regions Maximum length	435	387	2,382
Zero coverage regions Mean length	32.62	45.70	378.82
Zero coverage regions Standard deviation	69.68	73.79	390.11
Zero coverage regions Total length	1,207	1,371	106,071

**Table 5.3.** Number of conserved, variable, parsimony-informative, and singleton sites for aligned plastome matrix of 155,701 bp conducted in MEGA5.2.

<b>Region</b>	<b># conserved</b>	<b># variable</b>	<b># parsimony-informative</b>	<b># singletons</b>
plastome	148,994	3,948	1,090	2,772

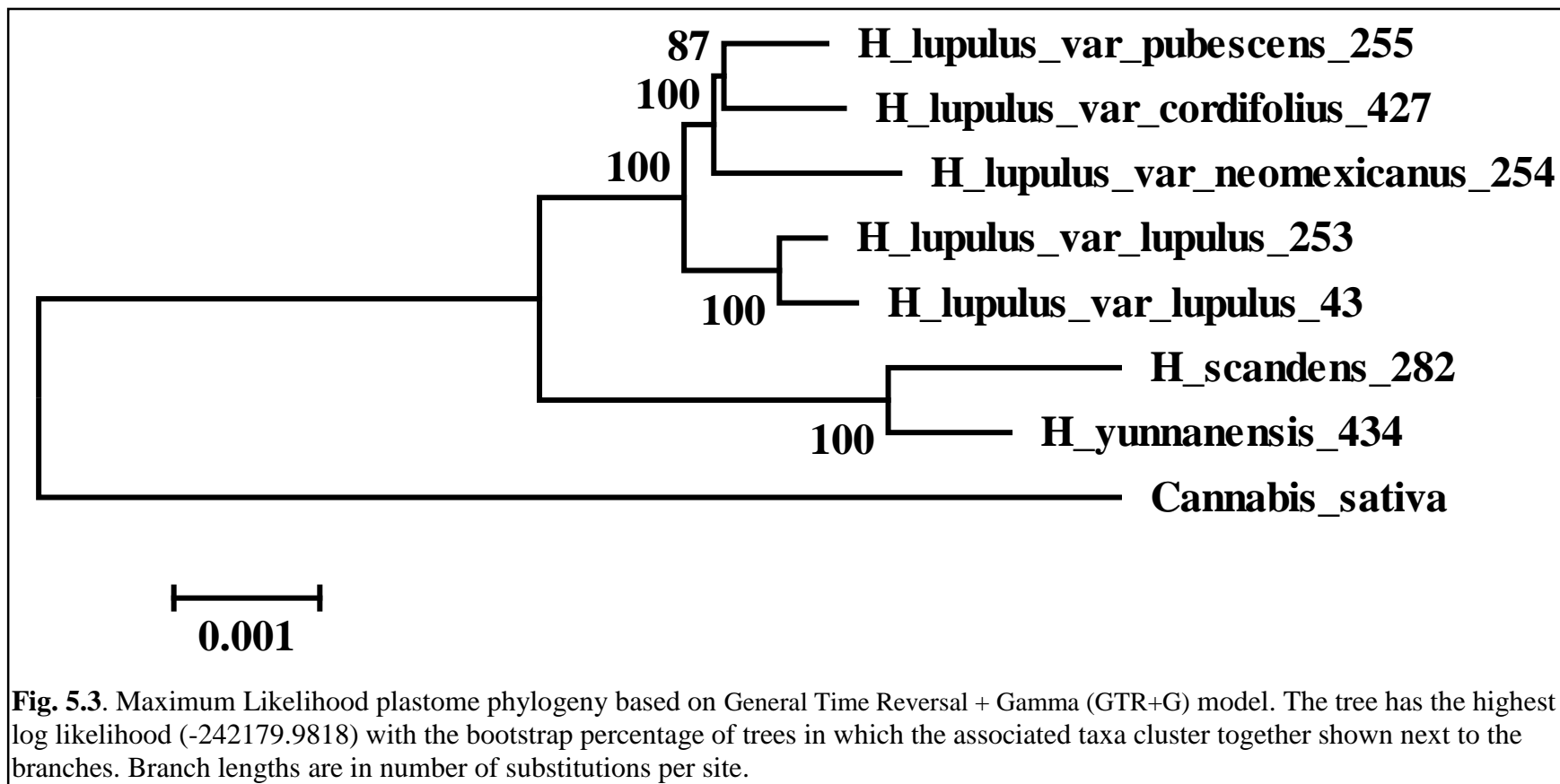
**Table 5.4.** Estimates of evolutionary divergence over sequence pairs between taxa conducted in MEGA5.2. The number of base differences per sequence and the mean distance estimates are on the bottom and top diagonals, respectively

Taxa	1	2	3	4	5	6	7	8	
H_lupulus_var_pubescens_255	1		0.002	0.002	0.002	0.002	0.006	0.005	0.013
H_lupulus_var_cordifolius_427	2	226		0.002	0.002	0.002	0.006	0.005	0.013
H_lupulus_var_lupulus_253	3	290	295		0.001	0.002	0.006	0.005	0.013
H_lupulus_var_lupulus_43	4	322	331	124		0.003	0.006	0.005	0.013
H_lupulus_var_neomexicanus_254	5	300	322	365	391		0.007	0.006	0.013
H_scandens_282	6	872	882	865	911	959		0.002	0.015
H_yunnanensis_434	7	766	776	757	787	845	356		0.014
Cannabis_sativa	8	1875	1904	1878	1902	1930	2161	2060	

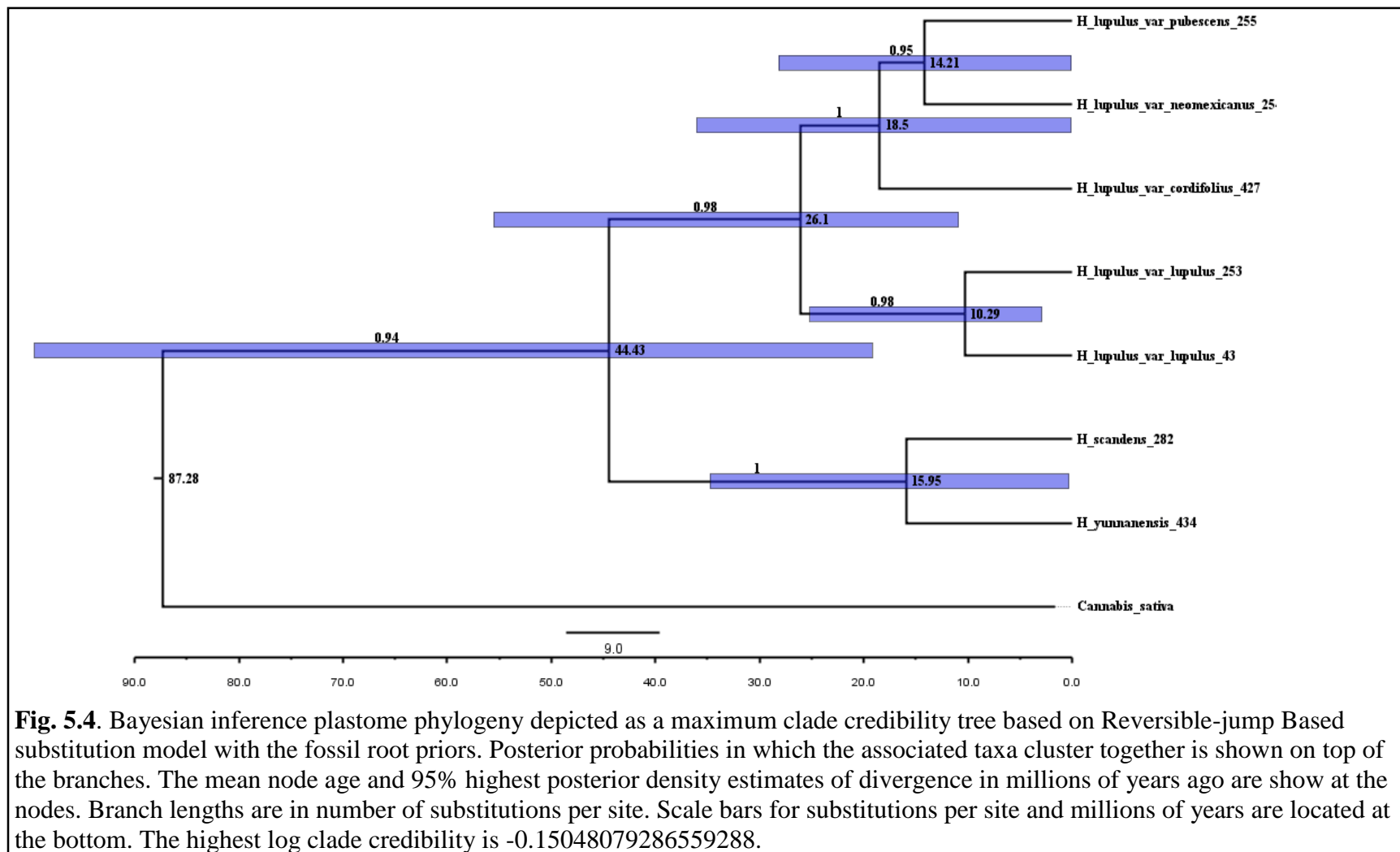
**Table 5.5.** Plastome matrix comparisons for *Humulus* and *Cannabis* samples with the number of gaps on the top and number differences on the bottom diagonals for the nucleotide data set. The table was generated in CLC Genomics Workbench 7.

Taxa	1	2	3	4	5	6	7	8	
Cannabis_sativa	1		2001	1848	2409	1460	1520	1493	1533
H_lupulus_var_pubescens_255	2	4131		1139	2022	1663	1687	2056	2026
H_lupulus_var_cordifolius_427	3	4146	1468		1911	1520	1566	1901	1925
H_lupulus_var_neomexicanus_254	4	4587	2430	2419		2117	2177	2496	2538
H_lupulus_var_lupulus_253	5	3643	2082	2001	2629		952	1477	1479
H_lupulus_var_lupulus_43	6	3749	2150	2106	2744	1195		1527	1513
H_scandens_282	7	3998	3177	3071	3661	2575	2707		984
H_yunnanensis_434	8	3908	2987	2974	3580	2445	2550	1451	

MLPP and BIPP phylogenies show two clades of *Humulus* with *Cannabis* as the outgroup (Figs. 5.3 and 5.4). The MLPP tree has high bootstrap support at the nodes for a *H. lupulus* clade (100%) and a *H. scandens/H.yunnanensis* clade (100%). Within the *H. lupulus* clade, two additional clades are composed of: 1) the samples H\_lupulus\_var\_lupulus\_253 and H\_lupulus\_var\_lupulus\_43 with high bootstrap support (100%) and 2) the samples H\_lupulus\_var\_pubescens\_255 and H\_lupulus\_var\_cordifolius\_427 with high bootstrap support (87%) that are sister to H\_lupulus\_var\_neomexicanus\_254 with high bootstrap support (100%). Similarly, the BIPP tree shows high posterior probability support at the main nodes for a *H. lupulus* clade (0.98) and a *H. scandens/H.yunnanensis* clade (1.0). Furthermore for the BIPP tree, high support was found for the interior nodes of: 1) the samples H\_lupulus\_var\_lupulus\_253 and H\_lupulus\_var\_lupulus\_43 with high posterior probability support (0.98) and 2) the samples H\_lupulus\_var\_pubescens\_255 and H\_lupulus\_var\_neomexicanus\_254 with high posterior probability support (0.95) that are sister to H\_lupulus\_var\_cordifolius\_427 with high posterior probability support (1.0). The mean node age and 95% highest posterior density (HPD) estimates of divergence in millions of years ago (mya) roots the Cannabaceae at 87.28 mya for the split between *Humulus* and *Cannabis* (HPD lower 66 mya; HPD upper 136.5 mya). The split at the node for a *H. lupulus* clade and a *H. scandens/H.yunnanensis* is at 44.43 mya. Within the *H. lupulus* clade, a node at 26.1 mya separates the samples H\_lupulus\_var\_lupulus\_253 and H\_lupulus\_var\_lupulus\_43 from the other three varieties. Divergence between the samples H\_lupulus\_var\_lupulus\_253 and H\_lupulus\_var\_lupulus\_43 occurred approximately 10.29 mya. In addition, H\_lupulus\_var\_cordifolius\_427 from East Asia split from the New World samples H\_lupulus\_var\_pubescens\_255 and H\_lupulus\_var\_neomexicanus\_254 approximately 18.5 mya, with an estimated split at 14.21 mya for the New World samples. *H. scandens* and *H.yunnanensis* diverged from one another approximately 15.95 mya.



