DISSERTATION

A STUDY OF PLANT DOMESTICATION AND EVOLUTION THROUGH THE TAXONOMIC REVISION OF WILD NORTH AMERICAN *HUMULUS*, A PHYTOCHEMICAL ASSAY FOR STIMULANT ALKALOIDS IN CELASTRACEAE, AND A PHYLOGEOGRAPHIC ANALYSIS OF *CATHA EDULIS* IN AREAS OF HISTORIC CULTIVATION

Submitted by

Luke Raymond Tembrock

Department of Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2015

Doctoral Committee:

Advisor: Mark P. Simmons

Amy L. Angert
Christopher M. Richards
Sarah M. Ward

ABSTRACT

A STUDY OF PLANT DOMESTICATION AND EVOLUTION THROUGH THE TAXONOMIC REVISION OF WILD NORTH AMERICAN *HUMULUS*, A PHYTOCHEMICAL ASSAY FOR STIMULANT ALKALOIDS IN CELASTRACEAE, AND A PHYLOGEOGRAPHIC ANALYSIS OF *CATHA EDULIS* IN AREAS OF HISTORIC CULTIVATION

The cultivation of plant species is essential to the survival of humans. The process of artificial selection that is used to modify wild individuals into improved cultivars results in genetic and morphological changes from wild progenitors. In order to understand the evolutionary patterns and processes involved with artificial selection both wild and cultivated populations must be thoroughly studied. Numerous methods are used to study the process of evolution under cultivation such as biology, chemistry, geography, history, linguistics and archeology. The understanding of evolution in a crop species is essential in current improvement programs to increase yield for a given crop.

I employed methods from the fields of taxonomy, analytical chemistry and phylogeography to study the process of evolution in cultivated plant species and/or their wild relatives. From a review of taxonomic, genetic, and phytochemical literature, as well as examination of morphological features I revised the wild North American *Humulus* (Cannabacae) in a manner that properly delimits the diversity found among the North American species. Using GC–MS and a forensics based derivatization method I assayed for the stimulant alkaloids cathinone, cathine, and similar compounds across the Celastraceae plant family. It was

found that that qat (*Catha edulis*) was the only species of those tested that biosynthesized cathinone and cathine. Using phylogeographic and population genetic techniques I inferred three wild regional origins, hybridization and numerous translocations out of the centers of origin for cultivated qat. From farmer interviews I examined what properties, genotype, phenotype, and/or geography explained the naming convention for qat cultivars among qat farmers. The character of stem color was found to highly plastic and thus genotype was not significantly correlated with the naming convention. Geographic patterns were confirmed for several cultivar names suggesting that anthropogenic factors are important in the naming conventions used among qat farmers. These four separate studies provide findings that not only clarify our understanding of evolutionary patterns among wild and cultivated species but provide a framework for breeding, conservation and forensic applications in the future.

ACKNOWLEDGEMENTS

I thank my major advisor Mark P. Simmons for his guidance and many insights during this project. Mark's continual encouragement and emphasis on quality ultimately made me a better researcher and scientist. I like thank my committee members Amy L. Angert, Christopher M. Richards, and Sarah M. Ward for all of their guidance and technical assistance. I especially thank Chris Richards for holding the necessary DEA permits, and allowing access to the NCGRP facility and equipment.

I thank the collections managers at BHSC, BUT, CSCN, ILLS, ISC, MONT, NEB, NY, RM, and US for loaning *Humulus* specimens for examination. In particular, I thank Jennifer Ackerfield at CS for her assistance in attaining and storing the loans as well as giving botanical advice whenever needed. I thank Corey Broeckling, Adam Heuberger, Sarah Lyons, Jessica Prenni, Frank Stermitz, and the many helpful staff members at the Proteomics and Metabolomics facility at Colorado State University for helping with my phytochemistry work. Assistance from Remi Bonnart, Adam Henk, Pat Reeves, and Ann Reilly at NCGRP was absolutely indispensable to the success of my Catha work and I thank them for this. I thank Terrence and Deena Walters of USDA for their access to imaging equipment and helping me with research assistantships during my graduate work. Funding from the American Institute of Yemeni Studies and Tom and Vernette Jirsa, donors to the H. D. Harrington Fellowship Fund was essential to the success of my project. I thank our many collaborators in Africa and Yemen, and in particular Manuel Curto from the University of Porto in developing qat SSR's.

I thank my colleagues Jennifer Cappa, Todd Gilligan, and Jenna McAleer for the many hours of assistance and discussion in completing my project. Lastly I thank my family for all of their support and understanding—especially Bill and Kathy Tembrock.

TABLE OF CONTENTS

CHAPTER 1: TAXONOMY, PHYTOCHEMISTRY, AND PHYLOGEOGRAPHY IN THE EMPIRICAL AND APPLIED STUDY OF EVOLUTION IN CROPS AND THEIR WILD RELATIVES

Introduction

The study of cultivated plants has been fundamental to our understanding of evolutionary processes. Human selection on plants in order to isolate desired traits has been regarded as one of humanity's largest biological experiments (Gepts 2004) from which much of the data have yet to be collected. Darwin (1876) introduced the concept of selection in the first chapter of *On the Origin of Species* by discussing how domesticated animals and plants are selected through human actions and how this process in some respects is analogous to natural selection. Researchers have continued to employ cultivated species to understand mechanisms underlying evolutionary patterns ever since (Ross-Ibarra 2007). Additionally, patterns of evolution in cultivated plants often parallel human historical and evolutionary patterns, thus providing an additional line of evidence in understanding human history (e.g., Verginelli et al. 2009; Hancock et al. 2010; Westengen et al. 2014).

Humans rely on plants for survival; with plants providing food, shelter, medicine, and fuel. Early humans consumed plant resources directly from natural environments with no evident conscious effort toward cultivating these plants outside their natural habitats. Beginning around 12,000 BP (before present) human societies began to cultivate plants, shifting from collecting plant resources from wild populations to growing plants in specially prepared environments (e.g. Kuijt & Goodale 2009; Zeder 2011). This shift toward an agrarian lifestyle altered human

societies to be more sedentary and dependent on plant cultivation (reviewed in Kuijt & Goodale 2009).

The artificial selection process by which wild plants are grown in cultivated settings involve morphologic and genetic changes from wild progenitors (Emshwiller 2006; Pickersgill 2009). These changes are related to the use of a given species. For instance grain size has increased since selection from wild progenitors in grass species such as wheat and barley (Purugannan & Fuller 2011). Morphological changes among cultivated plants often include increase in fruit and/or seed size, decrease in overall plant size, loss of seed dormancy, a shift toward annual life cycle, sterility, loss of defensive structures, increase or decrease of metabolites, changes in photoperiodic control, and changes in breeding system (Harlan 1992; Careau et al. 2010). A general consistency and predictability in the way that genetic and morphologic changes occur as a result of artificial selection is referred to as the "domestication syndrome" (Harlan 1992). Thus the definition of domestication encompasses the degree to which the domestication syndrome has altered a given cultivar, ranging from early cultivated individuals that are nearly identical to wild progenitors to highly selected individuals that have been so radically modified that they cannot persist from generation to generation without human manipulation (Miller and Gross 2011).

Understanding plant domestication often involves multiple different fields of study including biology, chemistry, geography, history, linguistics and archeology. As such multiple different data types can be employed in concert to answer questions regarding the history, origin, dispersal, and uses for different crop species. Data from historical records, archeological and linguistic studies are often used as the bases for a set of falsifiable hypotheses to be tested with genetic and/or chemical analyses. For example Morrel & Clegg (2007) tested historical

inferences generated from archeological and morphological data that suggested multiple origins for barley with genetic data to strongly support the multiple origin hypothesis. As technologies such as high–throughput genotyping, recovery of "fossil" DNA, remote sensing, and chemical assaying all improve and become less expensive, the resolution of inferences regarding the domestication process will also improve. Understanding the domestication process is important beyond historical curiosity; it is also an important step in documenting the standing genetic diversity found in wild and cultivated gene pools for developing improved crops. In fact, Warchefsky et al. (2014) proposed a breeding program that essentially replicates the early stages of domestication in which wild collected plants are allowed to freely cross with cultivated plants in order to generate novel cultivars. I examined the domestication process using three separate approaches at the family, genus and population levels.

Taxonomy as an important tool in delimiting diversity in crop wild relatives

I examined the wild diversity across the genus *Humulus* (Cannabaceae) in North American through a review of taxonomic, genetic, and phytochemical literature. From the review I proposed a taxonomic revision that reflects the evolutionary history and separates wild North American *Humulus* from Eurasian species of *Humulus,* as well as separating three distinct North American lineages. A taxonomic system that reflects evolutionary history and allows the straight forward diagnoses of each taxon is essential for the proper identification of wild source material from which cultivars and cultivar parents have been selected. Also an appropriate taxonomic delimitation of wild genetic diversity was needed in order to efficiently guide future breeding and conservation programs.

Phytochemical methods combined with phylogeny and ethnobotanical literature to assay for targeted and non–targeted metabolites in Celastraceae

I employed phylogenetic data (Simmons et al. 2008) and ethnobotanical accounts (Watt & Breyer-Brandwijk 1962; González et al. 1986) as a basis to test whether stimulant alkaloids are present in species of the Celastraceae aside from *Catha edulis* – a species cultivated for the stimulant alkaloids cathinone and cathine. Ethnobotanical accounts mention the Xhosa and Khoikhoi cultures of South Africa use of *Cassine schinoides* leaves to reduce fatigue, hunger, and thirst (Watt & Breyer–Brandwijk 1962), while accounts from Canary Island cultures mention chewing the leaves of *Gymnosporia cassinoides* to alleviate fatigue (González et al. 1986).

No prominent stimulant alkaloids were found in close relatives to *Catha edulis* or in species mentioned in the ethnobotanical literature as stimulants. But 26 separate compounds annotated as sterol– or terpene–like were annotated from the dataset, one of which might account for the stimulant properties of *Cassine schinoides*. Knowing that *Catha edulis* alone produces cathinone and cathine indicates that no other species within Celastraceae could be domesticated to produce this set of alkaloids. This finding demonstrates the need to preserve species and thus potentially unique biosynthetic pathways that could be the source of useful compounds such as medicines. In addition, the absence of cathinone or cathine in other Celastraceae species means that no drug scheduling outside of *Catha edulis* is necessary.

Phylogeography to study the origin and evolution of wild and cultivated qat (Celastraceae: *Catha edulis*)

I employed population genetic and phylogeographic methods to test historical accounts regarding the origin of cultivated *Catha edulis*. Historical and legendary accounts consistently assert three different origins for cultivated qat —Yemen, Ethiopia, and Kenya. From wild and cultivated collections across the three purported origins I tested the hypothesis that qat had a

single origin vs. multiple origins and the extent of translocation from the place(s) of origin. To test the origin hypotheses I had a dataset of 1561 individuals that were genotyped across 17 microsatellite loci. Additionally these data were employed to test the nature of the domestication process in *Catha edulis* against expectations for long lived perennials in general. In addition, clonal groups were identified and the extent to which they have been propagated and translocated from centers of origin was also assessed. From the clonal groupings the cultivar names used in Ethiopia, Kenya and Yemen were examined for which aspect (geography, genotype or phenotype) would explain the naming conventions used by qat farmers to refer to the cultivars that they cultivated.

Samples from across the native range of *Catha edulis* and within major areas of cultivation indicate that cultivated *Catha edulis* has three regional wild sources from which cultivated qat was selected. Southern Ethiopia contains one wild source area and two were delimited from Kenya on either side of the African Rift Valley. Within each of the three regional source areas further genetic sub–clustering was identified that followed a geographically rational pattern. For instance, subclusters were often found to be isolated to a single mountain or uplift formation. The genetic diversity among cultivated qat is equal to or greater than that found in the wild sources without cross–assignment to adjacent formations. Novel cultivated hybrids were found in areas where formerly allopatric wild genotypes have been brought into contact – most notably in Marsibit, Kenya. A separate Yemeni origin was not substantiated using several different genetic tests, but rather an Ethiopian origin for Yemeni cultivars was indicated. A moderate loss of genetic diversity was found due to the Yemeni translocation from Ethiopia; however a severe bottleneck was imposed in the translocation from Yemen to Madagascar. *Catha edulis* differs from other long lived perennial crops studied thus far in that the target of

selection is a secondary metabolite in the leaves and the initial time since cultivation is relatively recent. Despite these differences the evolutionary trends for cultivated qat are consistent with other long lived perennial species' domestication patterns.

Conclusions and future work

In total these three separate studies delimiting evolutionary patterns are useful findings in themselves to understand and evolution and cultivation. Further these studies are a starting point for further research and applied projects such as systematic breeding in *Humulus* with North American species, detailed structural characterization of terpenes and sterols in *Cassine*, and tracking of *Catha edulis* cultivation and illegal importation.

REFERENCES

Careau VD, Réale D., Humphries MM, Thomas DW (2010) The pace of life under artificial selection: Personality, energy expenditure, and longevity are correlated in domestic dogs. American Naturalist 175:753–758.

Darwin C (1876) The origin of species by means of natural selection, or the preservation of favoured races in the struggle for life. London: John Murray. 6th ed., with additions and corrections.

Ehmshiller E (2006) Genetic data and plant domestication. *In* Zeder MA, Bradley DG, Emshwiller E, Smith BD (eds.) Documenting Domestication, pp 99–122. University of California Press, Berkeley, California, USA.

Gepts P (2004) Crop domestication as a long‐term selection experiment. Plant Breeding Reviews 24:1–44.

González AG, López I, Ferro EA, Ravelo AG, Gutiérrez J, Aguilar AM, (1986) Taxonomy and chemotaxonomy of some species of Celastraceae. Biochemical Systematics and Ecology 14: 479–480.

Hancock AM, Witonsky DB, Ehler E, Alkorta–Aranburu G, Beall C, Gebremedhin A, Sukernik R, Uterman G, Pritchard J, Coop G, DiReinzo A (2010) Human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. Proceedings of the National Academy of Sciences of the United States of America 107: 8924–8930.

Harlan JR (1992) Crops & man. Madison, American Society of Agronomy, Inc.

Kuijt I, Goodale N (2009) Daily practice and the organization of space at the dawn of agriculture: a case study from the near east. American Antiquity 74: 403–422.

Miller, AJ, Gross BL (2011) From forest to field: Perennial fruit crop domestication. American Journal of Botany 98: 1389–1414.

Morrell PL, Clegg MT (2007) Genetic evidence for a second domestication of barley (*Hordeum vulgare*) east of the Fertile Crescent. Proceedings of the National Academy of Sciences of the United States of America 104: 3289–3294.

Pickersgill B (2009) Domestication of plants revisited – Darwin to the present day. Botanical Journal of the Linnean Society 161: 203–212.

Purugganan MD, Fuller DQ (2011) Archaeological data reveal slow rates of evolution during plant domestication. Evolution 65: 171–183.

Ross–Ibarra J, Morrel P, Gaut BS (2007) Plant domestication, a unique opportunity to identify the genetic basis of adaptation. Proceedings of the National Academy of Sciences of the United States of America 104: 8641–8648.

Simmons MP, Cappa JJ, Archer RH, Ford AJ, Eichstedt D, Clevinger CC (2008) Phylogeny of the Celastreae (Celastraceae) and the relationships of *Catha edulis* (qat) inferred from morphological characters and nuclear and plastid genes. Molecular Phylogenetics and Evolution 48: 745–757.

Verginelli F, Aru F, Battista P, Mariana–Costantini R (2009) Nutrigenetics in the light of human evolution. Journal of Nutrigenetics and Nutrigenomics 2: 91–102.

Westengen OT, Okonga MA, Onek L, Berg T, Upadhyaya H, Birkeland S, Khalsa SDK, Ring KH, Stenseth NC, Brysting AK (2014) Ethnolinguistic structuring of sorghum genetic diversity in Africa and the role of local seed systems. Proceedings of the National Academy of Sciences of the United States of America 11: 14100–14105.

Warchefsky E, Penmesta RV, Cook DR, von Wettberg EJB (2014) Back to the wilds: tapping evolutionary adaptations for resilient cropsthrough systematic hybridization with crop wild relatives. American Journal of Botany 101: 1791800.

Watt JM, Breyer–Brandwijk MG (1962) The medicinal and poisonous plants of southern and eastern Africa. Edinburgh, E. &. S. Livingstone Ltd.

Westengen OT, Okonga MA, Onek L, Berg T, Upadhyaya H, Birkeland S, Khalsa SDK, Ring KH, Stenseth NC, Brysting AK (2014) Ethnolinguistic structuring of sorghum genetic diversity in Africa and the role of local seed systems. Proceedings of the National Academy of Sciences of the United States of America 11: 14100–14105.

Zeder MA (2011) The origins of agriculture in the Near East. Current Anthropology 52: S221–S235.

CHAPTER 2: A REVISION OF NORTH AMERICAN *HUMULUS* (CANNABACEAE)

Summary

The taxonomic status of *Humulus* L. in North America has been debated for over 150 years. Recent molecular and morphological studies have positively identified three distinct evolutionary lineages among wild varieties of Nearctic *Humulus*. Here we review these studies, and present morphological evidence supporting the recognition of three North American species of *Humulus*. *Humulus lupuloides* (E. Small) Tembrock stat. nov., *Humulus neomexicanus* (A. Nelson & Cockerell) Rydb. stat. rev., and *Humulus pubescens* (E. Small) Tembrock, stat. nov. are elevated from their former positions as varieties of *Humulus lupulus*.

Introduction

"Our Netherlanders…can brew as good beer here [Manhattan Island, U.S.A.] as in our own Fatherland, for good hops grow in the woods" (De Vries 1655: 157).

Humulus L., commonly referred to as hops, is known to most as a necessary ingredient for brewing beer. Hops 'cones' are added to beer to provide bitterness and aroma. The hops 'cones' are a series of closely appressed stipular bracts and bracteoles of the female inflorescence. The bracteoles possess numerous lupulin glands that produce resins and oils (Small 1978). The resins contain α–acids and β–acids that provide the bitter taste of beer whereas the oils impart aromatic qualities (Neve 1991). Today, hops are used primarily in beer, yet *Humulus* species have been used as food, medicine, and fiber in European, Indian, Chinese and indigenous North American cultures for centuries (reviewed by Zanoli & Zavatti 2008).

Cultivation of hops is first documented in the records of the Freisingen abbey, Bavaria, Germany A.D. 859 (Wilson 1975). Thereafter hops cultivation spread throughout Western

Europe and arrived in New England sometime before 1629 (Tomlan 1992). Yet, prior to the introduction of European hops plants, Dutch settlers employed indigenous North American hops to brew beer (Tomlan 1992). Introduction of European hops cultivars resulted in occasional spontaneous hybridization between North American and European plants giving rise to the 'cluster' varieties (Neve 1991; Van Valkenburg 1995). Numerous intentional crosses between North American and European hops have been made in the last ~100 years (e.g. Salmon 1934; 1938; Zimmerman *et al.* 1975;) resulting in cultivars (e.g. 'brewers gold', 'comet', and 'sunshine') high in total α–acid content and resistance to serious hops diseases and pests (Van Valkenburg 1995).

Humulus is a small genus of dioecious, dextrose twining herbaceous plants native across the temperate latitudes of the Northern Hemisphere and has been naturalized in areas of the southern hemisphere. The number of species recognized in *Humulus* has varied from one to ten in the past (International Plant Name Index 2011), with three being currently recognized (Small 1978; Neve 1991). *Humulus lupulus* L., the type species of the genus, is distributed across the temperate regions of Europe, Western Asia, and North America. Past treatments of *H. lupulus* have divided it into multiple different taxonomic combinations of species, subspecies, and/or varieties based mainly on geographic and morphological differences (Small 1978; Neve 1991; IPNI 2011). North American *Humulus,* in particular, has been the subject of numerous taxonomic revisions in an attempt to appropriately subdivide the morphological diversity and broad geographic extent found within the group (Small 1978; Neve 1991; IPNI 2011). The other two species, *H. japonicus* Siebold & Zucc. and *H. yunnanensis* Hu, are native to Asia, although *H. japonicus* has been grown as an ornamental in some areas of North America.

Nuttall (1848) applied the name *Humulus americanus* Nutt. to all wild North American *Humulus*. He differentiated *H. americanus* from European *H. lupulus* based on the morphology of leaf denticulations and geography (Nuttall 1848; Small 1978). Nuttall (1848: 23) also made special note in his explanation of *H. americanus* that "most luxuriant specimens from the borders of streams in the Rocky Mountains" had been collected by Dr. Gambel in or near New Mexico. Yet the specimen he designated as the type for *H. americanus* was collected from Pennsylvania and not based on the western specimens collected by Gambel. From Nuttall's description, it appears that all North American wild hops should be assigned to the epithet *americanus* (Small 1978).

Nuttall seems to have inserted a discussion of Gambel's western specimen to declare the nativity of *Humulus* to North America. Release from cultivation could not be claimed for *Humulus* found in New Mexico as it could for *Humulus* in New England, as hops were widely cultivated throughout the northeast. Gray (1857) rejected the idea that eastern North American *Humulus* were distinct from European species by including only the description for *H. lupulus* in his revised *Manual of the Botany of the Northern United States* (Davis 1957). Yet, in correspondence with colleagues, Gray noted the existence of indigenous *Humulus* in western North America (Davis 1957). The discovery and description of fossil *Humulus* in late Eocene strata from Florissant, Colorado (MacGinitie 1969) provides evidence for an ancient dispersal of *Humulus* to North America at least ~34.1 million years before present. Because of discrepancies in the range, description, and type specimen, the binomial *H. americanus* was never broadly applied in the botanical literature.

Nelson & Cockerell (1903) applied the variety name *neomexicanus* to *H. lupulus* collected in northern New Mexico. They based this name on differences in geographic

distribution, leaf lobe, and bract morphology. Following Nelson & Cockerell's rationale for recognizing variety *neomexicanus,* Rydberg (1917) recognized *H. neomexicanus* at the species level in his *Flora of the Rocky Mountains and Adjacent Plains.*

Davis (1957), through scoring of morphological characters in cultivated and wild *Humulus*, found that wild North American *Humulus* were distinguishable from European *Humulus*. Yet, Davis (1957: 293) suggested that the name *H. lupulus* L. should be retained for all wild perennial hops until "the native American hop is more thoroughly understood". Small (1978) conducted a phenetic analysis using vegetative and geographic characters in which he recognized five varieties within *H. lupulus*. Small's (1978) work resulted in the recognition of three North American varieties: *H. lupulus* var. *neomexicanus* (synonymous with Nelson and Cockerell's *H. lupulus* var. *neomexicanus* and Rydberg's *H. neomexicanus*), *H. lupulus* var. *lupuloides* (which was assigned tentative synonymy with *H. americanus*) and *H. lupulus.* var. *pubescens.* The three varieties remain widely recognized in recent floras and catalogues (e.g. Gleason & Cronquist 1991; Small 1997).

Since the recognition of Small's three North American *Humulus* varieties (1978) the concept and implementation of botanical varieties has been reconsidered and updated (e.g. McNiell et al. 2012; Braby et al. 2012; Ellison et al. 2014). Ellison et al. (2014) recommend the abandonment of the term variety as a taxonomic rank and suggest only subspecies be used in infraspecific designations. Subspecies as described by Braby et al. (2012) and modified by Ellison et al. (2014: 946) are "…evolving populations that represent partially isolated lineages of a well–defined species that are either allopatric or sympatric, phenotypically distinct, have at least one fixed diagnosable character state, and that these character differences are, or are

assumed to be, correlated with, at least partial evolutionary independence according to population genetic structure."

Reeves & Richards (2011) employed molecular methods, natural history, and distributional data from the three North American *Humulus* varieties (Small 1978) to evaluate five distinct species criteria under the framework of the general lineage species concept (de Queiroz 1998, 2005, 2007). They found that all three taxa fulfilled the criteria for the monophyly (de Queiroz & Donoghue 1988), diagnosability (Nixon & Wheeler 1990), and genotypic clustering (Mallet 1995) species criteria. Additionally var. *pubescens* satisfied the reproductive isolation species criterion (Mayr 1924) while var. *neomexicanus* satisfied the ecological species criterion. As such, Reeves & Richards (2011) recommended the elevation of var. *pubescens* and var. *neomexicanus* to species under the general lineage concept but did not recommend the elevation of var. *lupuloides* based on limited genetic sampling in their study and the lack of any defined biological processes to explain the genetic divergence they observed.

We believe that Reeves and Richards (2011) as well as others (e.g. Small 1978; Patzak et al. 2010) presented sufficient data to diagnose var. *lupuloides* from its congeners. Although they could not identify the process by which var. *lupuloides* is maintained as distinct, it appears to be a separately evolving metapopulation lineage based on several pattern–based criteria, which is necessary to designate a species under the general lineage concept (e.g. de Queiroz 1998; 2007; Marshall et al. 2006; Light et al. 2008; Abbott & Judd 2011). Definition or discovery of the underlying biological processes causing these patterns is not a requirement of the general lineage concept (or many other species concepts), and this lack of evidence should not detract from the several species criteria that support the three varieties of North American *Humulus* as distinct lineages. In support of this, we examined 56 specimens of var. *lupuloides* from across its entire

range, and determined that the morphological character states separating it from its congeners are consistent. Variety *lupuloides* is diagnosable from other species and varieties of *Humulus* by having greater than 20 hairs per linear cm on the abaxial surface of the leaf midrib, and not possessing a combination of five–lobed leaves and pubescence between the veins on the abaxial surface of the leaves.

Unfortunately Ellison et al. (2014) did not explicitly propose a conceptual framework or set of recommendations for differentiating subspecies from species. Consideration of subspecies in the context of the general lineage species concept could be useful. As individual barriers evolve between populations that fulfill a species concept, the isolation between these two populations increases. Thus if a species concept or a set of species concepts are partially fulfilled in testing populations to be species then the designation subspecies is appropriate . That is, provided this "partial isolation" is fixed, diagnosable, and evolutionarily correlated with the independence of populations. Once a species concept or set of species concepts are completely fulfilled for a given set of taxa tested a species designation should be considered.

Here we provide a taxonomic revision of North American *Humulus* and formal elevation of the three North American varieties to species. Recognition of North American *Humulus* as distinct from European *Humulus* on a morphological and molecular level has been observed and noted by botanists, horticulturists, and brewers since European settlement of North America (e.g., Nuttall 1848; Salmon 1934; Small 1978; Tomlan 1992; Reeves & Richards 2011), but a taxonomic designation that appropriately differentiates North American *Humulus* from European species has not been completed. The limited acceptance of Nuttall's (1848) *H. americanus* and Rydberg's (1917) *H. neomexicanus* by many botanists is likely a result of the limited material presented in Nuttall's and Rydberg's treatments. The taxonomic changes we present here create

consistency across the observed morphological and genetic diversity within *Humulus* and clearly separate the North American taxa from those in Europe and Asia. Our findings are based on a review of recent and historical studies using molecular and ecological evidence (e.g. Bassil et al. 2008; Patzak et al. 2010; Reeves & Richards 2011), robust morphological analyses (Small 1978), and detailed comparisons and observations made from live plants and herbarium specimens.

Materials and methods

For this study we examined 160 herbarium specimens from across the native range of North American *Humulus*, with a focus on areas of geographic overlap in order to find potential hybrid zones. The samples included 12 var. *japonicus*, 56 var. *lupuloides*, 29 var. *lupulus*, 26 var. *neomexicanus*, and 37 var. *pubescens*. Plant material for this study was examined from the following herbaria: BHSC, BUT, CS, CSCN, ILLS, ISC, MONT, NEB, NY, RM, and, US. Type specimens were examined from the digital collections of PH and RM. This material, in addition to field collections (collected in August and September 2011 adjacent to the cities of Lyons and Boulder, Colorado, U.S.A.), were used to make observations and measurements of morphological features for the illustrations, diagnoses and the dichotomous key involved with this revision. The materials examined in this study encompass and expand the geographic range of previous systematic studies on North American *Humulus*. This expanded sampling allowed us to test the robustness of previous morphological diagnoses in Small (1978), and Reeves and Richards (2011). Illustrations were produced using Adobe Illustrator CS5 (Adobe Systems Inc., San Jose, Calif.) from photographs taken using a Canon EOS 40D digital SLR (Canon U.S.A., Lake Success, N.Y.) mounted on a Visionary Digital BK Lab System (Visionary Digital, Palmyra, Va.). Species descriptions for all taxa follow the order indicated in Judd et al. (2008).

In addition to our morphological observations and measurements we conducted a literature review of previous *Humulus* studies that either directly or indirectly tested the distinctness of *Humulus* taxa by collecting data at various taxonomic levels proposed for *Humulus*. The studies include a variety of data types including fossil, cytogenetic, molecular, phytochemical and ecological. The data from these studies were used to either contradict or support the recognition of three North American species and/or a North American lineage separate from a European lineage. The results from this review were compiled into Table 1 for comparison.

Results and discussion

Diagnoses of North American hops have been based mostly on vegetative characters (e.g. Small 1978). *Humulus pubescens* and *H. lupuloides* possess four or fewer leaf lobes whereas *H. neomexicanus* has five or more leaf lobes that are sometimes further subdivided (Figs. 2.3A, 2.3A, 2.3A). Some early season immature *Humulus* leaves are without lobes and cordate in shape, yet examination of the leaf venation pattern at an immature stage can reveal the future lobing pattern (Small 1978). Preceding formation of lobes, immature leaves of *H. neomexicanus* have four or more prominent side veins whereas those of *H. pubescens* and *H. lupuloides* have two or fewer. We have observed that venation patterns are rarely necessary for identification, as leaf–lobe formation often initiates soon after leaf expansion.

Trichomes on the abaxial leaf surfaces are useful in differentiating *H. pubescens, H. neomexicanus,* and *H. lupuloides*, with *H. pubescens* possessing a greater density of trichomes on the midrib (> 100 per linear cm of midrib) and between leaf veins (Figs. 2.3D, 2.4D, 2.5D). Rarely, *H. neomexicanus* individuals have been found to possess somewhat dense pubescence on their abaxial leaf surface, but remain distinguishable from *H. pubescens* and *H. lupuloides* in

having five–lobed leaves and a dense concentration of glands (> 30 per cm^2) on the abaxial leaf surface (Figs. 2.3D, 2.4D, 2.5D; Small 1978).

The phylogenetic analyses conducted by Reeves and Richards (2011) using 555 Amplified Fragment Length Polymorphism (AFLP) loci were consistent with the varietal designations (Small 1978) for North American *Humulus* (Fig 2.1). In addition to the recent inferences made by Reeves and Richards (2011) using AFLPs, evolutionary relationships within *Humulus* have been examined in the following studies: Murakami *et al.* (2006) using nuclear and plastid sequence data; Bassil *et al.* (2008), Stajner *et al.* (2008), and Jakše *et al.* (2011) using genic microsatellites; and Howard *et al.* (2011) using diversity arrays technology (DArT) markers. All of these studies found that wild North American accessions represented distinct lineage from European and Asian accessions of *Humulus*. Additionally, Bassil *et al.* (2008) found that var. *neomexicanus* and var. *lupuloides* formed distinct groupings in their neighbor–joining analyses. Insufficient sampling among var. *pubescens* did not allow the authors to make any conclusions regarding the distinctness of the var. *pubescens* lineage (Bassil et al. 2008). From their molecular–clock analyses, Murakami et al. (2006) inferred that ancestral migrants of North American hops moved from East Asia into North America between 0.46 ± 0.17 to 0.69 ± 0.21 million years ago. But note that Murakami et al. (2006) did not employ fossil calibration despite the existence of much older Eocene *Humulus* fossil specimens from North America (MacGinitie 1969) and Asia (Collinson 1989) bringing into question there dates for ancestral migration of hops into North America. The Florissant formation from which the fossil *Humulus* was found has been dated to 34.07 million years using $^{40}Ar/^{39}Ar$ radiometric dating methods (Meyer 2003).

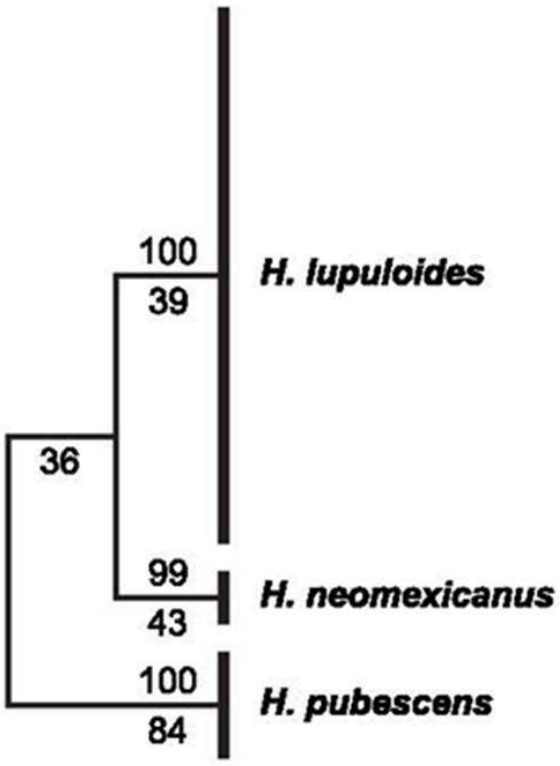


Figure 2.1. Cladogram of *Humulus* species from AFLP analysis, adapted from Reeves and Richards (2011). Numbers above the branches indicate bootstrap values for the clades from the strict consensus of 43 equally optimal trees using 369 parsimony informative AFLP characters. Root placement was determined using a stochastic Dollo substitution model. Numbers below the branches are cluster stability values determined from a hierarchical assignment after the results in a PCO–MC analysis. Vertical bars at branch terminals are proportional to the number of individuals sampled.

Cytogenetic data from sequence tagged sites (STS) markers has demonstrated that the molecular structure of the male sex chromosome in North American *Humulus* is distinct from that of European *Humulus* male sex chromosomes (Danilova & Henning 2005). Length differences for sex chromosomes have also been previously noted between *Humulus* of North American and European origin (Neve 1991; Haunold 1991; Shepard et al. 2000; Grabowska–Joachimiak et al. 2006).

Employing phytochemical data, Hummer et al. (2005) found that var. *lupuloides* and var. *pubescens* both have fixed presence of Xanthogalenol and 4'–O–methylxanthohumol. European

var. *lupulus* has been found to be fixed absent for these metabolites (Stevens et al. 2000), providing phytochemical evidence for the distinctness of North American from European hops. Patzak et al. (2010) expanded on previous phytochemical studies by simultaneously analyzing hop resins, polyphenols, and essential oils for wild and cultivated hops accessions from North America and Eurasia. Through a PCA analysis of all chemical characteristics, Patzak et al. (2010) were able to clearly differentiate between wild North American hops and wild Eurasian hops. Patzak et al. (2010) also employed STS and microsatellite markers in combination with the data from their chemical analysis to examine hops relationships at the molecular level and found that in their cluster analysis (UPGMA–based on Jaccard's similarity coefficient) of 257 individuals that wild North American hops clustered by geographic origin, which is consistent with Small's (1978) varieties.

Along with performing molecular analyses to resolve the monophyly of North American *Humulus* varieties, Reeves & Richards (2011) performed niche modeling to examine the degree to which varieties might differ in their ecological adaptations. Results from these analyses indicate that the niche occupied by var. *neomexicanus* is quantifiably different from that of var. *pubescens* and var. *lupuloides*. In addition, several other workers have made note of the marked difference in var. *neomexicanus* collection sites versus those of the other two varieties (e.g. Smith et al. 2006).

Neve (1991) suggested that the recognition of *H. americanus* in the past was rejected by later botanists because of the interfertility among hops taxa, implying adherence to the biological species concept (Mayr 1948) by taxonomists at that time. During cladogenesis, the emergence of characteristics that will satisfy a given lineage criteria have no universal order of emergence (de Queiroz 1998). Specifically, when considering *Humulus,* it appears that intrinsic barriers to

fertility built up slowly even across multiple cladogenic events; crosses with *Cannabis sativa* L., and the more distant *Urtica dioca* L. and *U. urens* L. to *H. lupulus* elicited the development of cones and abortive embryos (Neve 1991). Despite the apparently slow evolution of barriers to interfertility in *Humulus* and the Cannabaceae, Reeves & Richards (2011) in their AFLP analyses found no clear evidence of admixture between genetic clusters of wild var. *neomexicanus*, var. *lupuloides* and var. *pubescens*. Variety *neomexicanus* and var. *lupuloides* have been found to hybridize with European var. *lupulus* yet no evidence of hybridization between var. *pubescens* and European var. *lupulus* has been found, despite regions of sympatry (Small 1980; Fig 2.2A–C). Localized differences in phenology may explain the apparent reproductive isolation of *H. pubescens* from its congeners (Reeves & Richards 2011).

Additionally, as discussed by de Queiroz (2007) and Ellison et al. (2014), the formation of a sterile hybrid between two separate lineages will not provide a connection for persistent geneflow or continue as a distinct lineage in itself. Genetic incompatibilities in the sterile hybrid will ensure that this apparent linkage of the two lineages persists only for a single generation. As de Queiroz (2007) noted, species are lineages that persist across greater time scales than a single generation. Along with differences in phenology proposed by Reeves & Richards (2011), the sterility of hybrids between North American *Humulus* lineages could also be a mechanism by which sympatric populations remain isolated over longer time scales. Hybrid sterility has not been tested for wild North American *Humulus* lineages outside the context of agricultural hybrids. However, sterile *Humulus* hybrids have been found in nature and are especially common between wild North American and European crosses likely due to chromosomal interchange (Neve 1965).

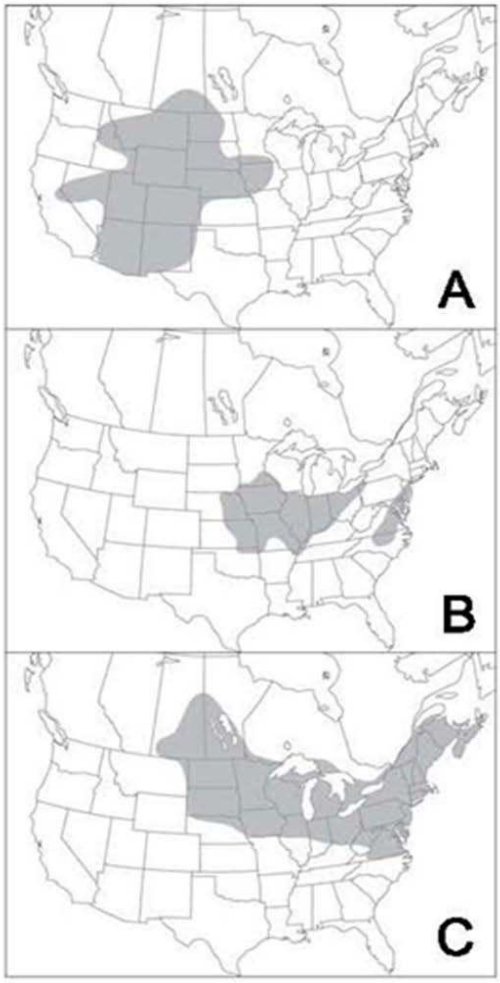


Figure 2.2. Range maps for the North American *Humulus* species. A. *H. neomexicanus* (This species has been infrequently collected from locations in northern Mexico, yet the extent of its range is not well documented and therefore not included in this map) B. *H. pubescens* C. *H. lupuloides*.

From this literature review no study was found that wholly refuted the separation of North American *Humulus* from Eurasian *Humulus*. Several studies (table 2.1) did not have conclusive results for separation of certain lineages because they did not collect the appropriate data and/or analyze it in a way that tested for separation of the lineages. Multiple different data types and analyses strongly support North American *Humulus* as a separately evolving lineage from the Eurasian lineages. From these results a taxonomic designation differing from Small's

Table 2.1. Summary of relevant studies and taxonomic revisions that provide evidence for or against separately evolving *Humulus* lineages.

