Investigations of the plant genus *Chenopodium* using both plastid and single-copy nuclear markers

By

Brian M. Walsh

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Botany)

at the

UNIVERSITY OF WISCONSIN-MADISON

2014

Date of final oral examination: April 29, 2014

The dissertation is approved by the following members of the Final Oral Committee:
Eve Emshwiller, Associate Professor, Botany
David A. Baum, Professor, Botany
Kenneth J. Sytsma, Professor, Botany
Johanne Brunet, Associate Professor, Entomology
Sissel Schroeder, Professor, Anthropology

SUMMARY

Investigations into the plant genus *Chenopodium* using both plastid and single-copy nuclear markers (under the direction of Eve Emshwiller)

The plant genus *Chenopodium* is poorly understood despite being domesticated as many as five times, and being one of the most noxious weeds known in agricultural and urban settings. Taxonomic problems at both the genus and species level have persisted for over 250 years. Due to the large number of species, nomenclatural inconsistencies, high degree of morphological trait plasticity, and trait similarities among relatively distantly related species, researchers have progressed very slowly in understanding *Chenopodium*. It wasn't until molecular studies were utilized, primarily plastid sequence data, that taxonomic inconsistency at the family, genus, and species level really became apparent. Work throughout the 1990s suggested Chenopodiaceae was not monophyletic and in 1998 was merged with Amaranthaceae. For the next 14 years *Chenopodium* infrequently appeared in phylogenetic analyses, but when species were included, *Chenopodium* was never the focus of the study and relatively few specimens were sampled. These studies demonstrated that the genus *Chenopodium* was not monophyletic by virtue of including other genera such as *Atriplex*, but the sparse sampling didn't allow further understanding.

I began research on *Chenopodium* with the intent of constructing the first comprehensive phylogenetic analysis, specifically addressing the polyphyly observed by other researchers using plastid DNA sequence data (Chapter 1). However, in 2012 a phylogenetic study of *Chenopodium* based on plastid DNA was published by other authors (Fuentes et al., 2012a; 2012b) who proposed dozens of nomenclatural changes based on their work. In response, I

modified the scope of my analysis to test the validity of their results using a different, more variable, intergenic plastid marker and different sampling. Our results were congruent with Fuentes et al. (2012a, 2012b) supporting all nomenclature changes except the inclusion of the Australian genera *Rhagodia* and *Einadia* within *Chenopodium*.

Chapter 2 of my research investigates the polyploid species of *Chenopodium* using a phylogeny constructed from the single-copy nuclear locus, *Salt Overly Sensitive 1* (*SOS1*). The polyploid (4-8*x*) species of *Chenopodium*, include the most widespread weeds in the genus, and at least three of the five domesticates, including *C. quinoa*. Previous researchers had identified the American polyploids as allotetraploid, but little else was known. Plastid markers are typically maternally inherited and are incapable of revealing the full genetic history of hybrid taxa. Likewise ITS is often problematic due to concerted evolution. Single/low-copy nuclear loci offer the best strategy for resolving allopolypoid ancestry and identifying each genome donor. Two distinct polyploidy lineages were identified, one American tertaploids, and the other Eurasian hexaploids. These lineages possess different combinations of homeologs indicating separate polyploidy origins.

There are five domesticated cultigens of *Chenopodium*, including the only eastern hemisphere domesticate, *C. giganteum* that originated in central Asia. The remaining four domesticates are native to the Americas. C*henopodium pallidicaule* (cañihua) is the only diploid domesticate and is cultivated in the Peruvian and Bolivian highlands. *Chenopodium quinoa* (quinoa) is the most widely known of domesticated *Chenopodium*, originating in Peru and Bolivia, but currently cultivated also in the United States of America, China, Pakistan, and New Zealand. *Chenopodium berlandieri* subsp. *nuttalliae* of Mesoamerica is morphologically very similar to *C. quinoa*. And, finally, *Chenopodium berlandieri* subsp. *jonesianum* is an extinct

cultigen of North America known from 800-3850 year old archaeological excavations (Fritz and Smith, 1988; Smith and Yarnell, 2009).