**Fig. 5.3.** Maximum Likelihood plastome phylogeny based on General Time Reversal + Gamma (GTR+G) model. The tree has the highest log likelihood (-242179.9818) with the bootstrap percentage of trees in which the associated taxa cluster together shown next to the branches. Branch lengths are in number of substitutions per site.



## **Discussion**

Within *Humulus*, there are two sister groups with high support from both the ML bootstraps and BI posterior probabilities: (1) the varieties of *H. lupulus* clade and (2) the *H. scandens* and *H. yunnanensis* clade. From the Cannabaceae plastome matrix, the nucleotide differences and mean distance estimates between the sister taxa show *H. scandens* and *H. yunnanensis* have slightly more base pair changes, possibly because *H. scandens* and perhaps *H. yunnanensis* are annual species that have undergone more mutations per generation than the perennial *H. lupulus*. Although, *H. yunnanensis* is putatively perennial, but the habit and ecology of *H. yunnanensis* remains to be confirmed. No herbarium specimens, living plants, or seeds have been collected for *H. yunnanensis* for almost two decades in the mountains of Yunnan Province.

In the *H. lupulus* clade, the basal taxon is either *H. lupulus* var. *neomexicanus* originally collected from the wild in Colorado (MLPP) or *H. lupulus* var. *cordifolius* from Yunnan Province (BIPP). In contrast to the derived placement of *H. lupulus* var. *pubescens* within the East Asia and New World clade of the common hop having varied support (87%:0.95), the other sister taxa relationships for *H. lupulus* var. *lupulus*, *H. scandens*, and *H. yunnanensis* are static. Overall for the MLPP and BIPP topology, the New World taxa are nested and polyphyletic within the Old World taxa to support an East Asian origin of *Humulus*. Therefore, the null hypothesis cannot be rejected now from a phylogeographic perspective:  $H_0$  = Phylogenomic methods support the genus *Humulus* originated in the Old World (i.e., China) and migrated to the New World (Murakami et al. 2006a, Neve 1991).

Further modeling in BEAST 2 with more selective priors (i.e., specific substitution rates for the complete annotated Cannabaceae plastome, the entire rDNA cistron along with additional fossil calibrations, and duplicate sampling) is a priority to revisit the hypothesis:  $H_1$  = After the K/T extinction event around 65 million years ago, open niches along riparian areas in the New World mountains (i.e., Colorado Rocky Mountains) allowed the perennial binning habit of *Humulus* to thrive in forest ecotones

and subsequently migrate to the Old World, then undergoing adaptive radiation into an two additional annual species that are putatively not interfertile within the *Humulus lupulus* complex. Based on the fossil record (Paleobiology Database 2013), the oldest classified *Humulus* leaf fossil was found near the time of the K/T event, 70.6 - 65.5 million years ago, at the Hell Creek and lower Fort Union Formations in the western Dakotas [see Johnson 2002 for *Humulus* (morphotype HC243 (n=93))], along with other Cannabaceae leaves [see Johnson 2002 for morphotypes HC81 (n=365) and HC225]. The second oldest *Humulus* leaf fossil (MacGinitie 1953; specifically see MacGinitie 1969:132 and Weber 2003 commenting that the taxa is not to be a member of *Vitis*) was found at 37.2 - 33.9 million years ago at the Florissant Beds in Colorado. Next in ancient geologic time, *Humulus* macrofossils are from the Russian Federation at 11.6 - 5.3 million years old (Dorofeev 1963) and then from Germany at 3.6 - 2.6 Ma years ago (Mai and Walther 1988). Confirming more macrofossils are truly *Humulus* and not another similarly leaved riparian taxon like *Vitis* for additional fossil constraints in BEAST 2 is a priority to conclude either a New or Old World origin of *Humulus*. In parallel, the annual erect habit of *Cannabis* further supports an ancient Central Asian origin for hemp along the common steppe environments of Eurasia where *Cannabis* and wild relatives thrive.

In this study, the mean root age of the Cannabaceae at 87.28 mya is similar to the mean root age for the sister family, the Moraceae, at 89.1 (72.6-110.0) mya (Zerega et al. 2005; also see Chen et al. 2012 for a discussion that *Humulus* is the best outgroup for the Moraceae). Zerega et al. (2005) clearly discusses the historical biogeography and a Laurasian migration hypothesis to explain the Moraceae distribution. The Laurasian migration hypothesis is as likely probable for the Cannabaceae (*s.s.*). For *Humulus* as well as *Cannabis*, the most ancient fossils are from the New World, and most Eurasia fossils are much younger. Furthermore, Zerega et al. (2005) found an estimated node age for the diversification of *Humulus* and *Cannabis* at 21 mya in the early-Miocene; however, in their study, the second outgroup sequences (i.e., *Celtis philippensis* Blanco, Cannabaceae *s.l.*) were pruned from the final fossil calibrated tree. With high support for

the placement of *Humulus* and *Cannabis* within the Cannabaceae but without high support for the precise placement of the sister genera in the Cannabaceae (Sytsma et al. 2002, Yang et al. 2013), the estimated node age for the diversification of *Humulus* and *Cannabis* at 21 mya in Zerega et al. (2005) may be too young. In addition, Murakami et al. (2006a) found much younger estimates for the split between the *H. lupulus* clade and *H. scandens*/*H. yunnanensis* clade at 6.55 mya, than 44.43 mya as found in this study. For further divergence comparisons of the Cannabaceae (*s.s.*) to groups within the sister family (e.g., Moraceae), *Ficus* radiation putatively began around 43.3 (40.1-51.0) mya and the Moraceae (*s.s.*) began around 58.6 (44.2-75.2) mya. Until macro- or microfossil evidence from Eurasia surfaces prior to the time of the K/T event 65 million years ago, at the very least *Humulus* and *Cannabis* (Cannabaceae *s.s.*) are supported in this study by a much more ancient divergence and a mid-Cretaceous origin. With the age and diversification of angiosperms now readily testable with complete chloroplast genome phylogenies (Chaw et al. 2004), the age and diversification of the Cannabaceae (*s.s.*) merits to be re-revisited (Bell et al. 2010). Specifically, duplicate samples for each taxa within *Humulus* added to the phylogeny presented here, particularly wild samples from East Asia and the New World, will give the crucial support for an ancient Laurasian origin of *Humulus* that putatively arose in what is now the New World.

## **Conclusion**

Based on total genomic DNA sequencing and comparisons of plastomes from the Cannabaceae (*s.s.*), the four main objectives of this study were accomplished to support the hypothesis (H<sub>5.1</sub>). An approach focusing on surfing a plant's total genomic DNA to capture the chloroplast genome (plastome) was successfully conducted with both fresh and herbarium material. As a result, the entire plastome should be the minimum DNA barcode for plant species (Boutain et al. 2013). High homology was found between the *Humulus* and *Cannabis* plastomes. The phylogenetic relationships for the intra- and interspecific taxa of *Humulus* were mostly resolved. A much more ancient age than previously proposed for the *Humulus* and *Cannabis* split in the Cannabaceae (*s.s.*) is reported here as a mid-Cretaceous and Laurasian origin. Compared to single or several DNA regions used to barcode plants, the phyloplastome model supports the unity of the *H. lupulus* complex. Further duplicated sampling of the recognized taxonomic species and varieties of *Humulus* is required to confirm the evolutionary relationships presented here.

### **Acknowledgements**

I thank Professor Jianchu XU, Ende LIU, Qiaohong LI, Mingming WANG, Jing-hau WANG, the Kunming Institute of Botany, Chinese Academy of Sciences, the Institute of Botany, Chinese Academy of Sciences, the Ministry of Science and Technology of the People's Republic of China, World Agroforestry Centre China Program, Jie LIU, Zhang-Yuan WU, Mei-qing YANG, Pei-Luen LU, Dake ZHAO, Yongtao ZHAO, Professor Arthur Eisenberg, Professor Michael Allen, David Visi, Richard Hale, Marnie Rout, Yan Zhang, Vicki Funk, Keri McNew Barfield, Tiana Franklin Rehman, Alison Sherwood, Seana Walsh, Mark Nickum, Tim Gallaher, Jason Cantley, Marjorie and Thomas Day, the University of North Texas Health Science Center, the University of North Texas at Denton, the Botanical Research Institute of Texas, the United States Department of Agriculture National Clonal Germplasm Repository in Corvallis, Kim Hummer, John Henning, Nahla Bassil, Bruce Bartlett, Ernest Small, the Agriculture and Agri-Food Canada, the National Science Foundation - East Asia and Pacific Summer Institutes for U.S. Graduate Students (Award ID: 1108575), the Graduate Student Organization at the University of Hawai'i at Mānoa (UH), the Beatrice Krauss Fellowship from the Botany Department at UH, the Anne S. Chatham Fellowship in Medicinal Botany from the Garden Club of America, and numerous office, lab, and herbarium colleagues and counterparts in China, Texas, and Hawaii. A version of this paper was present at the 5<sup>th</sup> International Barcode of Life Conference in Kunming, Yunnan, China in October 2013. Any errors are solely by the author.

## CHAPTER 6

### Outcomes, Applications, and Directions for Cannabaceae Research

#### Introduction

Preliminary studies of the taxonomic history of *Humulus* reveal that it is integrally intertwined with the much more extensive history of the Cannabaceae Martinov (1820). The Cannabaceae was recently expanded from two genera to 10 genera to include the sister taxa in the Celtidaceae (Tropicos.org 2013, Yang et al. 2013). In order to understand the full cultural and economic value of these plants, it is important to view the evolutionary history of the family as a whole. Before the proposed family revision that was based on molecular sequences of the Urticalean rosids (Sytsma et al. 2002) and a recent investigation of the genera in further detail (Yang et al. 2013), the only other widely accepted member in the Cannabaceae, *Cannabis*, has been used to propose parallel hypotheses for biogeographical, ecological, ethnobotanical, phytochemical, systematic, and taxonomic investigations (Emboden 1974, Hammond and Mahlberg 1973, Haney and Kutscheid 1975, Hillig 2005, Hood et al. 1973, Merlin 1972, Schultes et al. 1974, Small and Cronquist 1976, Small et al. 1975, 1976, for the current synthesis on *Cannabis* see Clarke and Merlin 2013, Clarke personal communication). However, *Humulus* does not present the parallel prohibition laws or negative social connotation that would impede hypothesis driven research.

Moreover and since the beginning of this study, scientific or non-scientific discussions about *Humulus* with both scientists and non-scientists typically began with a discourse about hops in beer and the closest sister taxon, *Cannabis*, which has been legalized recently in Colorado and Washington not only for medicinal but for recreational purposes. These two genera of the Cannabaceae seem quite obviously to have an ancient and significant conscious association with humans; therefore, *Cannabis* and *Humulus* provide parallel model systems that attempt to understand how, when, why, and with what effects plants were used in the past (Merlin 1972).

A high priority for genome research is aimed at improving germplasm resources of important economic plants that are vital for humanity. With many historical plant collections preserved in herbaria around the world, an ideal cost effective research program will implement compact, high-throughput DNA sequencers to study the systematic relationships and putative heirloom genetics of these economic plants awaiting the opportunity to be tapped for their knowledge. If extinction is a driver of biodiversity, then human impacts on biodiversity have significantly decreased, as well as increased, the number of cultivars available for germplasm development in crop species and their wild progenitors. Studying historical herbarium specimens, wild crop relatives, plants that have escaped home gardens, and populations of plants found in biodiversity hotspots will assist germplasm programs to secure and sustain food, fuel, fiber, and medicine for humans in a changing world.