Study	Data Type	Support for a seperately evolving North American Lineage	Support for a seperately evolving *H. neomexicanus*	Support for a seperately evolving *H. pubescens*	Support for a seperately evolving *H. lupuloides*	Recommend taxonomic revision
Nuttal 1848	Morphology	yes	yes	NA	NA	yes
Gray 1857	Morphology	yes/no	yes	no	no	no
Nelson and Cockerell 1903	Morphology	yes	yes	NA	NA	yes
Rydberg 1917	Morphology	yes	yes	NA	NA	yes
Davis 1957	Morphology	yes	NA	NA	NA	no
MacGinitie 1969 and Meyer 2003	Fossil/Radiometric	yes	NA	NA	NA	NA
Small 1978	Morphology and Geography	yes	yes	yes	yes	yes
Small 1981	Morphology and Geography	yes	NA	NA	NA	NA
Á. Löve & D. Löve 1982	Morphology	yes	NA	NA	yes	yes
Neve 1991	Cytogenetic (sex chromosome length difference)	yes	NA	NA	NA	NA
Haunold 1991	Cytogenetic (sex chromosome length difference)	yes	NA	NA	NA	NA
Shepard et al. 2000	Cytogenetic (sex chromosome length difference)	yes	NA	NA	NA	NA
Stevens et al. 2000	Phytochemical (HPLC-MS, 22 metabolites)	yes, except for *H. neomexicanus*	yes, from *H. pubescens* & *H. lupuloides*	yes, different from European	yes, different from European	NA
Danilova & Herning 2005	Cytogenetic (STS)	yes	NA	NA	NA	NA
Hummer et al. 2005	Phytochemical (HPLC, 5 metabolites)	yes in part	polymorphic	yes, different from European	yes, different from European	NA
Smith et al 2006	Ecology (observation)	NA	yes, from *pubesens* and *lupuloides*	yes, from *neomexicanus*	yes, from *neomexicanus*	NA
Grabowska-Joachimiak et al. 2006	Cytogenetic (sex chromosome length difference)	yes	NA	NA	NA	NA
Murakami et al. 2006	Molecular (Chloroplast and Nuclear DNA sequence)	yes	inconclusive (not directly tested)	inconclusive (not directly tested)	inconclusive (not directly tested)	NA
Bassil et al. 2008	Molecular (genic SSRs)	yes	yes	insufficient sampling	yes	NA
Stajner et al. 2008	Molecular (SSRs)	yes	NA	NA	NA	NA
Patzak et al. 2010	Phytochemical + Molecular (HPLC + SSRs)	yes	yes	yes	yes	NA
Reeves & Richards 2011	Molecular (AFLP)	NA	yes	yes	yes	yes
Reeves & Richards 2011	Morphology	NA	yes	yes	yes	yes
Reeves & Richards 2011	Ecology (niche modeling)	NA	yes, from *pubesens* and *lupuloides*	yes, from *neomexicanus*	yes, from *neomexicanus*	yes
Howard et al. 2011	Molecular (DArT markers)	yes	inconclusive (not directly tested)	inconclusive (not directly tested)	inconclusive (not directly tested)	NA
Jakše et al. (2011)	Molecular (EST-SSRs)	yes	NA	NA	NA	NA

(1978) arrangement in which intraspecific groupings are split from a European *H. lupulus* is clearly needed to properly delimit the genetic, geographic, morphological, ecological, cytogenetic, and phytochemical variation in the North American taxa from the Eurasian taxa. In addition, the studies that sampled the North American varieties found that multiple data types support the separation of these taxa as distinct lineages. Therefore the current system of *Humulus lupulus* varieties is incompatible with current taxonomic practices, the findings from multiple different studies, our own observations, and numerous different species concepts, including the general lineage concept. The elevation of these varieties to species will provide a taxonomic system that more accurately defines the separate *Humulus* lineages across their Holarctic, and within their Nearctic ranges. Further work is needed in understanding the diversity of the wild Eurasian *Humulus* lineages (Stajner et al. 2008) and applying the appropriate taxonomic designation thereof.

Taxonomic treatment

Humulus L., Sp. Pl. 1028. 1753. (Lectotype: BM Savage Catalogue entry 1178.1)

Lupulus Tourn. Ex Mill. Gard. Dict., ed. 6. 1752.

Vine–like, rightward twining, herbaceous perennials or annuals, rhizomatous or taprooted. Stems typically branched, 0.5–5+ m long (length often dictated by size of supporting vegetation), glabrate to densely pubescent with both pilose and cystolithic, two–forked hairs, sometimes tomentose at the nodes. Leaves opposite, petiolate; petioles typically twining, 2–12+ cm long, pubescent with cystolithic, two–forked hairs; laminae simple, entire or often palmately lobed, cordate, 3–15 cm long, with toothed margins, evidently veined, the abaxial surface glabrate to densely pubescent with both pilose and cystolithic hairs (either spinulose or two–forked), sometimes gland–dotted. Inflorescences on separate male and female plants (dioecious):

staminate inflorescences axillary and/or terminal, cymose panicles, 10–25 cm long, with (10–) 20–100+ small flowers; pistillate inflorescences axillary, spikes or short–pedicellate racemes, flowers paired or solitary, subtended by bracts and bracteoles. Infructescences pendulous, cone–like due to closely appressed bracts, (1–) 1.5–3 (–6) cm long, pale yellow to green; achenes enclosed in a lobeless, persistent calyx, terete or lenticular, 3–5 mm long, yellowish to yellow–brown.

Key to the native and naturalized North American species of *Humulus*: introduced species are indicated with an asterisk.

1. Petiole as long as or longer than lamina of leaf; pubescence on the abaxial side of the leaf midrib consisting of stiff spinulose trichomes…**H. japonicus*

1. Petiole shorter than lamina of the leaf; pubescence on the abaxial side of leaf veins made up of soft silvery trichomes… 2.

2. Glands on the abaxial leaf surface with fewer than 20 glands per cm^2; 15 to 25 hairs per linear cm on the abaxial surface of the leaf midrib; nodes weakly pubescent…**H. lupulus*

2. Glands on abaxial leaf surface greater than 25 glands per cm^2; 30 to 50+ hairs per linear cm on the abaxial surface of the leaf midrib; nodes with dense silvery pubescence… 3.

3. Most leaves possess 5 or more lobes; glands on abaxial leaf surface in excess of 30 per cm^2 … *H. neomexicanus*

3. Most leaves possess 3 lobes or remain unlobed and cordate, rarely leaves of 4 lobes are found; glands on abaxial leaf surface fewer than 30 per cm^2 … 4.

4. Abaxial surface of leaf midrib with more than 100 hairs per linear cm, these hairs spreading and silvery, abaxial surface of leaf evidently pubescent between veins… *H. pubescens*

4. Abaxial surface of leaf midrib with 20 to 75 hairs per linear cm; these hairs closely appressed to the midrib, abaxial surface of leaf sparsely pubescent or glabrous between veins… *H. lupuloides*

1. **Humulus neomexicanus** (A. Nelson & Cockerell) Rydb. stat. rev. Proc. Biol. Soc. Wash. *16: 45 1903*. (Figs. 2.1A–D, 2.5A) Type: U.S.A. New Mexico: Beulah, Aug. 1903, T.D.A. Cockerell 14 (Holotype RM, photograph!).

Perennials, without tendrils, rhizomatous and substantially taprooted. Stems sometimes > 5 m long, variously pubescent, with white to silver trichomes that sharply increase in density around the nodes and form a transverse line of tomentum subtending, and within the axil, of the interpetiolar stipules; interpetiolar stipules ± deciduous, splitting along central veins with age, 0.5–1.5 cm, apex acute or rarely acuminate, with parallel venation. Leaves petiolate; petioles 3–8 cm long, pubescent with both pilose hairs as well as prominent cystolithic, two–forked hairs, the latter often linearly arranged, 5–20 per linear cm, < 2 mm long, opaque and brown to dark green below, transparent to amber above; laminae mostly deeply 5–lobed and sometimes further subdivided, the central sinuses 5–8 cm deep, occasionally smaller laminae 3–lobed, or rarely without lobes, 7–15 cm long and 8–20 cm wide at widest point, bases deeply cordate, the margins dentate with 2–4 mm long teeth, apices acuminate, the abaxial surface gland–dotted with >30 glands per cm^2 and with 5 evidently raised primary veins, the midrib with ≥ 25 closely–appressed hairs per linear cm, secondary and tertiary veins terminating into acuminate, marginal denticulations. Staminate inflorescences axillary or rarely terminal, cymose panicles, 10–20 cm

long, with 50–100+ flowers subtended by light brown bracts 2–10 mm long, often present at the point of secondary panicle branching; pedicels 1–6 mm long; perianth comprised of a single whorl of 5 distinct lobes; lobes elliptic to ovate, the margins ciliate, 2–4 mm long, apices rounded to acuminate, light green to yellowish, becoming lighter near the margins, pubescent; anthers basifixed, with a distinct line of yellow glands along the abaxial surface of the connective, dehiscing along longitudinal sutures. Pistillate inflorescences axillary, short–pedicellate racemes, 1–2 cm long, bur–like at anthesis, with 10–50 paired flowers subtended by prominent yellow–green bracts; styles deciduous, 5–8 mm long. Infructescences 1–5 cm long with closely appressed bracts subtending the fruit; bracts ovate to elliptic, 1–2 cm long and approximately half as wide at fruiting, apices rounded to acuminate, browning with age, loosely enveloping the achene at the base and forming a wing about the fruit; achenes ovoid, terete to subterete, 3–5 mm long, cream to brown.

Phenology — Flowering and fruiting from the end of July to early October.

Distribution and habitat — Most closely associated with the Rocky Mountain Cordillera of western North America into northern Mexico, with some scattered collections from the north–central Great Plains (Fig. 2.5A). Mostly allopatric although its range overlaps with that of *H. pubescens* in eastern Nebraska and South Dakota and with that of *H. lupuloides* in Nebraska, North and South Dakota, Montana, and Wyoming (U.S.A.), Saskatchewan, Manitoba (Canada), and found in a few locations in northern Mexico.

Humulus neomexicanus are Found growing amongst rocky outcroppings and vegetation in mountain canyons or within riparian forest thickets, frequently co–occurring with *Acer negundo* L., *Clematis ligusticifolia* Nutt., *Prunus virginiana* L., and species of *Celtis* L., *Populus* L., *Prunus*, and *Salix* L. from 300 – 3000 meters above sea level (m.a.s.l.).

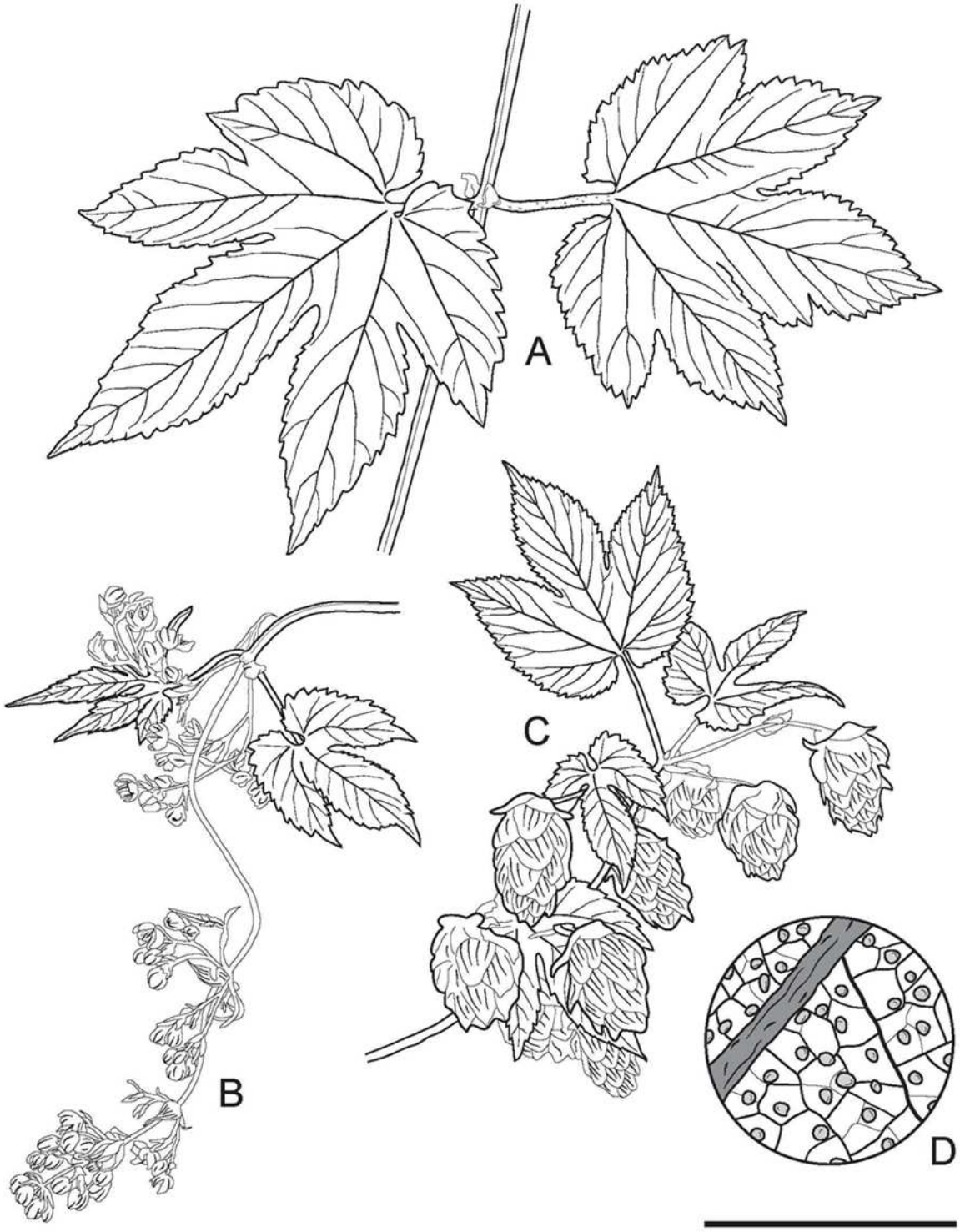


Figure 2.3. *Humulus neomexicanus* A. Stem and leaves B. Male inflorescence C. Female inflorescence D. Detail of abaxial leaf surface and central vein. Scale bar: A–C = 6 cm, D = 0.5 cm.

Representative specimens examined — U.S.A. Colorado: River flats, 23 August 1892, *R. Walter 803* (CS); Boulder County, 2 miles NW of Lyons along Hwy 36, 8 October 2011. *L. R. Tembrock 11–01* (CS); Boulder County, 80 meters E of South Boulder creek on S side of Baseline Rd., 8 October 2011, *L. R. Tembrock 11–02* (CS); Chaffee County, 3 miles S of Poncha Springs, 14 July 1961, *H. D. Harrington 9600* (CS); Chaffee County, near Hwy 285 3 miles S of Poncha Springs, 6 September 2004, *E. Sherman 24* (CS); Gunnison County, ~7 miles NE of Gunnison, 6 August 1960, *J. Barrell 185–60* (CS); Huerfano County, 10 miles west of Walsenburg along CO Highway 69, 17 July 1982, *R. G. Walter 0139* (ILLS); Jefferson County, east of Morrison, T 5 S, R 69 W, sec. 6, 14 July 1972, *Mooradian 72–269* (CS); Larimer County, 30 miles N of Fort Collins in Haigood Canyon, 11 August 1950, *H. D. Harrington 4768* (CS); Larimer County, Redstone canyon W of Fort Collins, 8 August 1944, *H. D. Harrington 608* (CS); Rio Blanco County, 0.75 miles ESE of junction of Deer & Davis gulches, 27 August 1982, *W. Baker & T. Naumann 82–350* (CS); Montana: Lewis & Clark County, 5 miles above Craig on Missouri river, 29 July 1900, *J. W. Blankinship 012003* (MONT); Petroleum County 34 miles N of Roundup in wooded ravines, 23 July 1948, *F. W. E. Booth 046120* (MONT); Stillwater County, Absarokee, 18 July 1914, *P.H. Hawkins 033418* (MONT); Nebraska: Hooker County, near forks of Dismal River, 3 July 1893, *P.A. Rydberg 1539* (NEB); Cherry County, Valentine National Wildlife refuge, S side of Dad's Lake, 28 June 2006, *R. F. Steinauer 2730* (NEB); Stanton County, Wood Duck Wildlife Management Area, 3 August 1999, *R. F. Steinauer 629* (NEB); Lancaster County, N of the entrance to Reller Natural History Research Area, 19 September 1990, *S. Rolfsmeier 8064* (NEB); New Mexico: Taos County, 7.8 miles S of Amalia on Costilla Creek, 15 August 1973, *N. H. & P. K. Holmgren 7208* (KANU); North Dakota: Bottineau County, near Strawberry Lake, 18 July 1973, *G. E. Larson 3134* (NEB); South

Dakota: Lawrence County, Little Spearfish Canyon, 19 August 1942, *F. L. Bennett 328* (BHSC); Lawrence County, SSW of Rod & Gun Campground along Forest Rd. 222, 15 September 2003, *S. J. & S. B. Rolfsmeier 549* (BHSC); Wyoming: Albany County, ~1.5 miles NW of Woods Landing, in Lake Owen Creek Canyon, 17 August 1978, *B.E. & L. Nelson 1915* (KANU); Washakie County, Foothills of Bighorn Mountains, in Tensleep Preserve, 22 Aug 2005, *D. Busemeyer et al. 2447* (ILLS); Washakie County, foothills of Bighorn Mountains, in Tensleep Preserve, union of Canyon & Cooks Creeks, 18 June 2004, *L. R. Phillippe et al. 36913* (ILLS); Washakie County , foothills of Bighorn Mountains, in Tensleep Preserve, union of Canyon & Cooks Creeks, 24 August 2005, *L. R. Phillippe et al. 37967* (ILLS).

Notes — Nearly all leaves on the plant are palmately lobed with 5 or more deeply cleft lobes (Fig. 1A). Leaf dissections nearest the midrib often surpass the midpoint of the leaf. Smaller leaves, less than 5 cm in length, sometimes only possess 3 lobes, however more than three veins branching from the midrib are evident in these cases. Few if any unlobed cordate leaves are present on plants. Abaxial surface of leaf midrib comprises at least 25 closely appressed trichomes per linear cm. Glands between veins on the abaxial surface occur at densities of 30 per cm^2 or often greater (Fig. 2.1D); other plant parts also possess a dense covering of glands.

2. Humulus pubescens (E. Small) Tembrock, stat. nov. Syst. Bot. 3:37–76 1978. (Figs. 2.2 A – 2.2 D & 2.5 B) Type: U.S.A. Missouri: Buckner, Low woods, 27 Sept 1912, B.F. Bush 6956 (Holotype GH, photograph!, isotypes MO, NY! and US!).

Perennials, without tendrils, rhizomatous and substantially taprooted. Stems sometimes > 5 m long, variously pubescent, with white to silver trichomes that sharply increase in density around the nodes and form a transverse line of tomentum subtending, and within the axil, of the

interpetiolar stipules; interpetiolar stipules ± deciduous, splitting along central veins with age, 0.5–1.5 cm long, apex acute or rarely acuminate, with parallel venation, pubescent. Leaves petiolate; petioles 3–10 cm long, pubescent with both pilose hairs as well as prominent cystolithic, two–forked hairs (especially abundant on upper half of petiole; Fig. 2.2D), the latter often linearly arranged, 5–20 per linear cm, < 2 mm long, opaque and brown to dark green below, transparent to translucent above; laminae unlobed and cordate or mostly 3– (4–) lobed, the central sinuses 3–7 cm deep, dark green with shallow, lighter green ridges, 6–18 cm long and 6–20 cm wide at widest point, bases evidently cordate, the margins dentate with 2–4 mm long teeth, apices acuminate, the abaxial surface evidently pubescent between the 3 prominently raised primary veins, the midrib with > 100 white to silvery spreading hairs per linear cm, secondary and tertiary veins terminating into acuminate, marginal denticulations. Staminate inflorescences axillary or rarely terminal, cymose panicles, 10–25 cm long, with 50–100+ flowers subtended by brown bracts 2–10 mm long, often present at the point of secondary panicle branching; pedicels 1–6 mm long; perianth comprised of a single whorl of 5 distinct lobes; lobes elliptic to ovate, the margins ciliate, 2–4 mm long, apices rounded to acuminate, light green to yellowish, becoming lighter near the margins, pubescent; anthers basifixed, with a distinct line of yellow glands along the abaxial surface of the connective, dehiscing along longitudinal sutures. Pistillate inflorescences axillary, short–pedicellate racemes, 1–2 cm long, bur–like at anthesis, with 10–50 paired flowers subtended by yellow–green bracts; styles deciduous, 5–8 mm long. Infructescences 1–5 cm long, with closely appressed bracts subtending the fruits; bracts ovate to elliptic, 1–2 cm long and approximately half as wide, apices rounded to acuminate, browning with age, loosely enveloping the achene at the base and forming a wing about the fruit; achenes ovoid, terete to subterete, 3–5 mm long, cream to brown.

Phenology —Flowering and fruiting from mid–August to early October.

Distribution and habitat — Mainly occurring in the central Midwestern U.S.A. (Fig. 2.5 B). The range of *H. pubescens* overlaps with that of *H. neomexicanus* in its western–most distribution and with that of *H. lupuloides* throughout the northern portion of its range.

Found growing on vegetation in moist lowland and riparian forest thickets, frequently co–occurring with *Sambucus nigra* L., *Viburnum lentago* L., and species of *Celtis*, *Clematis* L., *Fraxinus* L., *Salix*, and *Quercus* L., 0–1000 m.a.s.l.

Representative specimens examined — U.S.A. Illinois: Crawford County, 0.3 miles S of Illinois Rte. 33, *S. R. Hill 27588* (ILLS); Jo Daviess County, Floodplain of Mississippi River at the Savanna Army Depot, 23 August 1996, *L. R. Phillippe et al. 28124* (ILLS); Mason County, 2.5 miles W of Easton in Tomlin Timber Nature Preserve, 14 August 2003, *L. R. Phillippe 36019* (ILLS); Vermilion County, at Collison Seeps in Middle Fork Fish & Wildlife Area, 16 July 1991, *M. J. Morris et al. 892* (ILLS); Woodford County, 2 miles S of Spring Bay & 0.75 miles W of Illinois Rte. 26 in Spring Bay Fen Nature Preserve, 6 September 2006, *L. R. Phillipe & J. E. Ebinger 39174* (ILLS); Iowa: Adair County, NW corner of Pioneer Cemetery, 6 September 1999, *M. J. Leoschke & C. Kern 2024* (IA); Davis County, 1 mile W of Floris on the Hill Culture Exp. Farm, 1 October 1938, *A. Hayden, 9151* (IA); Hamilton County, public hunting area on Big Bear Rd., 1 September 2001, *J. D. Thompson 431052* (IA) Jefferson County, Center township, sec 25, 7 September 1935, *C. L. Gilly* 2887 (IA); Johnson County, Newport township, sec. 14, near old cemetery, 10 September 1956, *H. W. Pfeifer 64123* (IA); Lucas County, English township, SW ¼ of sec. 19, 4 September 1987, *J. Probasco 119* (IA); Page County, Wabash

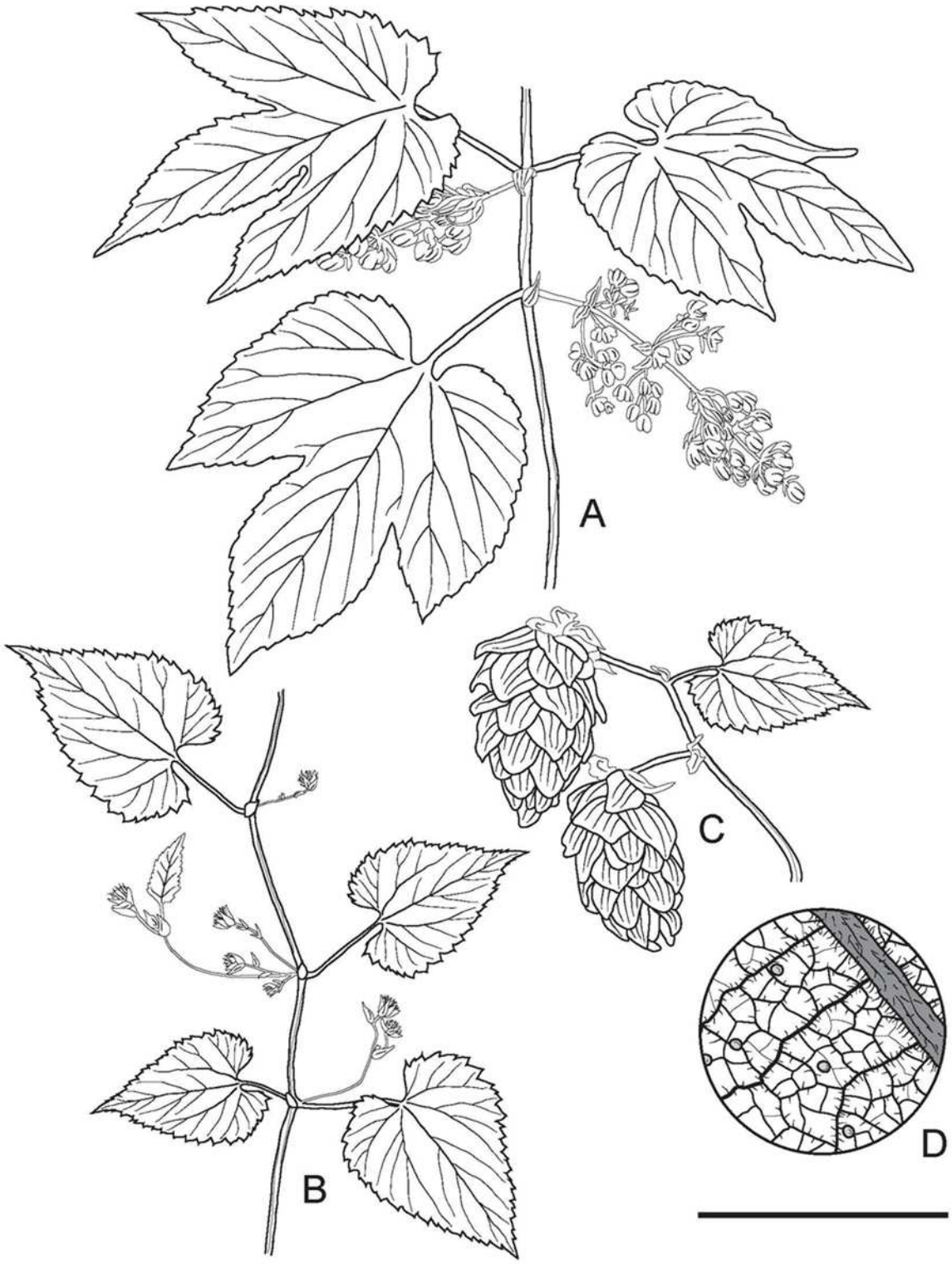


Figure 2.4. *Humulus pubescens* A. Stem and leaves with male inflorescence B. Female inflorescences in early antithesis C. Female infructescence D. Detail of abaxial leaf surface and central vein. Scale bar: A & B = 6 cm, C = 3 cm, D = 0.5 cm.

Trace Nature Trail, Grant Township, sec 21, 17 August 1989, *B. Wilson & J. Maher 3458* (IA); Palo Alto County, Highland township, sec 5, 30 September 1940, *A. Hayden 7358* (IA), Story County, moist ground near Ames, Franklin township, sec. 26, August 1907, *A. Hayden 2227 & 2228* (IA); Washington County, 4 miles W of Wellman, 15 September 1953, *B. L. Wagenknecht 1233* (IA); Winneshlek County, low roadside bordering Canoe Creek, 16 August 1959, *T. G. Hartley 8371* (IA); Indiana: Boone County, 10.4 miles NW of Royalton, 11 September 1943, *R. C. Friesner 17948* (BUT); Hancock County, 0.8 miles E of Maxwell, 5 August 1944, *R. C. Friesner 18368* (BUT); Parke County, 2 miles SW of Grange Corner, 16 August 1952, *F. B. Buser 1499* (BUT); Tippecanoe County, near state Rd. 5–6 miles N of Lafayette, 16 September 1941, *C. M. Ek 60123* (BUT); Michigan: Washtenaw County, October 1924, *S. H. Evenson 016767* (MONT); Missouri: Christian County, 1 August 1895, *J.W. Blankinship 012000* (MONT); Livingston County, fencerow 3.8 miles N & 7 miles W of Chillicothe, 3 September 1950, *S. Sparling 527* (IA); Livingston County, fencerow 4 miles N & 1.7 miles W of Chillicothe, 14 August 1950, *S. Sparling 433* & *438* (IA); Nebraska: Cass County, along railroad tracks near Cedar Creek, 16 July 1934, *J. L. Morrison 1244* (NEB); Richardson County, on Rulo–White Cloud Road, T 1 N, R 18 E, sec 26, 12 September 1974, *P. Shildneck C–7114* (ILLS); Richardson County, on west shoulder of Rulo–White Cloud Rd, 12 September 1974, *P. Shildneck C–7094* (KANU); Seward County, W of Branched Oak Lake, 4 September 1986, *R. B. Kaul 5510* (NEB); Stanton County, Wood Duck Wildlife Management Area, 3 August 1999, *R. F. Steinauer 630* (NEB); Washington County, 8 miles N of Arlington, in woods along Bell Creek, 8 October 1930, *A. L. Lallman 170842* (NEB); Ohio: Ashland County, 0.75 miles S of Mifflin near Rd., 15 September 1979, *G. T. Jones 80667* (MONT); Pennsylvania: mid–1800's, *A. Chapman* (US); Virginia: 1877, *L. F. Ward* (US); Fauquier County, Western slope of Bull run

Mountains, ½ mile north of Thoroughfare Gap, 10 September 1939, *H. A. Allard 7384* (US); Western slope of Bull run Mountains, thickets along railroad at Beverly Mill, 23 August 1942, *H. A. Allard 10351* (US); Loudoun County, vicinity of Purcellville on fence near creek, 10 August 1960, *E. J. Hambleton 209* (US); vicinity of Purcelville on fence near creek, June 1955, *E. J. Hambleton 305* (US).

Notes —Plants most often with a mixture of lobed and unlobed leaves, lobed leaves are palmately lobed with 3 or fewer lobes (Figs. 2.2 A, B). The dissections of the lobed leaves often do not surpass the midpoint of the leaf. Unlobed leaves possess no more than 3 veins branching from the midrib. Abaxial surface of leaf midrib with more than 100 trichomes per linear cm, these trichomes mostly spreading and silvery in appearance (Fig. 2.2 D). Leaves pubescent between leaf veins, glands present, most parts of the plant strongly pubescent.

3. Humulus lupuloides (E. Small) Tembrock stat. nov. Syst. Bot. 3:37–76 1978. (Figs. 2.3A–D, 2.5 C) Type: U.S.A. Maine, Aroostook County, along the St. John River, alluvial thicket, Allaguash Plantation, 11 Aug 1893, M.L. Fernald 96 (Holotype PH photograph!, Isotypes F, MO, NY!, UC, US!).

Humulus americanus Nutall Proc. Acad. Nat. Sci. Philadelphia 4 (Mar.–Apr.): 23 1848. Type: *W. Gambel, 0927462*, (Lectotype PH photograph!); lectotype designated by E. Small (1978).

Humulus lupulus subsp. *americanus* (Nutt.) Á. Löve & D. Löve Taxon 31: 121 1982.

Perennials, without tendrils, rhizomatous and substantially taprooted, with aerial branches sometimes arising from rhizomes. Stems sometimes > 5 m long, variously pubescent, with transparent to silver trichomes that increase in density around the nodes and form a transverse line of tomentum subtending, and within the axil of the interpetiolar stipules; interpetiolar stipules ± deciduous, splitting along central veins with age, 0.5–1.5 cm long, apices acute or

rarely acuminate, with parallel venation, pubescent. Leaves petiolate; petioles 2–8 cm long, pubescent with both pilose hairs as well as prominent cystolithic, two–forked hairs, the latter often linearly arranged, rarely exceeding 10 hairs per linear cm, < 1 mm long, opaque and brown to dark green below, transparent to translucent above; laminae unlobed and cordate or 3– (4–) lobed, the central sinuses 3–6 cm deep, dark green with shallow, lighter green ridges, 5–16 cm long and 5–18 cm wide at widest point, bases deeply cordate, the margins dentate with 2–4 mm long teeth, apices acuminate, the abaxial surface sparsely pubescent or glabrous between the 3 evidently raised primary veins, the midrib with 20–75 mostly appressed hairs per linear cm, secondary and tertiary veins glabrous and terminating into acuminate, marginal denticulations. Staminate inflorescences axillary or rarely terminal, cymose panicles, 10–20 cm long, with 50–100+ flowers subtended by brown bracts 2–10 mm long, often present at the point of secondary panicle branching; pedicels 1–6 mm long; perianth comprised of a single whorl of 5 distinct lobes; lobes elliptic to ovate, the margins ciliate, 2–4 mm long, apices rounded to acute, light green to yellowish, becoming lighter near the margins, weakly pubescent; anthers basifixed, with a distinct line of yellow glands along the abaxial surface of the connective, dehiscing along longitudinal sutures. Pistillate inflorescences axillary, short–pedicellate racemes, 1–2 cm long, bur–like at anthesis, with 10–50 paired flowers subtended by green to yellow bracts; styles deciduous, 5–8 mm long. Infructescences 1–5 cm long, with closely appressed bracts subtending the fruits; bracts ovate to elliptic, 0.7–2 cm long, apices rounded to acuminate, browning with age, loosely enveloping the achene at the base and forming a wing about the fruit; achenes ovoid, terete to subterete, 3–5 mm long, cream to brown.

Phenology — Flowering and fruiting from the end of July through September.

Distribution and habitat — Mainly found in the Great Plains of the northern U.S.A. and southern Canada with some occurrences in the northeastern U.S.A. and southeastern Canada (Fig. 2.5C).

Found growing on vegetation in moist lowland and riparian forest thickets, often co–occurring with *Ulmus americana* L., *Vitis riparia* Michx. and species of *Acer* L., *Celtis*, and *Salix*, 0–1000 m.a.s.l.

Representative specimens examined — Canada. Manitoba: Over a ruined building in The Pas, 26 August 1960, *W. Krivda 2232* (NY); Newfoundland: Osmund, mouth of little barachois brook, 1 September 1961, *E. Rouleau 7370* (US); Nova Scotia: Pictou County, East river, 25 August 1906, *C. B. Robinson 516* (NY); Frontenac County, Barrie Township, Concession 15, Mazinaw Lake, Bon Echo, in open woods, 12 September 1952, *J. M. Gillett 7230* (NY); Saskatchewan: District D' Assinibola, 10 miles south of Gainsborough, in forest, 5 September 1960, *B. Boivin 14018* (NY); U.S.A. Connecticut: The vicinity of Green's farm, 16 August 1894, *C. L. Pollard 228* (US); Illinois: Fayette County, Dean Hills Nature Preserve, in floodplain of Beck's Creek, 12 October 2000, *M. A. Feist 789* (ILLS); Marshall County, T 29 N, R 3 W, sec 35, along Cow Creek, 23 September 2000, *M. Brucker 78* (ILLS); Vermillion County, 1.3 miles NE of Collison near Kinney Ford Seep, 16 August 1991, *M. J. Morris et al. 917* (ILLS); Indiana: Cass County, along railroad near Galveston, 16 August 1942, *C. M. Ek 65060* (BUT); Elkhart County, along railroad west of Elkhart, 2 September 1942, *C. M. Ek 63867* (BUT); Howard County, 2 miles NW of Kokomo, 12 August 1936, *C. M. Ek 49369* (BUT); Huntington County, 8 September 1941, *C. M. Ek 57640* (BUT); Huntington County, 1 mile N & 1 mile E of Monument City, 18 September 1948, *R. C. Friesner 22510* (BUT); Marion County, 0.3 miles S of 59th St Indianapolis, 10 September 1943, *R. C. Friesner 18131* (BUT);

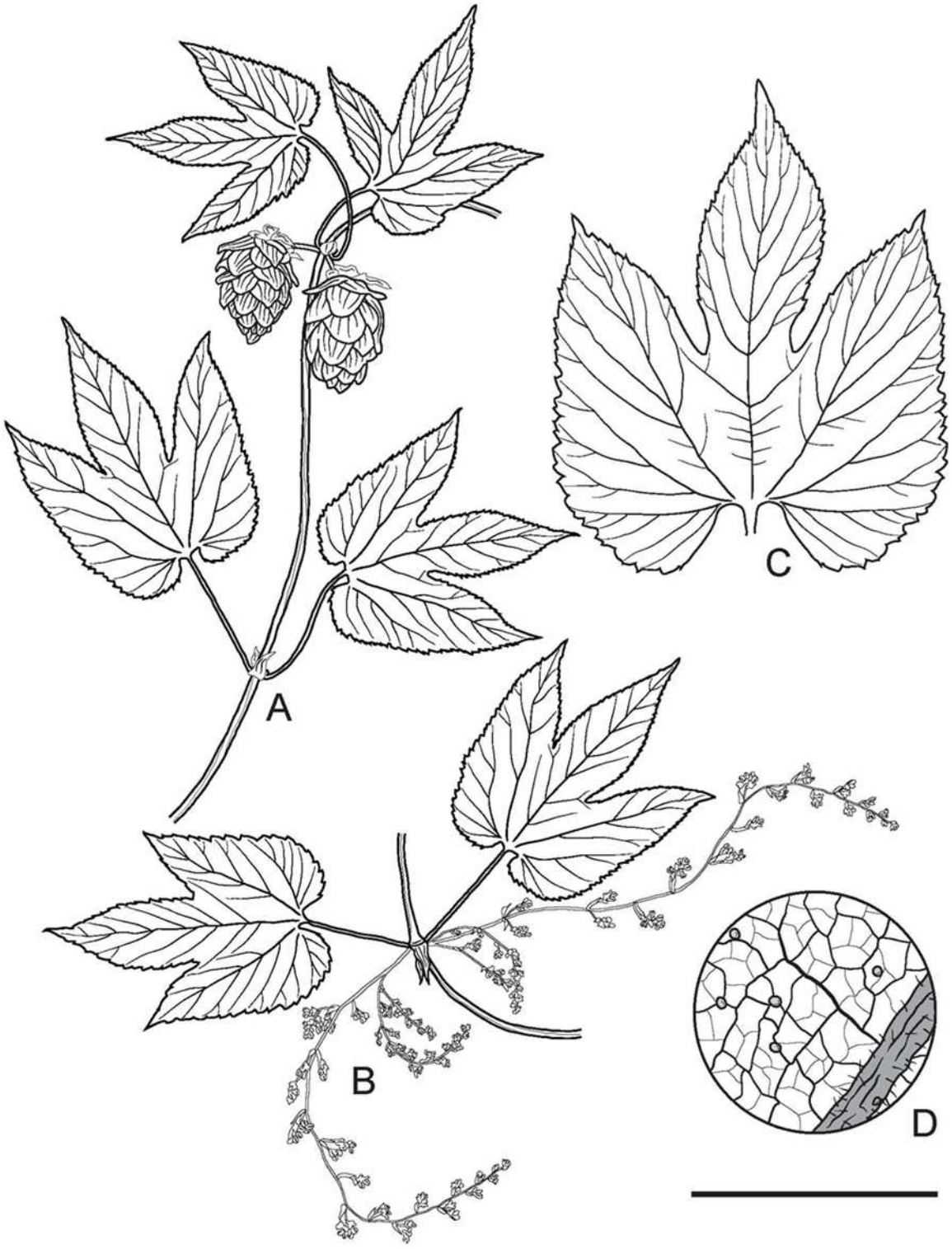


Figure 2.5. *Humulus lupuloides* A. Stem and leaves with female inflorescence B. Male inflorescences C. Typical leaf morphology (similar to *H. pubescens*) D. Detail of abaxial leaf surface and central vein. Scale bar: A & B = 3 cm, C = 5 cm, D = 0.5 cm.

Porter County, ditch along Hwy 49, 10 October 1987, *S. Heinemann 541* (BUT); Randolph County, 4 miles S of Farmland, 29 July 1944, *C. C. Deam 63030* (BUT); **Iowa**: Hamilton County, 100 yards S of the entrance to Boone Forks Public Hunting Area, 22 September 2001, *J. D. Thompson 439284* (IA); Hamilton County, Nesse Pioneer Cemetery, 2 August 2001, *J. D. Thompson 431228* (IA); Jones County, Jackson township, sec 16, NW ¼, 1948, *R. G. Brown 324* (IA); Lyon County, Sioux township, sec 18, 12 August 1954, *R. F. Thorne 14684* (IA); Monroe County, 5 miles SW of Albia, 27 September 1998, *M. J. Leoschke 1905* (IA); Sioux County, on Rock River, 21 August 1956, *J. L. Carter 1979* (IA); Maine: Aroostook County, Alluvial island, Seven Islands, Township xiii, Ranges 14 and 15, 24 July 1917, *H. St. John & G. Nichols 2277* (US); Alluvial thicket, Allaguash plantation, 11 Aug 1893, *M.L. Fernald 96* (NY); Massachusetts: Nantucket County, beyond outskirts of town, 11 August 1906, *E. P. Bicknell 3675* (NY); Maryland: Baltimore County, banks of Jones River, near Baltimore, apparently indigenous, 14 August 1873, *Morong* (NY); Michigan: Houghton County, on Otter Creek near Otter Lake, 6 September 1965, *E. A. Bourdo, Jr. 12256* (NEB); Kent County, NE corner of the intersection between Rusco St and Fruit Ridge Ave, 26 August 1999, *E. M. Smith 4413* (KANU); Wayne County, Chocolate river, Lake Superior, 13 August 1867, *H. Gillman* (NY); Minnesota: Clearwater County, Tannist camp, Itasca state park, 23 August 1929, *M. L. Grant 3134* (NY); Nebraska: Dawes County, 3.2 miles S & 4 miles E of Chadron on Bordeau Rd., 15 September 1977, *P. Theobald 107* (CSCN); Keya Paha County, on Turkey Creek 4.8 km E of Norden, 7 July 1983, *S. P. Churchill* & *J. Suttie 12465* (NEB); Madison County, Highland precinct, 31 August 1986, *J. J. Duhachek 168* (CSCN); New Hampshire: Grafton County, Sugar Hill, August 1888, *A. M. Vail* (NY); New York: Montgomery County, Amsterdam, 2 August 1864, *W. H. Leggett* (NY); Richmond County, Pleasant Plains, 14 September 1915, *A. H. & N. L. D.* (NY);

Sullivan County, Cochecton, August 1888, *N. L. Britton* (NY); Westchester County, White Plains, 1862, *O. R. Willis* (NY); Ohio: Cuyahoga County, Cleveland, *Alphonso Wood* (NY); Pennsylvania: Chester County, Banks of Brandywine Creek, *W. M. Canby* (NY); Lancaster County, In the vicinity of Smithville, 12 September 1891, *J. K. Small* (US); Susquehanna County, In rocky thickets at bridge over east branch of Lackawanna River, 13 July 1950, *S. L. Glowenke 11411* (NY); Washington County, Chartiers Creek bottom, 20 August 1859, *W. Brewer II* (US); South Dakota: Perkins County, T 13N, R 13W, sec 11, along Deep Creek, 6 September 1992, *J. McTighe 53* (BHSC); Lawrence County, 15 miles SW of Spearfish, near Roughlock Falls, 22 July 1966, *S. Stephens 7549* (KANU); Lawrence County, in Spearfish along fence, 10 September 1933, *F.L.B. 328* (BHSC); Vermont: Southern Vermont, 19 August 1898, *H. E. Day* (US); Virginia: Fauquier County, Western slope of Bull Run Mountains, scrambling over alders at bridge, 1 mile north of Hopewell gap, 29 August 1937, *H. A. Allard 3629* (US); Western slope of Bull Run Mountains, Thicket along railroad above Beverly Mill, 30 August 1937, *H. A. Allard 10365* (US); Wisconsin: Juneau County, Camp Douglas, 14 July 1890, *E. A. Mearns 605* (US); Manitowoc County, Two Rivers, 1 October 1866, *H. Gillman* (NY).