Chapters 1-2 address the relation among the extant cultigens, each apparently representing an independent domestication. Chapter 3 focuses specifically on testing hypotheses on the relationship among the extinct North American and the Mesoamerican domesticated *Chenopodium*, by comparing DNA sequence data obtained from domesticated archaeological seeds with modern populations.

Kistler and Shapiro (2011) published a very similar study to mine demonstrating that the North American and Mesoamerican domesticates were genetically distinct and the result of independent domestication events. Their experimental design was nearly identical to what I was preparing to do, coincidently using the same *trnQ-rsp16* plastid marker, and seeds from two of the same archaeological sites. Some fundamental differences existed between our studies, however: 1) their study included three archaeological sites while mine include seven; 2) they obtained sequence data from one seed per archaeological site while I sequenced as many as 4 seeds per site; and 3) I studied both plastid and single-copy nuclear markers. I found haplotype variation among seed assembeges, identifying three distinct genetic patterns: American wild-type-like, Eurasian wild-type, and sequences strikingly similar to Mesoamerican cultigen still still under cultivation. These results indicate that the domesrticated Chenoopodium of Eastern North America and Mesoamerica were derived from a single domestication. These results cannot determine in which region *Chenopodium* was domesticated, but the archaeological evidence of domesticated *Chenopodium* is 3000 years older in Eastern North America than in Mesoamerica. This striking disparity in archaeological evidence suggesting Chenopodium was domesticated in Eastern North America and much later introduced to Mesoamerica.

TABLE OF CONTENTS

CHAPTER 1

INVESTIGATIONS INTO SUBFAMILY CHENOPODIOIDEAE (AMARANTHACEAE) WITH EMPHASIS ON THE CORE *CHENOPODIUM* SPECIES, USING THE PLASTID INTERGENIC *trnQ-rsp16* MARKER

ABSTRACT

Premise of the study- Chenopodium (Amaranthaceae) is a taxonomically challenging group that remains poorly understood due to a paucity of phylogenetic analyses. One prior phylogenetic study with extensive sampling within *Chenopodium* has been published (Fuentes et al. 2012a, 2012b), which resulted in many taxonomic revisions being proposed. We use a different, more variable, plastid marker and increased sampling to test their results.

Methods- We sequenced the plastid *trnQ-rsp16* intergenic region for much of the Chenopodioideae subfamily and inferred phylogenetic relationships using Bayesian and maximum likelihood methods.

Key results- Our phylogeny is mostly congruent with Fuentes et al. (2012a, 2012b), however 18 American species are genetically very similar suggesting a recent, rapid radiation. Two Eastern Hemisphere species ally with American species suggesting a recent trans-oceanic dispersal..

Conclusions- The taxonomic revisions proposed near the base of Chenopodioideae appear warranted, however realignment of the Australian genera *Rhagodia* and *Einadia* is premature and needs further investigation.

Key Words:

Rhagodia, *Einadia*, *Chenopodium quinoa*, *Chenopodium pallidicaule*, *Chenopodium petiolare*, *Chenopodium bryonifolium*, *Chenopodium mucronatum*

INTRODUCTION

The plant genus, *Chenopodium* L. (Amaranthaceae), has been a taxonomically challenging group since the origin of plant taxonomy. Linnaeus (1753a) introduced *Chenopodium* along with 22 subordinate species. Based on the brief description Linnaeus provided, he must have recognizing that the species he named were quite morphologically distinct, yet it is not clear what traits Linnaeus used to place species within *Chenopodium*; certainly not leaf-shape despite the translation, "goose foot", presumably referring to the shape and dentition of the leaves. Overall these species shared enough traits in common that Linnaeus grouped them within a single genus. In the next volume of the same book series, Linnaeus introduced the genus *Atriplex* (Linnaeus, 1753b), whose range of variation overlaps what he would have observed among the *Chenopodium* species he named. Since this initial classification, researchers have recognized that *Chenopodium* is a large, complicated genus, resulting in diverse, often contradictory subgeneric classification schemes (Moquin-Tandon, 1840; Aellen and Just, 1943; Wahl, 1954; Mosyakin and Clemants, 1996). The generic lectotype for *Chenopodium* was decided upon at the 5th International Botanical Conference (Seward, 1931), only after years of debate. With the adoption of the lectotype, researchers finally had a foundation for generic consensus, yet even then researchers using only morphological methods had a difficult time making sense of the taxonomic groupings within the genus.