A genetic approach with DNA barcodes is successful at discovering the origin of unknown as well as known plants, particularly from cryptic herbarium specimens. Herbarium specimens housed at Harvard University Herbaria (HUH and associated collections), Herbarium of the Institute of Botany, Chinese Academy of Sciences, People's Republic of China, Beijing (PE), Herbarium of the Kunming Institute of Botany, Chinese Academy of Sciences, People's Republic of China, Kunming, Yunnan (KUN), Smithsonian Institution, United States National Herbarium, Washington, District of Columbia (US), and the Botanical Research Institute of Texas (BRIT and associated collections) were observed and studied for this dissertation. As a result, hundreds more *Humulus*, and even *Cannabis*, specimens await for further genomic research. With the costs of high-throughput sequencing dramatically dropping, the ease and success for phyloplastome analyses of a large number of *Humulus* herbarium specimens is now possible because both the hop and hemp plant genomes are underway.

### **The Cannabaceae Conundrum**

A taxonomy for the Cannabaceae (*s.s.* and *s.l.*) has been debated. For example, botanists disagree whether *Cannabis* is a monotypic or polytypic genus. With Linnaeus (1753, 2:1027) firmly ground on the taxonomy with one species, *C. sativa*, and Lamarck (1785) recognizing two species, *C. sativa* and *C. indica*, the detailed taxonomic investigation of wild *Cannabis* or even comprehensive and comparative studies of the range of variation found in cultivated hemp was totally lacking (Schultes et al. 1974). Until this gap in knowledge was brought to the attention of botanists, the taxonomic neglect in *Cannabis* was then systematically investigated (see Hillig 2005, Small and Cronquist 1976).

Furthermore, with advances in high-throughput sequencing becoming cheaper and more widely accessible to scientists in fields outside of biomedicine, forensics, ancient DNA, epidemiology, and cancer studies, the ‘next,’ ‘third,’ ‘fourth,’ and future generations of DNA sequencers are providing evidence that supports a dichotomy between *Cannabis* plants, which either do or do not show gene expression for tetrahydrocannabinolic acid (THCA) synthase and cannabidiolic acid (CBDA) synthase (i.e., marijuana and hemp, respectively). With the draft genome and transcriptome of *Cannabis sativa* by van Bakel et al. (2011), the separation of two hemp cultivars (‘Finola’ and ‘USO31’) from two marijuana strains (Purple Kush and Chemdawg) suggests additional analysis of diverse germplasm is warranted to investigate of the evolutionary history and the molecular impact of domestication and breeding on *Cannabis*. “Outstanding areas that might be addressed by further genomic investigation include whether the genus is composed of one or several species, the existence of ‘sativa’ and ‘indica’ gene pools, the relative contributions that wild ancestors have made to modern hemp and marijuana germplasm, and the process by which cannabis was first domesticated by humans (van Bakel et al. 2011).”

*Cannabis* is believed to have originated in Central Asia and adjacent regions to the south and west (Clarke and Merlin 2013). This vast area, which includes a great diversity of geographical zones and ecological conditions, makes it very difficult to pinpoint any specific area of origin or to determine how great the geographical distribution of wild hemp was before the advent of man (Anderson 1952, Schultes et al.

1974). Divergent populations of *Cannabis* could have arisen into distinct forms both morphologically and ecologically to be considered species, subspecies, and varieties (Schultes et al. 1974, Small and Cronquist 1976). As humans began to domesticate one or more of these putative species of *Cannabis* and transport them from place to place, hybridization occurred between the wild species and the early cultigens that originated from ‘dump heaps’ (Anderson 1952, Schultes et al. 1974) and hearths (Sauer 1952). Through continual introgressive hybridization with cultivated hemp, some of the original wild species of *Cannabis* may have gradually become extinct (Burger et al. 2008, Schultes et al. 1974). This process increased the variability in the gene pool of the cultivated plants and must have imparted some of the unique characters of the wild species to the cultivars (Schultes et al. 1974, also see Morrell and Clegg 2007). Further credence is given by the fact that great morphological variation exists between populations of cultivated hemp in various parts of Eurasia in characters which have not been selected for by humans (e.g., leaf size and shape; pigmentation of stem and fruit (Schultes et al. 1974, see Small 1978 for the vegetative features of *Humulus*); Boutain personal observation of hundreds of *Humulus* and *Cannabis* herbarium specimens).

The reproductive biology of different strains of cultivated *Cannabis* indicate that these plants are fully interfertile (Clarke 1993, Clarke and Merlin 2013, Schultes et al. 1974, Small 1972), which is a phenomena also seen in the *H. lupulus* complex (Small 1978, 1980, 1981). However, sterility barriers may not exist within *Cannabis* or *Humulus*, specifically in wild populations that have yet to be examined (Schultes et al. 1974). Geographically, varying degrees of reproductive isolation have been hypothesized for *Humulus* (Small 1978), but reproductive isolation can occur by other means than sterility barriers (Schultes et al. 1974).

Case in point, the combinations of morphological, anatomical, chemical, and genetic characters in *Cannabis* have maintained their integrity in spite of hybridization in a taxon of annual habit (Hillig 2005, Merlin 1972, Schultes et al. 1974). The maintenance of characters traits is a better indication of reproductive barriers than that resulting from experimentation with cultivated strains of doubtful and known origins (Schultes et al. 1974). Furthermore, well recognized species concepts vary from genus to genus and from

one family to another depending on the genetics of the group under consideration (Schultes et al. 1974; specifically see Chapter 4 in this dissertation for support of unique plastome structure in the Cannabaceae (*s.s.*)). Also, there is not the equivalence of units amongst plant families in the same sense of elements in chemistry, so species might be so rigorously defined to serve as the unit of evolution (Schultes et al. 1974). Taxonomists now hold that the population is the evolutionary unit, the biologically significant unit in plants (Funk et al. 2005, Hampton et al. 2001, Keeley et al. 2007, McClatchey 1998, Parker and Jørgensen 2003, Schultes et al. 1974, Small 1978, 1980, 1981, Steiger et al. 2002).

Outcomes of genetic, morphological, and chemotaxonomic analyses on *Cannabis*, accessions of known geographical origins showed a taxonomic revision is warranted (Hillig 2005, van Bakel et al. 2011). With this dissertation, a taxonomic revision for *Humulus* is also warranted; however, generating entire plastomes for two specimens of each putative *Humulus* taxa is a priority for future research to make final taxonomic conclusions. Furthermore, the required botanizing (Boutain and Gelderloos 2006) and collecting trips (Hampton et al. 2001, Hummer et al. 2002, 2003) to clarify the distribution of *Humulus* varieties in the proposed region of origin (i.e., China), as well as to understand the complete distribution of the putative endemic *H. yunnanensis* are vital to increase and conserve wild *Humulus* genetic resources (Hummer 2003, 2005, Khoury et al. 2013).

### **Main Research Hypotheses Revisited**

The main hypotheses in this dissertation were tested with the following conclusions:

Chapter 2) Rejecting H<sub>2.0</sub>, support was found for H<sub>2.1</sub> = The New World varieties of *Humulus* are clearly distinguishable using forensic, DNA barcode regions.

Chapter 3) Rejecting H<sub>3.0</sub>, support was found for H<sub>3.1</sub> = The Yunnan hop is a distinct species.

Chapter 4) Rejecting H<sub>4.0</sub>, support was found for H<sub>4.1</sub> = The entire plastome as a DNA barcode is ideal for genomic studies within the highly derived Cannabaceae (*s.s.*) compared to plastomes from the other Urticalean rosids and eudicots.

Chapter 5) Since New World taxa are polyphyletic within the Old World taxa, support was found for H<sub>5.0</sub> = Phylogenomic methods support the genus *Humulus* originated in the Old World (i.e., China) and migrated to the New World. Further modeling with selective priors and duplicated sampling is a priority to revisit the alternative (H<sub>5.1</sub>).

### **Unique Contributions to the Advancement of Cannabaceae Science**

Prior to conducting DNA barcode analyses of the putative species of *Humulus*, field observations of plant populations based on the locality information from herbarium specimens collected in the Yunnan Province of China indicated that much of the habitat where *Humulus* samples were once found through the early- to mid-20<sup>th</sup> century were no longer persisting. The current loss of native *Humulus* habitat by human influence on the landscape and climate change is similar in Eastern Asia as in North America (Hampton et al. 2001). With only one hop plant growing from 2010-2013 in Kunming Botanical Garden (Kunming Botanical Garden 2006; Boutain personal observation), both *H. lupulus* and *H. yunnanensis* were at one time part of the collection and even reproducing (Professor Shengji PEI personal communication). Using simple DNA barcodes (i.e., ITS2), the plant in Kunming Botanical Garden was confirmed to be *H. lupulus* var. *cordifolius*, not *H. yunnanensis* (Yang et al. 2013, Yang 005 (KUN), Yang personal communication).

Furthermore and as a result of short time in the field, the focus of this dissertation shifted from fresh field collections of *H. yunnanensis* to using historical botanical collections from KUN. A major contribution of this dissertation presents DNA extraction methods for herbarium material to be use with both Sanger and high-throughput sequencing technologies. Along with the simple DNA barcoding examples used to rediscover wild Michigan hops and support the endemic Yunnan hop as a unique species, an entire plastome barcoding approach brought better resolution to the plant group in question (i.e., Cannabaceae (*s.s.*)) with relative ease and low costs from compact high-throughput DNA sequencers. An additional contribution of this dissertation is the importance of public awareness of genomic technologies and how scientific knowledge should be disseminated to the public about crop conservation through phylogenomics. More specifically, the study of crop plants' evolutionary history with their wild relatives to secure the future of food, fuel, fiber, and medicine is vital for humanity (Anderson 1952, Kleinman 2003). Future surfers of genomes will have advanced software and hardware literally at their fingertips, but the first step is to conceptualize the complexity of DNA and genomics at any age level.

Lastly, photos and chemical analysis of hops grown in Hawaii supports the hypothesis that hops production by the plant does not require a period of vernalization. Therefore, hop cultivation and common garden experiments could be expanded to new regions (Boutain 2012b and 2012c, Seigner et al. 2009, Thomas and Schwabe 1969). An extensive study on the transcribed genes (transcriptomics) throughout the process of vernalization in *Humulus* spp. would be a fruitful research endeavor. Also, the conundrum and phylogenetic placement of *Humulus* and *Cannabis* within other members of the Cannabaceae (s.l.)/Celtidaceae is vital to support the phyloplastome conclusions presented here. To further educate surfers of genomes, this dissertation is disseminated on the world wide web at the URL: <<https://sites.google.com/site/jeffreymboutain/>>.

### **Overall Conclusion**

Simple DNA barcodes are quite useful for interspecific taxonomic questions in *Humulus*, while the complete plastome as a single DNA barcode shows promise for more resolved phylogenetic relationships in *Humulus*. The ITS2 barcode clearly resolves interspecies identification, and *petL-psbE* barcode further resolves inter- and intraspecific identities in *Humulus*. As with any phylogenetic study, the correct outgroup to test hypotheses about the ingroup is important. With a genome-scale visualization of the conserved and rearranged regions of the taxa under scrutiny, outgroup relationships for the Cannabaceae (s.s.), Urticalean rosids, and eudicots show the highly conserved plastome can indicate unique evolutionary histories via large inversions in the genome. The sequencing costs to surf a genome and capture the plastome barcode are dropping for compact high-throughput devices. As a result, the phylogenomic analyses presented here suggest the Cannabaceae (s.s.) is much more ancient than previously proposed (i.e., mid-Cretaceous). As theorized, the historical geographic location with the most number of species/varieties of cultivated plants within a given group is recognized as the center of origin for that group of cultivated plants (Vavilov 1992). For *Humulus*, results show East Asia as the mostly likely place of origin (Murakami et al. 2006a, Neve 1991, Small 1978, Wu et al. 2003). Lastly, a large scale phylogenomic study on the the Cannabaceae (s.l.)/Celtidaceae remains as a high priority for future research.