Notes — Plants often with a mixture of lobed and unlobed leaves, lobed leaves are palmately lobed with three or fewer lobes (Figs. 2.3A,B), very rarely leaves with four lobes are found. Abaxial surface of leaf midrib with 20 – 75 mostly appressed trichomes per linear cm. Pubescence lacking between primary leaf veins on the abaxial surface of the leaf, glands present but less dense than on *H. neomexicanus*. *Humulus lupuloides* shares both morphological and AFLP (Reeves and Richards 2011) character states of *H. pubescens* and *H. neomexicanus* but the character states are never found in the combinations present in *H.pubescens* or *H. neomexicanus*. Specifically, *H. lupuloides* is diagnosable from the combined absence of five lobed leaves and

pubescence between veins on the abaxial surface of the leaves. We found three specimens that we suspect might be hybrids between var. *lupulus* and var. *lupuloides* given they did not completelyadhere to our description of leaf midrib hair number for var. *lupuloides.* This finding is not surprising given the over 400 years of commercial hops production in the native range of var. *lupuloides* providing opportunities for chance hybridization (Small 1978). The putative hybrids are: Canada. Manitoba: Melita, along Graham Creek in thicket, 15 August 1950, *W. G. Dore & A. J. Breitung 12636A* U.S.A. Minnesota: St Louis County, Back fence of a cottage, Sec 19, uncommon, Duluth, 12 August 1936, *O.Lakela 1702* (NY); New York St. Lawrence County, roadside, 20 August 1914, *O. P. Phelps 383* (US).

REFERENCES

Abbott JR, Judd WS (2011) *Badiera subrhombifolia* (Polygalaceae), a new species from Hispaniola. Brittonia 63:161–170.

Bassil NV, Gilmore B, Oliphant JM, Hummer KE Hennig JA (2008). Genic SSRs for European and North American hops (*Humulus lupulus* L.). Genetic Resources and Crop Evolution 55:959–969.

Braby MF, Eastwood R, Murray N (2012) The subspecies concept in butterflies: has its application in taxonomy and conservation biology outlived its usefulness?. Biological Journal of the Linnean Society 106: 696–716.

Collinson ME (1989) The fossil history of the Moracaeae, Urticaceae (including Cecropiaceae), and Cannabaceae. In Crane, P.R. & Blackmore, S. (eds.) Evolution, systematics and fossil history of the hamamelidae, vol 2: 'higher' Hamamelidae. Clarendon Press, Oxford.

Danilova TV & Hennig JH (2000). Sex molecular markers reveal differences of Y chromosome molecular structure among hops of European and American origin. In: Hummer KE & Hennig, JH (eds.) Proceedings of the 1st International Symposium on Humulus, Acta Horticulturae 668: 85–92.

Davis EL (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 Queiroz K (2007) Species concepts and species delimitation. Systematic Biology 56: 879–886.

de Queiroz K (2005) Different species problems and their resolution. BioEssays 27: 1263–1269.

de Queiroz K (1998) The general lineage concept of species, species criteria, and the process of speciation: a conceptual unification and terminological recommendations. In: Howard, D. J. & Berlocher S. H. (eds.) Endless forms: species and speciation. Oxford University Press, New York, pp. 57–75.

de Queiroz K & Donoghue MJ (1988) Phylogenetic systematics and the species problem. Cladistics 4:317–338.

De Vries DP (1655) Voyages from Holland to America, A.D. 1632 to 1644, translated by Murphy HC (1853) James Lenox, New York, pp. 157.

Ellison AM, Davis CC, Calie PJ, Naczi, RFC (2014) Pitcher plants (Sarracenia) provide a 21st –century perspective on infraspecific ranks and interspecific hybrids: a modest proposal for appropriate recognition and usage. Systematic Biology 39: 939–949.

Gleason HA, & Cronquist A (1991) Manual of vascular plants of northeastern United States and adjacent Canada. The New York Botanical Garden, New York, 73–74 pp.

Grabowska–Joachimiak A, Sliwinska E, Piguła M, Skomra U, Joachimiak A (2006) Genome size in *Humulus lupulus* L. and *H. Japonicus* Siebold & Zucc. (Cannabaceae). Acta Societatis Botanicorum Polonae 3:207–214.

Gray A (1857) Manual of botany of the Northern United States revised edition; including Virginia, Kentucky, and all east of the Mississippi. G. P. Putnam & Co., New York, pp. 400.

Haunold A (1991) Cytology and cytogenetics of hops. In: Tsuchiya, T. & Gupta, P. K. (eds.) Chromosome engineering in plants: genetics, breeding, evolution. Elsevier, New York, pp. 551–563.

Howard EL, Whittock SP, Jakše J, Carling J, Matthews PD, Probasco G, Hennig JA, Darby P, Cerenak A, Javornik B, Kilian A, Koutoulis A (2011) High–throughput genotyping of hop (*Humulus lupulus* L.) utilizing diversity arrays technology (DArT). Theoretical and Applied Genetics 122: 1265–1280.

Hummer KE, Oliphant JM, Taylor AW, Deinzer ML, Smith JM (2005) Xanthogalenol and 4'–O–methylxanthohumol content in some American *Humulus.* In: Hummer, K. E. & Hennig, J. A. (eds.) Proceedings of the 1st International Symposium on Humulus, Acta Horticulturae 668: 229–232.

International Plant Names Index (IPNI) 2011. Published on the Internet http://www.ipni.org [accessed 6 December 2011].

Jakše J, Stajner N, Luthar Z, Jeltsch JM, Javornik B (2011) Development of transcript associated microsatellite markers for diversity and linkage mapping studies in hop (*Humulus lupulus* L.). Molecular Breeding 28: 227–239.

Judd WS, Campbell CS, Kellogg EA, Stevens PF, Donoghue MJ (2008) Plant systematics: a phylogenetic approach, third edition. Sinauer associates, Sunderland, MA, U. S. A., pp. 558–560.

Kubitzki K (1993) Cannabaceae. In: Kubitzki K, Rohwer JG, Bittrich V (eds.) The families and genera of vascular plants. II. Flowering plants: dicotyledons, magnoliid, hamamelid and caryophyllid families, eds. Kubitzki, K. Rohwer, J. G. & Bittrich, V. Springer, Berlin pp. 204–205.

Light JE, Toups MA, Reed DL (2008) What's in a name: the taxonomic status of human head and body lice. Molecular Phylogenetics and Evolution 47:1203–1216.

MacGinitie HD (1969) The Eocene Green River flora of northwestern Colorado and

northeastern Utah. University of California Publications in Geological Sciences 108:1–103.

Mallet J (1995) A species definition for the modern synthesis. Trends in Ecology and Evolution 10: 294–299.

Marshall JCE, Arévalo E, Benavides E, Sites JL, Sites Jr JW (2006) Delimiting species: comparing methods for Mendelian characters using lizards of the *Sceloporus grammicus* (Squamata: Phrynosomatidae) complex. Evolution 60: 1050–1065.

Mayr E (1942) Systematics and the origin of species. Columbia University Press, New York, pp. 334.

Meyer HW (2003) The Fossils of Florissant. Smithsonian Books, Washington. 69–70 pp.

Murakami A, Darby P, Javornik B, Pais MSS, Seigner E, Lutz A, Svoboda P (2006) Microsatellite DNA analysis of wild hops, *Humulus lupulus* L. Genetic Resources and Crop Evolution 53: 1553–1562.

Nelson A, Cockerell TDA (1903) Three new plants from New Mexico. Proceedings of the Biological Society of Washington16: 45–46.

Neve RA (1991) Hops. Chapman and Hall, London, pp. 10–25.

Neve RA (1965) Plant breeding section: The year's work. dept. of hop research annual report, Wye College, Univ. of London.

Nixon KC, Wheeler QD (1990) An amplification of the phylogenetic species concept. Cladistics 6:211–223.

Nuttall T (1848) Descriptions of plants collected by Mr. William Gambel in the Rocky Mountains and upper California. Proceedings of the Academy of Natural Sciences of Philadelphia 4: 7–26.

Patzak J, Nesvadba V, Krofta K, Henychova A, Marzoev AI, Richards K (2010) 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.

Reeves PA, Richards CM (2009) Accurate inference of subtle population structure (and other genetic discontinuities) using principal coordinates. PLoS ONE. 4:e4269.

Reeves PA, Richards CM (2011) Species delimitation under the general lineage concept: an empirical example using wild North American hops (Cannabaceae: *Humulus lupulus*). Systematic Biology 60: 45–59.

Rydberg PA (1917) Flora of the Rocky Mountains and adjacent plains, Colorado, Utah, Wyoming, Idaho, Montana, Saskatchewan, Alberta, and neighboring parts of Nebraska, South Dakota, and British Colombia. Published by the Author, New York.

Salmon ES (1934) Two new hops: 'Brewer's Favourite' and 'Brewer's Gold.' Journal of Southeast Agricultural College of Wye, Kent 34: 93–106.

Salmon ES (1938) 'Bullion Hop'; A new variety. Journal of Southeast Agricultural College of Wye, Kent 42: 47–52.

Shepard HL, Parker JS, Darby P, Ainsworth CC (2000) Sexual development and sex chromosomes in hop. New Phytologist 148: 397–411.

Small E (1978) A numerical and nomenclatural analysis of morphogeographic taxa of *Humulus*. Systematic Botany 3: 37–76.

Small E (1980) The relationship of hop cultivars and wild variants of *Humulus lupulus*. Canadian Journal of Botany 58: 676–686.

Small E (1997) *Humulus*. In: Flora of North America editorial committee, (eds). Flora of North America North of Mexico. New York and Oxford. vol 3: 384–387.

Smith JM, Oliphant JM, Hummer KE (2006) Plant exploration for native hop in the American southwest. Plant Genetic Resources Newsletter 147: 1–9.

Stajner N, Satovic Z, Cerenak A, Javornik B (2008) Genetic structure and differentiation in hop (*Humulus lupulus* L.) as inferred from microsatellites. Euphytica 161: 301–311.

Stevens JF, Taylor AW, Nickerson GB, Ivancic M, Henning J, Haunold A, Deizner ML (2000) Prenylflavonoid variation in *Humulus lupulus*: distribution and taxonomic significance of xanthogalenol and 4'–O–methylxanthohumol. Phytochemistry 53: 759–775.

Tomlan ML (1992) Tinged with gold: hop culture in the United States. The University of Georgia press, Athens pp. 9–41.

Van Valen L (1976) Ecological species, multispecies, and oaks. Taxon 25: 233–239.

Van Valkenburg D (1995) A question of pedigree–the role of genealogy in hop substitutions. Brewing Techniques 3: 54–59.

Wilson DG (1975) Plant remains from the Graveney Boat and the early history of *Humulus lupulus* L. in W. Europe. New Phytologist 75: 627–648.

Zanoli P, Zavatti M (2008) Pharmacognostic and pharmacological profile of *Humulus lupulus* L. Journal of Ethnopharmacology 116: 383–396.

Zimmermann C, Likens S, Haunold A, Horner C, Roberts C (1975) Registration of comet hop. Crop Science 15: 98.

CHAPTER 3: EMPLOYING GC-MS AND TWO-STAGE DERIVATIZATION TO ASSAY FOR CATHINE AND RELATED STIMULANT ALKALOIDS ACROSS THE CELASTRACEAE

Summary

Catha edulis (qat, khat, mirra) is a woody plant species that is grown and consumed in East Africa and the southern Arabian Peninsula for its stimulant alkaloids. The alkaloids responsible for the stimulant properties are cathinone, cathine, and norephedrine. These alkaloids are structurally and pharmacologically similar to amphetamines. Two Celastraceae species, in addition to qat, have been cited as stimulants in ethnobotanical literature. Leaves of all three of these species are chewed or made into a tea to alleviate fatigue and hunger. Recent phylogenetic reconstructions place the genera *Allocassine*, *Cassine*, *Lauridia*, and *Maurocenia* in a clade sister to *Catha edulis* and are thus primary candidates for possession of cathine and related alkaloids. Any species containing cathine and/or cathinone is of concern to drug–regulation authorities given the controlled status of cathine and cathinone in numerous countries. Here, leaf tissue from 43 *Celastraceae* species was extracted using an aqueous–alkaline solvent followed by an organic extraction in MBTE (methyl tert-butyl ether). The extract was derivatized in a two–stage process to provide accurate detection and quantification of trace levels of cathine and related alkaloids using gas chromatography coupled with mass spectrometry. Cathinone, cathine, and norephedrine were not detected in any of the 43 Celastraceae species assayed other than *Catha edulis*. However, we did identify the phenylalanine- or tyrosine-derived alkaloid phenylethylamine in *Brexia madagascariensis*, *Crossopetalum rhacoma*, *Lauridia tetragona*, and *Polycardia liberia*. An unknown compound with a similar profile was also found in

Crossopetalum rhacoma, Gymnosporia senegalensis, Lauridia tetragona, Polycardia liberia. Lastly nine species, including *Cassine schinoides,* were found to be enriched for numerous sterol and terpene-like compounds that in some instances (e.g., THC and Salvinorin A) can have psychotropic action.

Introduction

The plant family Celastraceae contains many phytochemicals with bioactivity including insecticides, anti–cancer agents, antibacterial compounds, anti–diabetic drugs, and central nervous system stimulants (e.g., Brünning & Wagner 1978; Yoshikawa et al. 1997; Spivey et al. 2002; López et al. 2011. *Catha edulis* is the only species in Celastraceae that is cultivated specifically for its stimulant alkaloids. Qat (as it is commonly known and correctly transliterated from Arabic; Varisco 2004) is mainly cultivated and/or collected from the wild and consumed in Ethiopia, Israel, Kenya, Somalia, Tanzania, Uganda, and Yemen. Qat is generally consumed by chewing the fresh young leaves and shoots followed by ingestion of the liquid constituents. The young shoot tips of qat contain the following stimulant alkaloids: (S)-cathinone (Fig. 3.1A), hereafter cathinone (United Nations Narcotics Laboratory 1975; Schorno & Steinegger 1979; Szendrei 1980) (1R,2S)-norephedrine, hereafter norephedrine; and (1S,2S)-norpseudoephedrine, hereafter cathine (Fig. 3.1B; Wolfes 1930). Cathinone is thought to be anywhere from 20% to equipotent in central–nervous–system–stimulant action to that of amphetamine (Mereu et al. 1983; Glennon 1986), while cathine is eight times weaker in pharmacological response than cathinone (Glennon et al. 1984). Cathinone is a schedule I controlled substance in the United States (U.S.) and cathine is schedule IV (Department of Justice 1988; 1993). The exportation of qat from East Africa to Europe and the U.S. has been on the increase. For instance, U.S. qat seizures in 1994 amounted to approximately 31,000 pounds, increasing to over 197,000 pounds

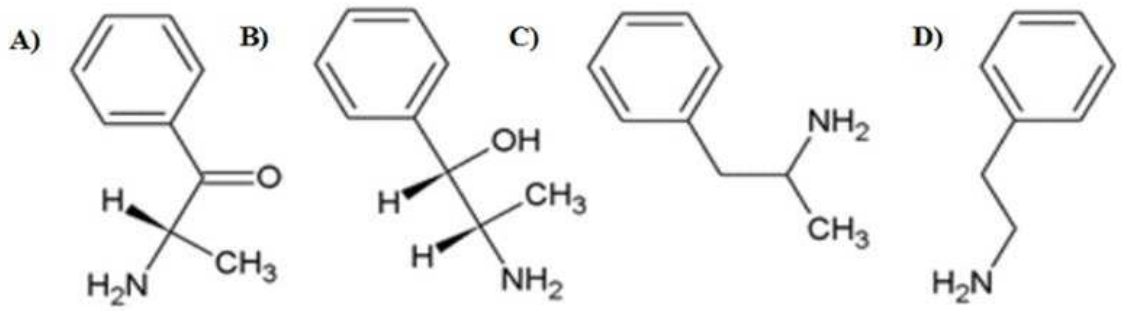


Fig 3.1**.** Molecular structures of A) cathinone, B) cathine, C) amphetamine and D) phenylethylamine.

by 2010 (Drug Enforcement Administration 2010).

The structural similarity of cathinone and cathine to amphetamine (Fig. 3.1 A–C) results in similar actions to the human central nervous system as amphetamine (Patel 2000; Hoffman & Al'Absi 2010). Physical reactions to qat ingestion include euphoria, increased motor stimulation, excitability, and increased energy (Wilder et al. 1994; Kalix 1996: Nencini et al. 1998). These effects of qat are attributed to cathinone's action on dopamine and serotonin receptors in the brain (Glennon & Liebowitz 1982; Babayan et al 1982; Kalix 1984; 1988). Cathinone is primarily concentrated in the young leaves and inflorescences of qat (Krizevski et al. 2007). As shoots develop, the cathinone is converted enzymatically into cathine, which is found in relatively high levels in mature leaves and stems (Krizevski et al. 2007). A similar pattern of alkaloid biosynthesis correlated with developmental stage of plant parts has been found in *Ephedra sinica* (ma huang) (Krizevski et al. 2010), *Mentha* x *piperita* (peppermint) (Davis et al. 2005), and *Papaver somniferum* (opium poppy) (Ziegler et al. 2006). Due to the rapid conversion of cathinone into the less potent cathine after harvest, qat is treated as a perishable commodity and is usually consumed within two –three days after harvest, with the highest value being given to the freshest leaves (Altabachew et al. 2013).

The only other species from which cathinone and cathine have been detected are *Ephedra gerardiana* and *E. sinica* (Krizevski et al. 2010; Grue-Sørensen & Spenser 1994). The presence of cathinone and cathine in the genus *Ephedra* is a clear demonstration of convergent evolution given that *Ephedra* is a gymnosperm rather than an angiosperm (Stevens 2012). In these *Ephedra* species the pathway has further steps in which cathine and norephedrine are N–methylated, yielding pseudoephedrine and ephedrine (Krizevski et al. 2010). A similar pathway is expected to occur in *Sida cordifolia* and *Pinellia ternata* given that ephedrine has been detected in these species (Oshio et al. 1978; Shoyama et al. 1983). The presence of cathine in relatively high levels in qat is likely due to the lack of N-methylase enzymes, resulting in the deposition of cathine in older qat tissues instead of ephedrine alkaloids, as in *Ephedra* (Krizevski et al. 2007). Although cathinone and cathine are currently only known from a single angiosperm species, similar phenethylamine alkaloids produced from comparable biosynthetic pathways have been found in approximately 40 vascular–plant families and to a lesser degree in some fungal and algal lineages (Smith 1977; Kuklin & Conger 1995; Kulma & Szopa 2007). Cactaceae and Fabaceae possess the greatest diversity of phenethylamines within the angiosperms, with *Lophophora williamsii* (peyote) being a well–known example from Cactaceae (Smith 1977; Kulma & Szopa 2007). The phenethylamine alkaloids are biosynthesized from the amino acid phenylalanine. As one of the 20 essential amino acids used in proteins, phenylalanine is present in all plants. Therefore the presence of phenethylamine alkaloids in plants indicates the evolution of enzymes able to modify phenylalanine into alkaloids. The evolution of enzymes able to convert amino acids into alkaloid precursors and alkaloids has been found to evolve from duplication of genes coding for primary metabolic enzymes (Ober & Hartmann 1999; Nakajima et al. 1993). Once these secondary metabolic pathways have evolved, diversification of end–

product alkaloids is thought to occur from minor changes to enzymes in the pathway. This diversification of alkaloid structures can be species– or even ecotype–specific (Ziegler et al. 2009). As such, the presence or absence of different types of alkaloids have been useful characters for chemosystematic analyses (e.g., Gomes & Gottlieb 1980).

Despite active phytochemical research in Celastraceae, a systematic investigation to identify cathinone and cathine as well as precursors and possible derivatives thereof has, to our knowledge, not been undertaken in this family. Much of the phytochemical research in Celastraceae has been focused on larger molecules, such as terpenoids, given their unique structures and potential as pharmaceutical products (reviewed in Shan et al 2013). We hypothesized that cathine could be present in Celastraceae species other than *Catha edulis* based on the following two reasons. First, previous studies (e.g., Grue-Sørensen & Spenser 1994; Caveney et al. 2001) have found alkaloids in species closely related to the original taxon from which an alkaloid was initially described. For example, ephedrine–type alkaloids were initially described from *Ephedra distachya* (Grue–Sørensen & Spenser 1988) and subsequently described from more *Ephedra* species as assays expanded to broader sampling within the genus (Grue-Sørensen & Spenser 1994; Caveney et al. 2001). The phylogenetic reconstruction of Simmons et al. (2008) places the genera *Allocassine*, *Cassine*, *Lauridia*, and *Maurocenia*, sister to *Catha edulis* and they proposed that species in these genera are the best candidates in which to assay for cathine and related alkaloids. The second basis for the presence of cathine or similar alkaloids in Celastraceae is the ethnobotanical accounts of species in Celastraceae aside from qat being used as stimulants. Accounts from Xhosa and Khoikhoi cultures of South Africa mention the use of *Cassine schinoides* leaves to reduce fatigue, hunger, and thirst (Watt & Breyer-Brandwijk 1962),

while accounts from Canary Island cultures mention chewing the leaves of *Gymnosporia cassinoides* to alleviate fatigue (González et al. 1986).

Materials and methods

Chemicals and reagents — The reagents and solvents used in this study were obtained from Sigma Aldrich (St. Louis, MO), Arcos Organics (Somerville, NJ), and Fisher Scientific (Fairlawn, NJ). The derivatization reagents used did not contain catalysts. Tests with the derivatizing reagent TMSI (trimethylsilylimidazole) containing the catalyst TMCS (trimethylchlorosilane) produced additional derivatization artifacts over the non–catalyst containing reagents. All water used in alkaloid extractions was purified using a Milli-Q (Billerica, MA) water filtration system.

Plant material — Samples for this study (Appendix 3.1) were selected from the Celastraceae such that two genera in each major clade (identified from recent phylogenetic work conducted by; Simmons et al. 2008) were extracted for alkaloids. The tissue used for the analysis was from collections made as part of ongoing Celastraceae phylogenetic studies (Simmons et al. 2008; McKenna et al. 2011). All samples with the exception of freshly harvested tissue for outgroup taxa were silica–gel preserved. Most tissue was from fully expanded leaves and had not been stored longer than 10 years. It is possible that some samples would not possess detectable levels of cathinone given their developmental stage. Krizevski et al. (2007) demonstrated that cathinone in qat was found at diminishingly lower levels as they progressively sampled older nodes from the apical meristem, while cathine levels increased. However, cathine should remain stable for relatively long periods (Chappell & Lee 2010) in older leaves, allowing for the inference of cathinone in young fresh leaves based on the pathway for cathinone biosynthesis presented in (Krizevski et al. 2010).

Alkaloid extraction — Our extraction method was adapted from (Krizevski et al. 2007) for extracting cathinone, cathine and 1-phenylpropane-1,2-dione (PPD) from fresh *Catha edulis* leaves. Leaf tissue was pulverized to a fine powder by applying high speed reciprocation for one minute to 2 mL plastic microcentrifuge tubes containing three 3.2 mm steel beads and approximately 100 mg silica dried leaf tissue. Ground leaf tissue was separated from the steel beads, weighed, and transferred to glass tubes; three mL of water containing 50 µg of caffeine per mL was added as an internal standard. The mixture was agitated for 45 minutes at room temperature and centrifuged at 55 rcf (relative centrifugal force) for 10 minutes. The supernatant was removed and transferred to glass tubes and 1.5 mL of 2N NaOH was added to de–protonate any alkaloids in the samples. Three mL of HPLC–grade MTBE were added to the aqueous extraction and vortexed for 30 seconds. The resulting biphasic mixture was centrifuged at 800 rcf for 8 minutes and the organic fraction removed and placed in a new glass tube. The organic–aqueous extraction step was repeated, resulting in an approximately 6 mL MTBE-based solution that was then dried under a gentle stream of nitrogen gas.

Derivatization — Our derivatization protocol was adapted from methods originally developed by (Spyridaki et al. 2001) for accurately detecting ephedrine alkaloids in the urine of athletes screened in accordance with anti-doping regulations, and Ranieri & Ciolino (2008), who extended these derivatization method for accurately quantifying ephedrine alkaloids in herbal medicines containing *Ephedra sinica* and *E. equisetina*. Derivatization was completed in two stages immediately after samples were dried. The first stage of derivatization used 75µl of MSTFA (N–methyl,N–trimethylsilyltrifluoroactetamide) and TMSI (trimethylsilylimidazole) at a 10:1 ratio and was heated for 15 minutes at 80° C. The second stage employed 30 µl of

MBTFA (methylbistrifluoroacetamide) with the same heating conditions as the first stage. Following derivatization the 105 µl solution was transferred to a 2 mL glass autosampler vial.

GC–MS conditions and instrumentation — Experiments were performed using a Thermo GC Trace Ultra coupled with a Thermo DSQ II mass spectrometer (MS). The column used was a Thermo 30 m TG 5 MS with an internal diameter of 0.25 mm and a film thickness of 0.25 µm. Helium was used as the carrier gas at a flow rate of 1.2 mL per minute. The injection–port temperature was 250° C using split mode at a ratio of 1:10. The GC–oven regime for each injection was 80°C for 2 minutes, then ramped at 5° C per minute until 190° C was reached, followed by ramping at 15° C per minute until 310°C was reached and then held for 1 minute. The mass spectrometer was operated using electron impact (EI) ionization and scanned between 50–650 m/z from 4 to 15 minutes. Mass spectra were searched against the National Institute for Standards and Technology (NIST), Massbank, and Golm MS databases for metabolite identification. Cathine and phenylethylamine identifications were determined by confirmation against commercially prepared chemical standards (Sigma Aldrich, St. Louis, MO).

Data analysis and statistics — For each sample, raw data files were converted to .cdf format, and a matrix of molecular features as defined by retention time and mass (m/z) was generated using XCMS 1.42.0 software in R (Smith et al. 2006) for feature detection and alignment. Raw peak areas were normalized to total ion signal in R, outlier injections were detected based on total signal and PC1 of principle component analysis, and the mean area of the chromatographic peak was calculated among replicate injections (n=2). Molecular features were clustered using RAMClustR (Broeckling et al. 2014) Metabolites were annotated by searching mass spectra from the RAMclustR–derived clusters against the NISTv14 database and assessed based on quality of the mass spectra match to known compounds or classes of compounds. For

all annotated compounds of interest, peak detection and quantification (using peak area) were conducted from the original dataset using Tracefinder v3.2 (Thermo Scientific, Waltham, MA, U.S.A.). PCA was conducted on mean-centered and UV-scaled data using SIMCA P+ v12 (Umetrics, Umea, Sweden).

Results and discussion

Presence of cathinone, cathine and related compounds outside Catha edulis — The presence of cathine was not detected in any species assayed (table 3.1) except *Catha edulis* (Fig. 3.2 A–D). We were unable to detect cathinone from our silica–dried samples of *Catha edulis,* which we attribute to the older age of the leaves when harvested. The amine group of cathinone is expected to be TFA–derivative having prominent peaks at m/z 51, 77, and 105 (Westphal 2011). Additionally none of the proposed precursors or end products in the cathinone–biosynthesis pathway as proposed by Krizevski et al. (2010) were found outside *Catha edulis.* Valeric acid was detected in *Catha edulis* and 10 additional species assayed in this study. Because valeric acid is an alkyl carboxylic acid, it may be an alkaloid precursor that has not been identified in the cathinone biosynthetic pathway. Benzoic acid has been identified as the precursor to PPD in previous studies (Krizevski et al. 2007). The immediate precursor of cathinone, PPD, was not found in any sample but may have escaped detection given that: 1) it would not have been derivatized with our method, which provides a less sensitive assay for PPD than for our target metabolites, and 2) PPD has been found at very low concentrations within *Catha edulis* and *Ephedra*; this pattern is also expected for other taxa producing PPD (Krizevski et al. 2007). Despite the reduced sensitivity for detecting PPD, our extraction method was able to recover the PPD chemical standard at 33 µg per 1 mL sample despite the lack of derivatization.

Table 3.1. Taxa in which compounds were annotated as sterols ,terpenes, and unknown (labeled as ?2), as well as the stimulant–type alkaloids conformed with standards, cathine, PEA in addition to the unconfirmed acetamide,_2,2,2–trifluoro_N_(2–phenylethyl) (labeled as ?1). ++ indicates a significant signal compared to background across multiple extractions, + indicates possible presence of the compound but not significant across extractions and – indicates a signal equal to background levels for a given species. Annotated compounds are labeled by retention time in minutes

	Stimulant-type			Sterol Annotations																	Terpenes/Terpenoids Annotations								?2
Species	Cathine	PEA	?1	12	12.2	12.9	13	13.2	13.3	13.3	14.1	15.2	15.7	15.9	17.7	18.1	18.2	18.3	18.4	19	20.1	18	18.5	18.5	18.7	18.9	19.6	19.8	22.3	19.1
Brexia madagascariensis	-	++	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cassine parviflora	-	-	-	-	-	+	++	-	-	-	++	-	-	-	++	++	++	-	++	-	-	++	-	+	-	+	+	-	++	-
Cassine peragua	-	-	-	+	-	-	-	-	-	-	-	+	++	++	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+
Cassine schinoides	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	++	+	++	++	++	++	-	++
Catha edulis	++	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Cheiloclinum hippocratedides	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Crossopetalum rhacoma	-	++	++	-	-	-	-	-	-	-	++	++	++	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-
Euonymus fortunei	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	++	-	-	-	-	-	-	-	-	-
Gymnospora heterophylla	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	++	-
Gymnosporia wallichiana	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	++	-	-	-	-	-	+	+	-	-	-
Gymnosporia buxifolia	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	++	-	-	-	-	-	++
Gymnosporia cassinoides	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	++	-	-	-	-	-	-	-	-	-	-	-	-
Gymnosporia senegalensis	-	-	++	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Halleriopsis cathoides	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-
Lauridia tetragona	-	++	++	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Maurocenia frangula	-	-	-	++	++	++	++	++	++	++	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-
Maytenus boaria	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-
Peripterygia marginata	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-	-	-
Polycardia libera	-	++	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Populus angustifolia	-	++	++	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Populus tremuloides	-	-	-	-	++	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-
Pseudocatha mandenensis	-	-	-	-	-	+	++	++	++	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	+	-	-

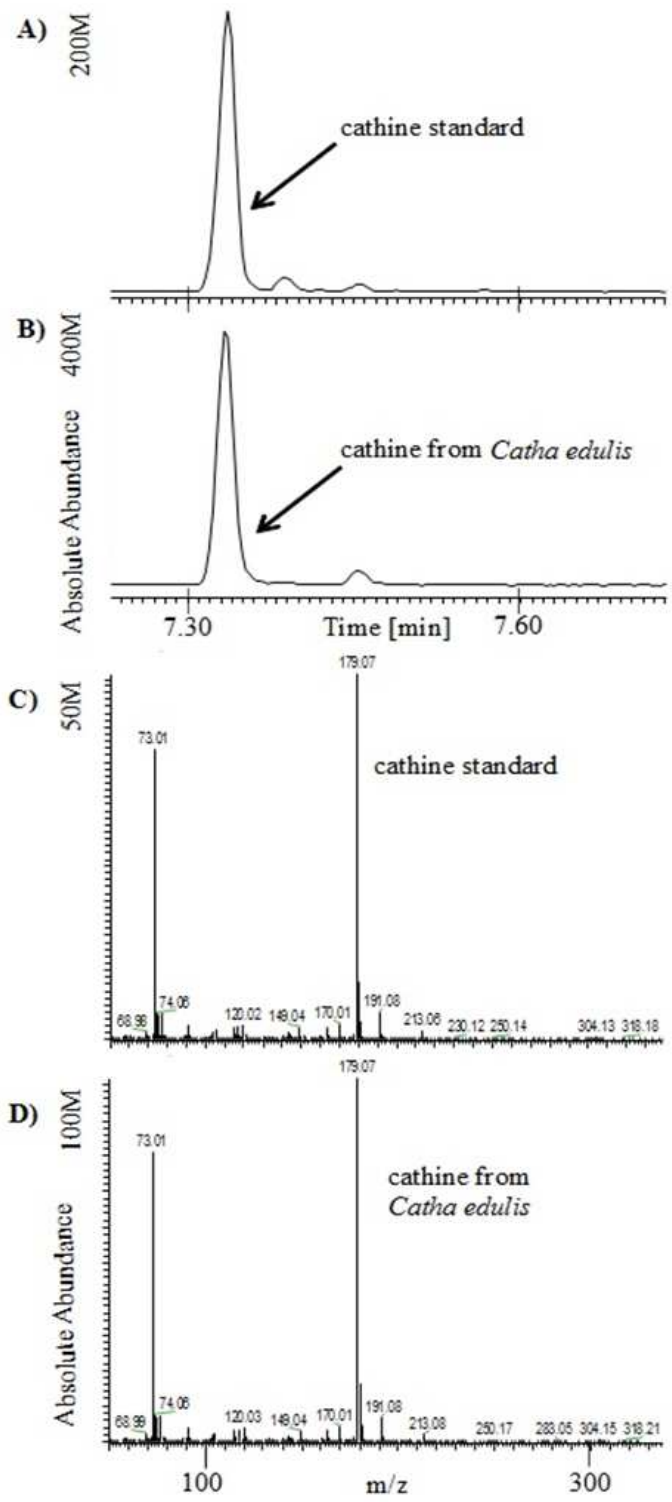


Fig 3.2. Chromatograms comparing chemical standards to plant organic extractions. A) cathine chemical standard, B) extract from *Catha edulis*. Electron impact spectra comparing chemical standards to plant organic extractions. C) cathine chemical standard, D) extract from *Catha edulis.* 'M' on the y axis is an abbreviation of million.

If cathine is further metabolized in other Celastraceae species, ephedrine and pseudoephedrine are possibly the next products from the action of N–methyltransferase (Okada et al. 2008; Krizevski et al. 2010). In the case of pseudoephedrine, the derivatization method we employed produced a single peak at 8.18 min. The derivatization method was verified for ephedrine alkaloids using both a chemical standard for pseudoephedrine and lyophilized root tissue of *Ephedra sinica*. Tests of both plant tissue and chemical standard produced repeatable and consistent results. No species in our study, including *Catha edulis*, was found to contain detectable amounts of pseudoephedrine or ephedrine. The absence of ephedrine and pseudoephedrine with a sensitive derivatization protocol further confirms earlier work by (Szendrei 1980) in which these compounds were not detected in *Catha edulis* using HPLC methods.

Much of the plant material used in this study was from fully expanded leaves. Therefore, if the pathway in other Celastraceae species follows that of *Catha edulis*, our expectation is that cathine might be the most abundant compound detectable in older leaves. If cathine were to be found in our assay then the presence of cathinone could be inferred in the younger leaves. Previous work (Chappell & Lee 2010) has demonstrated that cathinone and cathine could still be detected from air–dried leaves, without any desiccant, after 3 years stored at room temperature and the authors of that study inferred that cathinone and cathine could remain detectable after 10 years stored at room temperature. None of the plant material used in our study was older than 10 years, and it had been stored in silica gel at room temperature throughout that time. Therefore we conclude that age and condition of our plant material was not a major factor in the absence of cathine from other species .It can be expected for only a single species within a lineage to synthesize a given alkaloid, as is the case with morphine from *Papaver somniferum* (Ziegler et

al. 2009). In the case of *P. somniferum*, the closely related species *Papaver bracteatum* and *Papever arenarium* produce the precursors for morphine biosynthesis but not morphine itself (Ziegler et al. 2009). Given that only three enzymatic steps are needed to produce cathine from the common metabolite benzoic acid, it is possible that the entire pathway evolved after *Catha edulis* diverged from its most recent common ancestor with the extant sister group of *Allocassine, Cassine, Lauridia,* and *Maurocenia*. We cannot rule out the possibility that a species which is closely related to *Catha edulis* might have the ability to synthesize cathine or related alkaloids given that we were only able to sample four of the 13 species from the sister clade of *Catha edulis*.

Analysis of compounds detected by GC-MS — While our extraction and derivatization method for this study was specifically designed to detect stimulant alkaloids, numerous other compounds were also detected. Our metabolomics–analysis workflow yielded 296 features (defined as a cluster of m/z at a single retention time) from our dataset after alignment and clustering. Initial quality control steps and examination of PCA plots indicated that the most abundant compounds were derivatization reagents and related byproducts. This finding is not surprising given that the samples injected into the GC-MS instrument were suspended in derivatization reagents. The effect of this was to produce a grouping in the PCA plots by extraction batch. In order to find the biologically relevant compounds from our samples we conducted PCAs in the XCMS step of our workflow. The PCAs were organized by taxonomic grouping in order to find compounds that varied significantly across different levels of biological organization (family and genus). Using this approach we were able to detect 27 compounds (Table 1) that were of putative biological origin

and potentially relevant to our hypotheses regarding the stimulant properties for our species of interest. Of note, many more than 27 compounds were in the dataset but could not be assigned of biological origin due to low signal in the chromatogram, a lack of repeatability across extractions and/or low quality ion spectrum matches. Compounds not meeting these expectations could not be deemed of biological origin and were not included in further analyses. All 27 compounds were found to be significantly greater than the average background signal (MS "noise"). This processing and discovery method was confirmed by the detection of cathine, and PEA – which were confirmed by matching spectra and retention times to authentic standards.

Discovery of non–cathinone phenethylamine alkaloids outside Catha edulis — Two additional alkaloids were found in species other than *Catha edulis*. The alkaloid phenylethylamine (PEA; Fig. 3.1D) was found in detectable levels in the species *Brexia madagascariensis*, *Crossopetalum rhacoma*, *Lauridia tetragona*, *Polycardia liberia*, and an outgroup species, *Populus angustifolia*. The EI spectrum and retention time for PEA found in the plant samples was confirmed with 150 µg of chemical standard PEA (Fig. 3.3A–D) dissolved in 3 mL water that was processed through our extraction, derivatization, and GC–MS protocols. A compound similar to PEA was detected at RT 5.89 min (Fig. 3.3 A, B) and annotated as acetamide,_2,2,2-trifluoro_N_(2–phenylethyl) with a reverse match score of 907 when searched against the NIST database (Fig. 3.3 A). Why derivatization was impartial or incomplete in those samples containing PEA is uncertain. However MBTFA was found to be more competitive at derivatizing amine groups than MSTFA (Meng & Margot 2010). Therefore some interaction between the two reagents might account for the appearance of *N*–TFA PEA in some samples and not others. Evidence against this hypothesis is the fact that acetamide,_2,2,2-trifluoro_N_(2–phenylethyl) was not generated when the PEA chemical standard was derivatized. In either case,

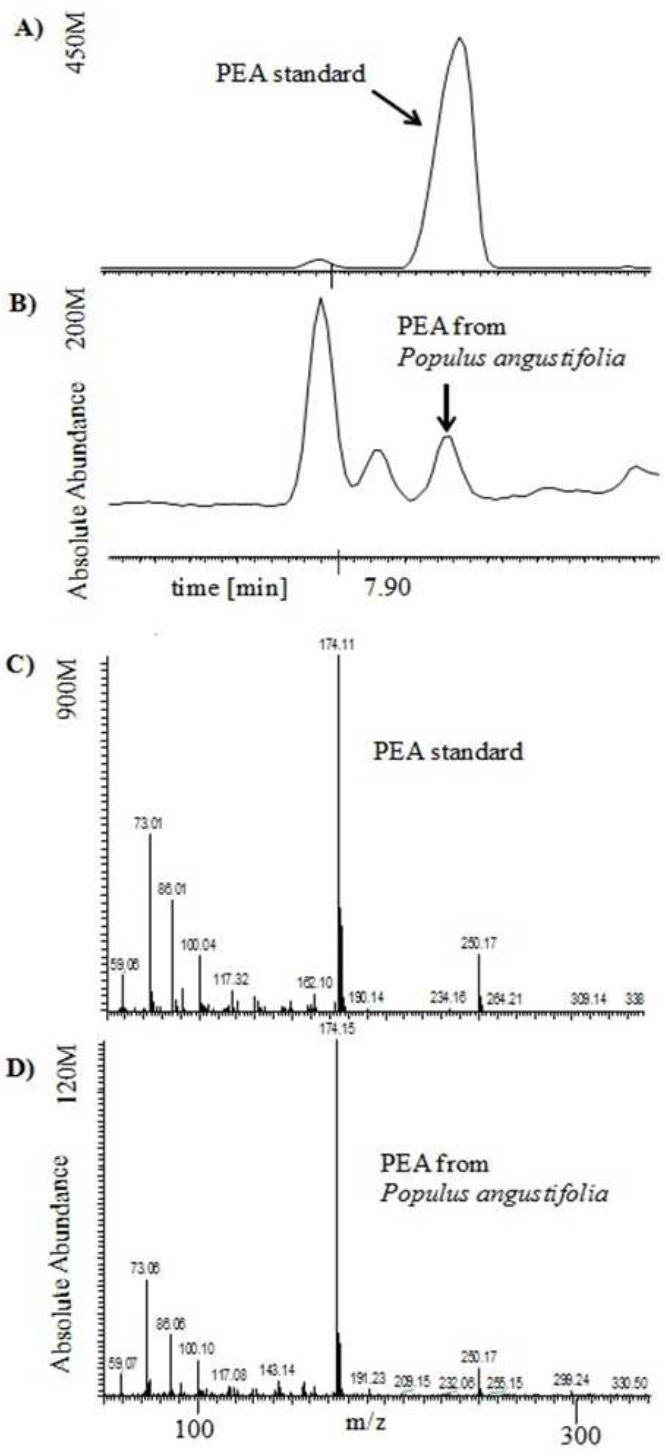


Fig 3.3. Chromatograms comparing chemical standards to plant organic extractions. A) phenylethylamine chemical standard, B) extract from *Populus angustifolia*. Electron impact spectra comparing chemical standards to plant organic extractions. C) phenylethylamine standard, D) extract from *Populus angustifolia.* 'M' on the y axis is an abbreviation of million.

the EI–spectrum for acetamide,_2,2,2-trifluoro_N_(2-phenylethyl) is consistent with a phenylalanine derived alkaloid and might be a byproduct, precursor or isomer of PEA given their similar spectra (Fig. 3.4 A). Acetamide,_2,2,2-trifluoro_N_(2–phenylethyl) was detected in *Crossopetalum rhacoma, Gymnosporia senegalensis, Lauridia tetragona, Polycardia liberia* and an outgroup species, *Populus angustifolia*.