There are two main reasons why researchers have been unable to agree upon *Chenopodium* taxonomy. The first major problem is that there appears to be abundant homoplasy in the genus, so different taxonomists obtain different results depending on which traits they based their classification upon. An extreme example of morphological characters being misleading is evident in *Chenopodiastrum murale* (syn. *Chenopodium murale*) and

Chenopodium album. These species are extremely similar for many morphological traits, and are most easily distinguished by subtle differences in pericarp texture, and by the somewhat less dentate leaf margins of *C. album*. Yet genetic data has lately confirmed that these similarities are due to convergent evolution (Fuentes et al., 2012a; 2012b).

The second major problem contributing to the taxonomic confusion of *Chenopodium* is that researchers often disagree on species delimitation. Some researchers recognized that certain species exhibit high degrees of phenotypic plasticity, while other researchers have interpreted this variability as indicative of distinct species. Partly as a result of those inconsistencies there are over 1000 proposed *Chenopodium* species, most of which are synonyms. However, the taxonomy is so cumbersome and confused that researchers can only guess how many species are included within *Chenopodium*; if one accepts recent revisions there are probably only 50-75 species.

Molecular analyses have had a massive impact on *Chenopodium* taxonomy. A growing body of broad phylogenetic studies (Hohmann et al., 2006; Kadereit et al., 2003, 2005, 2010), have demonstrated that *Chenopodium*, as traditionally delimited, is non-monophyletic, resulting in several alternate taxonomic schemes (Mosyakin and Clemants, 2002, 2008; Clemants and Mosyakin, 2003; Zhu et al., 2003; Mosyakin, 2012). These molecular studies did not, however, focus specifically on relationships within *Chenopodium* s. str. They often used the same 5-8 species, so their utility in addressing intrageneric taxonomic relationships was limited. Two *Chenopodium* phylogenies were recently published; one using plastid non-coding *trnL-trnF* and nuclear ITS sequence data for 142 taxa (Fuentes et al., 2012a), while the other (Fuentes et al., 2012b) included about half the taxa, most the same samples included in the former study, but with additional sequence data for plastid non-coding *matK-trnK*. These represent a leap forward

in our understanding of *Chenopodium*, and specifically address the extreme polyphyly observed by previous researchers. Fuentes et al. (2012a, 2012b) identified a highly supported core *Chenopodium* clade sister to *Atriplex*. The core *Chenopodium* clade includes the type species for the genus, *C. album*, but species relationships within this clade are poorly resolved. Fuentes et al. (2012a, 2012b) proposed transferring all genera within the core *Chenopodium* clade to *Chenopodium*, including the Australian genera *Rhagodia* and *Einadia*, where as all former *Chenopodium* species outside the core *Chenopodium* clade were transferred to other genera.

The phylogenetic analyses of Fuentes et al. (2012a, 2012b) used the two most commonly used plastid regions, *trnL-trnF* and *matK-trnK* (Shaw et al., 2005). Some non-coding plastid loci are more variable than others, but this is largely dependent upon the taxa sampled. We will test multiple plastid loci to identify a more variable region to better resolve species relationships among the poorly-resolved core-*Chenopodium* clade.

The object of this research is 1) to test the phylogenetic analyses of *Chenopodium* that served as the basis for many major taxonomic revisions (Fuentes et al. 2012a, 2012b) using a different plastid locus, 2) to find a more variable plastid region to better resolve species relationships within the core-*Chenopodium* clade, and 3) to add 11 species of *Chenopodium* that have never been included in previous analyses, including samples from under-studied geographic regions.

MATERIALS AND METHODS

Taxon sampling - In total, 234 samples were included in this study representing 77 species. This study expanded the sampling of the core *Chenopodium* clade by 25%. *Microtea debilis* of the Phytolaccaceae was included as outgroup. Additionally *Beta*, *Guilleminea*, *Amaranthus*, and *Salsola* taxa were all included as representatives of the broader Amaranthaceae. Each of these genera belongs to different subfamilies and are relatively distantly related to each other.