## REFERENCES

- Ahrends, A., C. Rahbek, M.T. Bulling, N.D. Burgess, P.J. Platts, J.C. ött, V.W. Kindemba, N. Owen, A.N. Sallu, A.R. Marshall, B.E. Mhoro, E. Fanning & R. Marchant. Conservation and the botanist effect. *Biological Conservation* 144:131-140.
- Almakarem, A.S.A., K.L. Heilman, H.L. Conger, Y.M. Shtarkman & S.O. Rogers. 2012. Extraction of DNA from plant and fungus tissues *in situ*. *BMC Research Notes* 5:266. Available at URL: <<http://www.biomedcentral.com/1756-0500/5/266>>.
- Anderson, E. 1952. *Plants, Man and Life*. University of California Press, Berkeley.
- Angiosperm Phylogeny Group. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161:105-121.
- Atherton, R.A., B.J. McComish, L.D. Shepherd, L.A. Berry, N.W. Albert & P.J. Lockhart. 2010. Whole genome sequencing of enriched chloroplast DNA using Illumina GAII platform. *Plant Methods* 6:22.
- Baker, H.G. 1978. *Plants and Civilization*. Wadsworth Publishing Company, Inc., Belmont, California.
- Barth, H.J., C. Klinke & C. Schmidt. 1994. *The Hop Atlas: The History and Geography of the Cultivated Plant, First English Edition*. Jon. Barth & Sohn., Nuremberg, Germany.
- Behr, J. & R.F. Vogel. 2009. Mechanisms of hop inhibition: hop ionophores. *Journal of Agricultural and Food Chemistry* 57:6074–6081.
- Behre, K-E. 1999. The history of beer additives in Europe – a review. *Vegetation History and Archaeobotany* 8:35-48.
- Bell, C.D., D.E. Soltis & P.S. Soltis. 2010. The age and diversification of the angiosperms: Re-revisited. *American Journal of Botany* 97:1296-1303.
- Bennett, B.C. 2007. Doctrine of Signatures: An explanation of medicinal plant discovery or dissemination of knowledge?. *Economic Botany* 61:246-255.
- Bouckaert, R., J. Heled, D. Kuehnert, T. Vaughan, C-H. Wu, D. Xie, M. Suchard, A. Rambaut & A.J. Drummond. 2013. BEAST 2: A software platform for Bayesian evolutionary analysis. Available at URLs: <<http://code.google.com/p/beast2/>> and <<http://beast2.cs.auckland.ac.nz>>.
- Boutain, J. 2012a. *On the Origin of Hops (Humulus, Cannabaceae): Ethnobotany, phylogeny, and next-generation hops in China*. Abstract and Poster Presented at the 4th Biennial Conference of International Association for Ecology & Health in Kunming, Yunnan, Kunming, P.R. China - October 15th-18th, 2012. Available at

URL:

<[http://www.ecohealth.net/pdf/Announcements/EcoHealth\\_2012\\_Final\\_Program\\_Oct15.pdf](http://www.ecohealth.net/pdf/Announcements/EcoHealth_2012_Final_Program_Oct15.pdf)>. Accessed 19 October 2013.

Boutain, J. 2012b. *Hopping into the Past, Present, and Future of Humulus (Cannabaceae)*. Oral presentation at Botanical Research Institute of Teaxs Visiting Researcher Seminar. Tuesday, December 18th, 2012. Available at URL: <<http://donations.brit.org/page.aspx?pid=443&srctid=1&erid=254615&trid=f4daf30f-945e-4013-92bf-5e93c125a0a9>>. Accessed 19 October 2013.

Boutain, J. 2012c. *Homegrown Hops in the Hawaiian Islands*. Abstract for poster presentation submitted to the 53rd Annual Meeting of the Society for Economic Botany in Frostburg, Maryland, USA - June 3rd-6th, 2012. Available at URL: <[http://www.econbot.org/pdf/2012\\_ScientificProgram.pdf](http://www.econbot.org/pdf/2012_ScientificProgram.pdf)>. Accessed 19 October 2013.

Boutain, J.R. & O.G. Gelderloos. 2006. The Jensen Landscape at Henry and Clara Ford's Home: An Eight-year comparison of canopy vegetation. *Meetings of Minds XIV Journal of Undergraduate Research* 8:135-140.

Boutain, J. & J. Xu. 2012. *On the Origin of Hops (Humulus, Cannabaceae)*. Abstract for oral presentation submitted to the 53rd Annual Meeting of the Society for Economic Botany in Frostburg, Maryland, USA - June 3rd-6th, 2012. Available at URL: <[http://www.econbot.org/pdf/2012\\_ScientificProgram.pdf](http://www.econbot.org/pdf/2012_ScientificProgram.pdf)>. Accessed 19 October 2013.

Boutain, J.R., T.J. Gallaher & B. Wadman. 2010a. *Spatial correlation of the regions of grain domestication with frequencies of human alcohol dehydrogenase haplotypes imply separate and independently derived fermentation technologies*. Abstract for poster presentation submitted to the 51st Annual Meeting of the Society for Economic Botany in Xalapa, Veracruz, Mexico - June 6th-10th, 2010. Available at URL: <[http://www.econbot.org/\\_organization\\_/07\\_annual\\_meetings/meetings\\_by\\_year/2010/seb\\_cuadernillo.pdf](http://www.econbot.org/_organization_/07_annual_meetings/meetings_by_year/2010/seb_cuadernillo.pdf)>. Accessed 19 October 2013.

Boutain, J.R., A.R. Brown, D.T. Webb & B.H. Toyofuku. 2010b. Simplified procedure for hand fracturing, identifying, and curating small macrocharcoal specimens. *IAWA Journal* 31:139–147.

Boutain, J.R., D.K. Visi, M.S. Allen, J. Xu, E. Small, K.E. Hummer & W.C. McClatchey. 2013. *New approach to DNA barcoding fresh plant and herbarium specimens with semiconductor sequencing: A phyloplastome model with hops (Humulus, Cannabaceae)*. Oral presentation at Fifth International Barcode of Life Conference, October 27-31th, 2013. Available at URL: <<http://www.dnabarcodes2013.org/dct/attach/Y2xiOmNsYjpwZGY6NjI2MzY=>>>. Accessed 31 October 2013.

Buck, E.J., C. Wiedow, C. Carlisle, D. Chagné & R. Beatson. 2009. Validation of new sex specific DNA markers for hops (*Humulus lupulus* L.). *Acta Horticulturae* 848:323-328.

Burger, J.C., M.A. Chapman & J.M. Burke. 2008. Molecular insights into the evolution of crop plants. *American Journal of Botany* 95:113-122.

Burgess, A.H. 1964. *Hops: Botany, Cultivation, and Utilization*. Interscience Publishers Inc., New York.

CBOL Plant Working Group, P.M. Hollingsworth, L.L. Forrest, J.L. Spouge, M. Hajibabaei, S. Ratnasingham, M. van der Bank, M.W. Chase, R.S. Cowan, D.L. Erickson, A.J. Fazekas, S.W. Graham, K.E. James, K-J. Kim, W.J. Kress, H. Schneider, J. van AlphenStahl, S.C.H. Barrett, C. van den Berg, D. Bogarin, K.S. Burgess, K.M. Cameron, M. Carine, J. Chacón, A. Clark, J.J. Clarkson, F. Conrad, D.S. Devey, C.S. Ford, T.A.J. Hedderson, M.L. Hollingsworth, B.C. Husband, L.J. Kelly, P.R. Kesanakurti, J.S. Kim, Y-D. Kim, R. Lahaye, H-L. Lee, D.G. Long, S. Madriñán, O. Maurin, I. Meusnier, S.G. Newmaster, C-W. Park, D.M. Percy, G. Petersen, J.E. Richardson, G.A. Salazar, V. Savolainen, O. Seberg, M.J. Wilkinson, D-K. Yi & D.P. Little. 2009. A DNA barcode for land plants. *Proceedings of the National Academy of Sciences* 106:12794-12797. doi: 10.1073/pnas.0905845106.

Carter, P.R., E.A. Oelke, A.R. Kaminski, C.V. Hanson, S.M. Combs, J.D. Doll, G.L. Worf & E.S. Oplinger. 1990. *Hop: Alternative field crops manual*. University of Wisconsin-Extension, Cooperative Extension, University of Minnesota: Center for Alternative Plant and Animal Products and the Minnesota Extension Service. Available at URL: <<http://www.hort.purdue.edu/NEWCROP/AFCM/hop.html>>. Accessed 10 October 2013.

Case, M.A., K.M. Flinn, J. Jancaitis, A. Alley & A. Paxton. 2007. Declining abundance of American ginseng (*Panax quinquefolius* L.) documented by herbarium specimens. *Biological Conservation* 134:22-30.

Cerenak, A., Z. Satovic, J. Jakse, Z. Luthar, K. Carovic-Stanko & B. Javornik. 2009. Identification of QTLs for alpha acid content and yield in hop (*Humulus lupulus* L.) *Euphytica* 170:141-154.

Chadwick, L.R., G.F. Pauli & N.R. Farnsworth. 2006. The pharmacognosy of *Humulus lupulus* L. (hops) with an emphasis on estrogenic properties. *Phytomedicine* 13:119-131.

Chaw, S.M., C.C. Chang, H.L. Chen & W.H. Li. 2004. Dating the monocot-dicot divergence and the origin of core eudicots using whole chloroplast genomes. *Journal of Molecular Evolution* 58:424-441.

Chen, R., Z. Zhang, W. Sun, Y. Fu, Q. Zhang, L. Wei, Y. Liang & R. Dong. 2012. System position and divergent time of based on ITS sequence *Humulus scandens*. *American Journal of Plant Sciences* 3:1470-1475.

China Plant BOL Group, D-Z. Li, L-M. Gao, H-T. Li, H. Wang, X-J. Ge, J-Q. Liu, Z-D. Chen, S-L. Zhou, S-L. Chen, J-B. Yang, C-X. Fu, C-X. Zeng, H-F. Yan, Y-J. Zhu, Y-S. Sun, S-Y. Chen, L. Zhao, K. Wang, T. Yang & G-W. Duan. 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. *Proceedings of the National Academy of Sciences* 108:19641-19646. doi: 10.1073/pnas.1104551108.

Clarke, R.C. 1993. *Marijuana Botany: An Advanced Study: The Propagation and Breeding of Distinctive Cannabis*. Ronin Publishing, Inc., Oakland.

Clarke R.C. & M.D. Merlin. September 2013. *Cannabis: Evolution and Ethnobotany*. University of California Press, Berkeley.

Collinson, M.E. 1989. The fossil history of the Moraceae, Urticaceae (including Cecropiaceae), and Cannabaceae. In Crane, P.R. & S. Blackmore, Eds. *Evolution, Systematics, and Fossil History of the Hamamelidae, Volume 2: 'Higher' Hamamelidae*. Clarendon Press, Oxford.

Crawford, J.M (Trans). 1888. *The Kalevala: The Epic Poem of Finland into English, Third Ed (1910)*. The Robert Blake Company, Cincinnati. Available at URL: <<http://www.sacred-texts.com/neu/kveng/index.htm>>. Accessed: April 2009.

Crepet, W.L., K.C. Nixon & M.A. Gandolfo. 2004. Fossil evidence and phylogeny: The age of major angiosperm clades based on mesofossil and macrofossil evidence from Cretaceous deposits. *American Journal of Botany* 91:1666-1682.

Cronn, R.C., R.L. Small, T. Haselkorn & J.F. Wendel. 2002. Rapid diversification of the cotton genus (*Gossypium*: Malvaceae) revealed by analysis of sixteen nuclear and chloroplast genes. *American Journal of Botany* 89:707-725.

Cronn, R., A. Liston, M. Parks, D.S. Gernand, R. Shen & T. Mockler. 2008. Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic Acids Research* 36:e122. doi: 10.1093/nar/gkn502.

Cronn, R., B.J. Knaus, A. Liston, P.J. Maughan, M. Parks, J.V. Syring & J. Udall. 2012. Targeted enrichment strategies for next-generation plant biology. *American Journal of Botany* 99:291-311.

Darling, A.E., B. Mau & N.T. Perna. 2010. progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss, and Rearrangement. *PLoS One* 5(6):e11147. doi:10.1371/journal.pone.0011147.