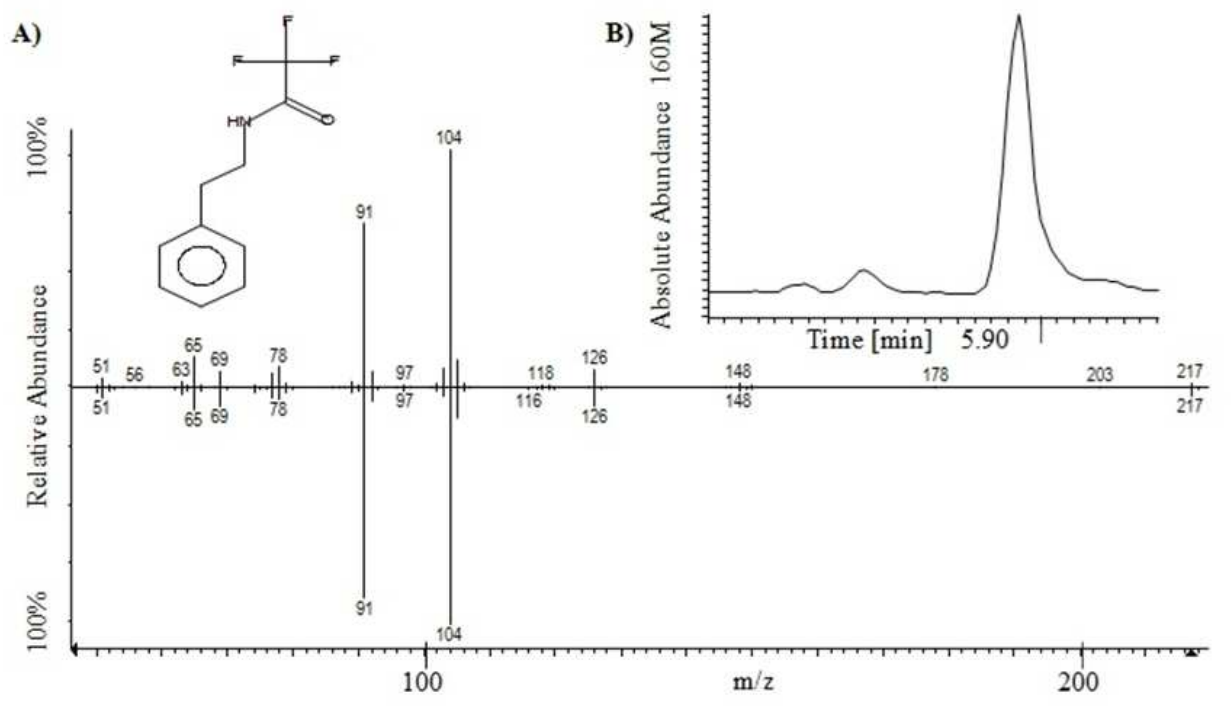


Fig 3.4. A) Electron impact spectra comparing unknown compound from *Crossopetalum rhacoma* (top spectrum) to NIST best match (reverse match 907) of acetamide 2,2,2–triflouro–N(2–phenylethyl) and chemical structure B) Chromatogram of unknown compound. 'M' on the y axis is an abbreviation of million.

Presence of acetamide,_2,2,2–trifluoro_N_(2–phenylethyl) is correlated with PEA presence in the species *Crossopetalum rhacoma, Lauridia tetragona, Polycardia liberia,* and *Populus angustifolia,* which suggests that the unknown could be an intermediate metabolite in PEA biosynthesis for those species. The apparent correlation of these two compounds was tested in a phylogenetic context using Pagel's (1994) test of correlated character evolution and found to

be significant ($p = 0.001$). Therefore if acetamide,_2,2,2-trifluoro_N_(2-phenylethyl) is not an artifact of PEA derivatization, then it seems probable that they are unique compounds from a single biosynthetic pathway.

PEA is similar to cathine in that few enzymatic steps are needed to produce PEA from phenylalanine or tyrosine (Smith et al. 1989). Therefore the dispersed pattern of PEA presence in Celastraceae and other plant lineages might be attributable to the few evolutionary steps needed to gain or lose enzymatic function for PEA biosynthesis. When PEA presence is optimized (Fitch 1971) onto the phylogeny for Celastraceae, the presence of PEA shows no relationship to the cathine pathway (Fig. 3.5). That is, PEA biosynthesis does not appear to have evolved first in the Celastraceae ancestors of *Catha edulis* and subsequently given rise to the cathine pathway. Rather, Fitch optimization (1971) indicates four independent origins of PEA.

Sterols, terpenes, and related compounds — Of the 27 compounds detected from our workflow 19 were annotated as sterols, seven were annotated as terpenes/terpenoids of some type, and one did not match well to any compounds in the NIST metabolite database. A vast diversity of terpenes, including numerous unique types, have been described from the Celastraceae (Shan et al 2013). The bioactivity of terpenes in the Celastraceae has been studied within certain species given their potential application as painkillers, cancer treatments, anti–HIV agents, diabetes treatments, antidepressants, and numerous other medical applications (Shan et al 2013).

From our analysis, we found that *Pseudocatha mandenensis, Maurocenia frangula, Cassine parviflora,* and *Halleriopsis* cathoides were enriched in the type and quantity of compounds annotated as sterols. Four compounds annotated as sterols were highly abundant only in *Pseudocatha mandenensis,* three in *Maurocenia frangula,* one in *Cassine parviflora,* and one

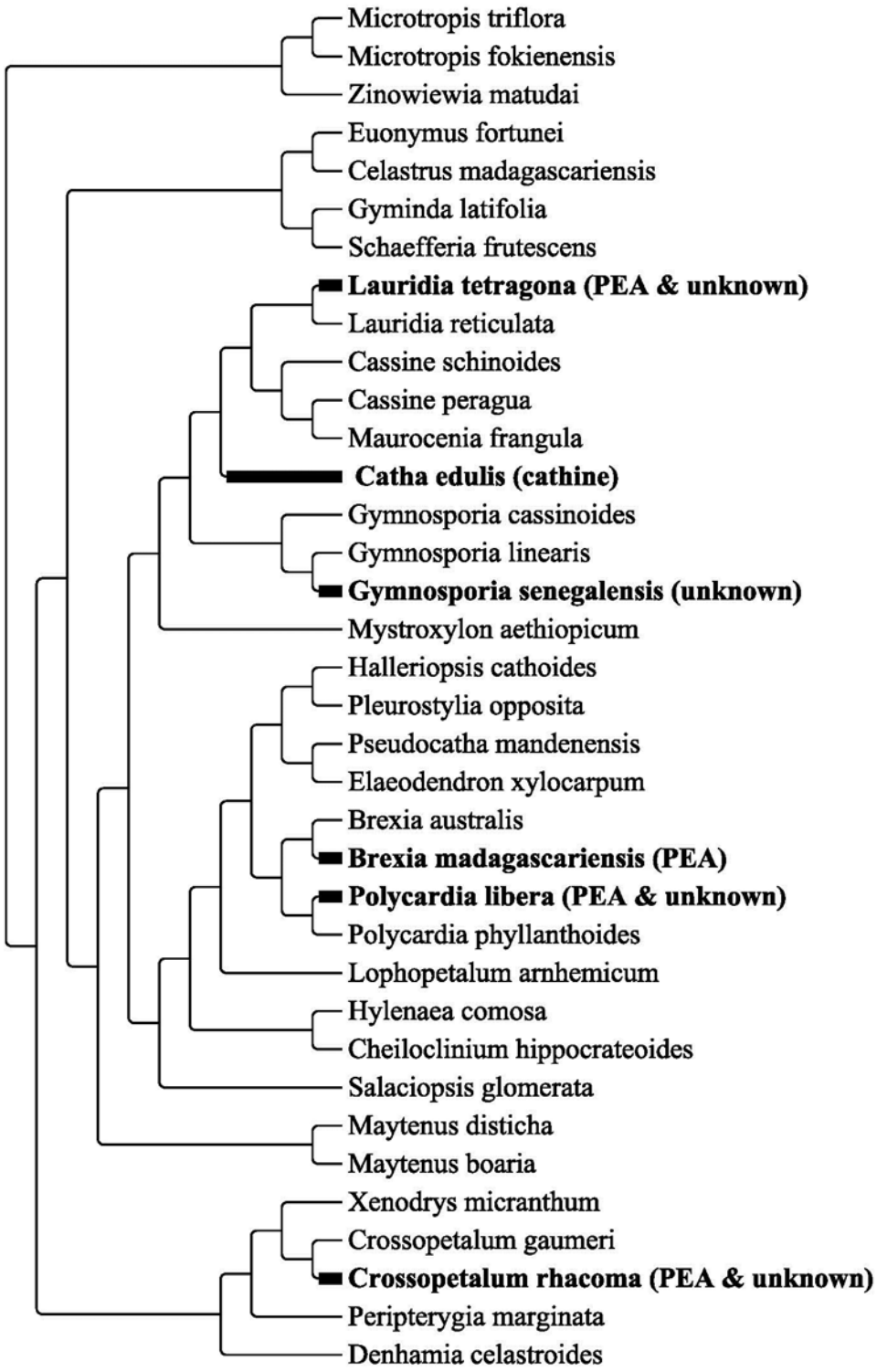


Fig 3.5. Phylogenetic relationships (from Simmons et al. 2008; McKenna et al. 2011) among the taxa sampled with cathine, PEA, and acetamide,_2,2,2–trifluoro_N_(2–phenylethyl) labeled as unknown. Occurrences of compounds are mapped onto the tree using Fitch (1971) optimization; bold branches represent independent derivations.

in *Halleriopsis cathoides* (Fig. 3.6A). Seven compounds annotated as sterols of relative high abundance were detected in two or more of the following three species: *Pseudocatha mandenensis, Maurocenia frangula,* and *Cassine parviflora* (Fig. 3.6A). While sterols are routinely discovered in phytochemical assays (e.g., Kaweetripob et al. 2013) in the Celastraceae, *Pseudocatha mandenensis*, *Maurocenia frangula,* and *Cassine parviflora,* to our knowledge have not sampled in previous investigations. Additional experiments utilizing other methods of chemical identification (e.g. MS^n or nuclear magnetic resonance, NMR) and chemical standards is needed to confirm our initial annotations and/or describe these potentially novel compounds. The sterol type compound from *Halleriopsis cathoides* is of particular interest for the following three reasons: 1) *Halleriopsis cathoides* has not been formally published taxonomically, thus it has almost certainly never been assayed; 2) the annotated compound (RT 20.1) is unique across the entire dataset, and 3) the highest match was to preganalone, a neurosteroid with sedative, anxiolytic, anesthetic, and anticonvulsant properties (Carl et al. 1994). Our conclusions regarding sterols from our sample set are provisional but provide a sound starting point for further research.

Regarding compounds annotated as terpene–like, we found eight such compounds with sufficient signal to be distinguished from background and annotated. Five compounds annotated as terpenes were found in high abundance in *Cassine schinoides,* one in *Pseudocatha mandenensis,* one in *Peripterygia marginata,* and one found in both *Pseudocatha mandenensis* and *Cheiloclinum hippocratedides* (Fig. 3.6B). As with sterols, terpenes are phytochemicals commonly found in the Celastraceae (Shan et al. 2013), however, so far as we are aware, the species *Cassine schinoides, Pseudocatha mandenensis,* and *Cheiloclinium hippocrateoides* have not been part of any published phytochemical work on terpenes. The most interesting finding

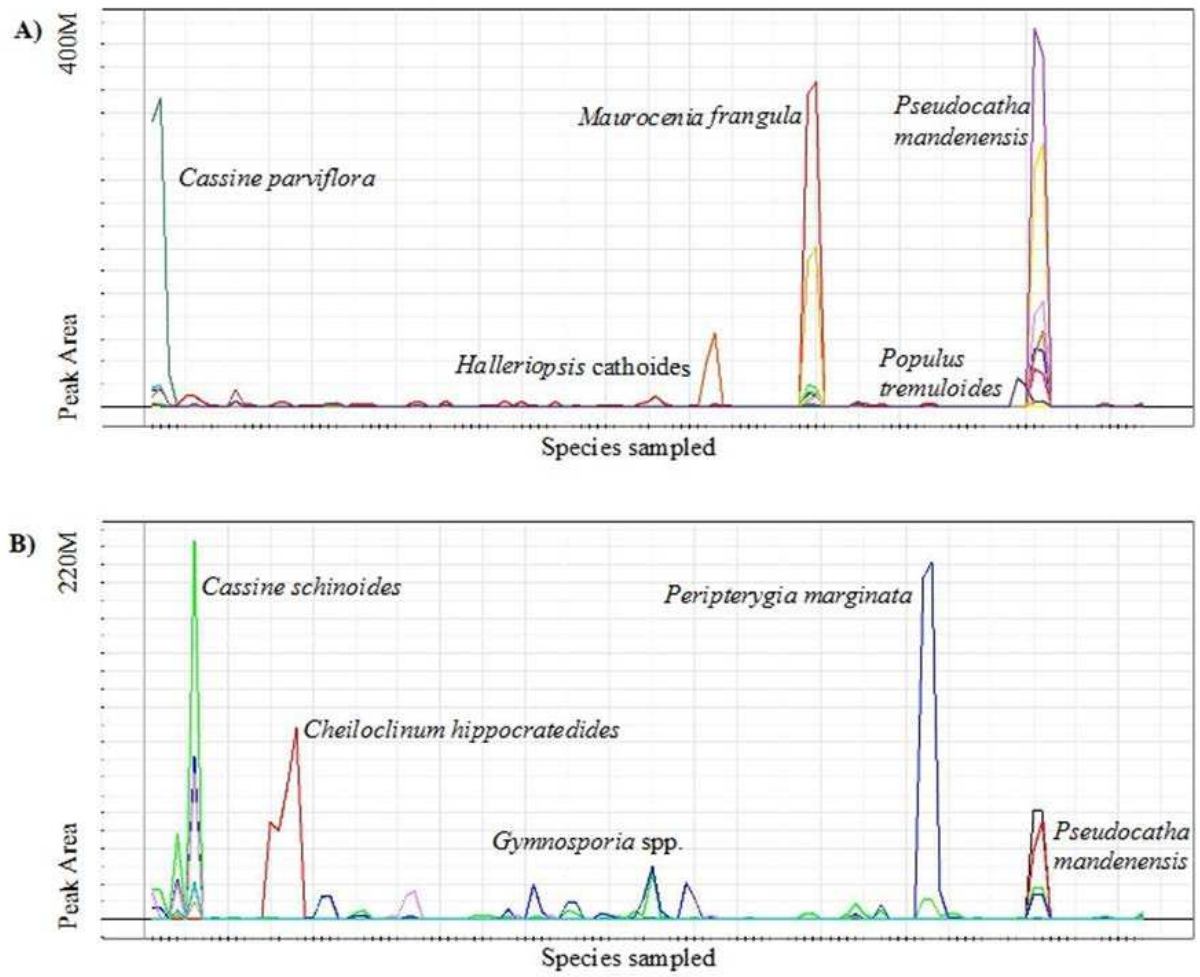


Fig 3.6. A) Compounds annotated as sterols from the data set; y axis is peak area and x axis are samples. The samples, including replicates, are placed in alphabetical order on the y axis; different compounds are indicated by different colored lines. B) Compounds annotated as terpenes from the data set; y axis is peak area and x axis are samples. The samples, including replicates, are placed in alphabetical order on the y axis; different compounds are indicated by different colored lines. 'M' on the y axis is an abbreviation of million.

regarding the terpene annotations is the five compounds annotated as terpenes that were found in significant abundance in *Cassine schinoides.* While terpenes and triterpenoids have been described from more distantly related species such as *Elaeodendron balae* (Fernando et al. 1989) and *Elaeodenron xylocarpa* (Núñez et al. 2013) (formerly *Cassine*), no phytochemical assay has been conducted in the more narrowly defined African *Cassine* genus, which is

distantly related to *Elaeodendron*. Two of our annotations for terpenes occurring in *Cassine schinoides* were friedelanone, a triterpene type broadly diversified in the Celastraceae and found in *Cassine balae* (Fernando et al. 1989). While friedelanones have not been found to be stimulants, little is known about the pharmacology of most of the friedelanones discovered from the Celastraceae (Shan et al. 2013). Triterpenes from *Maytenus gonoclada* and *Maytenus imbricata* were found to be central nervous system stimulants through inhibition of acetylcholinesterase (Rodrigues et al. 2014) Furthermore, some terpenoids such as tetrahydrocannabinol can have pronounced pharmacological action (Isbell et al. 1967). Whether the terpenes in *Cassine schinoides* have stimulant effects is not known, but the diversity and abundance of these compounds in *Cassine schinoides* compared to other species in the dataset certainly suggests that they might explain the stimulant effects attributed to this species. Additionally, given that *Pseudocatha mandenensis* and *Cheiloclinium hippocratedides* have not been heretofore assayed, our terpene annotations require further clarification using orthogonal analytical techniques and comparisons to authentic standards.

Conclusion

A two-stage derivatization procedure was used to assay for stimulant alkaloids in *Catha edulis* and 42 other species in the Celastraceae. No alkaloids that are sanctioned as controlled substances were found in species other than *Catha edulis*. Ethnobotanical and phylogenetic criteria were used to select which taxa are most likely to possess stimulant alkaloids. The two species other than *Catha edulis* mentioned in the ethnobotanical literature as stimulants, *Gymnopsoria cassinoides* and *Cassine schinoides*, did not possess detectable levels of the stimulant alkaloids cathine or cathinone. Nor did these species possess detectable levels of alkaloids in general. Two alkaloids aside from the cathinone–derived alkaloids were found in

species other than *Catha edulis*. PEA was found in *Brexia madagascariensis*, *Crossopetalum rhacoma*, *Lauridia tetragona*, *Polycardia liberia*, and a more distantly related species from another order of plants, *Populus angustifolia*. PEA is a common metabolite in numerous angiosperm species but has not been reported from the Celastraceae. PEA is not expected to have any pharmacological effects and was not found in Celastraceae species used as stimulants. An alkaloid (acetamide,_2,2,2–trifluoro_N_(2-phenylethyl)) that is similar in structure to PEA was detected in *Crossopetalum rhacoma*, *Gymnosporia senegalensis*, *Polycardia liberia*, and a more distantly related species from another order of plants, *Populus angustifolia.* To our knowledge, these species are not mentioned in the ethnobotanical literature as stimulants. The pattern of PEA and acetamide,_2,2,2–trifluoro_N_(2-phenylethyl), when mapped onto the inferred Celastraceae phylogeny (Fig. 3.5), have evolved independently multiple times and did not require the stepwise evolution of numerous enzymes as in alkaloids such as morphine (Ziegler et al.2009) The evolution of cathinone and cathine biosynthesis in *Catha edulis* appears to have evolved recently and rapidly as no direct precursors were found in any of the closely related species that we assayed.

Based on these data, it does not appear that cathinone or related alkaloids are responsible for the stimulant properties of *Cassine schinoides* and *Gymnosporia cassinoides*. From our data, *Cassine schinoides* was found to be enriched for five terpene–like compounds that may be responsible for the stimulant properties noted for this species. Our metabolomics workflow characterized eight compounds that were annotated as terpene–like and 18 compounds that were annotated as sterol–like. These compounds were variously enriched in the species *Cassine parviflora*, *Cassine schinoides*, *Cheiloclinium hippocratedides*, *Halleriopsis cathoides*, *Maurocenia frangula*, *Peripterygia marginata*, and *Pseudocatha mandenensis* —a group of

species heretofore not well characterized in the phytochemical literature. Follow–up studies using additional analytical analyses are warranted to elucidate the structure of these potentially novel compounds.

REFERENCES

Altabachew M, Chandravanshi BS, Redi–Abshiro M, Torto N, Chigome S, Pule BO (2013) Evaluation of the effects of various drying techniques on the composition of the psychoactive phenylpropylamino alkaloids of khat (*Catha edulis* Forsk) chewing leaves. Bulletin of the Chemical Society of Ethiopia 27: 347–358.

Babayan E, Barkov N, Machula M (1982) Mediator mechanisms of cathinone effects on animal behavior. Drug and Alcohol Dependence 12: 31–35.

Broeckling CD, Afsar FA, Neumann S, Ben–Hur A, Prenni JE (2014) RAMClust: a novel feature clustering method enables spectral–matching–based annotation for metabolomics data. Analytical Chemistry 86: 6812–6817.

Brünning R, Wagner H (1978) Übersicht über die Celastraceen–Inhalsstoffe: Chemie Chemotaxonomie, Biosynthesie, Pharmokologie. Phytochemistry 17:1821–1858.

Carl P, Høgskilde S, Lang–Jensen T, Bach V, Jacobsen J, Sorensen MB, Grälls M, Widlund L (1994) Pharmacokinetics and pharmacodynamics of eltanolone (pregnanolone), a new steroid intravenous anesthetic, in humans. Acta Anaesthesiologica Scandinavica 38: 734–741.

Caveney S, Charlet D, Freitag H, Maier–Stolte M, Starratt A (2001) New observations on the secondary chemistry of world *Ephedra* (Ephedraceae), American Journal of Botany 88: 1199–1208.

Chappell JS, Lee MM (2010) Cathinone preservation in khat evidence via drying, Forensic Science International 195: 108–120.

Davis EM, Ringer KL, McConkey ME, Croteau R (2005) Monoterpene metabolism, cloning, expression, and characterization of menthone reductases from peppermint, Plant Physiology 137: 873–881.

DEA (Drug Enforcement Agency), KHAT: AKA: *Catha edulis*, http://www.usdoj.gov/dea/pubs/pressrel/pr072606a.htmL (accessed 2 March 2015).

Department of Justice (1988) Schedules of controlled substances; temporary placement of cathine ((+)–norpseudoephedrine), fencamfamin, fenproporex and mefenorex into schedule IV, Federal Registry 53: 17459–17460.

Department of Justice (1993) Schedules of controlled substances: placement of cathinone and 2,5–Dimethoxy–4–ethylamphetamine into schedule I, Federal Registry 58: 4316–4318.

Fernando HC, Gunatilaka AAL, Tezuka Y, Kikuchi T (1989) Studies on terpenoids and steroids – 18[1] balaenonol, balaenol, and isobalaendiol: three new 14(15)–ene–quinone–methide triterpenoids from *Cassine balae*. Tetrahedron 18: 5867–5876.

Fitch WM, Toward defining the course of evolution: minimum change for a specific tree topology, Systematic Biology 4: 406–416.

Glennon RA, Liebowitz S, (1982) Serotonin affinity of cathinone and related analogues, Journal of Medicinal Chemistry 25: 393–397.

Glennon RA, Schechter MD, Rosecrans JA (1984) Discriminative stimulus properties of S (—)– and R (+)–cathinone, (+)–cathine and several structural modifications, Pharmacology Biochemistry and Behavior 21:1–3.

Glennon RA (1986) Discriminative stimulus properties of phenylisopropylamine derivatives, Drug and Alcohol Dependence 17: 119–134.

Gomes CMR, Gottlieb OR (1980) Alkaloid evolution and angiosperm systematics, Biochemical Systematics and Ecology 8: 81–87.

González AG, López I, Ferro EA, Ravelo AG, Gutiérrez J, Aguilar AM, (1986) Taxonomy and chemotaxonomy of some species of Celastraceae. Biochemical Systematics and Ecology 14: 479–480.

Grue–Sørensen G, Spenser ID (1988) Biosynthesis of ephedrine. Journal of the American Chemical Society 110: 3714–3715.

Grue–Sørensen G, Spenser ID (1994) Biosynthetic route to the *Ephedra* alkaloids, Journal of the American Chemical Society 116: 6195–6200.

Hoffman R, Al'Absi M (2010) Khat use and neurobehavioral functions: suggestions for future studies. Journal of Ethnopharmacology 132: 554–563.

Isbell H, Gorodetzsky CW, Jasinski D, Claussen U, Spulak FV, Korte F (1967) Effects of (–)Δ^9–Trans–Tetrahydrocannabinol in man. Psychopharmacology 11: 184–187.

Kalix P (1984) Recent advances in khat research. Alcohol and Alcoholism 19: 319–323.

Kalix P (1988) Khat an amphetamine–like plant material: effect on catecholamine release, Progress in catecholamine research part A., Basic aspects and peripheral mechanisms. Alan R Liss Inc.

Kalix P (1996) *Catha edulis,* a plant that has amphetamine affects. Pharmacy World Science 18: 69–73.

Kaweetripob W, Mahidol C, Prawat H, Ruchirawat S (2013) Lupane friedelane, oleanane, and ursane triterpenes from the stem of *Siphonodon celastrineus* Griff. Phytochemistry 96: 404–417.

Krizevski R, Dudai N, Bar E, Lewinsohn E (2007) Developmental patterns of

phenylpropylamino alkaloids accumulation in khat (*Catha edulis*, Forsk.), J. Ethnopharmacology. 114: 432–438.

Krizevski R, Bar E, Shalit O, Sitrit Y, Ben–Shabat S, Lewinsohn E (2010) Composition and stereochemistry of ephedrine alkaloids accumulation in *Ephedra sinica* Stapf. Phytochemistry 71: 895–903.

Kuklin AI, Conger BV (1995) Catecholamines in plants. Journal of Plant Growth Regulation 14: 91–97.

Kulma A, Szopa J (2007) Catecholamines are active compounds in plants. Plant Science 172: 433–440.

López MR, de León L, Moujir L (2011) Antibacterial properties of phenolic triterpenoids against *Staphylococcus epidermidis*, Planta Medica 77: 726–729.

McKenna MJ, Simmons MP, Bacon CD, Lombardi JA (2011) Delimitation of the segregate genera of *Maytenus* s. l. (Celastraceae) based on morphological and molecular characters. Systematic Botany 36 (2011) 922–932.

Meng P, Margot P (2010) Constituent profiling of ecstasy seizures by GC–MS after joint derivatization with MSTFA and MBTFA. Forensic Science International 28: 52–54.

Mereu GP, Pacitti C, Argiolas A (1983) Effect of (–)–cathinone, a khat leaf constituent, on dopaminergic firing and dopamine metabolism in the rat brain. Life Sciences 32: 1383–1389.

Nakajima K, Hashimoto T, Yamada Y (1993) Two tropinone reductases with different stereospecificities are short–chain dehydrogenases evolved from a common ancestor. Proceedings of the National Academy of Sciences of the United States of America 90: 9591–9595.

Nencini P, Fraioli S, Pascucci T, Nucerito C (1998) (−)–Norpseudoephedrine, a metabolite of cathinone with amphetamine–like stimulus properties, enhances the analgesic and rate decreasing effects of morphine, but inhibits its discriminative properties. Behavioral Brain Research 92: 11–20.

Núñez MJ, Ardiles AE, Martínez ML, Torres–Romero D, Jiménez IA, Bazzocchi IL, (2013) Triterpenoids from *Cassine xylocarpa* and *Celastrus vulcanicola* (Celastraceae), Phytochemistry Letters 6: 148–151.

Ober D, Hartmann T. (1999) Homopsermidine synthase, the first pathway–specific enzyme of Pyrrolizidine alkaloid biosynthesis, evolved from deoxyhypusine synthase. Proceedings of the National Academy of Sciences of the United States of America 96: 14777–14782.

Okada T, Mikage M, Sekita S (2008) Molecular characterization of the phenylalanine ammonia–lyase from *Ephedra sinica*. Biological and Pharmaceutical Bulletin 31: 2194–2199.

Oshio H, Tsukui M, Matsuoka T (1978) Isolation of l–ephedrine from "*Pinellia* tuber". Chemical and Pharmaceutical Bulletin 26: 2096–2097.

Pagel M (1994) Detecting correlated evolution on phylogenies: a general method for the comparative analyses of discrete characters, Proceedings of the Royal Society London B 255: 37–45.

Patel N (2000) Mechanism of action of cathinone: the active ingredient of khat (*Catha edulis*), East African Medical Journal 77: 329–332.

Ranieri TL, Ciolino LA (2008) Rapid selective screening and determination of ephedrine alkaloids using GC–MS footnote mark. Phytochemical Analysis 19: 127–135.

Rodrigues VG, Silva FC, Duarte LP, Takahashi JA, Matildes BLG, Silva GDF, Silva RR, Vieira–Filho SA (2014) Pentacyclic triterpenes from *Maytenus* genus as acetylcholinesterase inhibitors, International Journal of Pharmacy and Pharmaceutical Science, 6: 918–920.

Simmons MP, Cappa JJ, Archer RH, Ford AJ, Eichstedt D, Clevinger CC (2008) Phylogeny of the Celastreae (Celastraceae) and the relationships of *Catha edulis* (qat) inferred from morphological characters and nuclear and plastid genes. Molecular Phylogenetics and Evolution 48: 745–757.

Schorno X, Steinegger E (1979) CNS–active phenylpropylamines of *Catha edulis* FORSK, (Celastraceae) of Kenyan origin, Experientia 15: 572–574.

Shan W–G, Zhang L–W, Xiang J–G, Zahn Z–J, (2013) Natural Friedelanes. Chemistry and Biodiversity 10: 1392–1434.

Shoyama Y, Hatano K, Nishioka I (1983) Rapid and simple multiplication of *Pinellia ternata* by tissue culture. Planta Medica 47: 103–105.

Smith TA (1977) Phenylethylamine and related compounds in plants, Phytochemistry 16: 9–18.

Smith TH (1989) Plant amines in Encyclopedia of plant physiology 8: secondary plant products. eds. E.A. Bell, B.V. Charlwood pp. 433–460. Springer–Verlag, Berlin.

Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G (2006) XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Analytical Chemistry 78: 779–787.

Spivey AC, Weston M, Woodhead S (2002) Celastraceae sesquiterpenoids: biological activity and synthesis. Chemical Society Reviews 31: 43–59.

Spyridaki MHE, Tsitsimpikou CJ, Siskos PA, Georgakopoulos CG (2001) Determination of ephedrines in urine by gas chromatography–mass spectrometry. Journal of Chromatography B 758: 311–314.

Stevens PF (2012) Angiosperm Phylogeny Website. Version 12. [continuously updated] http://www.mobot.org/MOBOT/research/APweb/. (accessed March 2015).

Szendrei K (1980) The chemistry of khat, Bulletin on Narcotics 32: 5–35.

Thiers B [continuously updated]. Index Herbariorum: A global directory of public herbaria and associated staff. New York Botanic Garden's Virtual Herbarium. http://sweetgum.nybg.org/ih/ (accessed May 2015)

United Nations Narcotics Laboratory (1975) Studies on the chemical composition of khat: III, Investigations on the phenylalkylamine fraction, Geneva, United Nations Narcotics Laboratory.

Varisco DM (2004) Turning over a new leaf: the impact of qat (Catha edulis) in Yemeni Horticulture, Botanical progress, horticultural innovation and cultural change, eds., M. Conan and W. J. Kress. Washington, D.C., Dumbarton Oaks Research Library and Collection pp. 239–256.

Watt JM, Breyer–Brandwijk MG (1962) The medicinal and poisonous plants of southern and eastern Africa. Edinburgh, E. &. S. Livingstone Ltd.

Westphal F (2011) personal communication.

Wilder P, Mathys K, Brenneisen R, Kalix P, Fisch HU (1994) Pharmacodynamics and Pharmacokinetics of khat: a controlled study. Clinical Pharmacology and Therapeutics 55: 556–562.

Wolfes O (1930) Über das Vorkommen von *d*–nor–iso–Ephedrin in *Catha edulis*. Archiv der

Pharmazie 268: 81–83.

Yoshikawa M, Murakami T, Shimada H, Matsuda H, Yamahara J, Tanabe G, Muraoka O (1997) Salicinol, potent antidiabetic principle with unique thiosugar sulfonium sulfate structure from the Ayurvedic traditional medicine *Salacia reticulata* in Sri Lanka and India, Tetrahedron Letters 38: 8367–8370.

Ziegler J, Voigtlander S, Schmidt J, Kramell R, Miersch O, Ammer C, Gesell A, Kutchan TM, (2006) Comparative transcript and alkaloid profiling in *Papaver* species identifies a short chain dehydrogenase/reductase involved in morphine biosynthesis. Plant Journal 48: 177–192.

Ziegler J, Facchini PJ, Geißler R, Schmidt J, Ammer C, Kramell R, Voigtländer S, Gesell A, Pienkny S, Brandt W (2009) Evolution of morphine biosynthesis in opium poppy. Phytochemistry 70: 1696–1707.

CHAPTER 4: THE ORIGINS, DISPERSAL, AND EVOLUTION OF *CATHA EDULIS* (QAT, CELASTRACEAE) IN AREAS OF HISTORIC CULTIVATION

Summary

Qat (Celastraceae: *Catha edulis*) is a woody plant species cultivated for its stimulant alkaloids and is an essential part of the economy and culture in highland areas of Ethiopia, Kenya and Yemen. Despite the importance of this species, the wild origins and dispersal patterns have only been described in often contradictory historical documentation. We examined cultivated qat's wild origin, human dispersal, and genetic divergence relative to wild qat. We sampled 17 SSR markers and 1561 wild and cultivated individuals across the historic areas of qat cultivation. Based on genetic structure using Bayesian and non–parametric methods, two centers of origin in Kenya and one in Ethiopia were found for cultivated qat. The centers of origin in Ethiopia and northeast of Mt. Kenya are the primary sources of cultivated qat genotypes. From the results of Bayesian assignment methods and genetic distance based tests the cultivated qat genotypes grown in Yemen were founded from genotypes of Ethiopian origin rather than indigenous populations, as claimed by some sources. Translocations of cultivated qat with a Kenyan origin were not found outside Kenya, however 10 cultivated genotypes in Kenya were assigned to an Ethiopian origin. Hybrid genotypes of Ethiopian and Kenyan parentage were found in northern Kenya. Within Kenya the Mt Kenya/Meru center of origin has been the largest source of translocations with 35 genotypes from this area found dispersed to other parts of Kenya. Evolution under cultivation as defined by the divergence of cultivated from wild genotypes was found when comparing Ethiopian cultivated and wild genotypes. This pattern of

divergence may be caused by the extinction of wild qat populations in Ethiopia caused by loss of Ethiopian forests, sampling issues, and/or artificial selection for agronomic traits.

Introduction

Qat [Celastraceae: *Catha edulis* (Vahl) Forssk. ex Endl.] is a woody evergreen crop of major economic and cultural importance in East Africa and southwest Arabia, which is grown for its stimulant alkaloids. An estimated 20 million people consume qat on a daily basis in eastern Africa (Al–Motarreb 2002), qat is second only to coffee as a foreign–exchange commodity in Ethiopia (Bhalla 2002; Gebissa 2004; Anderson et al. 2007), and 6% of the gross domestic product in Yemen is generated from qat cultivation and sales (World Bank 2007). In addition to local and regional consumption of qat, international exportation has expanded since the 1950's (e.g., Goldsmith 1999; Gebissa 2004; Carrier & Godez 2009). For example, in 1997 the amount of legally imported qat to England was 43,671 kilograms and in 2013 the amount imported was over 2.26 million kilograms (Andersen et al. 2007; House of Commons 2013). Despite the cultural and economic importance of qat, the historical and oral accounts are often contradictory in regard to the wild origin(s) of cultivated qat. The question of qat origin has not previously been examined using genetic markers. As a consequence of our limited knowledge concerning the genetic diversity of wild and cultivated qat, the evolutionary patterns, such as which cultivated linages have been dispersed from centers of origin and the genetic divergence between cultivated qat groups and extant wild populations, are unknown.

In wild forest environments (Fig. 4.1 A), qat grows as a woody shrub or tree (Fig. 4.1 B) from 2 – 25 meters tall (Robson et al. 1994). The natural habitats of qat are evergreen submontane forests (1100 – 2400 m), growing at forest margins and amidst forests on rocky outcroppings (Robson et al. 1994). The native range of qat is throughout eastern Africa (and

possibly southwest Arabia) from Ethiopia to South Africa and west to Angola (Palgrave 1983; Robson et al. 1994). Flowers of qat are bisexual. Little is known about whether qat is self–compatible or strictly outcrossing; from the floral morphology (Fig 4.1 E) either or both breeding systems are possible. Clonal growth is limited in the wild, with moderately wind dispersed seeds largely accounting for the natural reproduction of qat (M. P. Simmons pers. obs.).

Cultivated qat is primarily propagated clonally via rooted cuttings (e.g., Kennedy 1987; Gebissa 2004; Carrier 2007) and trained into several forms from low growing shrubs resembling cultivated tea plants, to narrow upright trees, and open downward trained trees (Fig. 4.1 C; e.g., Kennedy 1987; Carrier 2007). The main purpose of the different training schemes is to encourage the growth of new stems, which are the part harvested and sold (Fig. 4.1 D), and to make harvesting easier. Qat cultivation takes place at scales from backyard gardens to mono–cropped orchards but it is most typically grown on small scale farms where it is often intercropped with maize, potatoes, beans, ensete, and other food crops (Fig. 4.1 C).

Qat is the only species in Celastraceae that is cultivated on a large scale specifically for its stimulant alkaloids. Qat is cultivated and/or collected from the wild and consumed in Ethiopia, Israel, Kenya, Madagascar, Rwanda, Somalia, Tanzania, Uganda, and Yemen (e.g., Kennedy 1987; Gebissa 2004; Anderson et al. 2007; Carrier 2007; Carrier & Godez 2009). Qat is generally consumed by chewing fresh young leaves and shoots followed by ingestion of the liquid constituents. The young shoot tips of qat contain the following stimulant alkaloids: (S)–cathinone (United Nations Narcotics Laboratory 1975; Schorno & Steinegger 1979; Szendrei 1980), (1R,2S)–norephedrine, and (1S,2S)–norpseudoephedrine, which is commonly referred to as cathine (Wolfes 1930). Because of the rapid conversion of cathinone into the less potent

Figure 4.1.Wild and cultivated qat, A) Ethiopian forest habitat in which wild qat was collected, B) Wild qat tree habit in southern Kenya (see arrow), C) Typical Ethiopian qat farm, intercropped with maize, D) Qat bundle of mixed varieties for sale in Ethiopian market, and E) Qat flowers.

cathine after harvest, qat is treated as a perishable commodity and is usually consumed within two to three days after harvest, with the highest value being given to the freshest leaves (Altabachew et al. 2013).