Samples were collected from various sources either as seeds, fresh leaf tissue, or herbarium samples (Table 1). Samples obtained as seeds were grown to maturity in the University of Wisconsin-Madison Walnut Street greenhouse to verify species identity. Voucher samples were obtained for all specimens and deposited in the WIS herbarium.

DNA extraction - DNA was extracted from either fresh or dried leaf material. Fresh leaf material was harvested either from greenhouse-grown specimens or onsite for wild populations. Tissue that was not promptly used was stored frozen at -80C until just prior to DNA extraction. Herbarium and otherwise dried samples were stored at ambient temperature in an herbarium quality cabinet. DNA was extracted from approximately 0.05-0.1 g of fresh or frozen leaf tissue or 0.01-0.05 g from dried samples, according to Alexander *et al.* (2006), using silica membrane columns (Epoch Life Science, Sugar Land, TX, USA).

Gene selection target amplification, and sequencing - 23 plastid regions were assessed for variability based on sequence data for four reference samples (*C. album*, two accessions of wild *C. belandieri*, and a domesticated *C. berlandieri* subsp. *nuttalliae*). The initial screen assessed variability between *C. berlandieri* and *C. album*, and between wild and domesticated *C.*

berlandieri by comparing the total number of polymorphic characters and introns. The *trnQ*-5'*rsp16* intergenic spacer was selected for further amplifications based on the initial screen.

Target amplification **-** PCR reactions were performed in 25 µl reactions (1 µl DNA, 0.2 µl *Taq* polymerase (5 units/µl, New England BioLabs, Ipswich, Massachusetts, USA), 1.0 µl of each primer pair (10 µM ea.), 1.25 µl BSA (10 mg/ml), 2 µl dNTP (reagent consists of 200 µM of each nucleotide), 2.5 µl 10x ThermoPol Reaction buffer (1.5 mM $MgCl_2$ included, New England BioLabs, Ipswich, Massachusetts, USA) as follows: initial denaturation was at 94°C for 5 m; followed by 35 cycles of 94°C for 30 s, 52°C for 1 m and 72°C for 1.5 m; and 72°C for 7 m.

PCR fragments were visualized with ethidium bromide staining after separation on a 1% Agarose gel, then purified using Agencourt Ampure beads (Beckman Coulter Genomics, Morrisville, North Carolina, USA). Purified samples were cycle sequenced (25 cycles of 96°C for 10 s and 50°C for 2 m) using either the trnQ(UUG) or rsp16x1 primers (Shaw et al. 2007) and ABI Big Dye Terminator and run on a ABI 3730 automated sequencer (Applied Biosystems, Carlsbad, California, USA) at the University of Wisconsin-Madison Biotechnology Center (UWBC).

Alignment assembly and indel scoring **-** Sequences were edited and contigs assembled using Sequencher 4.9 (Gene Codes Corporation, Ann Arbor, MI, USA). Contigs were manually aligned using MacClade 4.08a OS X (Maddison and Maddison, 2005).

The total dataset was divided into two smaller datasets, consisting of a core *Chenopodium* dataset and a non-core *Chenopodium* dataset. For the core *Chenopodium* dataset, all indels except autapomorphies were scored as binary characters and added to the end of the alignment. Indels were weighted equivalent to sequence data. For the highly-divergent non-core

Chenopodium dataset, indels were not included due to complex nested patterns that could not be scored with confidence.

Phylogenetic Analyses **-** The program jModelTest v2.1.3 (Darriba et al., 2012) was used to select the most appropriate model of sequence evolution for each dataset for both maximum likelihood and Bayesian analyses. All tree-files were visualized with the program FigTree v1.3.1 (Rambaut, 2009).

Maximum Likelihood Analyses - ML analyses with 1000 bootstrap replicates were performed using the CIPRES Science Gateway v3.1 (Miller et al. 2010), using the RAxML-HPC on BlackBox software (Stamatakis, 2006; Stamatakis et al. 2008) with the model GTR+gamma.