- Darling, A.C.E., B. Mau, F.R. Blatter & N.T. Perna. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research* 14:1394-1403.
- Davis, E.L. 1957. Morphological complexes in hops (*Humulus lupulus* L.) with special reference to the American race. *Annals of the Missouri Botanical Garden* 44:271-294.
- De Candolle, A.P. 1883. *Origine des Plantes Cultivées*. Librairie Germer Bailliere et Cie, Paris.
- De Keukeleire & A. Heyerick. 2005. Prenylfavonoids account for intriguing biological activities of hops. *Acta Horticulturae* 668:175-190.
- DeLyser, D.Y. & W.J. Kasper. 1994. Hopped beer: The case for cultivation. *Economic Botany* 48:166-170.
- Dewey, F.E., S.Pan, M.T. Wheeler, S.R. Quake & E.A. Ashley. 2012. DNA sequencing: Clinical applications of new DNA sequencing technologies. *Circulation* 125:931-944.
- Doležel, J., J. Bartoš, H. Voglmayr & J. Greilhuber. 2003. Nuclear DNA content and genome size of trout and human. *Cytometry Part A* 51A:127-128.
- Donoghue, M.J. 2008. A phylogenetic perspective on the distribution of plant diversity. *Proceedings of the National Academy of Sciences* 105(Suppl. 1):11549-11555.
- Dorofeev, P.I. 1963. *Tretichnye Flory Zapadnoi Sibiri (Tertiary Floras of Western Siberia)*. Izdat. Akad. Nauk, Moscow/Leningrad.
- Doyle, J.J. 1991. DNA protocols for plants. In Hewitt, G., A.W.B. Johnson & J.P.W. Young, Eds. *Molecular Techniques in Taxonomy, NATO ASI Series H, Cell Biology, Volume 57*. SpringerVerlag, Berlin.
- Doyle, J.J. & E.E. Dickson. 1987. Preservation of plant samples for DNA restriction endonuclease analysis. *Taxon* 36:715-722.
- Doyle, J.J. & J.L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19:11-15.
- Doyle, J.J. & J.L. Doyle. 1990. A rapid total DNA preparation procedure for fresh plant tissue. *Focus* 12:13-15.
- Egan, A.N., J. Schlueter & D.M. Spooner. 2012. Applications of next-generation sequencing in plant biology. *American Journal of Botany* 99:175-185.

- Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32:1792-1797.
- Elvin, M. 2004. *The Retreat of the Elephants: an environmental history of China*. Yale University Press, New Haven, Connecticut.
- Emboden, W.A. 1974. *Cannabis* - a polytypic genus. *Economic Botany* 28:304-310.
- Flannery, M.C. 2011. Moonstruck by herbaria. *The American Biology Teacher* 73:291-294.
- Fleet, C.M. M.F.N. Rosser, R.A. Zufall, M.C. Pratt, T.S. Feldman & P.P. Lemons. 2006. Hiring criteria in biology departments of academic institutions. *BioScience* 56:430-436.
- Fleming, M.P. & R.C. Clarke. 1998. Physical evidence for the antiquity of *Cannabis sativa* L. (Cannabaceae). *Journal of the International Hemp Association* 5:80-92.
- Flythe, M.D. 2009. The antimicrobial effects of hops (*Humulus lupulus* L.) on ruminal hyper ammonia-producing bacteria. *Letters in Applied Microbiology* 48:712-717.
- Foster, B.C., N. Kearns, J.T. Arnason, A. Saleem, C. Ogradowczyk & S. Desjardins. 2009. Comparative study of hop-containing products on human cytochrome P450-mediated metabolism. *Journal of Agricultural and Food Chemistry* 57:5100-5105.
- Fu, J.Q. 1992. *Humulus lupulus* var. *fengxianensis*. *Acta Botanica Boreali-Occidentalia Sinica* 12(6):70, 72, f.
- Fulton, T.L. 2012. Setting Up An Ancient DNA Laboratory. In Shapiro, B. & M. Hofreiter, Eds. *Ancient DNA: Methods and Protocols, Methods in Molecular Biology, Volume 840*. Springer Protocols, Springer Science+Business Media, Humana Press, New York.
- Funk, V.A., R.J. Bayer, S. Keeley, R. Chan, L. Watson, B. Gemeinholzer, E. Schilling, J.L. Panero, B.G. Baldwin, N. Garcia-Jagas, A. Susanna & R.K. Jansen. 2005. Everywhere but Antarctica: using a supertree to understand the diversity and distribution of the Compositae. *Biologische Skrifter* 55: 343-374.
- Gaertner, J. 1788. *De Fructibus et Seminibus Plantarum: accedunt seminum centuriae quinque priores cum tabulis Aeneis LXXIX*. Volume 1. Typis Academiae Carolinae, Stuttgart, Baden-Württemberg.
- Gigliano, G.S. 1998. Identification of *Cannabis sativa* L. (Cannabaceae) using restriction profiles of the Internal Transcribed Spacer II (ITS2). *Science & Justice* 38:225-230.

- Gigliano, G.S., P. Caputo & S. Cozzolino. 1997. Ribosomal DNA analysis as a tool for the identification of *Cannabis Sativa* L. specimens of forensic interest. *Science & Justice* 37:171-174.
- Gilibert, J.E. 1792. *Exercitia Phytologica, Volume 2*. Delamollière, Lyon.
- Gilibert, J.E. 1798. *Histoire des Plantes d'Europe, Volume 1*. Leroy, Lyon.
- Gledhill, D. 2002. *The Names of Plants, Third Edition*. Cambridge University Press, Cambridge.
- Glenn, T.C. 2011. Field guide to next-generation DNA sequencers. *Molecular Ecology Resources* 11:759-769.
- Grabowska-Joachimciak, A., E. Sliwinska, M. Pigula, U. Skomra & A.J. Joachimciak. 2006. Genome size in *Humulus lupulus* L. and *H. japonicas* Seibold & Zucc. (Cannabaceae). *Acta Societatis Botanicorum Poloniae* 75:207-214.
- Gray, A. 1867. *Manual of the Botany of the Northern United States*. Ivison, Phinney, Blakeman, & Co., New York.
- Grudzinskaja, I.A. 1988. K sistematike semeistva Cannabaceae. (On the taxonomy of Cannabaceae.) *Botaničeskij Žurnal* (Moscow & Leningrad) 73:589-593.
- Grumbine, R.E. 2011. *Where the Dragon Meets the Angry River: Nature and Power in the People's Republic of China*. Island Press, Washington, DC.
- Haglund, R. 2013. *Wave of craft beer washes over Michigan*. Available at URL: <<http://bridgemi.com/2013/06/wave-of-craft-beer-washes-over-michigan/>>. Accessed 27 June 2013.
- Haiminen, N., D.N. Kuhn, L. Parida & I. Rigoutsos. 2011. Evaluation of methods for de novo genome assembly from high-throughput sequencing reads reveals dependencies that affect the quality of the results. *PLoS ONE* 6: e24182. doi:10.1371/journal.pone.0024182.
- Hajibabaei, M., G.A.C. Singer, P.D.N. Herbert & D.A. Hickey. 2007. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *Trends in Genetics* 23:167-172.
- Hall, B.G. 2011. *Phylogenetic Trees Made Easy: A how to manual*, Fourth Edition. Sinauer Associates, Inc., Sunderland, Massachusetts.
- Hammond, C.T. & P.G. Mahlberg. 1973. Morphology of glandular hairs of *Cannabis sativa* from scanning electron microscopy. *American Journal of Botany* 60:524-528.

- Hampton, R., G. Nickerson, P. Whitney & A. Haunold. 2002. Comparative chemical attributes of native North American hop, *Humulus lupulus* var. *lupuloides* E. Small. *Phytochemistry* 61:855-862.
- Hampton, R., G. Nickerson, P. Whitney & A. Haunold. 2003. Erratum to “Comparative chemical attributes of native North American hop, *Humulus lupulus* var. *lupuloides* E. Small” [*Phytochemistry* 61 (2002) 855-862]. *Phytochemistry* 63:737-738.
- Hampton, R., E. Small & A. Haunold. 2001. Habitat and variability of *Humulus lupulus* var. *lupuloides* in upper midwestern North America: A critical source of American hop germplasm. *Journal of the Torrey Botanical Society* 128:35-46.
- Haney, A. & B.K. Kutscheid. 1975. An ecological study of naturalized hemp (*Cannabis sativa* L.) in east-central Illinois. *The American Midland Naturalist* 93:1-24.
- Haznedaroglu, B.Z., D. Reeves, H. Rismani-Yazdi & J. Peccia. 2012. Optimization of de novo transcriptome assembly from high-throughput short read sequencing data improves functional annotation for non-model organisms. *BMC Bioinformatics* 13:170. doi:10.1186/1471-2105-13-170.
- Heiser, Jr., C.B. 1990. *Seed to Civilization: The story of food*. Harvard University Press, Cambridge, Massachusetts.
- Herrmann, B. & S. Hummel, Eds. 1994. *Ancient DNA: Recovery and analysis of genetic material from paleontological, archaeological, museum, medical, and forensic specimens*. Springer-Verlag, New York.
- Hillig, K.W. 2005. *A systematic investigation of Cannabis*. Ph.D. Dissertation, Indiana University, Bloomington, Indiana.
- Hillig, K.W. & P.G. Mahlberg. 2004. A chemotaxonomic analysis of cannabinoid variation in *Cannabis* (Cannabaceae). *American Journal of Botany* 91:966-975.
- Holder, M. & P.O. Lewis. 2003. Phylogeny estimation: traditional and Bayesian approaches. *Nature Reviews Genetics* 4:275-284. doi:10.1038/nrg1044.
- Hollingsworth, P.M. 2011. Refining the DNA barcode for land plants. *Proceedings of the National Academy of Sciences* 108(49):19451-19452. doi: 10.1073/pnas.1116812108.
- Hood, L.V.S, M.E. Dames & G.T. Barry. 1973. Headspace volatile of marijuana. *Nature* 242:402-403.
- Hu, H.H. 1936. Notulae systematicae ad Florem Sinensium VII. *Bulletin of the Fan Memorial Institute of Biology* 7:211-218.

- Hummel, S. 2003. *Ancient DNA Typing: Methods, strategies and applications*. Springer-Verlag, Berlin Heidelberg, Germany.
- Hummer, K., J. Oliphant & J. Jackson, J. 2003. Expedition to collect genetic resources of wild American hops from montane islands of Arizona and New Mexico. *USDA ARS National Clonal Germplasm Repository Exploration Report*.
- Hummer, K., J. Oliphant, D. Cook & J. Jackson. 2002. Wild American hops expedition September 9-20, 2003. *USDA ARS National Clonal Germplasm Repository Exploration Report*.
- Hummer, K.E. 2003. Partnerships promoting the conservation of fruit genetic resources. *Acta Horticulturae* 623:225-259.
- Hummer, K.E. 2005. Wild *Humulus* genetic resources at the U.S. National Clonal Germplasm Repository. *Acta Horticulturae* 668:75-80.
- Jansen, R.K., C. Saski, S-B. Lee, A.K. Hansen & H. Daniell. 2011. Complete plastid genome sequences of three rosids (*Castanea*, *Prunus*, *Theobroma*): Evidence for at least two independent transfers of *rpl22* to the nucleus. *Molecular Biology and Evolution* 28:835-847.
- Jiang, H. & Z. Ding. 2008. A 20 Ma pollen record of East-Asian summer monsoon evolution from Guyuan, Ningxia, China. *Palaeogeography, Palaeoclimatology, Palaeoecology* 265:30-38.
- Jiang H.E., X. Li, Y.X. Zhao, D.K. Ferguson, F. Hueber, S. Bera, Y.F. Wang, L.C. Zhao, C.J. Liu & C.S. Li. 2006. A new insight into *Cannabis sativa* (Cannabaceae) utilization from 2500-year-old Yanghai Tombs, Xinjiang, China. *Journal of Ethnopharmacology* 108:414-422.
- Jiang, H., X. Mao, H. Xu, J. Thompson, P. Wang & X. Ma. 2011. Last glacial pollen record from Lanzhou (Northwestern China) and possible forcing mechanisms for the MIS 3 climate change in Middle to East Asia. *Quaternary Science Reviews* 30:769-781.
- Johnson, K.R. 2002. Megaflora of the Hell Creek and lower Fort Union Formations in the western Dakotas: Vegetational response to climate change, the Cretaceous-Tertiary boundary event, and rapid marine transgression. In Hartman, J.H., K.R. Johnson & D.J. Nichols, Eds. The Hell Creek Formation and the Cretaceous-Tertiary boundary in the northern Great Plains: An integrated continental record of the end of the Cretaceous. *The Geological Society of America Special Paper* 361:329-391.
- Jones, K. 1998. Robertsonian fusion and centric fission in karyotype evolution of higher plants. *The Botanical Review* 64:273-289.