The oldest known historical reference to qat is in a book by Nagib Ad–Din that was copied in 1237 A.D., with a note in the margin by a different writer who mentioned that qat was from Kilwa (Tanzania) and Yemen (El Mahi 1962). The earliest credible reference to qat cultivation in Ethiopia is the chronicles of King ‘Amda Syon I, who ruled Ethiopia from 1314–1344 A.D., wherein Sultan Sabra al–Din proclaimed that he would plant qat in the Ethiopian city of Mar’ade (Huntingford 1965). The first known historical mention of qat cultivation in Yemen is from the early 14th century in which Sultan Dāwād denounced qat cultivation (Varisco 1994). Based on these historical records, qat has been cultivated for > 700 years in Ethiopia and Yemen. In addition to the Ethiopian and/or Yemeni origin(s) of qat use, Goldsmith (1988, 1994, 1999) and Carrier (2007) suggested an independent origin of qat cultivation in central Kenya. Late 19th century European explorers noted that qat (known as mirra in Kenya) production was already well established in the Nyambene Hills northeast of Mt Kenya as evidenced by the presence of old trees in cultivation at the time of European contact (Carrier & Gezon 2009). Furthermore, qat use is culturally important to the Bantu–speaking Meru people in the Nyambene Hills, as the ancient qat trees are regarded as symbolic connections to their ancestors (Carrier & Gezon 2009).

From historical references regarding the origin of qat, the following three hypotheses of cultivated qat origin are plausible. First, qat cultivation originated in Yemen and dispersed (via humans trading cuttings and/or seeds) to Ethiopia and Kenya. Second, qat cultivation originated in Ethiopia and spread to both Yemen and Kenya. Third, qat cultivation had independent origins from distinct wild populations in Ethiopia, Kenya, and Yemen (with or without subsequent translocation). Extant wild populations of qat are known in Ethiopia and Kenya (Palgrave 1983; Robson et al. 1994) but have not been described from Yemen. There are suitable habitats in

Yemen for wild qat, however field studies in these areas have not reported wild qat there (Hall et al. 2008).

The process of artificial selection involved with plant cultivation over time will produce cultivars that diverge morphologically and genetically from their wild progenitors (Harlan 1992). Eventually this process of divergence will produce cultivars that cannot survive without human intervention; this continuum of divergence is known as domestication and those cultivars that cannot survive without human intervention are considered fully domesticated (Harlan 1992; Miller & Gross 2011). While the processes that produce this divergence are well known and studied among annual grain crops, they are far less well characterized among long lived perennial crops (Allaby et al. 2008; McKey 2010; Miller & Gross 2011). The degree, processes, and timing by which divergence occurs among long lived perennials have long been debated (e.g., Zohary & Spiegelroy 1975; Harlan 1992; Zohary 2004; Miller & Schaal 2006; Kislev et al. 2006; Coart et al. 2006; Perrier et al. 2011; Bouby et al. 2013).

The two opposing theories regarding divergence of cultivars in long lived perennial species are as follows. First, selection of wild plants occurs once or only a few times from rare individuals that possess desired traits which are stable through time via clonal propagation (e.g., Zohary & Spiegelroy 1975; Harlan 1992; Zohary 2004; Kislev et al. 2006). Alternatively, numerous individuals from several wild populations are selected, followed by hybridization among cultivars and between cultivars and wild populations; both natural and artificial selection act upon a large number of genotypes involved at the early stages of domestication (e.g., Coart et al. 2006; Miller and Schaal 2006; Perrier et al. 2011; Bouby et al. 2013). Two processes of divergence between wild and cultivated types are plausible based on these alternative theories of long–lived perennial domestication. First, divergence occurred by continued natural selection

and extinction acting on the wild populations only while the clonal cultivars remained essentially unchanged. Second, divergence occurred by extinction and selection (both artificial and natural in the cultivated populations) acting on both wild and cultivated gene pools.

The objectives of this study were to determine the wild origins of cultivated qat within the principle historic regions of cultivation as noted from historical writings and to deduce whether divergent cultivars have evolved since the initial cultivation(s) of qat. We applied SSR (simple sequence repeat) markers with model–based and non–parametric clustering methods in order to clarify the genetic structuring of wild collections from Ethiopia, Kenya and Tanzania. Using the best supported partitions among wild–collected individuals as a reference we then examined which wild populations represented the probable progenitors of cultivated qat. Finally, we analyzed the cultivated and wild genotypes together in order to quantify divergence among the two groups and determine whether the patterns of divergence supported either of the opposing theories for long–lived perennial domestication.

Materials and methods

Plant collections — Together with Ethiopian and Kenyan colleagues, Dr.Mark P. Simmons collected 1481 qat specimens in 2009 – 2010 in Ethiopia and Kenya. Collection trips were conducted in collaboration with Addis Ababa University and the National Museums of Kenya. A total of 659 silica–gel–preserved leaf specimens were collected and photographed from Ethiopia and 822 from Kenya. Of these 1481 specimens, 567 cultivated specimens were sampled from 256 Ethiopian farms and 525 cultivated specimens were collected from 217 Kenyan farms (including those in which the farmers reported cultivating wild qat that was either *in situ* or had been translocated). Four wild populations (29 – 37 specimens per population; average of 31) were sampled from Ethiopia (only three had been previously reported for the

entire country), and 20 wild populations (5 – 30 specimens per population; average of 13) were sampled from Kenya. Additional specimens from Ethiopia include those from one feral population (18 specimens), a farm in which wild plants had been brought into cultivation (three specimens), and 26 specimens purchased from three markets (including the largest in Ethiopia at Aweday, near Harar).

In 2012 and 2013 Abdul Wali al Khulaidi of the Yemeni Agricultural Research Authority in Ta'izz, and Mansoor Althobhani of the University of Sana'a, collected 55 cultivated qat specimens from major qat–production areas in Yemen. Additional existing specimens of qat included six cultivated specimens from Madagascar that Mark P. Simmons and colleagues collected in 2006 – 2007, as well as 16 specimens from Tanzania, one specimen from Malawi, one specimen from Zimbabwe and one specimen from South Africa that had been collected by the Missouri Botanical Garden since 1990. Although native to the country, qat is not legally cultivated in Tanzania (Carrier 2007). The complete set of 1561 specimens encompasses the regions in which qat is most intensively cultivated as well as all of the areas mentioned in historical literature as the historical places of origin for cultivated qat.

Mapping of individuals and populations was conducted in Google Earth 5.0 (Google Inc. Mountain View, CA).

DNA extraction, PCR, and microsatellite genotyping — DNA was extracted from herbarium specimens or leaf tissue that was dried in silica gel. We applied a customized DNA extraction process (Appendix 4.1) that combined methods from protocols in De la Cruz et al. (1995), Alexander et al. (2007), and Lemke et al. (2011). This specialized DNA extraction protocol was applied to reduce tannins and other extraction and PCR inhibiting metabolites present in qat. Prior to PCR, DNA extractions were quantified using Quant–iT Pico Green

dsDNA assay (Life Technologies, California, USA) and equilibrated to 100 – 150 ng/µL when initial concentrations allowed.

The 17 microsatellites used for genotyping were selected from Curto et al. (2013) on the basis of consistent amplification, allelic richness, and their ability to be multiplexed. From the 17 primer pairs, PCR bi/triplexes were designed using Multiplex Manager 1.2 (Table 4.1; Holley & Geerts 2009). PCR multiplexes were designed such that two plexed reactions could be combined before fragment analysis to reduce cost and increase efficiency. Forward primers were directly labeled with fluorophores while the reverse primers were synthesized with PIG–tails in order to reduce stutter artifacts when scoring (Brownstein et al. 1996). PCR was conducted in 5 µL reactions with 2.5 µL Qiagen Multiplex Master Mix (Qiagen, Venlo, Netherlands), 0.05 µL multiplex primer set, 1.45 µL molecular grade H_2O, and 1 µL DNA template. Thermocycler conditions were: an initial 15–minute 95°C cycle to activate the Taq polymerase followed by 40 cycles of 30 seconds at 94°C, 90 seconds at 55°C, and 60 seconds at 72°C, with a final extension of 30 minutes at 60°C.

After PCR, 1 µL of the reaction product was diluted 20 times with water, and 1 µL of the diluted product was mixed with 5 µL HiDi Formamide (Applied BioSystems,Waltham, Massachusetts, USA) containing 10 µL/mL of GeneScan 500LIZ Size Standard (Applied BioSystems). The samples were run on either a 96–capillary, 50–cm array 3730XL DNA Analyzer or a 16 capillary 50 cm array 3130 DNA Analyzer (Applied BioSystems). All raw ABI files where examined and binned initially in GeneMapper 5.0 (Applied BioSystems) using the default microsatellite settings. Binning calls were then manually examined and all identified mis–calls were re–scored before further analyses. Bin fidelity and mutant alleles that do not

Table 4.1. Primer sequences, SSR motif, and multiplex design.

Primers (from Curto et al. 2013)	PCR plex	Capillary plex	Motif	Forward Sequence (with florophore)	Reverse Sequence (with GTTTCTT PIG-tail)
CE56	1	A	GA	PET-acaaccaatgtcgtacagaga	gtttcttcagtatttgtatctgcagtacag
CE45	2	A	AG	6-FAM-ggtttagcctctcctgcaagt	gtttcttcaagtgcggactcaacaagt
CE50	2	A	CT	6-FAM-aactacccgccatttcgac	gtttcttagctgggcgattgactaaag
CE37	2	A	ATCT	VIC-actcgaaaaacatggtgcag	gtttctttgagcctcaatctggagaca
CE23	3	B	ATC	PET-agcagcagcaacaacaagaa	gtttcttcagcaagggaggccttatta
CE43	3	B	ATT	VIC-cagatcctcctcctccttca	gtttcttggatgccacaaatgctgat
CE24	4	B	AC	NED-tcttgctccttcaacctcaa	gtttcttatcttgccagcttccgtcta
CE29	4	B	AAT	6-FAM-gccaacctcttgttctggag	gtttctttaggtttggccattcgattc
CE40	5	C	GT	6-FAM-tggtatagcccatatcgtcag	gtttcttcacactacgcttcacgcttc
CE31	5	C	AC	VIC-ttcccaaaagtgttgctgag	gtttcttctttactagggcccgtcctt
CE34	5	C	CT	NED-cggatgccaaaacactatca	gtttcttatccaagaggttttggttgc
CE22	6	C	AG	PET-gctgcaagaggtaggtggat	gtttcttttctctctctcgcttggcta
CE42	6	C	GT	VIC-aaggggaaggagagagatgc	gtttcttcctcatcctgatgtggctaa
CE41	6	C	GA	6-FAM-ggacagaattcccaaaacgac	gtttctttttacattattgccagctcgatc
CE3	7	D	CT	6-FAM-ccttctatcaccctcccaca	gtttcttccctctgtattgcacggttt
CE57	7	D	TTA	VIC-tcctgtcctgataatatcctg	gtttcttagtcccactgattgttatgac
CE29	8	D	AAT	PET-gccaacctcttgttctggag	gtttctttaggtttggccattcgattc

follow the expected SSR–motif–length pattern) were checked using Flexibin 2.0 (Amos et al. 2007).

Total error rate across all loci was calculated with four redundant genotypes on each 384–well plate by dividing the number of incorrect calls by the number of total calls across all redundant genotypes and loci. The error rate across all genotypes was calculated by including the primers CE3 and CE42 in two different PCR multiplexes and comparing the allele calls across all genotypes and dividing the total number of different calls by the number of total calls. Any individual with 17% or more missing data was removed from the final data set. The total percentage of missing data was kept below 5% based on simulations demonstrating that levels of missing data above 5% can negatively influence genetic–clustering methods (Reeves et al. submitted).

Genetic data analysis — Clones were identified from the dataset and tested for chance genotypic redundancy from sexual reproduction using the P_{sex} calculation with 10,000 simulations in MLGsim 2.0 (Stenberg et al. 2003). The P_{sex} calculation is a test statistic that is used to assess whether redundant genotypes are the result of chance sexual reproduction or truly clonal based on simulated recombination from the observed allele frequencies. In addition, genotypes with missing data and/or differing by a single allele were added to the clonal groups identified through MLGsim. By including genotypes differing at a single allele we accounted for the error rate calculated in our dataset. Clonal groups were then given unique numeric designations as well as the codes ‘M’ if they included genotypes with missing data and ‘E’ if they included genotypes differing by a single allele. Data quality of the final set of wild genotypes was assessed using Global Linkage Disequilibrium tests and Hardy Weinberg

Equilibrium as implemented in FSTAT 2.9.3.2 (Goudet 1995) and GeneAlex 6.5 (Peakall & Smouse 2012).

Population divisions among the collected genotypes were assessed in STRUCTURE 2.3.4 (Pritchard et al. 2000; Falush et al. 2003, 2007; Hubisz et al. 2009). STRUCTURE is a model based approach that employs a Bayesian algorithm to assign genotypes to clusters that assume HW equilibrium and complete linkage between loci within clusters (K). It is known that non–sexual mating systems, such as clonal reproduction, can cause deviations from HW and thus inferences based on these assumptions. Therefore we primarily relied on STRUCTURE to assess wild population structuring in qat given that clonal propagation is common place among qat farmers.

A range of K values from 1 to 24 were assessed from the 24 Ethiopian and Kenyan wild collection sites, conducting 40 simulations per K value with 30,000 iterations per simulation (the first 10,000 as burnin) using the admixture model. The optimal value for K was determined using the ΔK method (Evanno et al. 2005) as implemented in Structure Harvester 0.6.94 (Earl & vonHoldt 2012). Optimal values of K from the ΔK method were further assessed through examination of simulation congruence in CLUMPAK 1.1 (Kopelman in press).

Substructuring within each of the clusters identified using the most highly supported value of K was assessed using Rosenberg et al.'s (2001) method for breed identification, which we extended to phylogeographic partitioning. This was performed by running STRUCTURE again for each individual cluster to identify any patterns of subclustering. STRUCTURE parameters were identical across all runs with the exception that a smaller range of K values was used to potentially identify substructuring within clusters.

The STRUCTURE partitionings of the wild populations were used as the reference clusters for which to assign the cultivated genotypes. Individual assignment tests were conducted in GeneClass 2.0 (Piry et al. 2004) using the Bayesian assignment method of Baudouin & Lebrun (2001). Assignment of cultivated genotypes with a cutoff of 0.95 posterior probability was first done with all cultivated genotypes in our dataset to the most highly supported set of genetic clusters, followed by intra–cluster assignment of cultivars using only the cultivated genotypes assigned to the initial cluster. In this way more specific geographic and genetic centers of origin within a cluster were identified.

Discovery of hybrid genotypes was conducted first by examination of STRUCTURE results in which cultivated genotypes were included with all wild progenitor populations. Those genotypes that had nearly equal assignment to two wild clusters or subclusters were treated as putative hybrids (Vähä & Primmer 2006). The putative hybrids from STRUCTURE were then checked for the presence of diagnostic alleles that might elucidate the parentage of a given hybrid. Genotypes geographically adjacent to putative hybrids were examined to check if they possessed sets of alleles that could produce the genotype of those putative hybrids. Genetic diversity and partitioning results were also assessed and visualized using the non–parametric approach of PCoA (Principle Coordinate Analysis) as implemented in GeneAlex.

Results

Origins of cultivated qat — The final dataset included 1561 individuals, genotyped for up to 17 microsatellite loci, with 2.5% missing data and an estimated error rate of 2.4% based on replicate genotypes and two replicate primer pairs. From the best supported partitioning of the wild qat genetic data three clusters were inferred (Fig 4.2). The three clusters are as follows: 1) genotypes from four wild collection sites in the highlands of southern Ethiopia, 2) genotypes

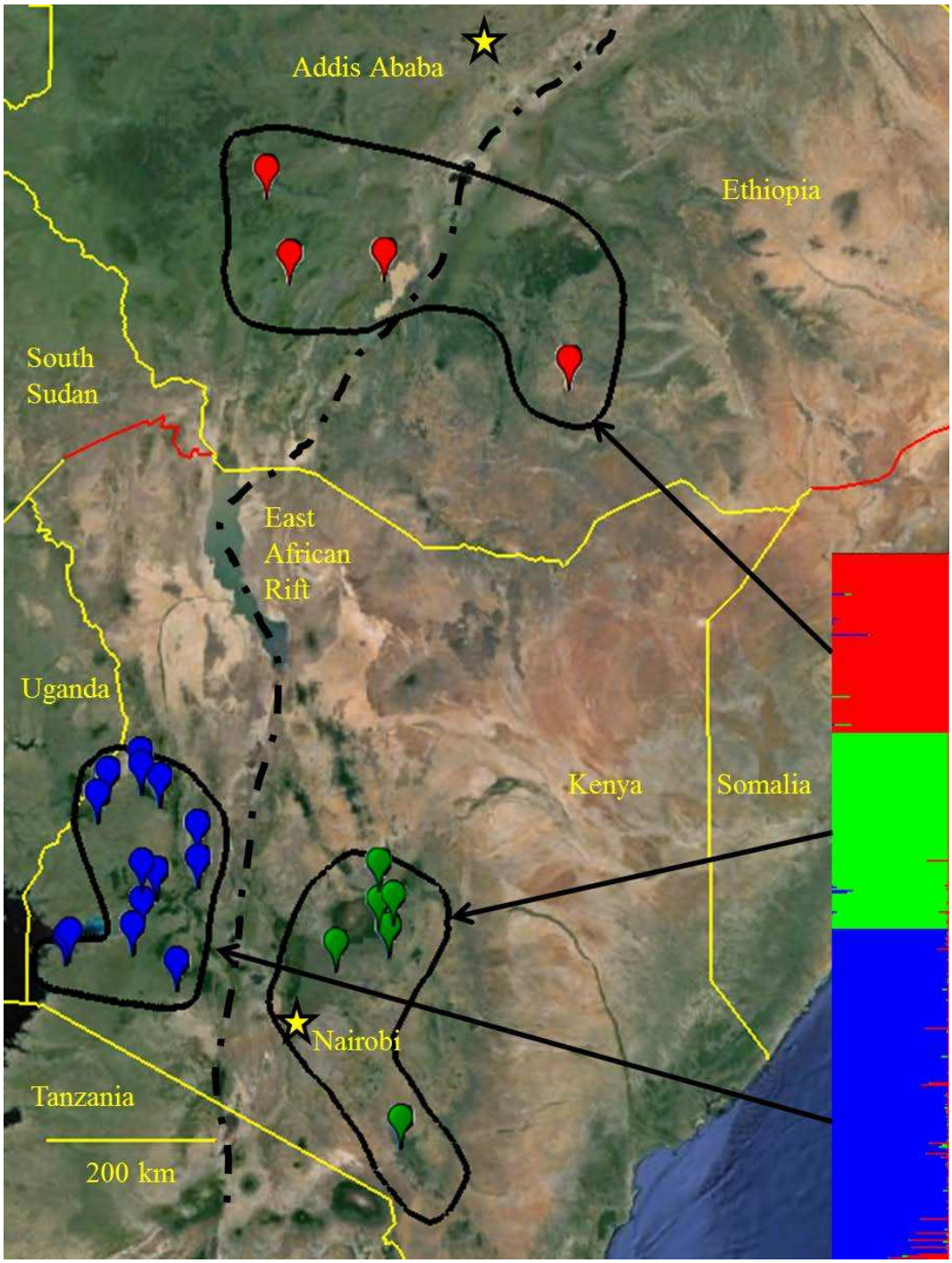


Figure 4.2. Wild qat collection sites from Ethiopia and Kenya with the highest supported genetic partitioning as determined through multiple STRUCTURE runs as indicated by bar plot in lower right. The East African Rift is indicated by the dashed line.

from six wild collection sites around Meru/Mt Kenya and adjacent to Mt Kilimanjaro on the eastern side of the EAR (East African Rift), and 3) genotypes from the 12 wild collection sites from the eastern slopes of Mt Elgon in the north across the Endebess Bluff south to the Mau Forest on the western side of the EAR. The Malawian, South African, Tanzanian and Zimbabwean genotypes were weakly supported as being clustered with the Mt Elgon cluster on the west side of the EAR. Because of the small number and dispersed geographic collection of these samples, the power to assign these genotypes to a separate cluster was limited.

From the ΔK analysis (Evanno et al. 2005) evidence for genetic substructuring was indicated by peaks at values of K = 5, 7, and 11 (Fig. 4.3). Given this we explored patterns of hierarchical substructuring within each of the three clusters. Each of the three clusters was found to have a pattern of substructuring consistent with geographic features. The Mt Elgon cluster west of the EAR was split into two subclusters that are geographically separated by a lower elevation area across the Endebess Bluff (Fig. 4.4). The Mt Kenya/Meru cluster east of the EAR was split into two subclusters, each occurring on higher elevation uplift features. The southern subcluster consisted of genotypes collected near Mt Kilimanjaro, and the northern subcluster consisted of genotypes from five collection sites around the southern flanks of Mt Kenya to the Meru area (Fig. 4.4). The Ethiopian cluster was split into two subclusters consisting of two collection sites on the west side of the EAR near Lake Abaya and two collection sites that span the EAR to the north (Fig. 4.4).

All but three cultivated genotypes assigned to one of the three genetic clusters at the 0.95 PP (posterior probability) cutoff. Of the three genotypes that did not assign, two where later inferred to be hybrids (see below). The remaining unassigned genotype was collected from the Nyambene area in Kenya. When the Tanzanian genotypes were included in the reference set this

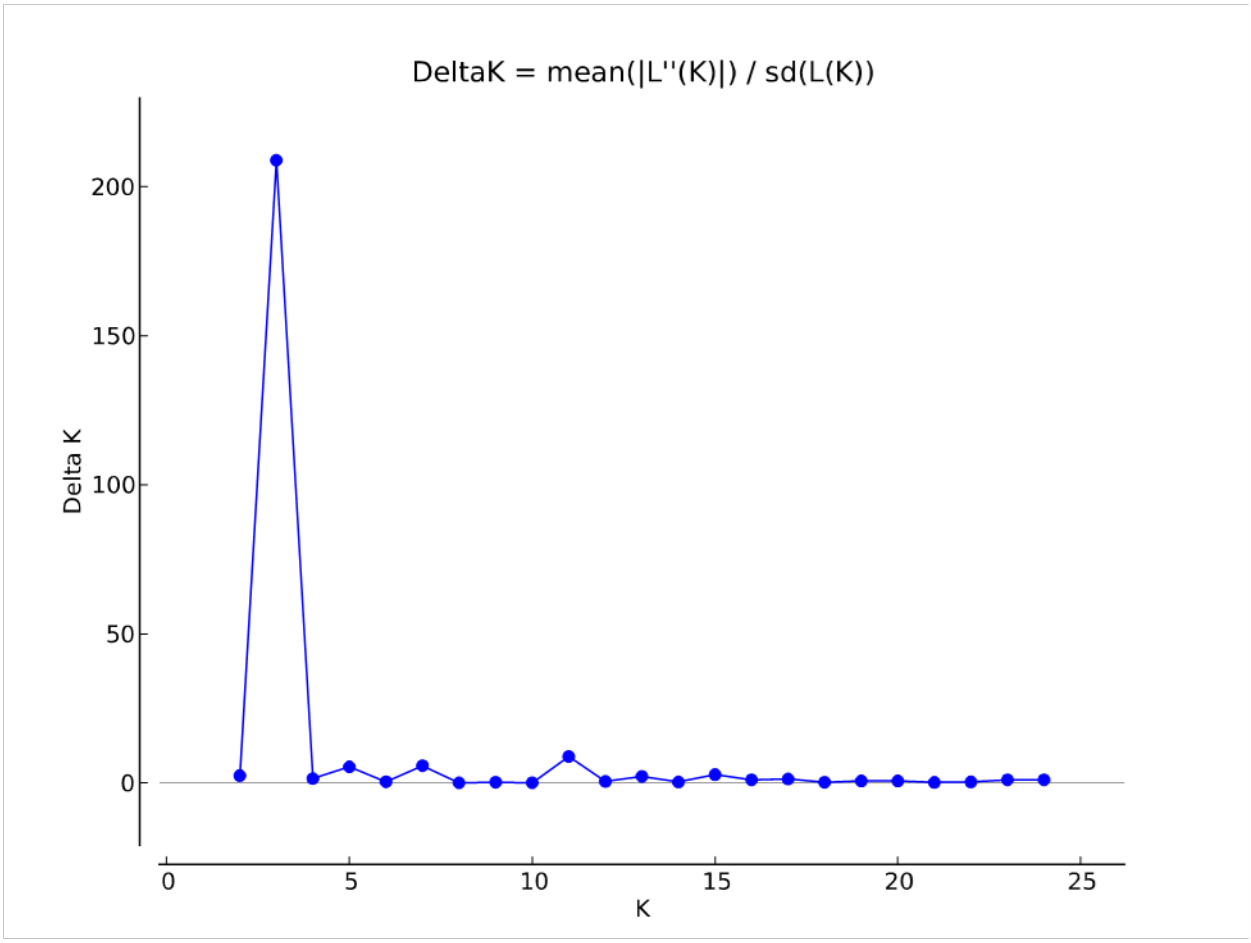


Figure 4.3. Delta K results for all wild samples indicating the best supported partition of K=3 delimiting the Ethiopian, Mt Kenya/Meru, and Mt Elgon collection sites. Other peaks are at K = 5, 7, and 11.

genotype assigned to the Tanzanian set with 0.99 PP. The number of unassigned genotypes increased when assignments were conducted at the subcluster level. This reduction in assignment percentage is expected as the range of allele frequencies is reduced due to the smaller size of the subpartitioned reference sets (e.g., Cain et al. 2000).

The three wild clusters have been differentially sampled from in the founding of qat cultivars. One hundred three cultivar genotypes were assigned to the cluster west of the EAR in Kenya, and within this cluster 34 were assigned to the northern subcluster and 58 to the southern

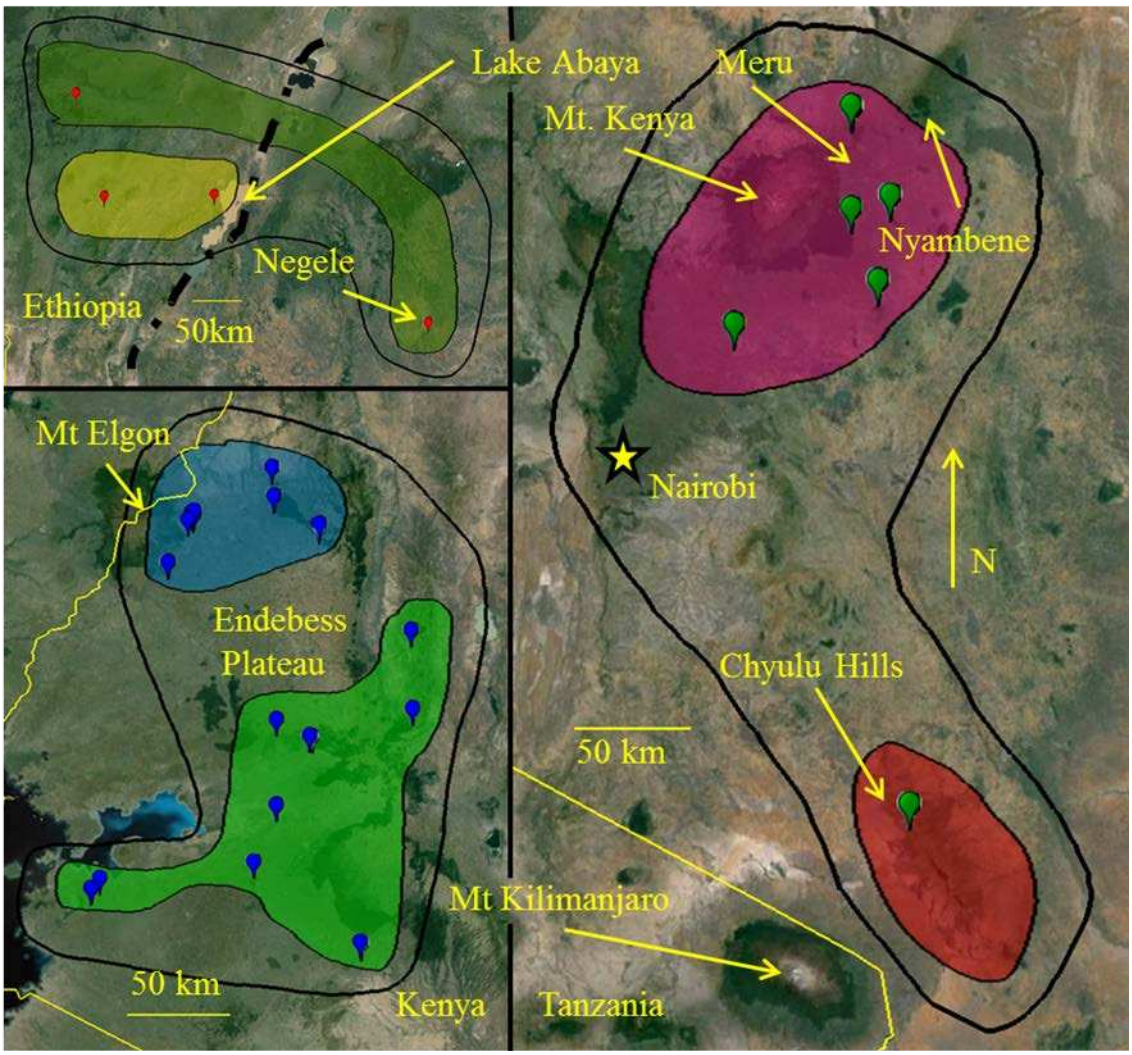


Figure 4.4 Genetic sub–clusters within each of the three most highly supported partitions.

subcluster, leaving nine cultivar genotypes that do not assign to either subcluster (Fig. 4.5). Two hundred five cultivar genotypes were assigned to the cluster east of the EAR in Kenya and within this cluster 185 were assigned to the northern subcluster around Mt Kenya/Meru, and 18 to a southern subcluster near Mt Kilimanjaro in the Chyulu Hills area, with one cultivar genotype not assigning to either subcluster (Fig. 4.5). Two hundred ninety three cultivar genotypes were assigned to an Ethiopian origin and within this cluster 54 were assigned to the northern subcluster, 198 to the southern subcluster west of Lake Abaya (Fig. 4.5) and 41 cultivar genotypes did not assign to either subcluster.

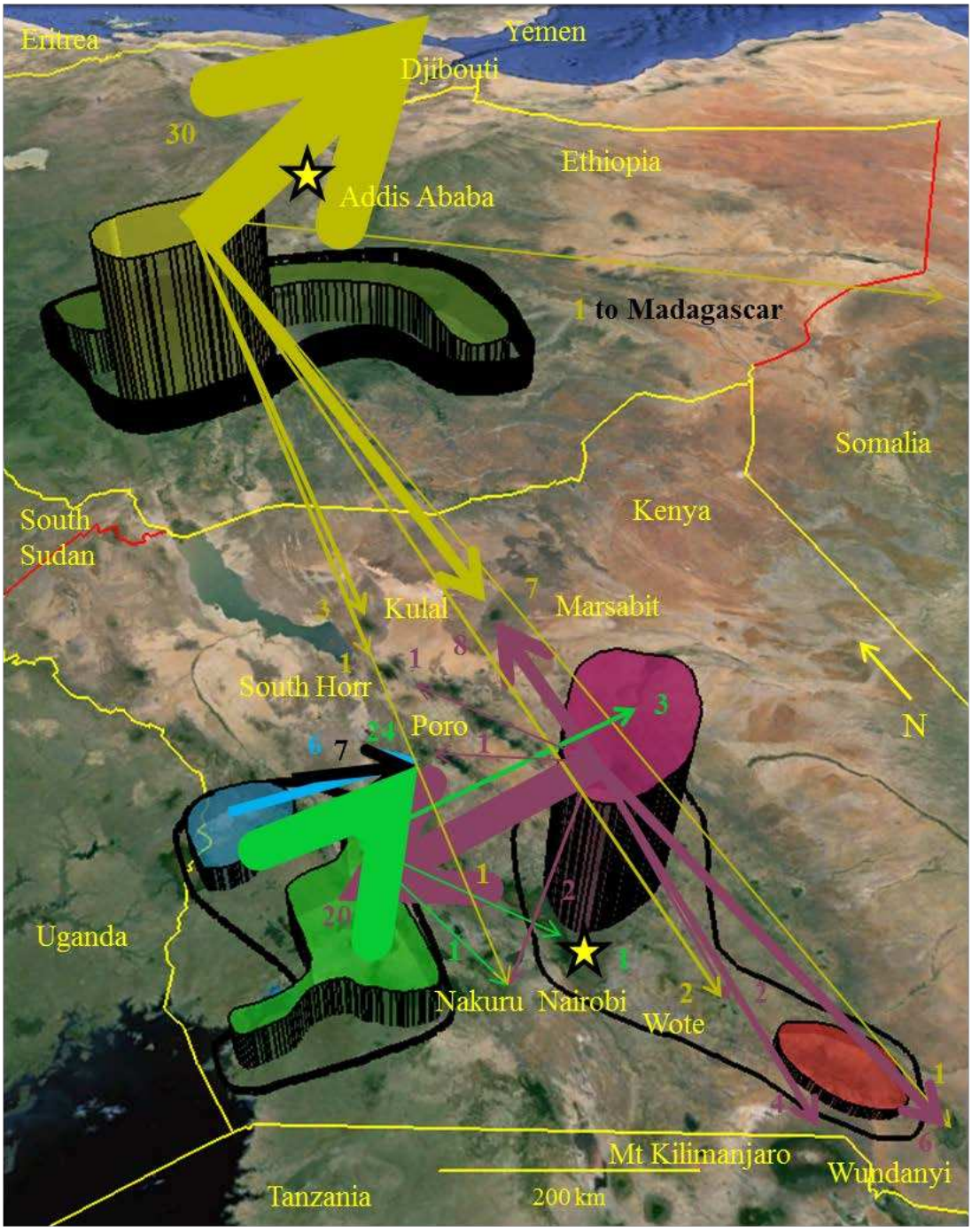


Figure 4.5 The contribution of each wild cluster and subcluster to the number of cultivated genotypes as well as translocations. Height of subclusters is proportional to the number of cultivated genotypes assigned to each. Height of the outline around subclusters is proportional to the number of genotypes assigned to the larger cluster but not assignable to any subcluster. Translocations are indicated by arrows originating within subclusters. Width of arrows is proportional to the number of translocated genotypes; numbers (color coded by origin) by arrows indicate the number of translocated genotypes.

Dispersal of cultivated qat — Ethiopia is the source of the greatest number of translocations to other regions, most notably Yemen (Fig. 4.5). Thirty Yemeni cultivar genotypes assigned to the Ethiopian cluster and all but one assigned to the southern subcluster [the unassigned genotype also had the highest PP (0.63) of assignment to the southern subcluster]. We applied four different methods to test for an independent Yemeni origin of Yemeni cultivated qat. First, as noted above, genetic clustering with assignment did not support a separate Yemeni origin. Second, we calculated Goldstein et al.'s (1995) linear genetic distance between the two southern progenitor populations in Ethiopia to the cultivated genotypes in Ethiopia that were assigned to this subcluster and to the cultivated genotypes in Yemen. The mean genetic distance between the Ethiopian wild source populations and Ethiopian cultivars is 6.1, whereas it is 5.0 between Ethiopian wild source populations and Yemeni cultivars. Thus, in the context of the genetic distances between groups, the Yemeni cultivated material is not clearly distinguishable from Ethiopian cultivated material. Third, in order to test the possibility that extant Yemeni cultivated qat had both a Yemeni origin and an Ethiopian origin, PCoA (Fig. 4.6) was applied to identify any outliers among the Yemeni genotypes. In the PCoA analysis the Yemeni genotypes clustered with Ethiopian cultivars, providing additional support for an Ethiopian origin of Yemeni cultivated qat. Fourth, no private alleles were identified in any of the Yemeni cultivars. But three Yemeni genotypes possessed allele 234 at locus CE22, which was absent among Ethiopian genotypes but present at 0.37 among all Kenyan genotypes. This shared allele may be the result of a convergent mutation between an Ethiopian lineage that was translocated to Yemen, introgression from Kenyan qat to Yemeni cultivated qat, or extinction of the allele from the genotypes currently in Ethiopia. In comparing allele size distributions between wild Ethiopian and wild Kenyan qat populations, allele 234 in Yemeni cultivated qat could be the

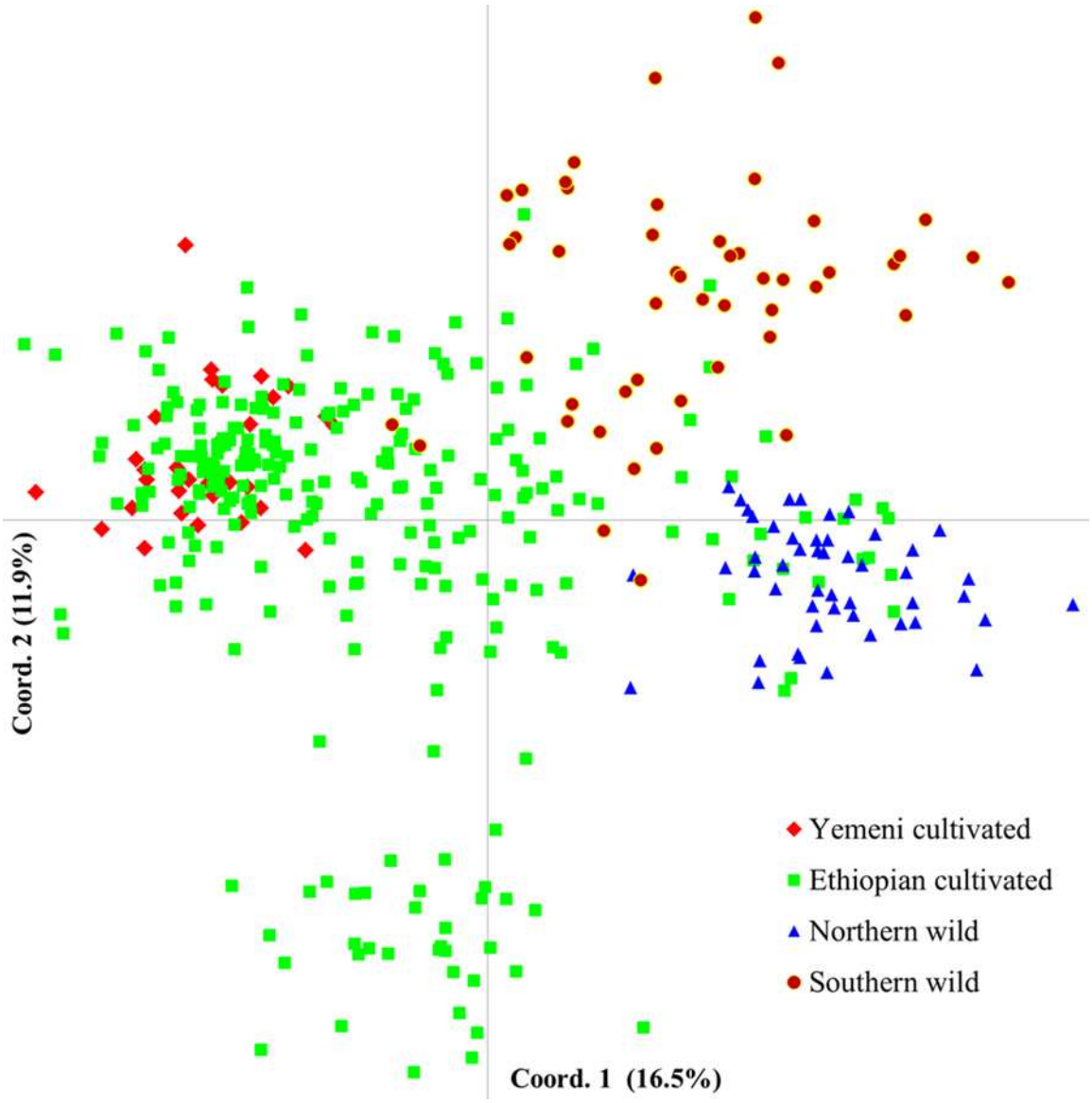


Figure 4.6. PCoA plot of Ethiopian wild and cultivated genotypes with Yemeni cultivated genotypes (red diamonds) included to demonstrate the Ethiopian origin of Yemeni qat.