Bayesian Analysis - Each Bayesian phylogenetic analysis was conducted using the CIPRES Science Gateway v3.1, using MrBayes 3.2.2 on XSEDE (Huelsenbeck and Ronquist, 2001). Each analysis consisted of two runs of 500 million generations, starting from a random tree and sampling every 50 000 generations. Analyses used four Metroplolis-coupled chaines. Based on preliminary analyses, the temperature was set to 0.14 for the non-core *Chenopodium* dataset, and 0.115 for the core *Chenopodium* dataset. The first 25% of trees from each run were discarded as burn-in.

RESULTS

Choice of plastid marker- Since plastid sequence markers are reported to be generally more conserved than nuclear loci (Clegg et al., 1994 ;Pillon et al., 2013; Turner et al., 2013), 23 markers were assessed to determine the most variable. Among the four reference taxa sequenced, all regions demonstrated high variability among *C. album* and *C. berlandieri* taxa. However, only four markers demonstrated intraspecific variability among the *C. berlandieri* taxa sampled, *trnQ-rsp16*, *trnL-trnT*, *ndhJ-trnF-trnL*, and *psbD-trnT*. The *trnQ-rsp16* region was selected because it demonstrated two variable characters among the *C. berlandieri* taxa compared to zero or one variable character observed among other plastid markers. Amplified sequences for *trnQ-rsp16* ranged between 369 and 678 bases, 569-678 bases for the core *Chenopodium* clade. There were 40 and 265 parsimony informative sites in the core *Chenopodium* and non-core *Chenopodium* datasets, respectively. Nine potentially informative indels were observed and scored for the core *Chenopodium* dataset.

Non-core *Chenopodium* results- For clarity, the nomenclatural reassessments proposed in Fuentes et al. (2012b) and Mosyakin and Clemants (2002) will be used in the text, tables, and figures to better discuss and visualize comparisons across multiple studies.

Here we briefly describe the major clades that will be discussed further below (Fig. 1).

Outgroup: *Beta trigyna*, representing subfamily betoideae, was found to be sister to the rest of Amaranthaceae. The relative position of this sample is inconsistent with most other studies. Two studies (Hohmann et al., 2006; Müller and Borsch, 2005) with substantially better sampling within subfamilies, observed *Beta* sister to *Salsola*. Our phylogeny, as well as that in Kadereit et

al. (2005), places *Beta* unresolved with reference to *Amaranthus*, but clearly not associated with *Salsola*. However, these nodes are poorly supported.

Guilleminea densa and *Amaranthus* taxa, representing subfamilies Gomphrenoideae and Amaranthoideae, unite in a moderately supported clade, consistent with other studies (Müller and Borsch 2005, Kadereit et al. 2003). Under the given sampling *Salsola* sp., representing subfamily Salsoloideae, is sister to the Chenopodioideae.

Blitum/Spinacia clade: *Blitum* taxa are united, with high support, in a well supported clade, sister to a well supported *Spinacia* clade. This result is fully congruent with Fuentes et al. (2012a, 2012b), but conflicts with other previously published studies (Kadereit et al., 2003, 2005, 2010; Müller and Borsch, 2005; Hohmann et al., 2006). This study and Fuentes et al. (2012a, 2012b) included a denser sampling within the Chenopodioideae subfamily than in previous studies and has resulted in an improved estimation of these relationships.

Dysphania/Teloxys clade: *Teloxys aristata* taxa unite in a subclade sister to a *Dysphania* subclade. These subclades together form a highly supported *Dysphania/Teloxys* clade. The close relationship of *Dysphania* and *Teloxys* is consistent among all studies that contain representatives of these genera (Fuentes 2012a, 2012b; Kadereit et al. 2003, 2005).

The relative positions of the *Blitum/Spinacia* and *Dysphania/Teloxys* clades are often not resolved, forming a polytomy (Fuentes et al., 2012b; Kadereit et al., 2003, 2005, 2010). The current study suggests, albeit weakly, the *Blitum/Spinacia* is more distantly related from the core *Chenopodium* clade than is the *Dysphania/Teloxys*.

Oxybasis/Lepandra clade: A moderate to weakly supported clade unites *Lipandra polyspermum*, *Chenopodium carnosulum*, and several species of *Oxybasis*. The monotypic genus *Lipandra*

forms a highly supported subclade sister to the *Oxybasis* clade, which itself is highly supported and further subdivided into subclades. The *Chenopodium carnosulum* sample was not included in previous phylogenetic studies and thus no nomenclature change had previously been proposed. This sample is sister to an *Oxybasis rubra* subclade.