- Judd, W.S., R.W. Sanders & M.J. Donoghue. 1994. Angiosperm family pairs—preliminary phylogenetic analyses. *Harvard Papers in Botany* 5:1-51.
- Karabín, M., L. Jelínek, T. Kinčl, T. Hudcová, B. Kotlíková & P. Dostálek. 2013. New approach to the production of xanthohumol-enriched beers. *The Journal of the Institute of Brewing* 119:98-102.
- Kavalier, A.R., H. Yang, E.J. Kennelly, M.C. Coles, T. Dinh, J. Koelling & P.D. Matthews. 2009. Application of newly-developed tools for high throughput analysis to assess the effects of prohexadione-Ca on the phytochemical, morphological, and targeted metabolic profile of hops (*Humulus lupulus*). *Acta Horticulturae* 848:21-32.
- Keeley, S.C., Z. H. Forsman & R. Chan. 2007. A phylogeny of the “evil tribe” (Vernonieae: Compositae) reveals Old/New World long distance dispersal: Support from separate and combined congruent datasets (*trnL*, *ndhF*, ITS). *Molecular Phylogenetics and Evolution* 44: 89-103.
- Khoury, C.K., S. Greene, J. Wiersema, N. Maxted, A. Jarvis & P.C. Struik. 2013. An inventory of crop wild relatives of the United States. *Crop Science* 53:1496-1508.
- Kim, K-J. & R.K. Jansen. 1998. A chloroplast DNA phylogeny of lilacs (*Syringa*, Oleaceae): Plastome groups show a strong correlation with crossing groups. *American Journal of Botany* 85:1338-1351.
- Kleinman, K. 2003. Revising Plants, man and life: Edgar Anderson reshapes a classic book. *Archives of Natural History* 30:85-96.
- Kress, W.J. & D.L. Erickson, Eds. 2012. *DNA Barcodes: Methods and Protocols. Methods in Molecular Biology, Volume. 858*. Springer Protocols, Springer Science+Business Media, Humana Press, NY, USA.
- Ku, C-S. & D.H. Roukos. 2013. From next-generation sequencing to nanopore sequencing technology: paving the way to personalized genomic medicine. *Expert Review of Medical Devices* 10:1-6.
- Kučera, J. & K. Krofta. 2009. Mathematical model for prediction of alpha acid contents from meteorological data for ‘Saaz’ aroma variety. *Acta Horticulturae* 848:131-140.
- Kunin, V., A. Copeland, A. Lapidus, K. Mavromatis & P. Hugenholtz. 2008. A bioinformatician's guide to metagenomics. *Microbiology and Molecular Biology Reviews* 72:557. doi: 10.1128/MMBR.00009-08.
- Kunming Botanical Garden. Kunming Institute of Botany, Chinese Academy of Sciences. 2006 *An Enumeration of Plants Growing in Kunming Botanical Garden*. Yunnan Publishing Group Corporation, Yunnan Publishing, Kunming.

- Kou, X., D.K. Ferguson, J. Xu, Y. Wang & C. Li. 2006. The reconstruction of paleovegetation and paleoclimate in the Late Pliocene of West Yunnan, China. *Climatic Change* 77:431–448.
- Lamarck, J-B. 1778. *Flore Française, Volume 2*. Agasse, Paris.
- Lamarck, J-B. 1785. *Encyclopédie Méthodique: Botanique, Volume 1*. Panckoucke, Paris.
- Lamy, V., S. Roussi, M. Chaabi, F. Gosse, A. Lobstein & F. Raul. 2008. Lupulone, a hop bitter acid, activates different death pathways involving apoptotic TRAIL-receptors, in human colon tumor cells and in their derived metastatic cells. *Apoptosis* 13:1232–1242.
- Lander E.S. & M.S. Waterman 1988. Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* 2:231-239.
- Lavoie, C. 2013. Biological collections in an ever changing world: Herbaria as tools for biogeographical and environmental studies. *Perspectives in Plant Ecology, Evolution and Systematics* 15:69-76.
- Lederberg, J. & A.T. McCray. 2001. 'Ome sweet 'omics-- a genealogical treasury of words. *The Scientist* 15, 8 (April 2, 2001).
- Lemey, P., M. Salemi & A-M. Vandamme, Eds. 2009. *The Phylogenetic Handbook: A practical approach to phylogenetic analysis and hypothesis testing, Second Edition*. Cambridge University Press, New York.
- Linnaeus, C. 1753. *Species Plantarum, Volume 2*. Impensis Laurentii Salvii, Stockholm.
- Li, S. & X. Luo. 2003. *Compendium of Materia Medica : bencao gangmu*. Foreign Languages Press, Beijing, China.
- Li, W., Q. Song, R.H. Brlansky & J.S. Hartung. 2007. Genetic diversity of citrus bacterial canker pathogens preserved in herbarium specimens. *Proceedings of the National Academy of Sciences* 104:18427-18432. Available at URL: <[www.pnas.org/cgi/doi10.1073/pnas.0705590104](http://www.pnas.org/cgi/doi/10.1073/pnas.0705590104)>.
- Liu, J., M. Möller, L-M. Gao, D-Q. Zhang & D-Z. Li. 2011. DNA barcoding for the discrimination of Eurasian yews (*Taxus* L., Taxaceae) and the discovery of cryptic species. *Molecular Ecology Resources* 11:89-110.
- Liu, J., M. Möller, J. Provan, L-M. Gao, R.C. Poudel & D-Z. Li. 2013. Geological and ecological factors drive cryptic speciation of yews in a biodiversity hotspot. *New Phytologist* 199:1093-1108.

- Liu, C., L. Shi, Y. Zhu, H. Chen, J. Zhang, X. Lin and X. Guan. 2012a. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genomics* 13:715. doi:10.1186/1471-2164-13-715.
- Liu, L., Y. Li, S. Li, N. Hu, Y. He, R. Pong, D. Lin, L. Lu & M. Law. 2012b. Comparison of next-generation sequencing systems. *Journal of Biomedicine and Biotechnology*. Article ID 251364. doi:10.1155/2012/251364.
- Loman, N.J., R.V. Misra, T.J. Dallman, C. Constantinidos, S.E. Gharbia, J. Wain & M.J. Pallen. 2012. Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology* 30:434-439.
- Lou, X. 2005. *Chinese hop growing*. Hop Growing Association of Gansu Province. Proceedings of the IHGC Technical Commission, Lanzhou-China. Available at URL: <<http://www.czhops.cz/tc/pdf/chinesehop.pdf>>.
- Loureiro, J. de. 1790. *Flora Cochinchinensis, Volume 2*. Ulyssipone, Lisbon.
- Löve, Á. & D. Löve. 1982. *Humulus lupulus* L. spp. *americanus*. In Löve, Á., Ed. 1982. IOPB chromosome number reports LXXIV. *Taxon* 31:119-128.
- Ma, J. & S. Clemants. 2006. A history and overview of the Flora Reipublicae Popularis Sinicae (FRPS, Flora of China, Chinese edition, 1959-2004). *Taxon* 55:451-460.
- MacGinitie, H.D. 1953. Fossil Plants of the Florissant Beds of Colorado. *Carnegie Institute of Washington, D.C., Contributions to Paleontology Publication* 599:1-198.
- MacGinitie, H.D. 1969. The Eocene Green River Flora of northwestern Colorado and northeastern Utah. *University of California Publications in Geological Sciences* 83:1-203.
- Magalhães, P.J., P. Dostalek, J.M. Cruz, L.F. Guido & A.A. Barros. 2008. The impact of a xanthohumol-enriched hop product on the behavior of xanthohumol and isoxanthohumol in pale and dark beers: a pilot scale approach. *Journal of the Institute of Brewing* 114:246-256.
- Mai, D.H. & H. Walther. 1988. Die pliozaenen Floren von Thueringen, Deutsche Demokratische Republik. *Quartaerpalaeontologie* 7:55-297.
- Manchester, S.R. 2001. Update on the megafossil flora of Florissant, Colorado. *Denver Museum of Nature & Science* 4:137-161.
- Marafioti, R.L. 1970. The meaning of generic names of important economic plants. *Economic Botany* 24:189-207.

- Marks, R.B. 2012. *China: Its environment and history*. Rowman & Littlefield Publishers, Inc., Lanham.
- Martinov, I.I. 1820. Cannabaceae, 99. *Tekhno-Botanicheskii Slovar*. Rossilska Akademia, St. Petersburg.
- Matvienko, M., A. Kozik, L. Froenicke, D. Lavelle, B. Martineau, B. Perroud & R. Micheltmore. 2013. Consequences of normalizing transcriptomic and genomic libraries of plant genomes using a duplex-specific nuclease and tetramethylammonium chloride. *PLoS ONE* 8:e55913. doi:10.1371/journal.pone.0055913.
- Maximowicz, C.J. 1879. Humulus. In Franchet, A. & L. Satavier, Eds. *Enumerato Plantarum in Japonia, Volume 2*. Savy, Paris.
- McAdam, E.L. 2013. *Molecular and quantitative genetic analyses of hop (Humulus lupulus L.)*. Ph.D. Dissertation, University of Tasmania, Australia.
- McClatchey, W. 1998. A new species of *Metroxylon* (Arecaceae) from Western Samoa. *Novon* 8:252-258.
- Merlin, M.D. 1972. *Man and Marijuana*. Fairleigh Dickinson University Press, Rutherford-Madison-Teaneck, New Jersey.
- Merlin, M.D. 2003. Archaeological evidence for the tradition of psychoactive plant use in the Old World. *Economic Botany* 57:295-323.
- Merrill, E.D. 1935. A commentary on Loureiro's Flora Cochinchinensis. *Transactions of the American Philosophical Society, New Series* 24:1-445.
- Miller, P. 1768. *Gardeners Dictionary, Edition 8*. London.
- Miquel, F.A. 1866. Prolusio Florae Japonicae. *Annales Musei botanici lugduno-batavi* 2:69-212.
- Moerman, D.E. 1998. *Native American Ethnobotany*. Timber Press, Portland.
- Moerman, D.E. 2009. *Native American Medicinal Plants: An Ethnobotanical Dictionary*. Timber Press, Portland.
- Moerman, D.E. 2010. *Native American Food Plants: An Ethnobotanical Dictionary*. Timber Press, Portland.
- Moldenke, H.N. 1935. Nomenclatural and taxonomic notes. *Phytologia* 1:167-171.
- Morrell P.L. & M.T. Clegg. 2007. Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. *Proceedings of the National Academy of Sciences* 104:3289-3294.

- Mozny, M., R. Tolasz, J. Nekovar, T. Sparks, M. Trnka & Z. Zalud. 2009. The impact of climate change on the yield and quality of Saaz hops in the Czech Republic. *Agricultural and Forest Meteorology* 149:913-919.
- Mukherjee, A., S. Chandra Roy, S. De Bera, H.-E. Jiang, X. Li, C.-S. Li & S. Bera. 2008. Results of molecular analysis of an archaeological hemp (*Cannabis sativa* L.) DNA sample from North West China. *Genetic Resources and Crop Evolution* 55:481-485.
- Murakami, A. 2000. Comparison of sequence of *rbcL* and non-coding regions of chloroplast DNA and ITS2 region of rDNA in genus *Humulus*. *Breeding Science* 50:155-160.
- Murakami, A. 2001. Structural differences in the intergenic spacer of 18S-26S rDNA and molecular phylogeny using partial external transcribed spacer sequence in hop, *Humulus lupulus*. *Breeding Science* 51:163-170.
- Murakami, A., P. Darby, B. Javornik, M.S.S. Pais, E. Seigner, A. Lutz & P. Svoboda. 2003. Molecular evolution of hops, *Humulus lupulus*. In: E. Seigner, Ed. *Proceedings of the Scientific Commission of the International Hop Growers Convention*. Wolnzach, Germany.
- Murakami, A., P. Darby, B. Javornik, M.S.S. Pais, E. Seigner, A. Lutz & P. Svoboda. 2006a. Molecular phylogeny of wild hops, *Humulus lupulus* L. *Heredity* 97:66-74.
- Murakami, A., P. Darby, B. Javornik, M. S. S. Pais, E. Seigner, A. Lutz & P. Svoboda. 2006b. Microsatellite DNA analysis of wild hops, *Humulus lupulus* L. *Genetic Resources and Crop Evolution* 53:1553-1562.
- Neafsey, D.E. & B.J. Haas. 2011. 'Next-generation' sequencing becomes 'now-generation.' *Genome Biology* 2011, 12:303. Available at URL: <<http://genomebiology.com/2011/12/3/303>>.
- Nelson, A. & T.D.A. Cockerell. 1903. Three new plants from New Mexico. *Proceedings of the Biological Society of Washington* 16:45-46.
- Neve, R.A. 1991. *Hops*. Chapman and Hall, London, UK.
- Ni, X., D.L. Gebo, M. Dagosto, J. Meng, P. Tafforeau, J.J. Flynn & K.C. Beard. 2013. The oldest known primate skeleton and early haplorhine evolution. *Nature* 498:60-64.
- Niedringhaus, T.P., D. Milanova, M.B. Kerby, M.P. Snyder & A.E. Barron. 2011. Landscape of next-generation sequencing technologies. *Analytical Chemistry* 83:4327-434.
- Nuttall, T. 1848a. Descriptions of plants collected by Mr. William Gambel in the Rocky Mountains and Upper California. *Proceeding of the Academy of Natural Sciences of Philadelphia* 4:7-26.