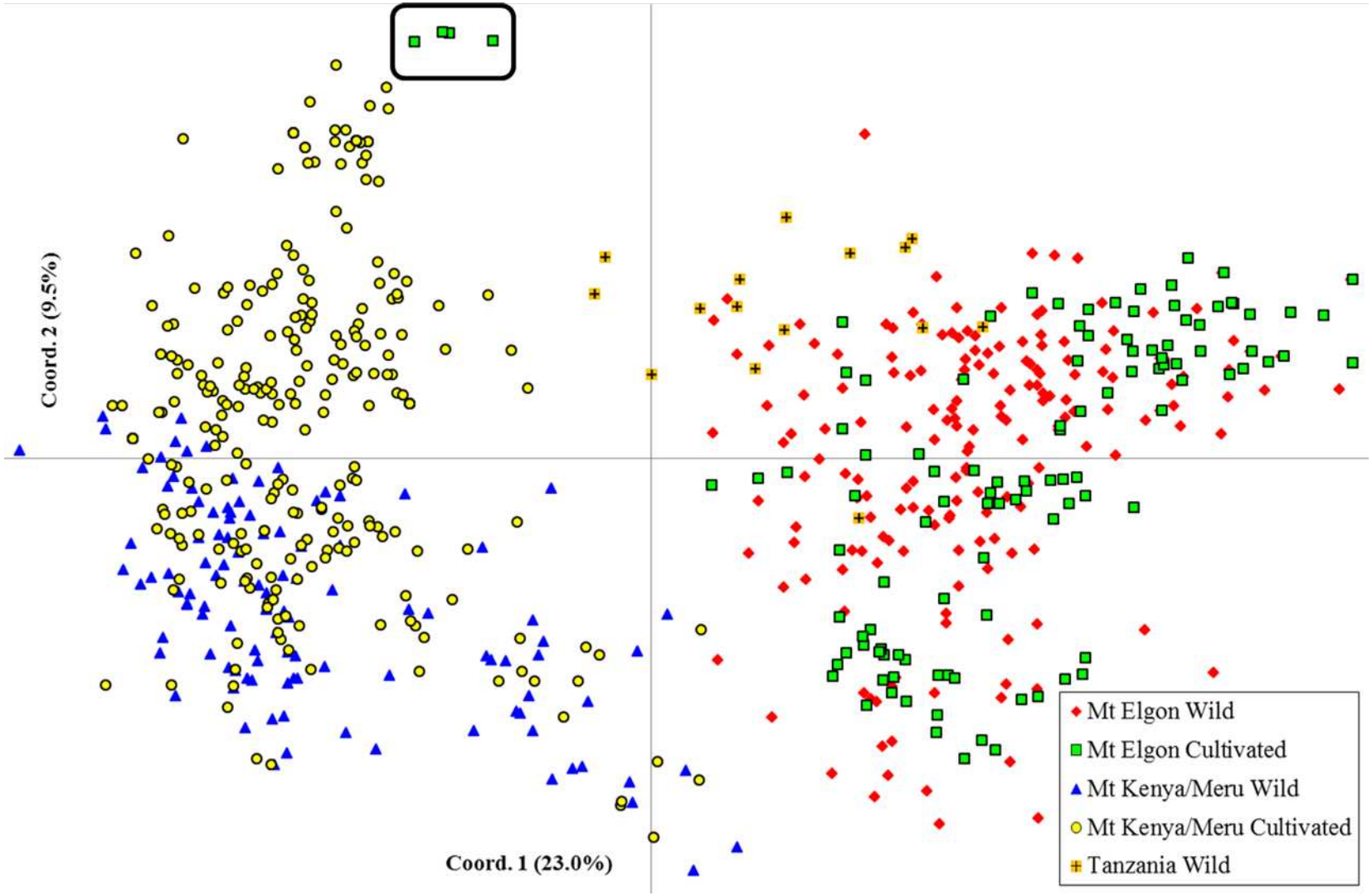
result of a convergent stepwise mutation from allele 232, which is present in Ethiopia, to allele 234 (Appendix 4.2). In any case the shared allele does not provide substantial evidence of a Kenyan origin for Yemeni cultivated qat. Translocations of cultivated qat from Ethiopia into Kenya were identified. Eleven cultivar genotypes found in Kenya were assigned to the southern

Ethiopian subcluster. In Madagascar all six individuals sampled were from a single clonal genotype that assigned to the southern Ethiopian subcluster (Fig. 4.5).

We did not identify any Kenyan cultivated genotypes that had been translocated to other countries that we sampled based on the assignment tests. But translocations across the EAR and to other parts of Kenya were rather common (Fig. 4.5). From the southern subcluster of the Mt Elgon cluster, two genotypes were found in the Nyambene area northeast of Meru, one genotype near Nairobi, one genotype in the city of Nakuru in the EAR, and 24 near the village of Poro in Samburu County. Two Poro farmers stated that cultivated individuals were translocated from unsampled wild populations from regions near Poro. The Poro genotypes assigned to the Mt Elgon cluster because the wild populations from which they were selected were not part of the wild reference set and the wild Mt Elgon genotypes are most closely related to the Poro genotypes than the Mt Kenya/Meru genotypes. The three genotypes translocated to Nyambene from the southern subcluster may actually be from an unsampled or distant population given that they do no group closely with any Kenyan clusters in PCoA plots (Fig. 4.7). No long distance translocations were identified as originating from the northern subcluster of the Mt Elgon cluster except for seven genotypes found in Poro, which as mentioned above may be the result of locally translocated wild collections.

Only localized cultivation was found for the 18 cultivated genotypes that were assigned to the southern subcluster on the eastern side of the EAR near Mt Kilimanjaro. The northern subcluster near Mt Kenya/Meru was the source of the greatest number of translocated genotypes in Kenya with 20 cultivated genotypes collected from the western side of the EAR (Fig 4.5). Eight genotypes from the northern subcluster were collected in the Mt Marsabit area, where farmers noted that cuttings had been obtained from Meru for the establishment of their orchards.

Figure 4.7. PCoA plot of the Mt Elgon and Mt Kenya/Meru centers of origin, and Tanzanian wild genotypes. Boxed points are genotypes collected near Meru but assigning to the Mt Elgon cluster.



In addition, one translocated genotype was found in Poro, one in South Horr, and one in Wamba north of Mt Kenya. Translocation into southern Kenya from the northern subcluster was represented by four genotypes near Mt Kilimanjaro, six genotypes near the town of Wundanyi east of Mt Kilimanjaro and two genotypes near the town of Wote ~150 kilometers north of Mt Kilimanjaro (Fig. 4.5). Translocation into the EAR from the northern subcluster was represented by two genotypes collected in the city of Nakuru. The northern subcluster within the Mt Kenya/Meru area has the greatest sampled genotypic diversity in Kenya and is the source of nearly all Kenyan translocated material. This pattern of translocation corroborates Kenyan farmer interviews wherein Meru was mentioned as the most common source of qat for starting their farms.

Hybridization among cultivated qat genotypes — Clear examples of hybridization were found among cultivated genotypes in the isolated highland areas of Marsabit and Kulal in northern Kenya, between the southern edge of the Ethiopian highlands and the northern edge of the Mt Kenya uplift area. Hybrids from Marsabit and Kulal were inferred based on STRUCTURE results in which genotypes had a near 50% proportioning of assignment to an Ethiopian and Kenyan origin together with an intermediate placement between Ethiopian and Kenyan genoptypes in a PCoA plot (Fig. 4.8). Furthermore, when considering only the genotypes found in Marsabit, 19 private alleles were found among the genotypes of a Kenyan origin and 15 private alleles among the genotypes of an Ethiopian origin. These 34 private alleles are from the loci CE22, CE37, CE41, CE42, and CE43. All hybrids from Marsabit and Kulal were heterozygous at these five loci, possessing one of each private allele (Appendix 4.3). Three hybrid genotypes were also found in Kulal but we hypothesize that they did not originate from

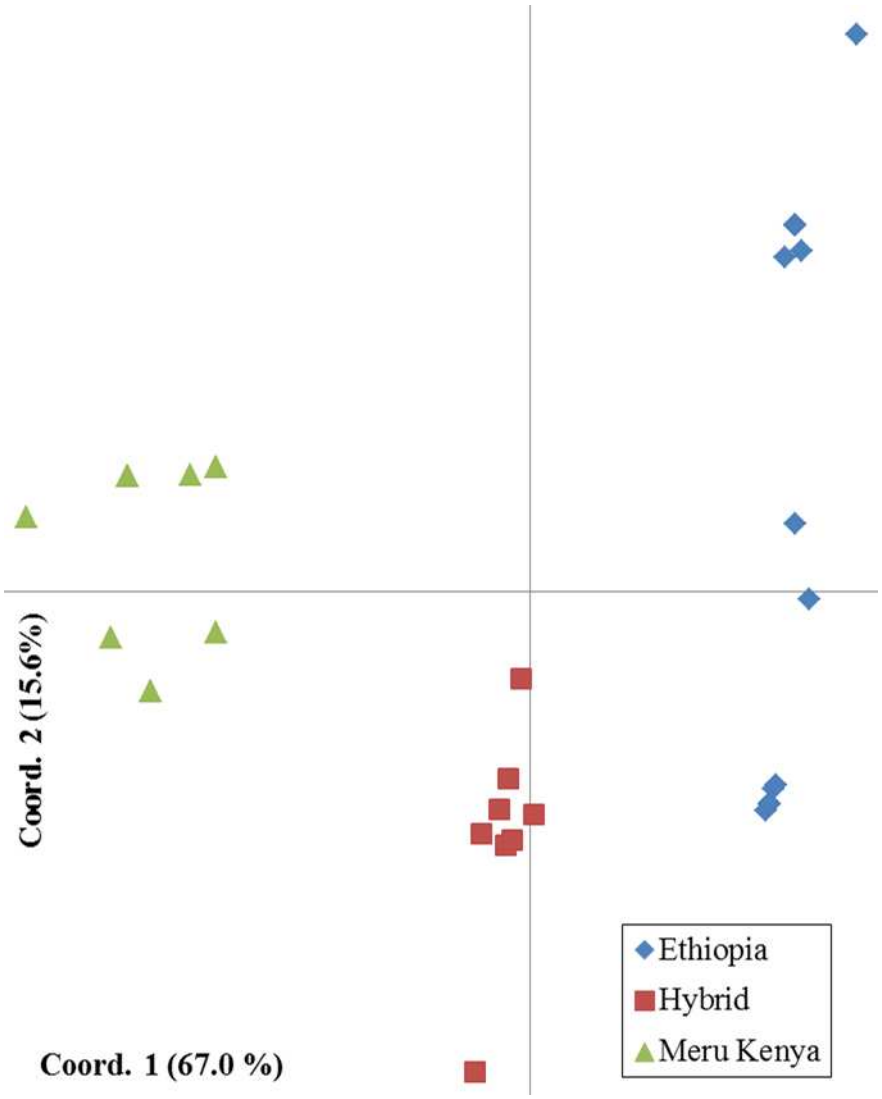


Figure 4.8. PCoA plot of all genotypes collected from Marsabit and Kulal. Hybrid genotypes are inferred from their placement between the Ethiopian and Kenyan genotypes.

Kulal given that only hybrid genotypes and Ethiopian genotypes were collected there. Farmer interviews also provided insights into the origin of hybrid genotypes (see below).

Cultivated qat in Marsabit was represented by seven Ethiopian genotypes, eight Mt Kenya/Meru genotypes, and four hybrid genotypes. In Kulal the cultivated qat was represented by three Ethiopian genotypes and three hybrid genotypes (Fig 4.8).

Divergence between wild and cultivated qat — In order to assess the divergence of cultivated populations from wild progenitors we combined cultivated genotypes (minus clonal repeats to avoid bias among the model based methods) with wild progenitors in an assessment of population structure, both across the entire sampling area and within each of three origins of qat cultivation. When analyzing the cultivated and wild genotypes together across the entire sampling area the most highly supported partitioning of the data from the STRUCTURE analysis is K = 3, which is similar to that for the wild–only analysis of population structuring (Fig. 4.2).

When analyzing only the wild and cultivated genotypes assigning to the Ethiopian center of origin, a pattern of divergence between cultivated and wild populations is clear at the highest supported partitioning of K = 2 (Fig. 4.9 A). At K = 2, 114 (0.95 proportion of genotype assigning to a given cluster) cultivated genotypes are separated from the remaining wild and cultivated genotypes. From the STRUCTURE results a weaker pattern of divergence is found with the cultivated and wild genotypes from the two Kenyan centers of origin. In these instances the cultivated genotypes cluster with the wild genotypes of origin at the highest supported K value of two (Fig. 4.9 B and C). However evidence for the divergence of cultivated from wild genotypes is present in the PCoA plot for the Mt Kenya/Meru cluster (Fig. 4.9 B).

In exploring a range of K values, 99 of the 205 cultivated genotypes from the Mt Kenya/Meru origin form a novel cluster in all simulations at K = 3, whereas the clustering of cultivated and wild genotypes together in the PCoA plots from the Mt Elgon origin did not occur until K = 4 and at that point differing solutions among the simulations did not always resolve cultivated genotypes as a separate cluster (Fig. 4.10). This pattern indicates that the genetic signal of divergence between the wild and cultivated genotypes in Ethiopia is strongest, moderate in Mt Kenya/Meru and weakest in Mt Elgon.

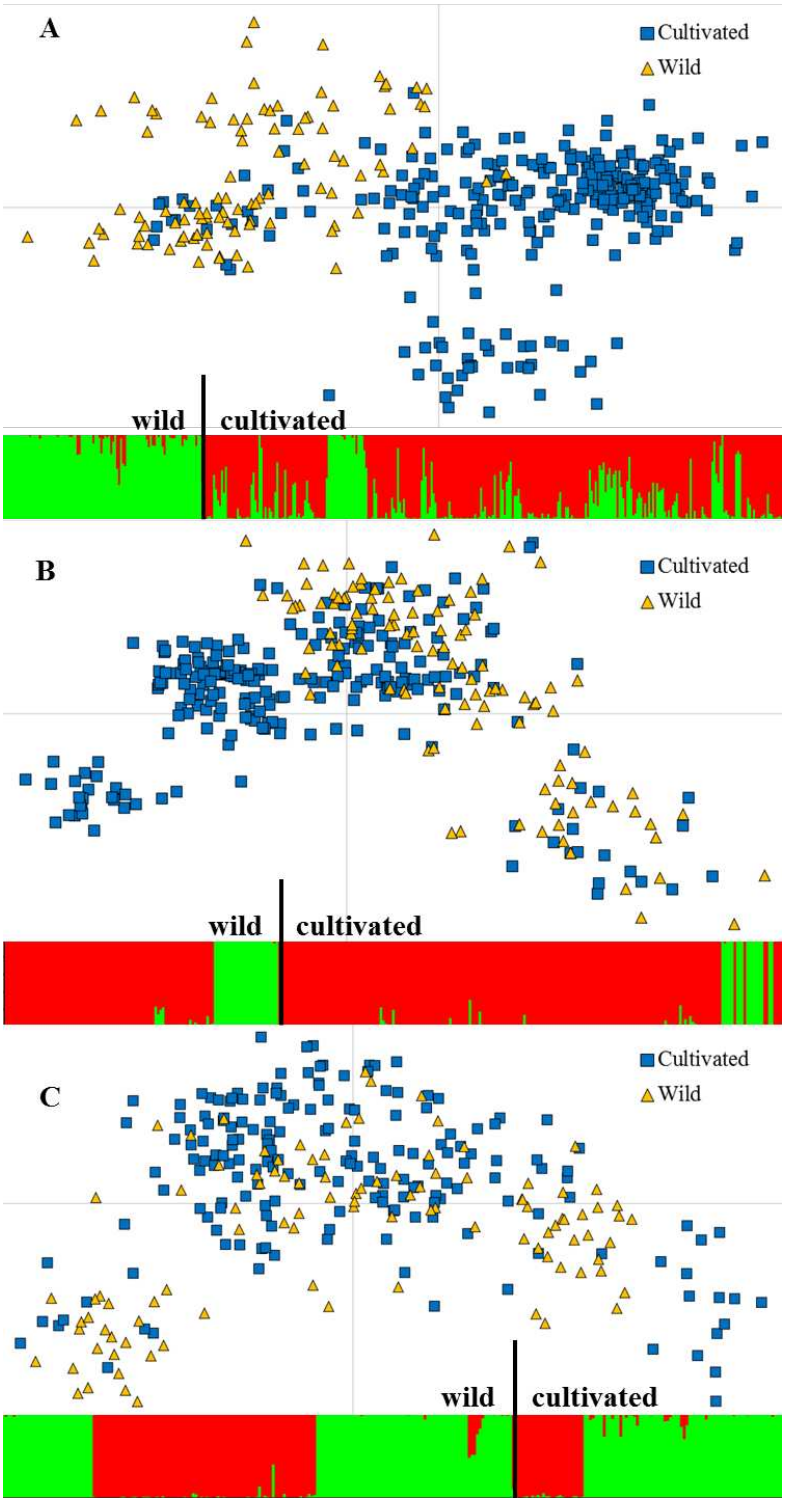
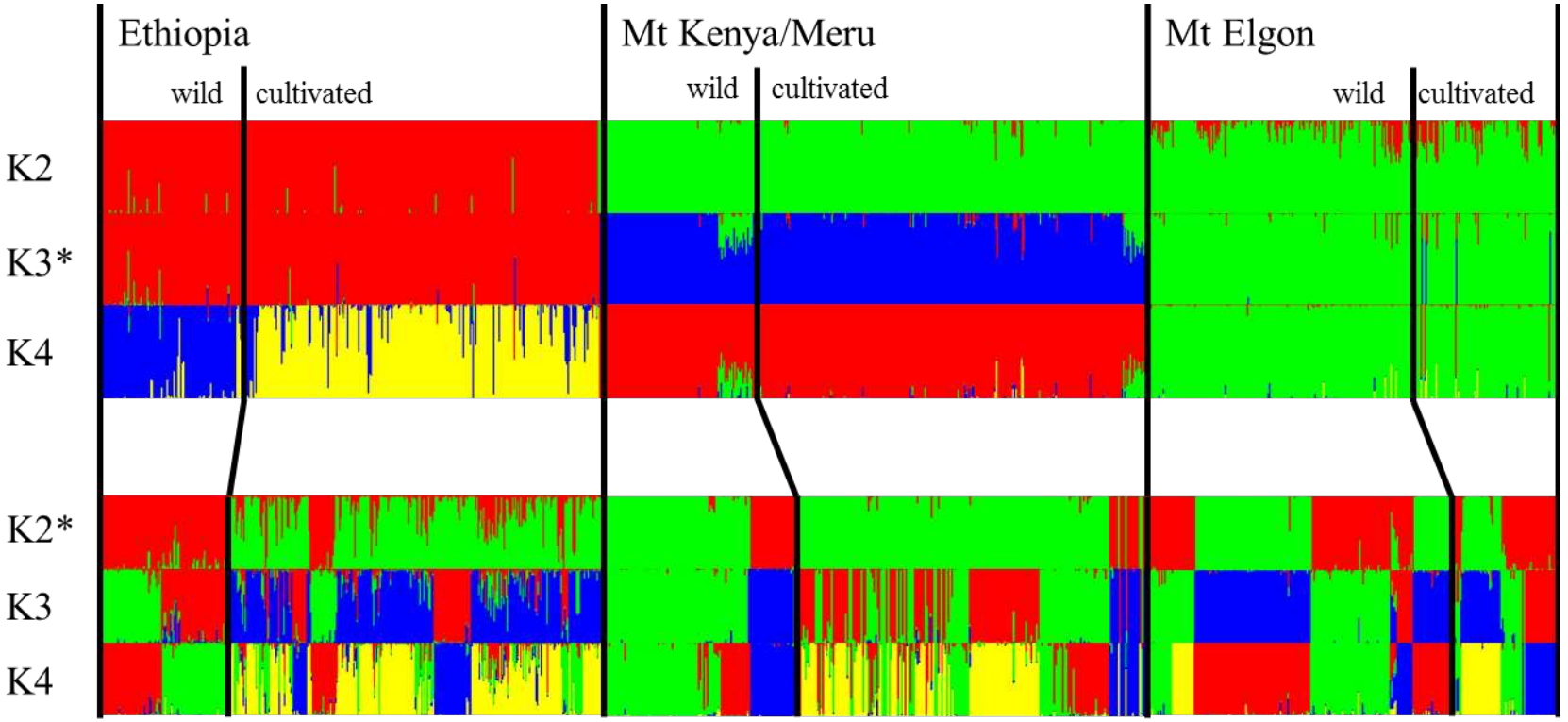


Figure 4.9. PCoA and STRUCTURE barplots for each of the three centers of origin with wild samples as well as cultivated samples that were assigned to that center of origin. PCoA plots are color coded by cultivated and wild status, and STRUCTURE bar plots are color coded by cluster membership. A) The Ethiopian center of origin with both methods indicating divergence, B) Mt Kenya/Meru center of origin with PCoA indicating divergence, and C) Mt Elgon center of origin with neither method indicating a strong pattern of divergence.

Figure 4.10. STRUCTURE barplots for wild and cultivated genotypes for the entire dataset (upper set of barplots) and within each center of origin (lower set of barplots). Each level indicates multiple different partitioning solutions (best supported solution indicated by *) such that the support for the divergence of cultivated from wild genotypes can be assessed.



Discussion

Genetic structuring reveals multiple centers of origin for cultivated qat — The phylogeographic patterns for wild qat populations are similar to the findings for other studies examining genetic patterns within the Afromontane flora (e.g., Kebede et al. 2007; Kadu et al 2013; Mairal et al. 2015). For instance, Kebede et al.'s (2007) study of genetic structuring in *Lobelia giberroa* Hermsl. (Campanulaceae) and palynological work (e.g., Vincens 2005) indicate that forest habitat "bridges" that connected high altitude areas across the EAR in Kenya began to contract with interglacial drying around 5500 calibrated years before present. Similarly, this trend of glacial–period–habitat–bridge recession was also inferred to affect plant species population structuring between the Ethiopian highlands and the mountainous areas of south central Kenya (e.g., Kebede et al. 2007). Thus a similar biogeographic explanation probably applies to the broad scale pattern in which wild qat populations are not connected by gene flow during warmer, dry periods.

The population structure found among wild qat genotypes indicates that the EAR and other geographic features are important barriers in limiting gene flow between populations. The four Kenyan subclusters delimited by STRUCTURE each consist of genotypes associated with a given uplift feature, indicating persistent isolation effects between these uplift areas. The exception to this biogeographic patterning among subclusters is two collection sites in Ethiopia that form a subcluster but are separated by the EAR (Fig. 4.4). This pattern might be attributable to more recent dispersal and gene flow across the EAR in Ethiopia and/or fewer wild populations available for collection. Whatever the case the northern subcluster was not sampled as intensively in the founding of cultivated qat, with 68% of the applicable cultivar genotypes being assigned to the southern subcluster. Furthermore, when using individual wild collection sites as

reference sets for assignments, the Negele site to the east of the EAR was not inferred to be an origin for any Ethiopian cultivated genotypes. Based on this evidence, we infer that Ethiopian cultivated qat was mainly derived from wild populations to the west of the EAR.

The Ethiopian and Mt Kenya/Meru genetic clusters have historically been the primary sources for cultivated genotypes based on genetic analyses of cultivated individuals that we sampled. Wild genetic diversity, long term cultivation and the system of farming (e.g., degree of trade between farmers and propagation methods) are thought to be key factors in producing genotypically diverse cultivars (Harlan 1992). This dynamic between wild diversity and cultivation intensity is demonstrated by comparison of the Mt Elgon and Mt Kenya/Meru centers of origin. Wild genotypic diversity is very high in the Mt Elgon center of origin, yet twice as many cultivar genotypes (205 vs. 103) originated from the Mt Kenya/Meru center of origin, which indicates longer and/or more intensive cultivation from that area. Our genetically based finding of multiple origins for qat in Kenya and Ethiopia corroborates earlier anthropologically–based assertions (e.g., Huntingford 1965; Carrier 2007).

Wild qat genotypes that overlap with cultivated genotypes in PCoA plots are evidence for a relatively recent initial cultivation or ongoing introgression with wild genotypes. By contrast, ‘ancient’ crops often cannot be clearly matched to extant wild populations because the wild progenitor populations are either extinct or the cultivated and wild populations have diverged to a degree that confident assignments are not possible (e.g., Ladizinsky 1999; Denham et al. 2003). Based on this expected pattern, Ethiopia and the Mt Kenya/Meru areas of qat cultivation are the oldest areas of cultivation in that they are the source of the greatest number of cultivated genotypes and divergence was found between many of the cultivated and wild genotypes (Fig. 4.9 A and B).

Differing patterns of translocation from wild centers of origin— We found no evidence for an independent Yemeni origin of cultivated qat derived from Yemeni wild populations. Rather the evidence from our STRUCTURE results, assignment tests, genetic distance, and PCoA analyses support an Ethiopian wild origin of Yemeni cultivated qat. By way of comparison our methods of genetic cluster identification separated genetic groups on either side of the EAR in Kenya that may have been connected by gene flow as recently as 5000 years ago (e.g., Kebede et al. 2007). The six–million–year separation between Africa and the Arabian Peninsula (Fernandes et al. 2006) would have resulted in numerous allelic differences between Ethiopian and any wild Yemeni qat populations such that individual genotypes from a possible wild Yemeni origin would be distinct from genotypes of African origin. Yet no consistent differences were found among the cultivated Yemeni genotypes (Fig. 4.6). Additionally, if Yemeni cultivated material was sourced from native Yemeni populations, then we expect the mean genetic distance to be distinctly higher between the wild Ethiopian and cultivated Yemeni genotypes, which was not the case. However, in Yemen the majority of cultivated genotypes were not collected in Ethiopia or Kenya and thus Yemen is an important source of cultivated qat genotypic diversity. A similar situation applies to Ethiopian and Yemeni coffee genotypes (e.g., Silvestrini et al. 2007).

Identifying the origin of Yemeni qat is relevant to the importance of qat to the economy and cultural identity of Yemen. In no other country is qat as important to the economy, or have as high a percentage of qat consumers, as in Yemen (World Bank 2007). Legendary accounts about Yemeni origins such as Awzulkernayien, the Yemeni goat herder, learning about wild qat's effects from observing the stimulating effect that it had on his goats (e.g., Getahun & Krikorian 1973) are not supported from our findings. Yet accounts such as Shaikh Ibrāhīm Abū

Zaharbūi introducing qat into Yemen from Ethiopia, for which he was immortalized in Yemen (e.g., Burton 1856), are consistent with our results.

The source of Malagasy qat is thought to be a translocation shortly after 1920 from Yemen (Carrier & Gozon 2009). The six individuals attained from Madagascar are all from a single clonal genotype that was the most broadly distributed and commonly encountered clonal group in our dataset. Our findings are consistent with a Yemeni origin (Carrier & Gozon 2009) but are equally consistent with the possibility that Malagasy qat was founded directly from Ethiopia. This pattern of reduced genetic diversity associated with distant translocation from wild centers of origin in qat is very similar to the translocation pattern in coffee. Coffee was native to the highlands of Ethiopia, numerous cultivar genotypes were brought to Yemen and very few were taken from Yemen to found coffee production in other parts of the world (Simpson & Ogorzaly 2000).

Distant translocation from either of the Kenyan centers of origin was not found in our dataset, though translocation within Kenya was quite common. The Meru area is the primary source from which cultivated qat is translocated. Translocations to Mt Marsabit led to the formation of novel hybrid genotypes from genetically distant parents (Fig. 4.8). This pattern was not documented in other areas founded by translocated genotypes. Unlike the Ethiopian and Mt Kenya/Meru centers of origin, few translocated qat genotypes were assigned to the Mt Elgon center of origin. This pattern indicates recent and/or less intensive cultivation from the Mt Elgon center of origin.

Hybridization between translocated genotypes — The pattern of hybridization among qat cultivars (Fig. 4.8) is similar to the pattern reported for other long–lived perennial crops such as apple (e.g., Cornille et al. 2015; Volk et al., 2015) and almond (e.g., Delplancke et al. 2011)

wherein novel hybridization occurred after genotypes from disparate centers of origin were brought together. However, translocation followed by hybridization of distantly related genotypes appears to be atypical when considering the progressively reduced genotypic and genetic diversity among qat translocated from Ethiopia to Yemen and onto Madagascar. Furthermore in Wote and Wundanyi in southern Kenya where both Ethiopian and Kenyan genotypes are grown together hybrid genotypes were not found.

Given that only Ethiopian genotypes and hybrids were found in Kulal, it seems unlikely that the hybridizations occurred in Kulal. Three farmers growing hybrids in Kulal mentioned that these plants were more bitter and potent than other genotypes grown in the area. All six interviewed farmers in Kulal named Marsabit, Meru and/or neighboring farms as their sources of cuttings to establish their orchards.

It is unclear exactly why farmers in Marsibit and Kulal preserved hybrid genotypes. Only one qat hybrid from Marsabit was found to possess a trait of interest – the production of high quality (desirable effects and flavor) qat. In Kulal it is not clear why hybrid cultivars have been retained given that the only comments from Kulal farmers regarding their hybrids was that they produced bitter and potent qat. Thus does the lack of wild qat in Marsibit and Kulal from which to select new genotypes provide cause farmers to preserve chance hybrids or is the preservation of hybrid genotypes incidental? The often cited reason of hybrid vigor does not appear to be an obvious reason for the retention of these hybrids given that farmers mentioned the difficulty in growing hybrid genotypes. It is interesting that the hybrids are noted as having a ‘strong’ taste profile, thus hybridization may have created a unique and desirable chemical profile.

Patterns of divergence between cultivated and wild qat —Separation between cultivated and wild genotypes is not detectable at K = 2 when analyzing the dataset encompassing all wild

and cultivated collections from Ethiopia and Kenya (Fig. 4.10). Thus broad scale biogeographical processes at this level appear to override the signal of divergence created through artificial selection. However, when examining higher levels of K across the entire dataset or examining divergence within each center of origin, the next partitions to emerge are the cultivated genotypes from Ethiopia (Figs. 4.9 and 4.10). Thus the processes of artificial selection and cultivation in the areas of historic qat cultivation may have resulted in similar levels of divergence to the natural processes that created patterns of subclustering within each of the three wild centers of origin.

Two sampling issues are important to take into consideration when explaining the patterns of divergence seen in the different centers of origin. First, and unrelated to artificial selection, is the issue of sampling wild populations while conducting fieldwork. The wild populations from which a cultivated lineage was sourced may not have been sampled during fieldwork, which has been shown to influence individual assignments to populations (e.g., Cain et al. 2000). Second, the wild population(s) from which a cultivar was sourced could be extinct. This latter scenario may explain the pattern found among Ethiopian wild and cultivated samples as the vast majority of primary Ethiopian forests have been cleared and/or converted into agricultural uses (Dessie & Kleman 2007; Dessie and Kinlund 2008; Tadesse et al. 2014; Guillozet et al. 2015).

If extinction of wild populations is causing the observed patterns of divergence in the Ethiopian center of origin, then not only is farming an important factor in the preservation of genetic diversity in general, but also the preservation of genetic diversity from wild source populations. The observed trend among long–lived perennials is for relatively high genetic diversity to be retained among the cultivated genepool in relation to wild source populations

(Miller and Gross 2011). However, the degree to which this relative measure of diversity has been influenced by wild–population extinction versus on–farm preservation of cultivars has not been broadly assessed.

In regards to evolution under cultivation as defined by the divergence of cultivated from wild genotypes, the Ethiopian cultivated genotypes are the most divergent from extant wild qat compared to the other centers of origin. The strength of this pattern breaks down by center of origin with the Ethiopian center having a well–supported pattern of divergence, the Mt Kenya center weakly supported and, the Mt Elgon center has no clear evidence of divergence (Fig. 4.9). The differing patterns of divergence may be the result of different times since initial cultivation with the Ethiopian center of origin being the oldest and Mt Elgon the youngest. Early initial cultivation in Ethiopia and the Meru area of Kenya is supported in the historical documentation (e.g., Huntingford 1965; Carrier 2007).

Lastly which theory of domestication applies to cultivated qat (Figs 4.9, 4.10)? The instant domestication theory (Zohary & Spiegelroy 1975; Harlan 1992; Zohary 2004; Kislev et al. 2006) is not supported based on the number of cultivated genotypes within different areas of cultivation. However, only three clonal genotypes made up nearly half of all clonal genotypes thus it does appear that while a large number of genotypes are cultivated only a small number of genotypes become predominate in cultivation. From our data, hybridization between distantly related genotypes does not appear to be a major factor in the early stages of qat domestication but maybe important in later stages of domestication outside centers of origin as seen in Marsabit and Kulal.

Conclusion

Qat has been an important plant in the history, culture, and economy of East Africa. We have added not only to our understanding of qat origins but also provided insight into the early stages of long–lived perennial domestication through the analysis of phylogeographic patterns. We identified three separate centers of origin for cultivated qat, two of which (Meru and Ethiopia) have provided the vast majority of cultivars grown throughout the region. Of these two sources of diversity Ethiopia is clearly the original country of origin for exported genotypes to Madagascar and Yemen, while Kenya has a greater wild and cultivated genotypic diversity. From these results we found no evidence for Yemen being a separate center of origin but rather it appears to be a secondary center of diversity in which cultivated genotypes no longer found in Ethiopia are preserved. In addition to the unique genotypes discovered in Yemen the areas of Marsabit and Kulal in northern Kenya generated novel diversity through hybridization of genotypes from Kenyan and Ethiopian parentage. Lastly, evidence for divergence between cultivated and wild genotypes in the historical areas of cultivation in Ethiopia and Mt Kenya/Meru indicates a dynamic process of selection and extinction has taken place among cultivated and wild qat populations and/or indicates potentially unsampled wild populations. In any case, examining divergence in concert with individual assignment as well as hybrid discovery further clarifies the history of qat cultivation and provides a set of research questions for further study. Two such questions are 1) Do wild populations persist in the northern Ethiopia or Nyambene that more accurately represent the wild progenitors of cultivated qat in those areas? and 2) Has hybridization between distantly related genotypes produced novel and desirable chemical traits? Both questions could be examined through further field work and growth chamber experiments respectively.

REFERENCES

Achtak H, Ater M, Oukabli A, Santoni S, Kjellberg F, Khadari B (2010) Traditional agroecosystems as conservatories and incubators of cultivated plant varietal diversity: the case of fig (*Ficus carica* L.) in Morocco. BMC Plant Biology 10: 1471–2229.

Alexander PJ, Rajanikanth G, Bacon CD, Bailey CD (2007) Recovery of plant DNA using a reciprocating saw and silica–based columns. Molecular Ecology Notes 7: 5–9.

Allaby RG, Fuller DQ, Brown TA. (2008) The genetic expectations of a protracted model for the origins of domesticated crops. Proceedings of the National Academy of Sciences of the United States of America 105: 13982–13986.

Al–Motarreb A, Baker K, Broadly KJ (2002) Khat: pharmacological and medical aspects and its social use in Yemen. Phytotherapy Research 16: 403–413.

Altabachew M, Chandravanshi BS, Redi–Abshiro M, Torto N, Chigome S, Pule BO (2013) Evaluation of the effects of various drying techniques on the composition of the psychoactive phenylpropylamino alkaloids of khat (*Catha edulis* Forsk) chewing leaves. Bulletin of the Chemical Society of Ethiopia 27: 347–358.

Amos W, Hoffman JI, Frodsham A, Zhang L, Best S, Hill AVS (2007) Automated binning of microsatellite alleles: problems and solutions. Molecular Ecology Notes 7: 10–14.

Anderson D, Beckerleg S, Hailu D, Klein A (2007) *The khat controversy: stimulating the debate on drugs*. Oxford, Berg.

Baudouin L, Lebrun P (2001) An operational Bayesian approach for the identification of sexually reproduced cross–fertilized populations using molecular markers. Proceedings

of the International Symposium on Molecular Markers Eds. Doré, Dosba & Baril Acta Hort. 546: 81–94, Leuven, Belgium.

Bhalla, N (2002) Ethiopia's khat dilemma. BBC News.

Bouby L, Figueiral I, Bouchette A, Rovira N, Ivorra S, Lacombe T, Pastor T, Picq S, Marinval P, Terral J–F (2013) Bioarchaeological insights into the process of domestication of grapevine (*Vitis vinifera* L.) during roman time in southern France. PLoS ONE 8(5): e63195. doi:10.1371/journal.pone.0063195.

Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non–templated nucleotide addition by *Taq* DNA polymerase: primer modifications that facilitate genotyping. BioTechniques 20: 1004–1010.

Burton RF (1856) *First footsteps in East Africa: or, an exploration of Harar*. London, Longman, Brown, Green, and Longmans.

Cain ML, Milligan BG, Strand AE (2000) Long–Distance seed dispersal in plant populations. American Journal of Botany 87: 1217–1227.

Carrier NCM (2007) *Kenyan khat: the social like of a stimulant*. Leiden, Brill.

Carrier NCM, Gezon L (2009) Khat in the Western Indian Ocean. Plantes et Sociétés, Études Océan Indien 43: 271–297.

Chakraborty R, Jin L (1993) A unified approach to study hypervariable polymorphisms: statistical considerations of determining relatedness and population distances. EXS 67: 153–175.

Coart ES, VanGlabeke M, DeLoose, AS, Larsen I, Roldán–Ruiz A (2006) Chloroplast diversity in the genus *Malus*: New insights into the relationship between European wild

apple (*Malus sylvestris* (L.) Mill.) and the domesticated apple (*Malus domestica* Borkh.). Molecular Ecology 15: 2171–2182.

Cornille A, Feurtey A, Gélin U, Ropars J, Misvanderbrugge K, Gladieux P, Giraud T (2014) Anthropogenic and natural drivers of gene flow in a temperate wild fruit tree: a basis for conservation and breeding programs in apples. Evolutionary Applications 8: 373–384.

Curto MA, Tembrock LR, Puppo P, Nogueira M, Simmons MP, Meimberg H (2013) Evaluation of microsatellites of *Catha edulis* (qat: Celastraceae) identified using pyrosequencing. Biochemical Systematics and Ecology 49: 1–9.

De le Cruz R, Whitkus R, Mota–Bravo L (1995) Tropical tree DNA isolation and amplification, Molecular Ecology 4: 787–789.

Denham TP, Haberle SG, Lentfer C, Fullagar R, Field J, Therin M, Porch N, Winsborough B (2003) Origins of agriculture at Kuk swamp in the highlands of New Guinea. Science 301: 189–193.

Deplancke M, Alvarez N, Espíndola A, Joly H, Brouck E, Arrigo N (2012) Gene flow among wild and domesticated almond species: insights from chloroplast and nuclear markers. Evolutionary Applications 5: 317–329.

Dessie G, Kleman J (2007) Pattern and magnitude of deforestation in the south central rift valley region of Ethiopia. Mountain Research Development 27: 162–168.

Dessie G, Kinlund P (2008) Khat expansion and forest decline in Wondo Genet Ethiopia. Geografiska Annaler: Series B, Human Geography 90: 187–203.

Earl, Dent A. and vonHoldt, Bridgett M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4: 359–361

El Mahi T (1962) A preliminary study on khat (*Catha edulis* Forskal) together with the institutional history of coffee as a beverage in relation to khat, World Health Organization, Regional Office for the Eastern Mediterranean, WHO document EM/RC 11/10.

Evanno, G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14: 2611–2620.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164: 1567–1587.

Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. Molecular Ecology Notes 7: 574–578.

Fernandes CA, Rohling JE, Siddall M (2006) Absence of post–miocene red sea land bridge: biogeographic implications. Journal of Biogeography 33: 961–966.

Gao H, Williamson S, Bustamante CD (2007) A Markov Chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. Genetics 176: 1635–1651.

Gebissa E. 2004. *Leaf of Allah: khat & agricultural transformation in Harerge, Ethiopia*. Oxford, James Currey Ltd.

Getahun A, Krikorian AD (1973) Chat: coffee's rival from Harar, Ethiopia, botany, cultivation and use. Economic Botany 27: 353–377.

Goldsmith, P. 1988. The production and marketing of miraa in Kenya. *Satisfying Africa's food needs: food production and commercilization in African agriculture*, ed., R. Cohen. Boulder, Lynne Rienner Publishers: 121–152.

Goldsmith, P. 1994. Symbiosis and transformation in Kenya's Meru District. (Doctoral dissertation) Retrieved from Univeristy of Florida institutional repository. (system ID AA00025717:00001).

Goldsmith, P. 1999. The political economy of miraa. East African Alternatives 1999: 15–19.

Goldstein DB, Linares AR, Cavalli–Sforza LL, Feldman MW (1995) An Evaluation of genetic distances for use with microsatellite loci. Genetics 139: 463–471.

Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F–statistics. Journal of Heredity 86: 485–486.

Guillozet K, Bliss JC, Kelecha TS (2015) Degradation in an Afromontane forest in highland Ethiopia, 1969–2010. Small–Scale Forestry 14: 121–137.

Harlan JR (1992) *Crops & man*. Madison, American Society of Agronomy, Inc.

Hillman GC, Davies MS (1990) Domestication rates in wild–type wheats and barley under primitive cultivation. Biological Journal of the Linnean Society 39: 39–78.

Holleley CE, Geerts PR (2009) Multiplex Manager 1.0: a cross–platform computer program that plans and optimizes multiplex PCR. BioTechniques 46: 511–517.

House of Commons, Home Affairs Committee (2013) Khat. Eleventh Report of Session 2013–2014, HC 869.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources 9: 1322–1332.

Huntingford GWB, (1965) *The glorious victories of 'Amda Seyon, king of Ethiopia*. Oxford, Clarendon Press.

Kadu CAC, Konrad H, Schueler S, Muluvi GM, Eyog–Matig O, Muchugi A, Williams VL, Ramamonjisoa L, Kapinga C, Foahom B, Katsvanga C, Hafashimana D, Obama C, Geburek T (2013) Divergent patterns of nuclear genetic diversity across the range of the Afromontane *Prunus africana* mirrors variable climate of African highlands. Annals of Botany 111: 47–60.

Kebede M, Ehrich D, Taberlet P, Nemomissa S, Brochmann C (2007) Phylogeography and conservation genetics of a giant lobelia (*Lobelia giberroa*) in Ethiopia and Tropical East African mountains. Molecular Ecology 16: 1233–1243.

Kennedy JG (1987) *The flower of paradise: the institutional use of the drug qat in North Yemen*. Dordrecht, D. Reidel Publishing Company.

Kislev ME, Hartmann A, Bar–Yosef O (2006) Early domesticated fig in the JordanValley. Science 312: 1372–1374.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA (in press) CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources.

Ladizinsky G (1999) On the origin of almond. Genetic resources and crop evolution 46: 143–147.

Lemke L, Rex M, Zyprian E, Töpfer R (2011) A simple, inexpensive and environmentally friendly method for high throughput DNA extraction from grapevine (*Vitis* spp.) Vitis 50: 7–10.