Chenopodiastrum clade: A highly supported clade unites *Chenopodiastrum hybridum*, *C. simplex*, and *C. murale*. The former two species are characterized by large leaves reminiscent of oak leaves, while *C. murale* has leaves very similar to species in the core *Chenopodium* clade. *Chenopodiastrum murale* unite in a maximally supported subclade, sister to alternating *C. simplex*, and *C. murale* samples.

Atriplex clade: *Atriplex* species unite in a highly supported clade sister to the core *Chenopodium* clade. This phylogenetic analysis, as well as others (Fuentes et al., 2012a, 2012b), support *Atriplex* as sister to the core *Chenopodium* clade. As such, an *Atriplex* species was included in the core *Chenopodium* dataset as an outgroup.

Core Chenopodium clade- Here we briefly describe the major clades of the core-Chenopodium clade that will be discussed further below (Fig. 2).

Australian clade: The Australian taxa unite, sister to the rest of the *Chenopodium* clade, although support for this clade is very weak. The Australian clade was also observed by other researchers and is composed of mixed genera, including *Rhagodia*, and in other studies *Einadia* (Fuentes et al. 2012a, 2012b).

Fuentes et al. (2012a, 2012b) also observed the Australian species uniting within a mixed generic clade, although the clade was unresolved within the core-*Chenopodium* clade.

Chenopodium clade: Other than *C. vulvaria* and *C. carnosulum* all *Chenopodium* taxa form a well supported clade with much internal resolution, similar to prior studies (Fuentes et al. 2012a, 2012b). There are two major subclades, one composed almost entirely of Eurasian taxa, and the second composed mostly of American taxa, but also including *C. bryonifolium* and *C. mucronatum* from China and South Africa, respectively.

Eurasian major subclade: This grouping of taxa will be referred to as the Eurasian major subclade, despite the fact that it includes *C. oahuense* endemic to the Hawaiian Islands. This subclade consists of two additional subclades, one moderate-to-weakly supported consisting entirely of diploid species, and the other a highly supported clade of polyploid species.

Predominantly American major subclade: The well-supported predominantly American major subclade consists of four subclades of note: Old World Ally, Petiolare/pallidicaule, Mixed *Chenopodium*, and Mesoamerican domesticate subclade. The only non-American species in this major subclade are in the Old World Ally subclade.

Old World Ally subclade: This subclade is sister to all other species within the Premominantly American major subclade, and unites the only two non-American taxa, *C. bryonifolium* from China, and *C. mucronatum* from South Africa. This clade is not resolved in likelihood analyses.

Petiolare/pallidicaule subclade: The second subclade of note is the well supported union of wild *C. petiolare* and domesticated *C. pallidicaule*, both endemic to Peru and Bolivia.

Mixed *Chenopodium* subclade: The third and largest subclade, based on this sampling, unites 18 species among a series of nested polytomies. There are small subclades within the polytomies that tend to be composed of odd mixtures of species with no discernable pattern.

Mesoamerican domesticate subclade: Nested within the Mixed *Chenopodium* subclade is a smaller subclade uniting the nine accessions of the Mesoamerican domesticate, *C. berlandieri* subsp. *nuttalliae*.

DISCUSSION

Non-core Chenopodium clades- In this study there is essentially no support for the relative position of the *Blitum/Spinacia* clade relative to the core *Chenopodium* clade and the *Dysphania/Teloxys* clade. Other studies that included representative taxa for these clades could not resolve the relative positions for these clades either (Fuentes et al. 2012b; Kadereit et al. 2003, 2005, 2010). The only exception was the study of Fuentes et al. (2012a), which had strong support for *Dysphania-Teloxys* sister to *Spinacia-Blitum* and the core-*Chenopodium* clade. Despite this, and given the conflicting resolution obtained with almost the same dataset (Fuentes et al., 2012b), we conclude that the relationship among the *Blitum/Spinacia* and *Dysphania/Teloxys* clades cannot be ascertained with confidence at this time.