- Nuttall, T. 1848b. Descriptions of plants collected by William Gambel, M.D., in the Rocky Mountains and Upper California. *Journal of the Academy of Natural Sciences of Philadelphia, Series Two* 1:149-189.
- Padial, J.M., A Miralles, I. De la Riva & M. Vences. 2010. The integrative future of taxonomy. *Frontiers in Zoology* 7:16. doi:10.1186/1742-9994-7-16. Available at URL: <<http://www.frontiersinzoology.com/content/7/1/16>>.
- Paleobiology Database. 2013. The data on 08 July 2013 using the quick search for *Humulus*. Available at URLs: <<http://paleodb.org/>> and <<http://fossilworks.org>>.
- Parker K.C. & S. Jørgensen. 2003. Examples of the use of molecular markers in biogeographic research. *Physical Geography* 24:378-398.
- Parks, M., R. Cronn & A. Liston. 2009. Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biology* 7:84. Available at URL: <<http://www.biomedcentral.com/1741-7007/7/84>>.
- Patzak, J., V. Nesvadba, A. Henychová & K. Krofta. 2010a. Assessment of the genetic diversity of wild hops (*Humulus lupulus* L.) in Europe using chemical and molecular analyses. *Biochemical Systematics and Ecology* 38:136-145.
- Patzak, J., V. Nesvadba, K. Krofta, A. Henychova, A.I. Marzoev & K. Richards. 2010b. Evaluation of genetic variability of wild hops (*Humulus lupulus* L.) in Canada and the Caucasus region by chemical and molecular methods. *Genome* 53:545-557.
- Pavlovič, M., X. Luo, I.J.Košir, M. Virant & F. Gu. 2006. Expansion of the Chinese hop and brewing industry. *Hmeljarski bilten –Hop Bulletin* 13:61-69.
- Peredo, E.L., M. Ángeles Revilla, B.M. Reed, B. Javornik, E. Cires, J. Antonio Fernández Prieto & R. Arroyo-García. 2010. The influence of European and American wild germplasm in hop (*Humulus lupulus* L.) cultivars. *Genetic Resources and Crop Evolution* 57:575-586.
- Perelman, P., W.E. Johnson, C. Roos, H.N. Seuánez, J.E. Horvath, M.A.M. Moreira, B. Kessing, J. Pontius, M. Roelke, Y. Rumpler, M.P.C. Schneider, A. Silva, S.J.O'Brien & J. Pecon-Slattery. 2011. A molecular phylogeny of living primates. *PLoS Genetics* 7(3): e1001342. doi:10.1371/journal.pgen.1001342.
- Pillay, M. & S.T. Kenny. 1994. Chloroplast DNA differences between cultivated hop, *Humulus lupulus* and the related species *H. japonicus*. *Theoretical and Applied Genetics* 89:372-378.
- Pillay, M. & S.T. Kenny. 1996. Structure and inheritance of ribosomal DNA variants in cultivated and wild hop, *Humulus lupulus* L. *Theoretical and Applied Genetics* 93:333-340.

- Pillay, M. & S.T. Kenny. 2006. Structural organization of the nuclear ribosomal RNA genes in *Cannabis* and *Humulus* (Cannabaceae). *Plant Systematics and Evolution* 258:97-105.
- Pollan, M. 2001. *The Botany of Desire: A Plant's-eye View of the World*. Random House, Inc., New York.
- Prather, L.A. O. Alvarez-Fuentes, M.H. Mayfield & C.J. Ferguson. 2004. The decline of plant collecting in the United States: A threat to the infrastructure of biodiversity studies. *Systematic Botany* 29:15-28.
- Prescott-Allen, D. & R. Prescott-Allen. 1986. *The First Resource: Wild species in the North American economy*. Yale University Press, New Haven.
- Pyke, G.H. & P.R. Ehrlich. 2010. Biological collections and ecological/environmental research: a review, some observations and a look to the future. *Biological Reviews* 85:247-266.
- Quail, M.A., M. Smith, P. Coupland, T.D. Otto, S.R. Harris, T.R. Connor, A. Bertoni, H.P. Swerdlow & Y. Gu. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 13:341. Available at URL: <<http://www.biomedcentral.com/1471-2164/13/341>>.
- Ravi, V., J.P. Khurana, A.K. Tyagi & P. Khurana. 2006. The chloroplast genome of mulberry: complete nucleotide sequence, gene organization and comparative analysis. *Tree Genetics & Genomes* 3:49-59.
- Reeves, P.A. & C.M. Richards. 2011. Species delimitation under the general lineage concept: An empirical example using wild North American hops (Cannabaceae: *Humulus lupulus*). *Systematic Biology* 60:45-59.
- Roemer, F. 1892. The variegated hop. *American Florist* 8:489.
- Rogers, S.O. & A.J. Bendich. 1985. Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Plant Molecular Biology* 5:69-76.
- Rogers, S.O. & A.J. Bendich. 1994. Extraction of total cellular DNA from plants, algae and fungi. *Plant Molecular Biology Manual* D1:1-8.
- Rösch, M. 2008. New aspects of agriculture and diet of the early medieval period in central Europe: waterlogged plant material from sites in south-western Germany. *Vegetation History and Archaeobotany* 17(Suppl 1):S225-S238.
- Rozas, J. 2009. DNA Sequence Polymorphism Analysis using DnaSP. In Posada, D., Ed. *Bioinformatics for DNA Sequence Analysis, Methods in Molecular Biology, Volume 537*. Springer Protocols, Springer Science+Business Media, Humana Press, New York.

- Russo, E.B., H-E. Jiang, X. Li, A. Sutton, A. Carboni, F. del Bianco, G. Mandolino, D.J. Potter, Y-X. Zhao, S. Bera, Y-B. Zhang, E-G. Lü, D.K. Ferguson, F. Hueber, L-C. Zhao, C-J. Liu, Y-F. Wang & C-S. Li. 2008. Phytochemical and genetic analyses of ancient cannabis from Central Asia. *Journal of Experimental Botany* 59:4171-4182. doi:10.1093/jxb/ern260.
- Rydberg, P.A. 1917. *Flora of the Rocky Mountains and Adjacent Plains*. Rydberg, New York.
- Salisbury, R.A. 1796. *Prodromus Stirpium in Horto ad Chapel Allerton Vigentium*. London.
- Salmon, E.S. 1934. Two new hops: Brewer's Favourite and Brewer's Gold. *Journal Southeast Agricultural College, Wye, Kent, England* 34:93-105.
- Sauer, C.O. 1952. *Agricultural Origins and Dispersals*. American Geographical Society, New York.
- Sattarian, A. 2006. *Contribution to the biosystematics of Celtis L. (Celtidaceae) with special emphasis on the African species*. Ph.D. Dissertation, Wageningen University, Wageningen, Netherlands.
- Särkinen, T., M. Staats, J.E. Richardson, R.S. Cowan & F.T. Bakker. 2012. How to open the treasure chest? Optimising DNA extraction from herbarium specimens. *PLoS ONE* 7: e43808. doi:10.1371/journal.pone.0043808.
- Schmidt-Lebuhn, A.N., N.J. Knerr & M. Kessler. 2013. Non-geographic collecting biases in herbarium specimens of Australian daises (Asteraceae). *Biodiversity and Conservation* 22:905-919.
- Schultes, R.E., W.M. Klein, T. Plowman & T.E. Lockwood. 1974. *Cannabis*: an example of taxonomic neglect. *Harvard University Botanical Museum Leaflets* 23:337-367.
- Scopoli, G.A. 1772. *Flora Carniolica, Second Edition*. Krauss, Vienna.
- Scotese, C.R. 2002. PALEOMAP website. Available at URL: <<http://www.scotese.com>>. Accessed December 2012.
- Seigner, E., A. Lutz, K. Oberhollenzer, R. Seidenberger, S. Seefelder and F. Felsenstein. 2009. Breeding of hop varieties for the future. *Acta Horticulturae* 848:49-58.
- Shapiro, B. & M. Hofreiter, Eds. 2012. *Ancient DNA: Methods and Protocols. Methods in Molecular Biology, Volume. 840*. Springer Protocols, Springer Science+Business Media, Humana Press, NY, USA.

- Shaw, J., E.B. Lickey, J.T. Beck, S.B. Farmer, W. Liu, J. Miller, K.C. Siripun, C.T. Winder, E.E. Schilling & R.L. Small. 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* 92:142-166.
- Shaw, J., E.B. Lickey, E.E. Schilling & R.L. Small. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany* 94:275-288.
- Shen, C. & J.N. Sofos. 2008. Antilisterial activity of hops beta acids in broth with or without other antimicrobials. *Journal of Food Science* 73:M438-M442.
- Shokralla, S., J.L. Spall, J.F. Gibson & M. Hajibabaei. 2012. Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology* 21:1794-1805.
- Siebold, P.F.von & J.G. Zuccarini. 1846. *Florae Japonicae*. Akademie der Wissenschaften, München.
- Sirrione, R.J., N. Rothwell, E. Lizotte, R. Goldy, S. Marquie & D.E. Brown-Rytlewski. 2010. Sustainable Hop Production in the Great Lakes Region. *Extension Bulletin* E-3083:1-12. Available at URL: <<http://www.uvm.edu/extension/cropsoil/wp-content/uploads/Sirrione-Sustainable-Hop-Production-in-the-Great-Lakes-Region.pdf>>. Accessed 18 October 2013.
- Siushih, C. 1998. Cannaboideae Engl. In Siushih, C. & C. Wu, Eds. *Flora Reipublicae Popularis Sinicae, Volume 23*. Science Press, Beijing.
- Small, E. 1972. Interfertility and chromosomal uniformity in *Cannabis*. *Canadian Journal of Botany* 50:1947-1949.
- Small, E. 1978. A numerical and nomenclatural analysis of mopho-geographic taxa of *Humulus*. *Systematic Botany* 3:37-76.
- Small, E. 1980. The relationships of hop cultivars and wild variants of *Humulus lupulus*. *Canadian Journal of Botany* 58:676-686.
- Small, E. 1981. A numerical analysis of mopho-geographical groups of cultivars of *Humulus lupulus* based on samples of cones. *Canadian Journal of Botany* 59:311-324.
- Small, E. & A. Cronquist. 1976. A practical and natural taxonomy for *Cannabis*. *Taxon* 25:405-435.
- Small, E., H.D. Beckstead & A. Chan. 1975. The evolution of cannabinoid phenotypes in *Cannabis*. *Economic Botany* 29:219-232.
- Small, E., P.Y. Jui & L.P. Lefkovitch. 1976. A numerical taxonomic analysis of *Cannabis* with special reference to species delimitation. *Systematic Botany* 1:67-84.