Mairal M, Pokorny L, Aldasoro JJ, Alarcón M, Sanmartín I (2015) Ancient vicariance and climate–driven extinction explain continental–wide disjunctions in Africa: the case of the Rand Flora genus *Canarina* (Campanulaceae). Molecular Ecology 24: 1335–1354.

McKey D, Benoit Punjol ME, Duputie A (2010) The evolutionary ecology of clonally propagated domesticated plants. New Phytologist 186: 318–332.

Miller AJ, Gross BL (2011) From forest to field: Perennial fruit crop domestication. American Journal of Botany 98: 1389–1414.

Miller AJ, Schaal BA (2006) Domestication and the distribution of genetic variation in wild and cultivated populations of the Mesoamerican fruit tree *Spondias purpurea* L. (Anacardiaceae). Molecular Ecology 15: 1467–1480.

Palgrave, KC (1983) *Trees of Southern Africa*. Second revised edition. Struik, Cape Town.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel, population genetic software for teaching and research – an update. Bioinformatics 28, 2537–2539.

Perrier X, De Langhe E, Donohue M, Lentfer C, Vrydaghs L, Bakry F, Carreel F, Hippolyte I, Horry J–P, Jenny C, Lebot V, Risterucci A–M, Tomekpe K, Doutrelepont H, Ball T, Manwaring J, de Maret P, Denham TP. 2011. Multidisciplinary perspectives on banana (*Musa* spp.) domestication. Proceedings of the National Academy of Sciences, USA 108: 11311–11318.

Piry S, Alapetite A, Cornuet J–M, Paetkau D, Baudouin L, Estoup A (2004) GENECLASS2: A software for genetic assignment and first–generation migrant detection. Journal of Heredity 95: 536–539.

Pritchard JC, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155: 945–959.

Reeves PA, Bowker CL, Fettig CE, Tembrock LR, Richards CM, (submitted 2015) Effect of error and missing data on population structure inference using microsatellite data, Molecular Ecology Resources.

Robson NKB, Hallé N, B. Mathew NB, Blakelock R (1994) *Celastraceae. Flora of tropical East Africa*, ed., R. M. Polhill. Rotterdam, A.A. Balkema. 108: 1–78.

Rosenberg NA, Burke T, Elo K, Feldman MW, Freidlin PJ, Groenen MAM, Hillel J, Mäki–Tanila A, Tixier–Boichard M, Vignal A, Wimmers K, Weigand S (2001) Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. Genetics 159: 699–713.

Schorno X, Steinegger E (1979) CNS–active phenylpropylamines of *Catha edulis* FORSK, (Celastraceae) of Kenyan origin, Experientia 15: 572–574.

Silvestrini M, Junqueira MG, Favarin AC, Guerreiro–Filho O, Maluf MP, Silvarolla MB, Colombo CA (2007) Genetic diversity and structure of Ethiopian, Yemen and Brazilian *Coffea arabica* L. accessions using microsatellites markers. Genetic Resources and Crop Evolution 54: 1367–1379.

Simpson B, Ogorzaly M (2000) Economic botany: plants in our world. McGraw Hill, New York.

Stenberg P, Lundmark M, Saura A (2003) MLGsim: a program for detecting clones using a simulation approach. Molecular Ecology Notes 3: 329–331.

Szendrei K (1980) The chemistry of khat, Bulletin on narcotics 32: 5–35.

Tadesse G, Zavaleta E, Shennan C, FitzSimmons M (2014) Policy and demography factors shape deforestation patterns and socio–ecological processes in southwest Ethiopian coffee agroecosystems. Applied Geography 54: 149–159.

Trimingham JS, (1965) *Islam in Ethiopia*. New York, Barnes & Noble, Inc.

United Nations Narcotics Laboratory (1975) Studies on the chemical composition of khat: III, Investigations on the phenylalkylamine fraction, Geneva, United Nations Narcotics Laboratory.

Vähä J-P, Primmer CR (2006) Efficiency of model–based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. Molecular Ecology 15: 63–72.

Vavilov N (1992) *Origin and Geography of Cultivated Plants* (translated by Löve D). Cambridge University Press, Cambridge.

Vincens A, Buchet G, Williamson D, Taieb M (2005) A 23,000 yr pollen record from Lake Rukwa (8°S, SW Tanzania): New data on vegetation dynamics and climate in Central Eastern Africa. Review of Palaeobotany and Palynology 137: 147–162.

Wolfes O (1930) Über das Vorkommen von *d*–nor–iso–Ephedrin in *Catha edulis*. Archiv der Pharmazie 268: 81–83.

World Bank (2007) Yemen, towards qat demand reduction, World bank document report no. 39738–YE.

Zohary D (2004) Unconscious selection and the evolution of domesticated plants. Economic Botany 58: 5–10.

Zohary D, Spiegel–Roy P (1975) Beginnings of fruit growing in the Old World. Science 187: 319–327.

CHAPTER 5: CLONAL DIVERSITY AND THE ETHNOTAXONOMY OF CULTIVATED *CATHA EDULIS* (QAT, CELASTRACEAE): HOW COLOR, GENOTYPE, AND GEOGRAPHY CORRESPOND IN THE FORMATION OF CULTIVAR IDENTITY

Summary

Like other crops that are widely cultivated, farmers and traders have developed a system of names to refer to different cultivars of qat, using stem color as the primary trait to differentiate these cultivars. Previous work has primarily used morphological and chemical features in an effort to find an association between cultivar names and a set of stable traits, but has not yielded any clear association between them. We used SSR genotypes and collections from across the major qat growing regions to test whether names and traits associated with a genotype were consistent across areas of cultivation. We found that there was no clear association between genotype and stem color, and thus a single name was often applied to numerous genotypes creating widespread homonymy in the ethnotaxonomy of qat cultivars. Despite widespread homonymy a subset of individuals from a single clonal group was generally associated with the majority of individuals collected under a single cultivar name. The East African Rift in central Ethiopia was found to be a barrier to both the exchange of both clonal genotypes and certain cultivar names. This finding does not have a clear ethnographic explanation given that the Oromic and Ahharic culture and language in central Ethiopia spans the East African Rift. In Kenya, both cultivar names and clonal genotypes were broadly dispersed with nearly all of the clonal genotypes originating from the Mt Kenya/Meru area in central Kenya. Understanding what factors influence the cultivar name used by qat farmers and traders is important given that different monetary value is associated with different cultivars. We infer that genotype is only one

of at least three factors including horticultural practices and environmental conditions that contribute to the attributes associated with certain qat cultivars.

Introduction

Qat [Celastraceae: *Catha edulis* (Vahl) Forssk. ex Endl.] is a woody evergreen crop of major economic and cultural importance in East Africa and southwest Arabia. An estimated 20 million people consume qat daily in eastern Africa (Al–Motarreb 2002), qat is the second largest foreign exchange commodity in Ethiopia (Bhalla 2002; Gebissa 2004; Anderson et al. 2007), and 6% of the total GDP in Yemen is generated from qat cultivation and sales (World Bank 2007). As with other crops that are widely cultivated, farmers and traders have developed a system of names to refer to different cultivars of qat. The color of the young stems and leaves is the primary trait by which farmers and traders distinguish different cultivars. Additionally the flavor, potency, and agronomic attributes, such as tolerance to different growing conditions and growth habit, are often used to differentiate qat cultivars (e.g., Lemissa 2001; Carrier 2007). Researchers have repeatedly questioned whether named qat cultivars reflect true biological varieties or are simply a system of referring to convergent traits among distantly related genotypes (e.g., Getahun & Krikorian 1973; Lemissa 2001).

Wild qat grows as a woody shrub or tree from 2 – 25 meters tall (Robson et al. 1994). The native habitats of qat are evergreen submontane forests (1100 – 2400 m), growing at forest margins and among the forest on rocky outcroppings (Robson et al. 1994). The range of qat is throughout eastern Africa (and possibly southwest Arabia) from Ethiopia to South Africa and west to Angola (Palgrave 1983; Robson et al. 1994). Flowers of qat are bisexual. Little is known about whether qat is self–compatible or strictly outcrossing; from the floral morphology either or both breeding systems are possible. Clonal growth appears to be uncommon in the wild with

modestly wind dispersed seeds largely accounting for the natural dispersal of qat (M. P. Simmons pers. obs).

Qat farmers generally apply clonal propagation methods, such as stem or root–sucker cuttings, to start new orchards (e.g., Kennedy 1987; Gebissa 2004; Carrier 2007). While rare, instances of seed propagation have been observed as a means of propagating cultivated qat (e.g., Mbuya et al. 1994; Anderson et al. 2007). In cultivation qat is trained into numerous forms from low growing shrubs to narrow upright trees, and open downward trained trees (e.g., Kennedy 1987; Carrier 2007). Different training methods are generally used to encourage the growth of new shoots, which are the parts harvested and sold, and to make harvesting easier. Qat cultivation occurs at scales from small backyard gardens to mono–cropped orchards, but is primarily grown on small–scale subsistence farms and often intercropped with maize, potatoes, beans, ensete, and other food crops.

Within the Celastraceae, *Catha edulis* is the only species that is cultivated on a large scale specifically for its stimulant alkaloids. Qat is cultivated and/or collected from the wild and consumed in Ethiopia, Israel, Kenya, Rwanda, Madagascar, Somalia, Tanzania, Uganda, and Yemen (e.g., Kennedy 1987; Gebissa 2004; Anderson et al. 2007; Carrier 2007; Carrier & Godez 2009). Qat is generally consumed by chewing fresh young leaves and shoots. The young shoot tips and leaves of qat contain the stimulant alkaloids: (S)–cathinone (United Nations Narcotics Laboratory 1975; Schorno & Steinegger 1979; Szendrei 1980), (1R,2S)–norephedrine, and (1S,2S)–norpseudoephedrine, commonly referred to as cathine (Wolfes 1930). Given the rapid breakdown of cathinone into the less potent cathine after harvest, qat is treated as a perishable commodity and is consumed within two to three days of harvest, with the higher prices associated with the freshest shoot tips (Altabachew et al. 2013).

As with other crops that are clonally propagated such as fig (Acktak et al. 2005), grape (Carimi et al. 2011; Meneghetti et al. 2012), and yam (Scarcelli et al. 2013), varietal names are often given to clones that are widely cultivated and popular among farmers. The cultivation practices, extent of geographic dispersal, age, and the degree to which a crop species has been studied often determine the degree to which a cultivar name matches a specific genotype (e.g., Acktak et al. 2005; Carimi et al. 2011; Meneghetti et al. 2012; Scarcelli et al. 2013). For instance, in traditional–fig–growing areas in Morocco multiple cultivar names are associated with a single lineage (synonomy) or a single cultivar name is associated with several lineages (homonymy) (e.g., Acktak et al. 2005; Carimi et al. 2011). The cause of cultivar synonymy or homonymy can be related to biological processes or human cultural and linguistic differences among different growers. It is possible that plants with different genotypes may have similar phenotypes because of phenotypic plasticity and thus be given the same cultivar name. Alternatively, phenotypic plasticity may result in individuals of the same genotype possessing different morphological character states and thus be given different cultivar names (e.g., Volk et al. 2004; Antonius et al. 2012). Further complicating the issue of cultivar names is the finding that single somatic mutations can change phenotypes within clonal lines, which are then given new cultivar names (Carrier et al. 2012). In cases where clonal cultivars have been dispersed across regions where people speak different languages or dialects, cultivars of the same genotype are often referred to in the local language and multiple names, sometimes with the same meaning, can be used to refer to a single genotype (e.g., Acktak et al. 2005). These issues of clonal–cultivar synonomy and homonymony are also found among intensively studied crops such as grape (e.g., Meneghetti et al. 2012; Scarcelli et al. 2013). We expected that issues of

synonomy and homonymny are common among qat cultivar given its' genotypic diversity and widespread cultivation among different cultures.

The main cultivars of qat recognized in Ethiopia are 'addi' white, 'dallota' (same as 'dalecha') which is white or light green, 'dimma' red leaved, and 'hamercot,' which is intermediate in appearance between 'dimma' and 'dalota'(Getahun & Krikorian 1973; Lemissa 2001). In Kenya cultivars are also recognized by color but of equal or greater importance is the age of qat, with the name 'mbaine' being reserved for qat that is older than 40 years and often well over 100 years old (Carrier 2007). Some of the common cultivars grown in Kenya are 'miraa miiru' dark or black in color, 'mirra meru' white, and 'kithara' red. In Yemen the following four cultivars have been mentioned in the literature: 'abyad' white, 'ahmar' red, 'aswad' black, and 'azraq' blue (Revri 1983; Kennedy 1987). In Ethiopia red cultivars of qat are often reported by farmers as being strong in flavor and potent based on their effect on the chewer whereas the white varieties are more palatable and subtle in effect (Getahun & Krikorian 1973). In Kenya darker cultivars are considered superior over the white or green varieties (Carrier 2007), but Krizevski et al. (2007) found that there was no major difference between red and green qat with respect to the alkaloid content. In numerous accounts the conditions under which qat are grown (elevation in particular) are cited as important factors in determining appearance, potency, and flavor (e.g., Revri 1983; Goldsmith 1994; 1999; Carrier 2006). Knowing which genotypes correspond to named cultivars is an essential first step in determining whether geography (both human and physical), morphology, genotype or a combination of these factors determine the naming of cultivars.

For the first time qat cultivar names and stem color will be assessed against genotypic data across the historic areas of cultivation in order to evaluate the degree to which synonymy

and homonymy are found among cultivated qat. We assessed patterns of geographic dispersal for clonal groups and cultivar names in order to determine whether geographic factors are correlated with cultivar name or clonal–genotype dispersal. We also applied association tests between genotype, elevation range, cultivar name and clonality to determine if any significant genetic associations are present among named cultivars. A naming system for qat cultivars that is consistent and encompasses the main factors determining a set of stable traits should be a helpful for qat farmers and traders. Cultivar names are an important factor in determining the monetary value of harvested qat in the marketplace (e.g., Hill 1965; Almedom & Abraham 1994; Carrier 2007).

Materials and methods

Plant collections — Together with Ethiopian and Kenyan colleagues, Mark P. Simmons collected 1481 qat specimens in 2009 – 10 in Ethiopia and Kenya. Collection trips were conducted in collaboration with Addis Ababa University and the National Museums of Kenya. A total of 659 silica–gel–preserved leaf specimens were collected from Ethiopia and 822 from Kenya. Of these 1481 specimens, 567 cultivated individuals were sampled from 256 Ethiopian farms and 525 cultivated individuals were collected from 217 Kenyan farms (including those in which the farmers reported cultivating wild qat that was either *in situ* or had been translocated). Four wild populations (29 – 37 specimens per population; average of 31 were sampled from Ethiopia (only three had been previously reported for the entire country), and 20 wild populations (5 – 30 specimens per population; average of 13) were sampled from Kenya. Additional specimens from Ethiopia include those from one feral population (18 specimens), a farm in which wild plants had been brought into cultivation (three specimens), and 26 specimens purchased from three markets (including the largest in Ethiopia at Aweday, near Harar).

In 2012 and 2013 Abdul Wali al Khulaidi of Yemeni Agricultural Research Authority in Taiz and Mansoor Althobhani of the University of Sana'a collected 55 cultivated qat specimens from across the major production areas in Yemen. Additional existing specimens of qat included six cultivated specimens from Madagascar that Mark P. Simmons collected in 2006 and 2007, and 16 specimens from Tanzania, one sample from Malawi, and one sample from Zimbabwe that were collected by the Missouri Botanical Garden since 1990 as part of their floristic studies in Africa. Although native to the country, qat is not legally cultivated in Tanzania. The complete set of 1561 specimens encompasses the regions in which qat is most intensively cultivated as well as all of the areas mentioned in historical literature as the places of origin for cultivated qat.

DNA extraction, PCR, and microsatellite genotyping — DNA was extracted from herbarium specimens or leaf tissue that was dried in silica gel. We applied a customized DNA extraction process as outlined in Tembrock et al. (in prep). The 17 microsatellites used for genotyping were selected from Curto et al. (2013) on the basis of consistent amplification, allelic richness, and their ability to be multiplexed. PCR and fragment analyses where multiplexed as per Tembrock et al. (in prep). The dataset analyzed here for cultivar identification is the same one analyzed by Tembrock et al. (in prep) for inferring the wild origins and subsequent dispersal of cultivated qat genotypes.

Genetic data analysis — Clones were identified from the dataset and then tested to assess whether genotypic repeats may have arisen from sexual reproduction using the P_{sex} calculation in MLGsim 2.0 (Stenberg et al. 2003). Individuals with missing data and differing by a single allele were added to the clonal groups identified through MLGsim. This step was conducted using visual examination of NJ (neighbor joining; Saitou & Nei 1987) trees as rendered in DARwin 6.0 (Perrier et al. 2003; Perrier & Jacquemond–Collet 2006). By including individuals differing at a

single allele we accounted for the error rate of 2.4% calculated for our dataset. Clonal groups were then given unique numeric designations as well as the codes ‘M’ if they included individuals with missing data and ‘E’ if they included individuals differing by a single allele. Data quality of the final set of clonal individuals was assessed using global linkage disequilibrium tests, and HWE (Hardy Weinberg Equilibrium) as implemented in GDA1.0 (Lewis & Zaykin 2001) and GeneAlex 6.5 (Peakall & Smouse 2012). Clonal repeats were removed before we conducted these test to ensure that replicate genotypes did not bias the results

The assignments of cultivars to geographic centers of origin were based on the results in Tembrock et al. (in prep). The relationship among clonal genotypes was explored using NJ as implemented in DARwin 6.0 (Perrier et al. 2003; Perrier & Jacquemond–Collet 2006). Robustness of groupings was assessed based on 1000 bootstrap replicates (Felsenstein 1985). While all individuals in our dataset were initially assessed via this method in order to group genotypes differing by a single allele, later analyses used only a single exemplar from each clonal group in order to simplify interpretation.

Association tests were conducted in GEPHAST 1.0 (Amos & Acevedo–Whitehouse 2009) in order to explore the possibility that elevation range, cultivar name or clonality were associated with a given locus or genotype. GEPHAST uses an extension of heterozygote–fitness correlation test that considers the entire genotype as well as each locus in correlation tests to multiple different phenotype categories including continuous and binary. Conducting tests without first accounting for population structure can yield false positives, thus all tests were conducted within genetically predetermined clusters to circumvent this problem. Continuous chisquare tests were used for elevation range and cultivar name comparisons, and the binary chisquare test was used for the trait of clonality. The trait of clonality was assessed by including

one genotype from each clonal, compiling those genotypes across all groups and comparing that set against the wild populations to examine if any allele is significantly correlated with clonal propagation. The initial number of randomizations was set to 1000 and if a p value lower than 0.02 (per default settings) was found 100,000 randomizations were conducted for that locus in order to more robustly test the association significance of this locus.

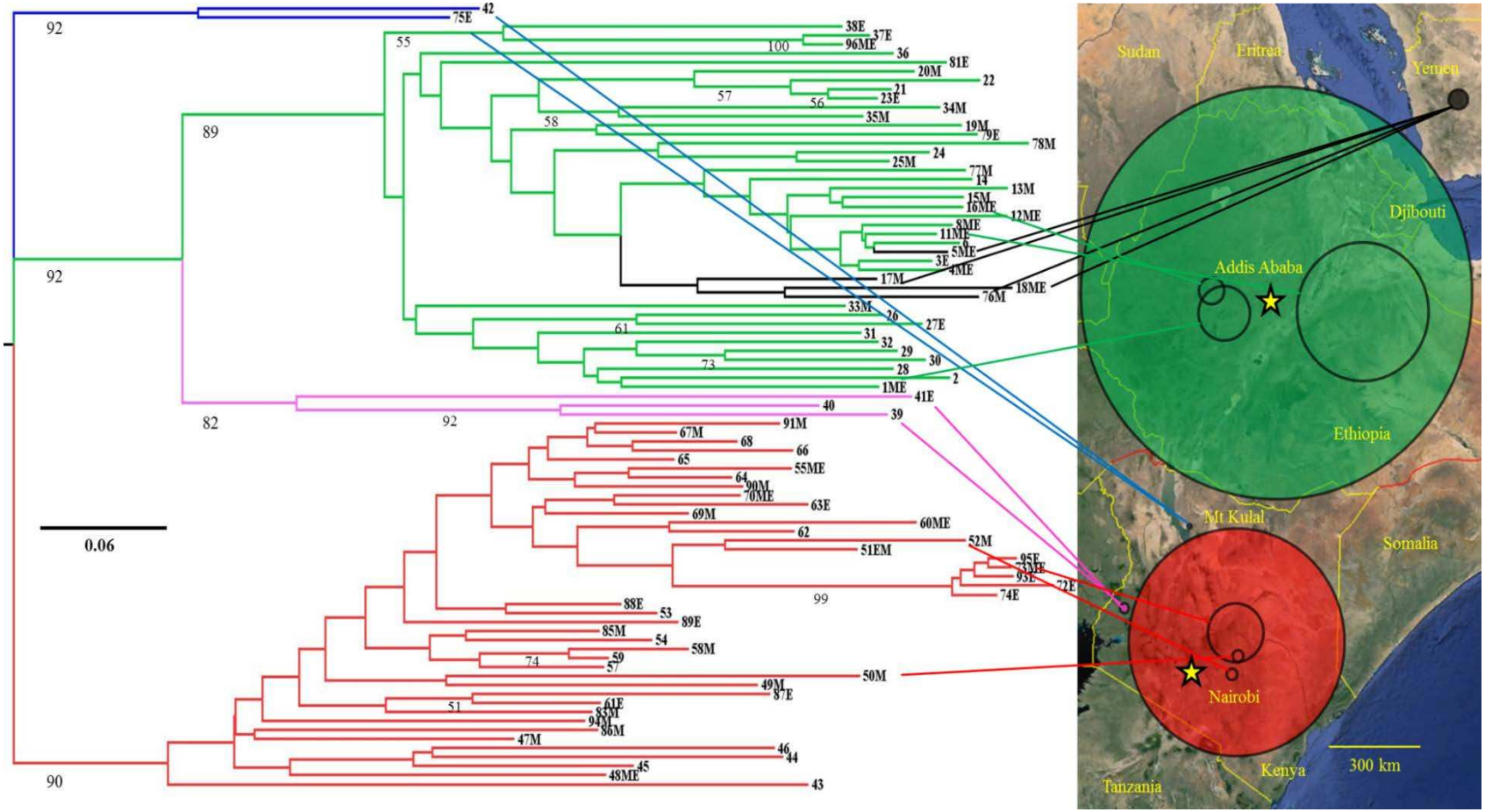
Geographic mapping of clones and cultivar names was conducted in Google Earth 5.0 (Google Inc. Mountain View, CA).

Results

Clonality among cultivated qat— Our method of identifying clones using the P_{sex} calculation (Stenberg et al. 2003) followed by grouping through pairwise deletion of loci with missing values and single allele differences in NJ indicated that 41% of the total dataset, or 638 individuals, belong to one of 86 clonal groups. Clonal groups ranged in size from two to 130 individuals. Nearly all clonal individuals were cultivated; only 2.2% of the putative clones were from wild populations. Forty two of the 86 clonal groups consisted of only two individuals, and 13 clonal groups consisted of three individuals. Only 13 clonal groups were represented by ten or more individuals, and of these only three groups were represented by 50 or more individuals. (Figure 5.1).

Clonality in different areas of qat cultivation — Clonal reproduction was found to be an important means by which to establish qat orchards. In Ethiopia 64.5% of the cultivated qat collected was part of a clonal group, 53.7% in Yemen, 56.8 % in the Mt Kenya/Meru area, and 100% in Madagascar. In the Mt Marsabit, Mt Kulal, and South Horr areas between the Ethiopian highlands and the volcanic uplift areas in central Kenya 63.5% of the individuals collected were identified as clonal groups. All of the clonal individuals from these three areas were of

Figure 5.1. Neighbor joining tree showing the relationships between clonal genotypes with branches color coded by center of origin. Each terminal of the tree represents a clonal group of two or more individuals, bootstrap values of 50% or greater indicated below branches. Circumference of circles on map is proportional to the number of clonal individuals associated with each center of origin. Small circles within large circles are proportional to the number of individuals from the three largest clonal genotypes within the Ethiopian and Mt. Kenya/Meru centers of origin. All circles are centered on the geographic area associated with the most intensive cultivation for that clonal genotype. Lines connecting the circles to the tree indicate the relationships between the most frequently collected clonal genotypes.



Ethiopian or Mt Kenya/Meru origin with the exception of four individuals around Mt. Kulal that were found to be of hybrid origin from parents of Ethiopian and Mt Kenya/Meru origin (Tembrock et al. in prep).

In areas of recent qat cultivation (save Madagascar and the volcanic uplift areas in northern Kenya) clonal individuals represent a smaller percentage of the cultivated qat. In the Chyulu Hills region near Mt Kilamanjaro 39.4% of the cultivated qat was clonal, and of those clonal individuals only 15.4% were inferred to be locally sourced with the remaining clones being sourced from the Mt Kenya/Meru area or Ethiopia. The majority of qat grown in the Chyulu Hills area appears to be locally sourced with only a single clonal genotype originating from this area. Similarly cultivated qat from around Mt Elgon in Kenya south to the Tanzanian border on the west side of the EAR (East African Rift) was found to contain only 25.7% clonal individuals. Of those clonal individuals 32.1% were sourced from wild populations on the west side of the EAR, with the remaining clonal individuals originating from the Mt Kenya/Meru area. In northern Ethiopia, qat cultivation near the southeast shores of Lake Tana was found to consist of only 6.7% clonal individuals – the lowest percentage of any qat–cultivation area.

Ethnotaxonomy, synonymy, and homonymy — Across Ethiopia, Kenya and Yemen 129 names were used to refer to different types of cultivated qat. The most often commonly used names among farmers in Ethiopia were ‘dimma,’ ‘addi,’ ‘kye,’ ‘nech,’ ‘dalecha,’ and ‘magala.’ In Kenya the most commonly encountered names were ‘githara,’ ‘magoca,’ ‘mbaine,’ ‘miira miiru,’ and ‘mitune.’ In Kenya, outside the Meru area, the naming convention for qat was not as consistently employed among qat farmers with numerous farmers referring to their qat as wild or of unknown type. In Yemen the names ‘abyadh,’ ‘ahmar,’ ‘ashraj,’ ‘azraq,’ and ‘safi’ were used alone or in conjunction with another word to modify the meaning. In Yemen the names ‘aswad,’

‘bayadh,’ ‘ghubri,’ ‘habashi,’ and ‘sini’ were used without modifiers. None of these most commonly found names corresponded to a single genotype. Instead all of the commonly used names were applied to both several clonal groups as well as non–clonal genotypes. In Ethiopia all six of the most common cultivar names were used to refer to the clonal group 11ME on at least one or more farms. While the cultivar names are used to refer to numerous genotypes, a high percentage of the individuals from each cultivar is represented by a single genotype. For instance, 62.8% of the individuals referred to as ‘dalecha’ were in the clonal group 11ME, 58.2% of the individuals referred to as ‘mitune’ were in the clonal group 74E, and 50% of the individuals referred to as ‘hamar’ were in the clonal group 5ME.

Qat farmers generally use the trait of stem color in the naming system of cultivars, although several instances of names being applied contrary to the color of the stem were noted. In Kenya the naming system considers the age class of plants by applying the designation ‘mbaine’ to older individuals, though 62.7% of ‘mbaine’ individuals collected in our study were assigned to the 73ME clonal group. In Ethiopia and Kenya clonal genotypes were found to be highly plastic for the trait of stem color (Figs. 5.2, 5.3).

The geography of clonal groups—The geographic range of cultivation among clonal groups varied depending on the clonal group and the place of origin. Clonal group 11ME included the largest number of individuals and had the largest geographic range of cultivation, extending from northwest Yemen to northern Madagascar (Fig 5.4). In Ethiopia, where 11ME is thought to originate (Tembrock et al. in prep), 11ME is cultivated on both sides of the EAR, extending from Dessie in the north to Finchawa in the south. But 11ME was found to be most intensively cultivated around the city of Harar and was the predominate cultivar grown in the area. Clonal group 1ME was the second most collected clonal cultivar in Ethiopia (51

Figure 5.2. Examples of clonal genotype 11ME (all genotypes shown are exact matches) from Ethiopia and Madagascar demonstrating the plasticity of stem color and synonymy. A) ‘addi’, B) ‘dimma’ albeit with unusually green stems for ‘dimma’, C) ‘nech’ D) ‘kye’, and E) green and red qat from Madagascar.

individuals) and was restricted to the west side of the EAR. The range of cultivation for 1ME was centered around the city of Jimma, ranging approximately 100 kilometers in all directions. The most frequently collected clonal group in Kenya, 73ME (55 individuals), was collected

Figure 5.3. Examples of phenotypic plasticity, synonymy, and homonymy among cultivated qat in Kenya. Where applicable the genotypic matches are exact. A) black ‘mbaine’ of clonal genotype 73ME, B) red ‘mbaine’ of clonal genotype 73ME, C) ‘kieru’ of clonal genotype 63E, D) ‘nthaara’ of clonal genotype 63E (growing on the same farm and the same age as C), E) ‘magoca’ from a non–clonal genotype with green stems and, F) ‘magoca’ of clonal genotype 50M with red stems.

primarily around the Nyambene area NE of Mt Kenya in central Kenya. Four 73ME individuals were collected on the west side of the EAR, approximately 350 kilometers from the its primary area of cultivation. In Yemen the most frequently collected clonal group was 11ME. But the clonal group 5ME, which was the largest clonal group unique to Yemen and represented by eight individuals, and was collected once near Sana’a with all other collections from an area SW of Taizz.

The range of cultivation among clonal groups varied depending on the center of origin in

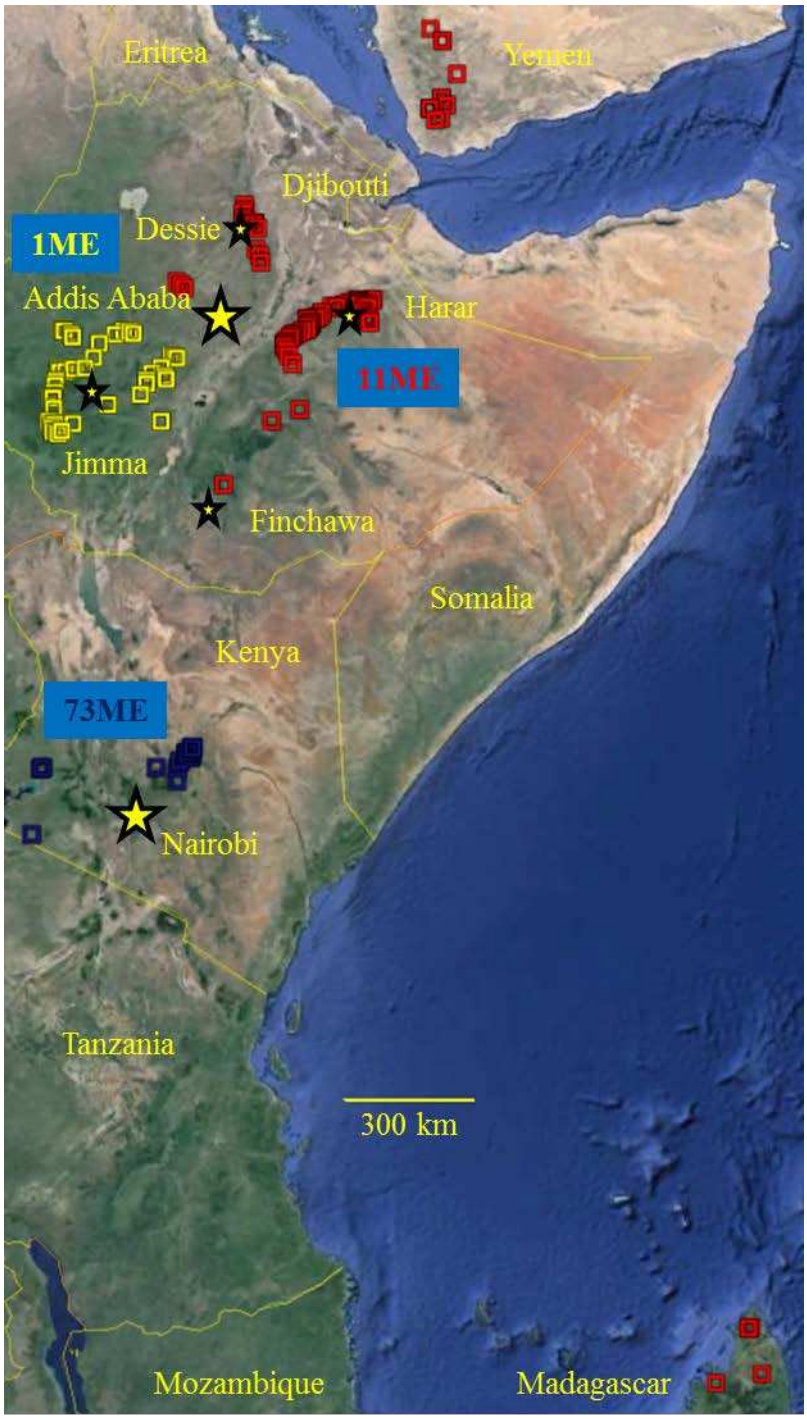


Figure 5.4. Geographic range of the 3 most commonly collected clonal genotypes 1ME, 11ME, and 73ME.

which they were cultivated. In Ethiopia three main areas of clonal diversity were found with little to no exchange of clonal genotypes between them. The three areas in which clonal groups are

concentrated are centered around the cities of Dilla and Harar on the east side of the EAR, and Jimma on the west side of the EAR. Seventeen clonal genotypes were found in the area surrounding Jimma, 11 around Dilla, and five around Harar. The only instances of exchange of clonal groups between these areas in Ethiopia are the cultivation of 11ME and 4ME individuals on the west side of the EAR around Addis Ababa and Dessie. Oddly the clonal group 35M, which was inferred to have an Ethiopian origin (Tembrock et al. in prep), was never collected in Ethiopia but was collected 20 times in five distinct areas of Kenya. Despite this atypical pattern, the origin of 35M to Ethiopia is supported by its nested placement in the NJ tree (Fig. 5.1) and its 1.0 PP assignment to the Ethiopian cluster (Tembrock et al. in prep). The other clonal group (20M) assigning to an Ethiopian origin and dispersed to Kenya was also collected in southern Ethiopia around Dilla.

From 45 clonal groups in Kenya all but six clonal groups originated from the Mt Kenya/Meru area. Of those that did not originate from the Mt Kenya/Meru area, two clonal groups were of hybrid origin from the Mt Kulal area, one group from the Chyulu Hills area in southern Kenya, and three from the Mt Elgon uplift area on the West side of the EAR (Tembrock et al. in prep). Thus the origin of nearly all the cultivated clonal groups in Kenya is the Mt Kenya/Meru area. As a consequence nearly all the cultivated clonal material found outside of the Mt Kenya/Meru areas is a result of translocation. That being stated, 164 cultivated genotypes were not found to be part of any clonal group in the Mt Kenya/Meru area, and 99 within the Mt Elgon area.

The geography of qat cultivar names—As with clonal groups the geographic patterns for qat–cultivar names vary depending on the region in which the name is used. In Ethiopia three patterns emerge in regards to the geographic range of commonly used cultivar names. The three

patterns found among qat cultivar names in Ethiopian 1) ‘hamarcot’ is used only around the city of Harar, 2) ‘dalecha’ is only used on the east side of the EAR 3) ‘addi’, ‘dimma’, and ‘migala’ are used on both sides of the EAR, and 4) ‘nech’ and ‘kye’ are used only on the west side of the EAR. Cultivar names in Kenya generally have a more limited geographic range. However the names ‘magoca’ and black ‘mbaine’ are used outside the core area of use in the Mt Kenya/Meru area in the Mt Marsabit area and on the west side of the EAR. In Yemen the common cultivar names ‘bayadh’, ‘hamar’ and ‘sini’, were only noted around Ta’izz whereas the name ‘abyadh’ was only noted in areas of cultivation north of the state of Ibb. The names ‘ahmer’ and ‘azraq’ were noted throughout areas of Yemeni cultivation however the practice of applying modifier terms to cultivar names was more prevalent in northern Yemen (Fig. 5.5, 5.6, 5.7).

Genetic diversity and association tests—Genetic diversity of the clonal groups from Mt Elgon, Mt Kenya/Meru, and Ethiopia was very similar to wild populations from the same areas. The F_{is} values for the clonal groups from Mt Elgon, Mt Kenya/Meru, and Ethiopia were 0.38, –0.11, and 0.17 respectively. The mean number of alleles per locus for the clonal groups from Mt Elgon, Mt Kenya/Meru, and Ethiopia were 2.59, 4.24, and 4.94. The average across all wild collection sites for F_{is} was 0.11 and 3.81 for the mean number of alleles per locus.

From the association tests none of the traits examined except for clonality yielded any significant results. From the binary chi–square test both Kenyan and Ethiopian but not Mt Elgon clones had loci that where significantly associated with clonality when compared against their wild progenitors. Despite the lack of a significant finding the elevation range of clonal group 11ME is worth noting as it is grown from near sea level in Madagascar to nearly 2600 meters in Ethiopia.

Figure 5.5. Geographic range of the cultivar names A) ‘addi’ and B) ‘dimma’.

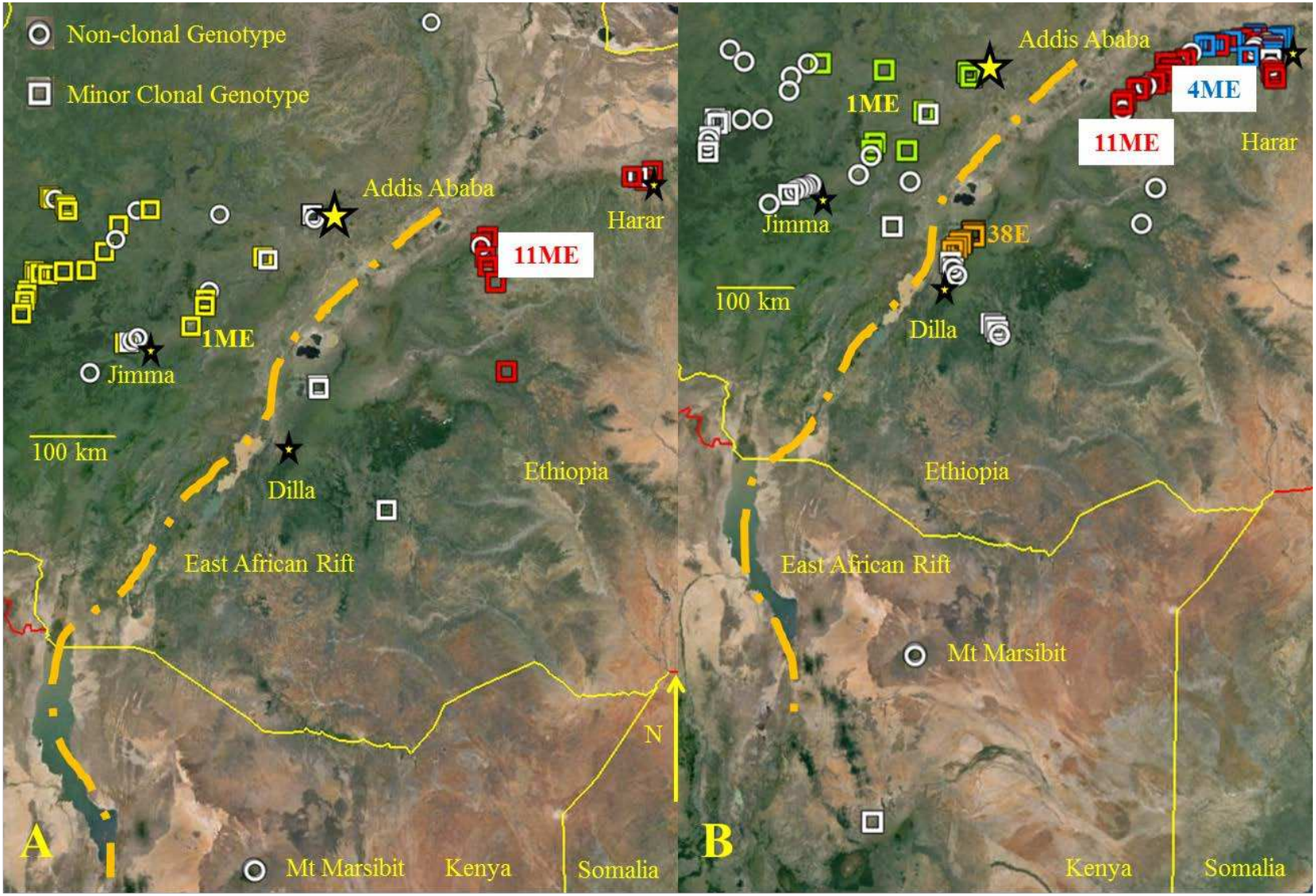


Figure 5.6. Geographic range of the cultivar names A) ‘kye’ and B) ‘dalecha’.

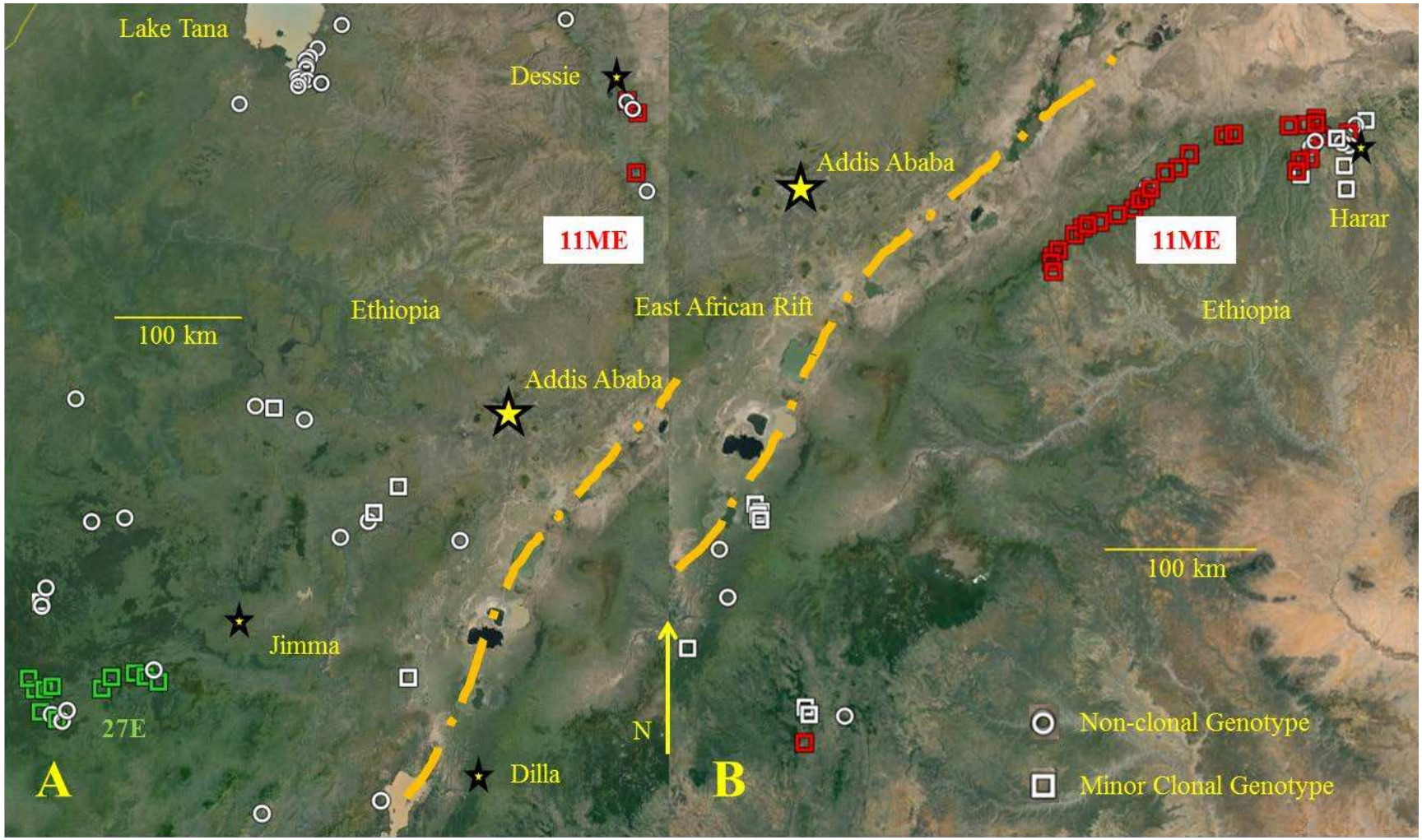
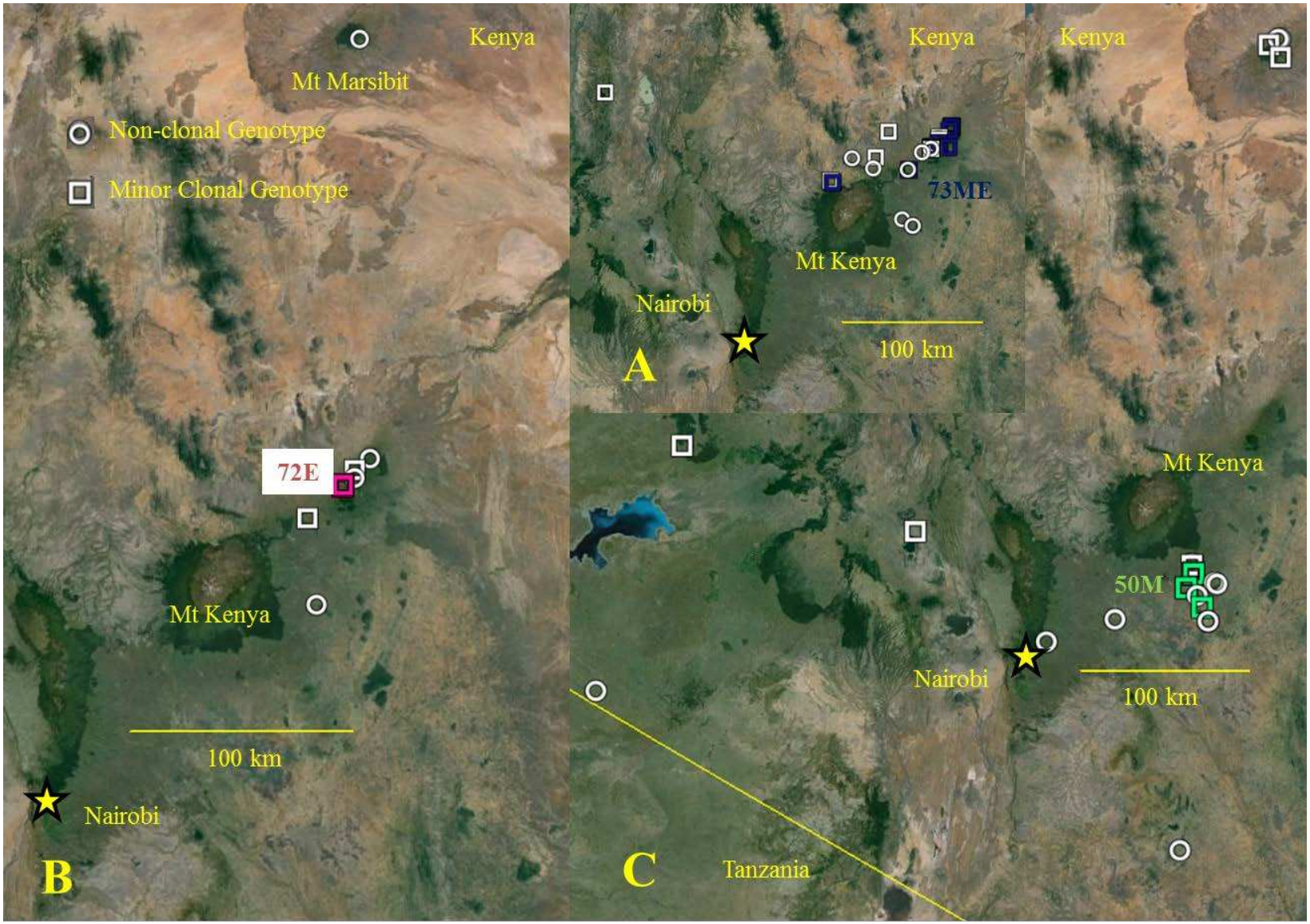


Figure 5.7. Geographic range of the cultivar names A) black ‘mbaine’, B) ‘githara’, and C) ‘magoca’.



Discussion

Clonality, wild selections, seed based propagation, translocation, and hybridization in the formation of cultivated qat diversity— The use of clonal reproduction was found to be an important means of propagation for cultivated qat as previously noted (e.g. Kennedy 1982; Carrier 2007). We found that clonal propagation is most intensively practiced in central Ethiopia around the cities of Dilla, Harar, and Jimma. Interestingly a pattern of restricted dispersal in each of these three areas was found for the clonal groups. That is, all clonal groups cultivated around Jimma where not found cultivated across the EAR around Dilla or Harar. The same was true for cultivated clones around Dilla. However the clonal group 4ME collected most often around Harar was also collected across the EAR near the city of Dessie. The clonal group 11ME most often collected around Harar was also collected near the city of Dessie, Addis Ababa, in Yemen, and was the only clonal group collected in Madagascar.

No evident population structure was found among these three groups. Thus is at this level of geographic and genetic sampling an initial selection from the Jimma area with translocation to Dilla is as likely as an initial selection from Jimma with subsequent cultivation in Jimma. However it is possible that these three separate areas of cultivation were founded from local wild populations and only locally traded for propagation. This explanation is problematic in that the Oromic peoples have historically moved across these three regions (e.g., Bulcha 1997; Zewde 2001). Our own findings of cultivar names being shared across the EAR and these three regions further demonstrate the free flow of qat naming conventions across theses areas of cultivation. While factors associated with human culture have been shown to be associated with the genetic structuring of sorghum in parts Africa (e.g. Westengen et al. 2014) this does not appear applicable here. Therefore further work utilizing more detailed cultural, linguistic and

environmental variables is needed to explain why only localized translocation of cultivated clonal genotypes are found in these areas.

The area southeast of Lake Tana in northern Ethiopia stands out from other areas of qat cultivation in that only a single clonal group was collected in this area. Farmers in this area made note that they often started their plants via seeds collected from their cultivated plants. The farmers also mentioned that they started plants via cuttings but as little evidence of this was found it would appear that clonal propagation is uncommon. This practice of intentionally propagating cultivated qat from seed was rarely mentioned throughout Ethiopia and Kenya except for instances of chance seedlings such as those in Mt Marisbit. Interestingly the only cultivar names used in this area were ‘nech’ and ‘kye’ despite not sharing any of the clonal genotypes associate with these names.

In Kenya clonal propagation is most intensive within the Mt Kenya/Meru area. In fact all other Kenyan areas of production we sampled, were found to possess clonal individuals from the Mt Kenya/Meru area. Thus as discussed in the historical literature (e.g. Carrier & Gezon 2009) the Mt Kenya/Meru area appears to be original site for intensive qat cultivation. As such the cultural notoriety associated with this area may have led to this area becoming the primary source of translocated qat to other areas. After 1981 the Kenyan government officially recognized the importance of qat production and began to provide material support to qat farmers (Goldsmith 1988). Thus, after 1981 qat production expanded outside the Mt Kenya/Meru area. From our results and farmer interviews qat production in areas like the Chyulu hills and Mt Elgon relied both on adjacent wild populations and importation of clones from the Mt Kenya/Meru area to establish their farms. One farmer in the Chyulu area made note that he had tried to utilize the wild plants from the area without success but that plants from Meru and

Ethiopia were able to prosper. Assignment tests conducted in Tembrock et al. (in prep) confirmed the origin of these plants

In the Mt Kulal and Mt Marsabit areas of Kenya the situation is similar to the Chyulu hills in that cultivars from Mt Kenya/Meru and Ethiopia were used to found qat cultivation in these areas. However no wild populations were thought to be involved in sourcing for cultivated qat despite the mention of wild populations among some farmers. These areas are important in that they provided the necessary conditions (i.e. translocation of genotypes from different origins into close proximity and the preservation of chance seedlings) to allow for the production of hybrid sourced clonal cultivars. This process of hybridization and propagation could have occurred in Chyulu but was not noted among the collected individuals.

Translocation of qat genotypes between different centers of origin has generally resulted in increasing genotypic diversity through introgression and/or increasing the number of genotypes in an area over strictly local selections. While translocation to more distant areas of cultivation has occurred genotypic diversity has often been reduced due to genetic bottleneck. The most extreme case of this being Madagascar in which only a single genotype has been collected. Unlike areas in which wild populations can be mined for genetic diversity qat in Madagascar does not have any wild qat thus translocation is the only source of new genetic material. The translocation of qat to Yemen has resulted in a less extreme reduction of cultivated genotypes. Rather Yemen stands as a secondary center for cultivated qat diversity with four clonal genotypes found only in Yemen as well as 26 non–clonal genotypes. This level of genotypic diversity is similar to areas of cultivation in Ethiopia of a similar geographic size. This suggests that multiple introductions and/or crossbreeding between introduced genotypes has

maintained a higher level of genotypic diversity in Yemen despite the apparent lack of any wild populations in Yemen.

Ethnotaxonomy in the identity of cultivated qat types—The Ethnotaxonomic system developed to recognize cultivated qat does not directly correspond to any single genotype, in addition multiple names are applied to a single genotype. Thus both synonymy and homonymy are common in the naming convention for cultivated qat. Moreover the established traits used to delimit cultivar names are sometimes not followed, adding further confusion to what traits are important in recognizing certain qat types.

In regards to stem color, the primary trait by which different types of qat are recognized, numerous examples of phenotypic plasticity were found. One of the most striking examples of this was two individuals grown on the same farm with identical genotypes, and age classes (as identified by the farmer) and similar phenotypes except for stem color. Due to this difference in stem color these two individuals were associated with two different cultivar names (Fig. 5.3, C, D).

It is difficult to speculate as to what environmental factor or factors might be causing a stem color change between individuals in such close spatial proximity. However light intensity and temperature have been shown to be important in influencing the expression of anthocyanins and might be relevant for qat coloration (e.g., Lightbourn et al. 2007). Although not related to stem color qat has been shown to be phenotypically plastic in regards to leaf arrangement (Krikorian 1985). Therefore qat appears to be phenotypically plastic in general as leaf arrangement is often a fixed trait used in identification of plant species. From our findings stem color does not appear to be a useful trait by which to identify cultivated qat types. However with

the current set of markers, studies associating genotype with phenotype can now be carried out in an effort to find stable characters to identify cultivated types.

Despite the problems associated with the naming system in qat, a trend among the most commonly used names provides a reference point from which to develop a more consistent naming system. For instance, the name ‘dalecha’ is made of 62.8% individuals from clonal group 11ME, ‘mitune’ 58.2% 74E, and ‘hamar’ 50% 5ME. Thus a core set of clonal individuals appear to be the original ‘hamar’ or ‘mitune’ for example. Thus these core sets or original clonal individuals could be used as the initial set in finding what traits constitute a true ‘mitune’.

Geography also plays an important role in the application of certain names – for example the name ‘kye’ is used to refer to red qat and ‘nech’ white qat (although sometimes inaccurately) on the west side of the EAR. In this way the names might be acknowledging the effect of ‘terroir’ on the quality (or even phenotype) of qat much as ‘champagne’ or ‘cognac’ do with carbonated wine and aged brandy respectively. In support of this hypothesis is the fact that the west side of the EAR in Ethiopia where qat is cultivated is mainly associated with volcanic geology whereas the area around Harar is associated with limestone geology (Kazmin 1972).

Lastly the age class of qat is thought to be an important factor in determining the quality or cultivar of qat. A similar claim is made for wine grapes – for instance older grape vines in Germany are given the designation ‘alte Reben’ (Robinson 2006). The ‘mbaine’ qat of Kenya appears to be 81% composed of individuals from 5 closely related clonal genotypes. Therefore the ‘mbaine’ designation is very much associated with genotype and the quality of ‘mbaine’ qat may be the result of genotype as well as age. No qat older the 60 years was found in Ethiopia whereas qat of 200 years in age was claimed from Yemen. Claims in regard to age should be tested to determine whether genotype or age is more important in producing high quality qat.

What qualities are associated with cultivated qat genotypes and loci?—We tested whether elevation range, cultivar name, and clonality could be significantly associated with a genotype or locus. We did this order to explore whether any attribute of those we routinely recorded could be underlying the designation of qat cultivars. It was hoped that a more stable character (molecular, ecological, or agronomic) than stem color could be found to designate qat cultivars. The only trait with a significant result was that of clonality when comparing clonal genotypes to wild genotypes. However this finding may be artefactual as Tembrock et al. (in prep) found that cultivated types in Ethiopia and in the Mt Kenya/Meru area may have been divergent through wild progenitor extinction or incomplete sampling. As noted by Amos and Acevedo–Whitehouse (2009) such population structure can result in false significant results. However it is also reasonable to suggest that a set of traits (such as suckering, and rapid root formation from stems) would be associated with individuals that are more readily propagated clonally. A more detailed examination of geographic, molecular, and morphologic variables need to be examined in order to more fully understand what factors are most important in driving the recognition of cultivated varieties throughout areas of qat cultivation.

Conclusion

We found that cultivar names do not strictly follow genotypes but rather cultivar names in some instances are associated with certain geographic regions. Such a geographic pattern points to a linguistic or other ethnographic quality in determination of qat cultivar names. The main character of color that farmers and traders use to distinguish cultivars was found to be plastic. Plasticity in stem color likely contributed to the lack of strict association between cultivar name and genotype. Clonal genotypes were not found associated with any set of attributes and

only three genotypes were found to be widespread. Suggesting that localized selections from wild populations are the primary means by which cultivated qat is established.

REFERENCES

Achtak H, Ater M, Oukabli A, Santoni S, Kjellberg F, Khadari B (2010) Traditional agroecosystems as conservatories and incubators of cultivated plant varietal diversity: the case of fig (*Ficus carica* L.) in Morocco. BMC Plant Biology 10: 1471–2229.

Almedom, AM Abraham S (1994) Women, moral virtue and *tchat*–chewing. *Gender, drink and drugs*,ed., M. McDonald. Oxford, Berg: 249–259.

Al–Motarreb A, Baker K, Broadly KJ (2002) Khat: pharmacological and medical aspects and its social use in Yemen. Phytotherapy Research 16: 403–413.

Altabachew M, Chandravanshi BS, Redi-Abshiro M, Torto N, Chigome S, Pule BO (2013) Evaluation of the effects of various drying techniques on the composition of the psychoactive phenylpropylamino alkaloids of khat (*Catha edulis* Forsk) chewing leaves. Bulletin of the Chemical Society of Ethiopia 27: 347–358.

Amos W, Acevedo–Whitehouse K (2009) A new test for genotype – fitness associations reveals a single microsatellite allele that strongly predicts the nature of tuberculosis infections in wild boar. Molecular Ecology Resources 9**:** 1102–1111.

Anderson D, Beckerleg S, Hailu D, Klein A (2007) *The khat controversy: stimulating the debate on drugs*. Oxford, Berg.

Antonius K, Aaltonen M, Uosukainen M, Hurme T (2012) Genotypic and phenotypic diversity in Finnish cultivated sour cherry (*Prunus cerasus* L.), Genetic Resources and Crop Evolution 59: 375–388.

Bhalla, N (2002) Ethiopia's khat dilemma. BBC News.

Brownstein MJ, Carpten JD, Smith JR (1996) Modulation of non–templated nucleotide addition by *Taq* DNA polymerase: primer modifications that facilitate genotyping. BioTechniques 20: 1004–1010.

Bulcha, M (1997) The politics of linguistic homogenization in Ethiopia and the conflict over the status of *Afaan Oromoo*, African Affairs 96: 325–352.

Carimi F, Mercati F, De Michele R, Fiore MC, Riccardi P, Sunseri F (2011) Intra–varietal genetic diversity of the grapevine (*Vitis vinifera* L.) cultivar 'Nero d' Avola' as revealed by microsatellite markers, Genetic Resources and Crop Evolution 58: 967–975.

Carrier G, Cunff LL, Dereeper A, Legrand D, Sabot F, Bouchez O, Audeguin L, Boursiquot J-M, This P (2012) Transposable elements are a major cause of somatic polymorphism in *Vitis vinifera* L., PLoS ONE 7: e32973.

Carrier NCM (2006) Bundles of choice: variety and the creation and manipulation of Kenyan khat's value. Ethnos 71: 415–437.

Carrier NCM (2007) *Kenyan khat: the social like of a stimulant*. Leiden, Brill.

Carrier NCM, Gezon L (2009) Khat in the Western Indian Ocean. Plantes et Sociétés, Études Océan Indien 42–43.

Curto MA, Tembrock LR, Puppo P, Nogueira M, Simmons MP, Meimberg H (2013) Evaluation of microsatellites of *Catha edulis* (qat: Celastraceae) identified using pyrosequencing. Biochemical Systematics and Ecology 49: 1–9.

De le Cruz R, Whitkus R, Mota–Bravo L (1995) Tropical tree DNA isolation and amplification, Molecular Ecology 4: 787–789.

Gebissa E. 2004. *Leaf of Allah: khat & agricultural transformation in Harerge, Ethiopia*. Oxford, James Currey Ltd.

Goldsmith, P. 1999. The political economy of miraa. East African Alternatives 1999: 15–19.

Google Inc. (2009) Google Earth (Version 5.1.3533.1731) https://www.google.com/earth/

Hill BG. (1965) Cat (*Catha edulis* Forsk). Journal of Ethiopian Studies 3: 13–23.

House of Commons, Home Affairs Committee (2013) Khat. Eleventh Report of Session 2013–2014, HC 869.

Kazmin V (1972) Geologic map of Ethiopia, Geological Survey of Ethiopia, Ministry of Mines, Energy and Water Resources, Addis Ababa.

Kennedy JG (1987) *The flower of paradise: the institutional use of the drug qat in North Yemen*. Dordrecht, D. Reidel Publishing Company.

Krikorian AD, (1985) Growth mode and leaf arrangement in Catha edulis (Kat), Economic Botany 39: 514–521.

Krizevski R, Dudai N, Bar E, Lewinsohn E (2007) Developmental patterns of phenylpropylamino alkaloids accumulation in khat (Catha edulis, Forsk.). Journal of Ethnopharmacology 114: 432–438.

Lemessa D (2001) Khat (Catha edulis): botany, distribution, cultivation, usage and economics in Ethiopia. Addis Abbaba, UN–Emergencies Unit for Ethiopia.

Lewis PO, Zaykin, D (2001) Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d16c).

Lightbourn GJ, Stommel JR, Griesbach RJ (2007) Epistatic interactions influencing anthocyanin gene expression in *Capsicum annuum*. Journal of the American Society for Horticultural Science 132: 824–829.

Mbuya LP, Msanga HP, Ruffo CK, Birnie A, Tengnäs B (1994) Useful trees and shrubs for Tanzania. Nairobi, Swedish International Development Authority.

Meneghetti, S, Costacurta A, Morreale G, Calò A (2012) Study of intra–varietal genetic variability in grapevine cultivars by PCR–derived molecular markers and correlations with geographic origins. Molecular Biotechnology. 50: 72–85.

Palgrave, KC (1983) *Trees of Southern Africa*. Second revised edition. Struik, Cape Town.

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel, population genetic software for teaching and research – an update. Bioinformatics 28, 2537–2539.

Perrier X, Flori A, Bonnot F (2003). Data analysis methods. In: Hamon P, Seguin M, Perrier, X, Glaszmann J C, Eds., Genetic diversity of cultivated tropical plants. Enfield, Science Publishers. Montpellier. pp 43–76.

Perrier X, Jacquemoud–Collet JP (2006). DARwin software http://darwin.cirad.fr/darwin.

Revri R. 1983. Catha edulis Forsk.: geographical dispersal, botanical, ecological and agronomical aspects with special reference to Yemen Arab Republic. Institut für Pflanzenbau und Tierhygiene in den Tropen und Subtropen. Göttingen, Universität Göttingen: 157.

Robinson J (2006) *The oxford companion to WINE*, 3rd edition, Oxford university press.

Robson NKB, Hallé N, B. Mathew NB, Blakelock R (1994) *Celastraceae. Flora of tropical East Africa*, ed., R. M. Polhill. Rotterdam, A.A. Balkema. 108: 1–78.

Saitou N, Nei M (1987) The neighbor–joining method: a new method for reconstructing phylogenetic trees, Molecular Biology and Evolution, 4: 406–425.

Scarcelli N, Couderc M, Baco MN, Egah J, Vigouroux Y (2013) Clonal diversity and estimation of relative clone age: application to agrobiodiversity of yam (*Dioscorea rotundata*), BMC Plant Biology 13: 178–188.

Schorno X, Steinegger E (1979) CNS–active phenylpropylamines of *Catha edulis*

FORSK, (Celastraceae) of Kenyan origin, Experiemtia 15: 572–574.

Stenberg P, Lundmark M, Saura A (2003) MLGsim: a program for detecting clones using a simulation approach. Molecular Ecology Notes 3: 329–331.

Tembrock, LR, et al. (in preparation) The origins, dispersal, and evolution of *Catha edulis* (Qat, Celastraceae) in areas of historic cultivation.

Volk GM, Henk AD, Richards CM (2004) Genetic diversity among U.S. garlic clones as detected using AFLP methods. Journal of the American Society for Horticultural Science. 129: 559–569.

Westengen OT, Okongo MA, Onek L, Berg T, Upadhyaya, Birkeland S, Khalsa SDK, Ring KH, Stenseth NC, Brysting AK (2014) Ethnolinguistic structuring of sorghum genetic diversity in Africa and the role of local seed systems. Proceedings of the National Academy of Sciences of the United States of America 111: 14100–14105.

World Bank (2007) Yemen, towards qat demand reduction, World bank document report no. 39738–YE.

Zewde B (2001) *A history of modern Ethiopia 1855–1991*, second edition, Jame Currey, Oxford.

APPENDICES

Appendix 3.1 List of Taxa sampled for alkaloid extraction with taxonomic authorities and collection information (upper case abbreviations in parentheses follow the Index Herbariorum (Thiers 2015) format and indicate the herbarium in which a sample has been deposited, if applicable).

Celastraceae: *Brexia australis* G.E. Schatz & Lowry— *R.H. Archer et al. 2989* (PRE); *Brexia madagascariensis* (Lam.) Ker Gwal.— *R.H. Archer et al. 2971*, Madagascar (PRE); *Cassine parvifolia* Sond.— *E. van Jaarsveld 2011–03*,Western Cape, South Africa; *Cassine peragua* L.— *E. van Jaarsveld 2011–05*, Western Cape, South Africa; *Cassine schinoides* (Spreng.) R.H.Archer— *E. van Jaarsveld 2011–04*, Western Cape, South Africa; *Catha edulis* (Vahl) Forssk. ex Endl— *R.H. Archer et al. 3016*, Madagascar (PRE); *Celastrus madagascariensis* Loes.— *R.H. Archer 3795*, Madagascar (PRE); *Cheiloclinium hippocrateoides* (Peyr. ex Mart.) A.C.Sm.— *J.A. Lombardi 6532*, Brazil (HRCB); *Crossopetalum gaumeri* (Loes.) Lundell— Fairchild Tropical Garden Acc. #FTBG–941055A; *Crossopetalum rhacoma* Hitchc.— *M. Islam 2009–20*, Bahia de Las Aguila, Dominican Republic; *Denhamia celastroides* (F. Muell.) Jessup— *M.W. Chase 2050*, cult. Bogor, Indonesia (K); *Elaeodendron xylocarpum* DC.— *M. Islam 2009–32a*, Puerto Escondido, Dominican Republic; *Euonymus fortunei* (Turcz.) Hand— *M.P.Simmons 1913*, cult. Illinois,USA (CS); *Gyminda latifolia* (Sw.) Urb.— *M. Islam 2009–33*, Puerto Escondido, Dominican Republic; *Gymnosporia buxifolia* (L.) Szyszył.— *M.P.Simmons 2429*, Kenya (CS); *Gymnosporia cassinoides* Masf.— *J A Reyes–Betancort 41716*, Tenerife: Barranco de Badajoz, Canary Is.; *Gymnosporia cassinoides* Masf.— *J. A. Reyes–Betancort* TFC 49.720, Canary Is.; *Gymnosporia cryptopetala* (female) Reyes–Bet. & A.Santos— *S. Scholz 39426*, Fuerteventura, Canary Is.; *Gymnosporia cryptopetala* (male) Reyes–Bet. & A.Santos— *S. Scholz 41096*, Fuerteventura, Canary Is.; *Gymnosporia divaricata*

Baker— *R.H. Archer et al. 2891*, Madagascar (PRE); *Gymnosporia gracileps* Loes.— *M.P.Simmons 2237*, Ethiopia (CS); *Gymnosporia heterophylla* Loes.— *M.P.Simmons 2075*, Ethiopia (CS); *Gymnosporia linearis* (L.f.) Loes.— *R.H. Archer et al. 2935*, Madagascar (CS); *Gymnosporia senegalensis* Loes.— *M.P.Simmons & G.W. Ngugi 2387*, Kenya (CS); *Gymnosporia senegalensis* Loes.— *R.H. Archer et al. 3039*, Madagascar (PRE); *Halleriopsis cathoides* R.H. Archer (ined.)— *R.H. Archer et al. 3035*, Madagascar (PRE); *Hylenaea comosa* (Sw.) Miers— *J.A. Lombardi 6400*, Brazil (HRCB); *Lauridia tetragona* (L.f.) R.H.Archer— *E. van Jaarsveld 2011–01*, Western Cape, South Africa; *Lophopetalum arnhemicum* Byrnes— *W. Price s.n.*, Australia (BRI); *Maurocenia frangula* Mill.— *E. van Jaarsveld 2011–02*,Western Cape, South Africa; *Maytenus grenadensis* Urb.— *M. Islam 2009–51*, Grenada; *Maytenus boaria* Molina— *M.P.Simmons & G. Smick 1908*, cult. San Francisco, California (CS); *Microtropis fokienensis* Dunn— *Xiong Weizhong 229*, China (MO); *Microtropis triflora* Merr. and Freeman— *H. He 09*, Chongqing, China (CTC); *Mystroxylon aethopicum* (Thunb.) Loes.— *R.H. Archer et al. 2921*, Madagascar (CS); *Peripterygia marginata* Loes.— *M.P.Simmons 1793*, New Caledonia (BH); *Pleurostylia opposita* (Wall.) Merr. & F.P.Metcalf— *A. Ford 2318*, Australia (BRI); *Polycardia libera* O. Hoffm— *R.H. Archer et al. 2899*, Madagascar (CS); *Polycardia phyllanthoides* Lam.— *R.H. Archer et al. 2904*, Madagascar (CS); *Pseudocatha mandenensis* ined.— *R.H. Archer et al. 2996*, Madagascar (CS); *Salacia* sp. nov. (?) not applicable— *R.H. Archer et al. 3040*, Madagascar (CS); *Salaciopsis glomerata* Hürl.— *M.P.Simmons 1895*, Sud Province, New Caledonia; *Schaefferia fructescens* Jacq.— Fairchild Tropical Garden Acc. #72611; *Siphonodon pendulus* F. M. Bailey— *A. Ford 4529*, Australia (BRI); *Xenodrys micranthum* R.H. Archer (ined.)— *R.H. Archer et al. 2902*, Madagascar (CS); *Zinoweieia matudai* Lundell— *M. P. Simmons & G. Smick 1906*, cult. San Francisco, California

(CS); Ephedraceae: *Ephedra sinica* Stapf— *L.R. Tembrock*, lyophilized powder purchased as Ma Huang; Papaveraceae: *Macleaya cordata* R.Br.— Denver Botanic Gardens Acc. # 0.02386; Rubiaceae: *Galium aparine* L.— *L.R. Tembrock 11–003*, cult. Colorado, USA; Salicaceae: *Populus angustifolia* James— *L.R. Tembrock 11–004*, cult. Colorado, USA; *Populus tremuloides* Michx.— *L.R. Tembrock 11–005*, cult. Colorado, USA.

Appendix 4.1. DNA extraction from *Catha edulis*

Chemical preparation — 1) Extraction/Lysis buffer: 2% SDS (w/v)*, 2% PVP 40 (w/v), 1% DIECA (w/v), 250 mM NaCl, 40 mM Ascorbic acid, 200 mM Tris HCl, and 5 mM EDTA, pH to 8.0. All compounds can be stored in concentrated aqueous solutions at room temperature until time of use, however due to the volatility of DIECA and the instability of ascorbic acid in aqueous solution, it is recommended that these compounds are added in powder form immediately before use. Enzymes proteinase K and RNase A should also be added immediately before use, as SDS will denature these enzymes over time. Below is an example of a 120 mL (enough for two 96 well plate extractions) preparation:

Stock Concentration	Working Concentration
39.6 mL 6% SDS	2%
39.6 mL 6% PVP 40	2%
6 mL 5M NaCl	250 mM
24 mL 1M Tris HCl, pH 8.0	200 mM
1.2 mL 0.5M EDTA, pH 8.0	5 mM
1.2g DIECA	1%
0.85 g ascorbic acid	40 mM
8.8 mL sterile water	NA

2) Enzymes: Proteinase K 1 mg/mL (store at 4°C) and RNase A 10 mg/mL or if stock is denoted in units then dilute such that 8 µL of working stock contains ~30 units of RNase A (store at –20°C).

3) Precipitation buffer: 5M potassium acetate (to make a total of 50 mL of solution add 24.54 g of anhydrous potassium acetate to 30 mL sterile water, agitate, slowly add 10 mL glacial acetic acid, with the remaining 10 mL of volume add glacial acetic acid and/or sterile water until the final volume of 50 mL and pH of 6.0 are attained (store at room temperature).

4) Binding buffer: 2M guanidine hydrochloride in 95% ethanol (store at room temperature).

5) Wash buffer I: 20% 10mM Tris–HCl pH 6.5, 80% ethanol (v/v).

6) Wash buffer II: 95% ethanol.

7) Elution buffer: 10mM Tris in H_2O, pH to 8.0.

*The weight–to–volume measurements assume a 1% mixture to be 1 gram of solute to 100 mL of solvent. For example a 5% (w/v) NaCl solution would contain 5 g of NaCl and 100 mL of water.

Procedure — 1) Place two grinding beads and 0.02 grams of silica dried leaf tissue into a 2 mL tube (stronger tubes are recommended to sustain the impact of grinding).

2) Grind samples through rapid reciprocating action for 30 seconds (more fibrous leaves may require longer grinding times, but avoid grind times longer than 1 minute as shearing damage can reduce DNA quality).

3) Open tubes carefully and add 600 µL of lysis/extraction buffer, 8 µL of proteinase K, and 8 µL of RNase A to the leaf tissue powder, invert tubes several times to thoroughly mix. All buffers and enzymes can be combined and vortexed immediately preceding their addition to the ground leaf material.

4) Allow mixture to incubate at 65°C for 30 minutes, mix by gentle inversion every 10 minutes.

5) Add 150 µL potassium acetate to above mixture, mix by inversion and incubate for 30 minutes (or as long as overnight) at –20°C.

6) Place tubes into centrifuge and spin for 15 minutes at 14,000 rpm.

7) Carefully remove tubes from centrifuge, pipette the supernatant being careful to leave the precipitate and plant particles at the bottom of the tube.

8) Place the extracted supernatant from step 7 into a 1.5 mL tube, add 1.5 volumes of guanidine hydrochloride (for instance if 500 µL was recovered from step 7 add 750 µL of guanidine hydrochloride) and mix by inversion, incubate at room temperature for 5–10 minutes.

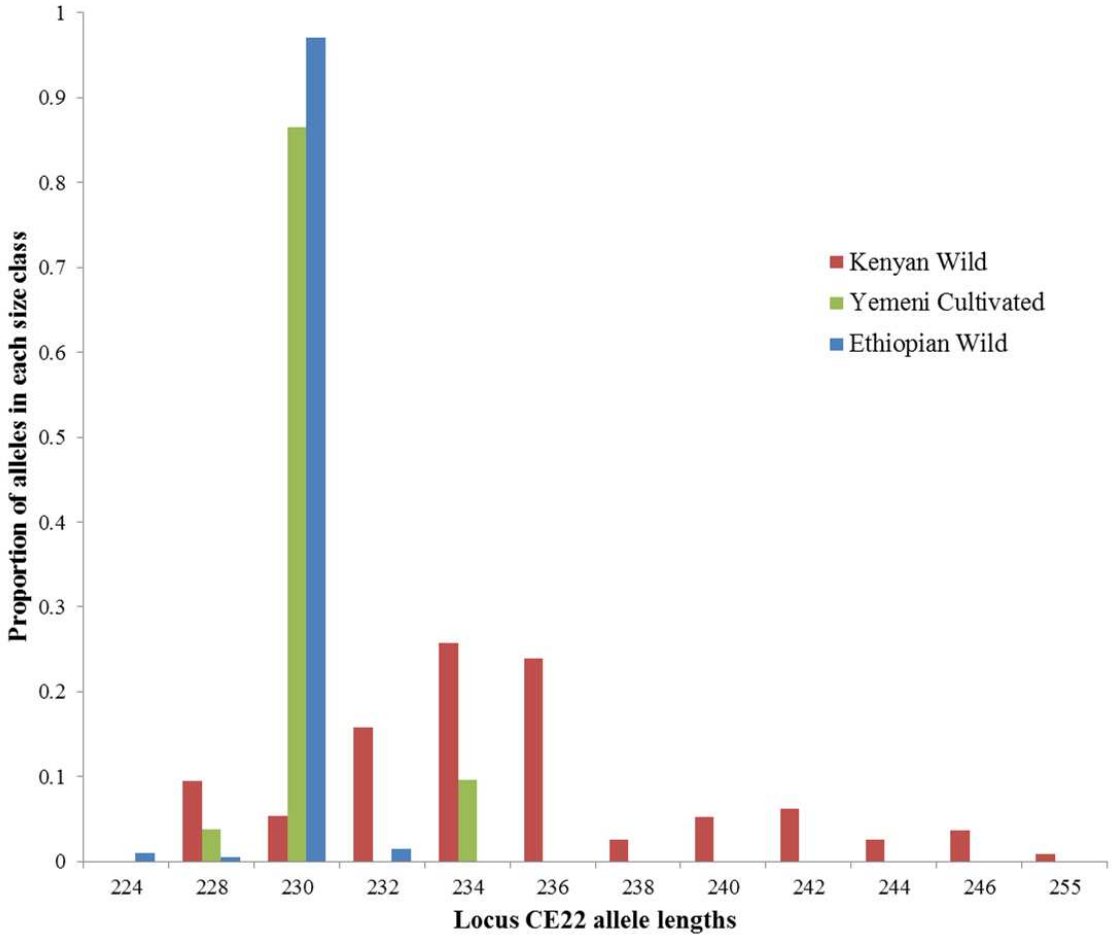
9) Place 900 µl (this volume is dependent on column type; overloading can result in shearing damage to the DNA) of mixture from step 8 onto column, force mixture across column by spinning in centrifuge for 3 minutes at 6,000 rpm, repeat this step until all of the mixture has been past across the column. Dispose of flow through product at the end of each centrifuge cycle. Caution: guanidine hydrochloride is a strong protein disrupting salt and should be disposed of as hazardous waste.

10) Clean the column by twice passing 500 µL of wash buffer I across the column with 1 minute of centrifugation at 6000 rpm. Heating the wash buffers to 60°C will increase the efficiency of salt removal during the wash steps.

11) Clean the column by twice passing 500 µL of wash buffer II across the column with 1 minute of centrifugation at 6000 rpm.

12) To remove residual ethanol from the column spin a final time at 6000 rpm for five minutes without adding ethanol. Dispose of the entire collection tube/plate and residual ethanol, being careful not to splash the underside of the column with ethanol.

13) Place column into a clean 1.5 mL tube or collection plate, add 100 µL of (60° C) elution buffer, let sit for 5 minutes to elute DNA, spin column for 1 minute and repeat this step once more if a second elution is needed.



Appendix 4.2 The distribution of allele sizes by geographic origin indicating a possible convergence among Yemeni cultivated qat at locus CE 22.

Appendix 4.3. Private alleles among cultivated genotypes collected in Marsabit and their subsequent presence in Mt Kulal and Mt Marsabit hybrids.

Origin	Accession	CE22	CE22	CE37	CE37	CE41	CE41	CE42	CE42	CE43	CE43
Mt Kenya/Meru	2413b	234	236	206	218	216	216	234	250	147	147
Mt Kenya/Meru	2414b	236	236	202	218	216	234	248	250	143	147
Mt Kenya/Meru	2414c	234	246	202	218	216	216	230	250	147	151
Mt Kenya/Meru	2416c	234	234	202	218	216	234	228	250	147	147
Mt Kenya/Meru	2417c	234	246	202	218	216	232	230	250		
Mt Kenya/Meru	2419b	236	236	202	218	216	234	248	250	143	147
Mt Kenya/Meru	2419c	234	246	202	218	216	216	228	250	143	147
Mt Kenya/Meru	2419d	234	234	202	218	216	216	228	250	143	147
Mt Marsabit/Hybrid	2416a	230	246	202	226	216	245	240	250	147	163
Mt Marsabit/Hybrid	2419e	230	232	202	226	234	245	240	250	147	163
Mt Marsabit/Hybrid	2419f	230	232	218	226	216	251	240	248	147	163
Mt Marsabit/Hybrid	2419g	230	230	202	222	216	251	240	248	147	163
Mt Kulal/Hybrid	2425a	230	234	218	222	232	245	230	242	151	163
Mt Kulal/Hybrid	2425c	230	234	202	226	232	245	230	240	151	163
Mt Kulal/Hybrid	2426d	230	234	218	222	232	245	230	242	151	163
Mt Kulal/Hybrid	2427b	230	246	218	226	216	251	242	250	147	163
Mt Kulal/Hybrid	2428a	230	234	202	226	232	245	230	240	151	163
Ethiopia	2412a	228	228	222	222	245	251	238	242	163	163
Ethiopia	2412b	230	230	222	226	245	251	240	242	163	163
Ethiopia	2413a	230	230	222	226	245	251	240	242		
Ethiopia	2414d	228	230	222	222	245	251	238	240	163	163
Ethiopia	2414e	228	228	222	222			238	242	163	163
Ethiopia	2416b	228	230	222	222	251	251	242	242	163	163
Ethiopia	2417b	228	228	222	222	245	251	238	242		