- Smith, J.M., J.M. Oliphant & K.E. Hummer. 2006. Plant exploration for native hop in the American southwest. *Plant Genetic Resources Newsletter* 147:1-9.
- Song, B.-H., X.-Q. Wang, F.-Z. Li & D.-Y. Hong. 2001. Further evidence for paraphyly of the Celtidaceae from the chloroplast gene *matK*. *Plant Systematics and Evolution* 228:107-115.
- Srećec, S., V. Zechner-Krpan, V. Petravić-Tominac, I.J. Košir & A. Čerenak. 2012. Importance of medical effects of xanthohumol, hop (*Humulus lupulus* L.) bioflavonoid in restructuring of world hop industry. *Agriculturae Conspectus Scientificus* 77:61-67.
- Srinivasan, V., D. Goldberg & G.J. Hass. 2004. Contributions of the antimicrobial spectrum of hop constituents. *Economic Botany* 58(Supplement):S230–S238.
- Stajner, N., Z. Satovic, A. Čerenak & B. Javornik. 2008. Genetic structure and differentiation in hop (*Humulus lupulus* L.) as inferred from microsatellites. *Euphytica* 161:301-311.
- Steele, P.R. & J.C. Pires. 2011. Biodiversity assessment: State-of-the-art techniques in phylogenomics and species identification. *American Journal of Botany* 98:415-425.
- Steele, P.R., K.L. Hertweck, D. Mayfield, M.R. McKain, J. Leebens-Mack & J.C. Pires. 2012. Quality and quantity of data recovered from massively parallel sequencing: Examples in Asparagales and Poaceae. *American Journal of Botany* 99:330-348.
- Steiger, D.L., C. Nagai, P.H. Moore, C.W. Morden, R.V. Osgood & R. Ming. 2002. AFLP analysis of genetic diversity within and among *Coffea arabica* cultivars. *Theoretical and Applied Genetics* 105:209-215.
- Stephens, M. & P. Donnelly. 2003. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics* 73:1162-1169.
- Stephens, M., N. Smith & P. Donnelly. 2001. A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics* 68:978-989.
- Stevens, J.F. & J.E. Page. 2004. Xanthohumol and related prenylflavonoids from hops and beer: to your good health! *Phytochemistry* 65:1317-1330.
- Straub, S.C.K., M. Fishbein, T. Livshultz, Z. Foster, M. Parks, K. Weitemier, R.C. Cronn & A. Liston. 2011. Building a model: Developing genomic resources for common milkweed (*Asclepias syriaca*) with low coverage genome sequencing. *BMC Genomics* 12:211. doi:10.1186/1471-2164-12-211.

- Straub, S.C.K., M. Parks, K. Weitemier, M. Fishbein, R.C. Cronn & A. Liston. 2012. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. *American Journal of Botany* 99:349-364.
- Stull, G.W., M.J. Moore, V.S. Mandala, N.A. Douglas, H-R. Kates, X. Qi, S.F. Brockington, P.S. Soltis, D.E. Soltis & M.A. Gitzendanner. 2013. A targeted enrichment strategy for massively parallel sequencing of Angiosperm plastid genomes. *Applications in Plant Sciences* 1:1200497. doi: <http://dx.doi.org/10.3732/apps.1200497>.
- Sundberg, M.D., P. DeAngelis, K. Havens, K. Holsinger, K. Kennedy, A.T. Kramer, R. Muir, P. Olwell, K. Schierenbeck, L. Stritch & B. Zorn-Arnold. 2011. Perceptions of strengths and deficiencies: Disconnects between graduate students and prospective employers. *BioScience* 61:133-138.
- Suzuki, K., S. Asano, K. Iijima & K. Kitamoto. 2008. Sake and beer spoilage lactic acid bacteria – a review. *Journal of the Institute of Brewing* 114:209-223.
- Sytsma, K.J., J. Morawetz, C. Pires, M. Nepokroeff, E. Conti, M. Zjhra, J.C. Hall & M.W. Chase. 2002. Urticalean Rosids: Circumscription, Rosid ancestry, and phylogenetics based on *rbcL*, *trnL-F*, and *ndhF* sequences. *American Journal of Botany* 89:1531-1546.
- Takoi, K., M. Degueil, S. Shinkaruk, C. Thibon, K. Maeda, K. Ito, B. Bennetau, D. Dubourdieu & T. Tominaga. 2009. Identification and characteristics of new volatile thiols derived from the hop (*Humulus lupulus* L.) cultivar Nelson Sauvin. *Journal of Agricultural and Food Chemistry* 57:2493-2502.
- Tamura, K., D Peterson, N. Peterson, G. Stecher, M. Nei & S. Kumar. 2011. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Method. *Molecular Biology and Evolution* 28: 2731-2739.
- Tautz, D., P. Arctander, A. Minelli, R.H. Thomas & A.P. Vogler. 2003. A plea for DNA taxonomy. *TRENDS in Ecology and Evolution* 18:70-74.
- The Plant List*. 2013. Version 1.1. Available at URL: <<http://www.theplantlist.org/>>. Accessed March 2014.
- Thomas, G.G. & W.W. Schwabe. 1969. Factors controlling flowering in the hop (*Humulus lupulus* L.) *Annals of Botany* 33:781-93.
- Tiffney, B.H. 1986. Fruit and seed dispersal and the evolution of the Hamamelidae. *Annals of the Missouri Botanical Garden* 73:394-416.

- Timme, R.E., J.V. Kuehl, J.L. Boore & R.K. Jansen. 2007. A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: Identification of divergent regions and categorization of shared repeats. *American Journal of Botany* 94:302-312.
- Tomlan, M.A. 1992. *Tinged with Gold: Hop culture in the United States*. University of Georgia Press, Athens.
- Townsend, M.S. & J.A. Henning. 2009. AFLP discrimination of native North American and cultivated hop. *Crop Science* 49:600-607.
- Turner, S.F., C.A. Benedict, H. Darby, L.A. Hoagland, P. Simonson, J.R. Sirrine & K.M. Murphy. 2011. Challenges and opportunities for organic hop production in the United States. *Agronomy Journal* 103:1645-1654.
- Tropicos.org. 2013. Missouri Botanical Garden. Available at URL: <<http://www.tropicos.org/>>. Accessed 18 October 2013.
- Van Cleemput, M., K. Cattoor, K. De Bosscher, G. Haegeman, D. De Keukeleire & A. Heyerick. 2009. Hop (*Humulus lupulus*)-derived bitter acids as multipotent bioactive compounds. *Journal of Natural Products* 72:1220-1230.
- van Bakel, H., J.M. Stout, A.G. Cote, C.M. Tallon, A.G. Sharpe, T.R. Hughes & J.E. Page. 2011. The draft genome and transcriptome of *Cannabis sativa*. *Genome Biology* 12:R102. doi:10.1186/gb-2011-12-10-r102.
- Vavilov, N.I. 1992. *Origin and Geography of Cultivated Plants*. English Edition, Doris Löve (Trans.). Cambridge University Press, Cambridge.
- Wandeler, P., P.E.A. Hoeck & L.F. Keller. 2007. Back to the future: museum specimens in population genetics. *Trends in Ecology and Evolution* 22:634-642.
- Wang, Q., Z.H. Ding, J.K. Liu & Y.T. Zheng. 2004. Xanthohumol, a novel anti-HIV-1 agent purified from Hops *Humulus lupulus*. *Antiviral Research* 64:189-194.
- Wang, H., M.J. Moore, P.S. Soltis, C.D. Bell, S.F. Brockington, R. Alexandre, C.C. Davis, M. Latvis, S.R. Manchester & D.E. Soltis. 2009. Rosid radiation and the rapid rise of angiosperm-dominated forests. *Proceedings of the National Academy of Sciences* 106:3853-3858.
- Weber, W.A. 2003. The Middle Asian element in the southern Rocky Mountain flora of the western United States: a critical biogeographical review. *Journal of Biogeography* 30:649-685.
- Whittall, J.B., J. Syring, M. Parks, J. Buenrostro, C. Dick, A. Liston & R. Cronn. 2010. Finding a (pine) needle in a haystack: chloroplast genome sequence divergence in rare and widespread pines. *Molecular Ecology* 19(Suppl. 1):100-114.

- Whittington, G. & A.D. Gordon. 1987. The differentiation of the pollen of *Cannabis sativa* L. from that of *Humulus lupulus* L. *Pollen Spores* 29:111-120.
- Whittock, S., G. Leggett, J. Jakše, B. Javornik, J. Carling, A. Kilian, P.D. Matthews, G. Probasco, J.A. Henning, P. Darby, A. Čerenak & A. Koutoulis. 2009. Use of diversity array technology (DArT) for genotyping of *Humulus lupulus* L. *Acta Horticulturae* 848:59-64.
- Williams, B.A., R.F. Kay & E.C. Kirk. 2010. New perspectives on anthropoid origins. *Proceedings of the National Academy of Sciences* 107:4797-4804.
- Wilson, D.G. 1975. Plant remains from the Graveney Boat and the early history of *Humulus lupulus* L. in W. Europe. *New Phytologist* 75:627-648.
- Winkler, H. 1920. *Verbreitung und Ursache der Parthenogenesis im Pflanzen – und Tierreiche*. Verlag Von Gustav Fischer, Jena. Available at URL <<http://www.biodiversitylibrary.org/title/1460#177>>. Accessed 18 October 2013.
- Wolf, A., W.R.L. Anderegg, S.J. Ryan & J. Christensen. 2011. Robust detection of plant species distribution shifts under biased sampling regimes. *Ecosphere* 2:115. doi:10.1890/ES11-00162.1. Available at URL <<http://www.esajournals.org/doi/pdf/10.1890/ES11-00162.1>>.
- Wu, Z., Z. Zhou & B. Bartholomew. 2003. 1. *Humulus*. Pp. 74-75 in Flora of China. Volume 5 (Ulmaceae through Basellaceae). Wu, Z.Y., P.H. Raven & D.Y. Hong, Eds. Science Press, Beijing, China and Missouri Botanical Garden Press, St. Louis, USA. Available at URL: <<http://flora.huh.harvard.edu/china/PDF/PDF05/Humulus.PDF>>. Accessed 18 October 2013.
- Wyman, S.K., R.K. Jansen & J.L. Boore. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20:3252-3255.
- Yamaguchi, N., K. Satoh-Yamaguchi & M. Ono. 2009. In vitro evaluation of antibacterial, anticollagenase, and antioxidant activities of hop components (*Humulus lupulus*) addressing acne vulgaris. *Phytomedicine* 16:369–376.
- Yang, M-Q., R van Velzen, F.T. Bakker, A. Sattarian, D-Z. Li & T-S. Yi. 2013. Molecular phylogenetics and character evolution of Cannabaceae. *Taxon* 62:473-485.
- Yeloff, D. & B. van Geel. 2007. Abandonment of farmland and vegetation succession following the Eurasian plague pandemic of AD 1347-52. *Journal of Biogeography* 34:575-582.
- Yesson, C., S.J. Russell, T. Parrish, J.W. Dalling & N.C. Garwood. 2004. Phylogenetic framework for *Trema* (Celtidaceae). *Plant Systematics and Evolution* 248:85-109.

- Xu, J., D.K. Ferguson, C. Li & Y. Wang. 2008. Late Miocene vegetation and climate of the Lühe region in Yunnan, southwestern China. *Review of Palaeobotany and Palynology* 148:36-59.
- Xu, Q., F. Tian, M.J. Bunting, Y. Li, W. Ding, X. Cao & Z. He. 2012. Pollen source areas of lakes with inflowing rivers: modern pollen influx data from Lake Baiyangdian, China. *Quaternary Science Reviews* 37:81-91.
- Zanoli, P. & M. Zavatti. 2008. Pharmacognostic and pharmacological profile of *Humulus lupulus* L. *Journal of Ethnopharmacology* 116: 383-396.
- Zapałowicz, H. 1908. *Conspectus Florae Galiciae criticus, Volume 2*. Sumptibus Academiae Literarum Cracoviensis, Cracoviae.
- Zerega, N.C., W.L. Clement, S.L. Datwler & G.D. Weiblen. 2005. Biogeography and divergence times in the mulberry family. *Molecular Phylogenetics and Evolution* 37:402-416.
- Zhang, T., X. Zhang, S. Hu & J. Yu. 2011. An efficient procedure for plant organellar genome assembly, based on whole genome data from the 454 GS FLX sequencing platform. *Plant Methods* 7:38.