Other than *Chenopodiastrum murale*, the *Chenopodiastrum* clade includes *C. hybridum* found throughout temperate Eurasia, and *C. simplex* found across temperate North America. Morphologically these species are nearly indistinguishable, with identification typically based on provenience rather than inherent traits. In this study, *C. hybridum* was non-monophyletic because one of the two Russian accessions were found to be sister to the single *C. simplex* from Wisconsin. These results suggest *C. hybridum* and *C. simplex* should perhaps be considered the same species, however the relationship among these samples is poorly resolved.

Core-Chenopodium clades- Despite attempts to optimize analysis parameters and enhance resolution, the base of the core clade could not be resolved with any confidence. This is consistent with the findings of Fuentes et al. (2012a, 2012b). These plastid regions do not appear to have the necessary genetic variability to distinguish the relationships among the Australian taxa (*Rhagodia* and *C. auricomum*) and the pan-European *C. vulvaria*. However, the Australian

taxa and *C. vulvaria* are individually or collectively highly supported as sister to the rest of the core *Chenopodium* clade.

Fuentes et al. (2012a) included one sample from each of the Australian genera, *Rhagodia*, *Einadia*, and *Chenopodium*, and observed that Australian samples, regardless of genera, unite in a highly supported clade. Despite the lack of resolution among the Australian clade and the core-Chenopodium clade, Fuentes et al. (2012a) proposed subsuming *Rhagodia* and *Einadia* into *Chenopodium*. Furthermore, both *Rhagodia* and *Einadia* produce red fleshy fruit unlike any species of *Chenopodium*. In light of these findings, subsuming *Rhagodia* and *Einadia* into *Chenopodium* is premature and warrants further research. A single-copy nuclear phylogeny based on DNA sequences of *Salt Overly Sensitive 1* (*SOS1*; Ch. 2), demonstrated considerably better resolution at the base of the core-*Chenopodium* clade than has been observed using plastid sequence data, and this phylogeny found a highly supported clade uniting *C.vulvaria* with the Australian *C. auricomum* as sister to the core-*Chenopodium* clade. Neither *Rhagodia* nor *Einadia* samples were included in the *SOS1* analysis, but assuming the Australian genera unite in a clade as they do in plastid analyses (Fuentes et al., 2012a, 2012b), the Australian genera may form a subclade distinct from the core-*Chenopodium* clade. Again, without further research the relationship of Australian genera to *Chenopodium* is currently too uncertain to warrant subsuming *Rhagodia* and *Einadia* into *Chenopodium*.

Core-Chenopodium clade: Fuentes et al. (2012a, 2012b) barely addressed this clade and its subclades, despite observing some of the same subclade patterns as shown in our study. This well-supported clade is composed of two major subclades, one consisting almost exclusively of Eurasian taxa, and the other of predominantly of American taxa.

Eurasian subclade: This major subclade includes several particularly weedy species which are occasionally referred to by researchers as the *C. album* complex, named after the most successful weed species in the genus, and the only one with a global distribution. This subclade can be further divided into two subclades, one consisting of diploid species and the other of polyploid species. These results are consistent with the phylogeny based on nuclear *SOS1* data (Ch. 2). However, the *SOS1*-based results found two additional clades composed of allopolyploid homeologs that have never been observed in plastid based studies (Ch. 2).

Predominantly American subclade: This large subclade is dominated by the highly supported 'Mixed *Chenopodium* subclade' consisting of 118 taxa spanning 18 species. The 'Mixed *Chenopodium* subclade' is largely unresolved as a series of polytomies including several small moderately supported subclades including 2-5 accessions. Taxon groupings do not appear to ally by species or geography. Some of the small subclades form sensible associations, such as uniting two *C. subglabrum* taxa collected in neighboring states. However, there are several small interspecific subclades that unite odd associations of species (e.g., *C. fremontii*-5 endemic to Southwestern United Sates groups with *C. hircinum* and *C. quinoa* from South America), as well as an intraspecific subclade (*C. berlandieri*-1 & *C. berlandieri* var. *bushianum*-1) that unites samples collected 750 miles apart, from South Dakota and Wisconsin, but does not associate with conspecific samples from Iowa. The patterns observed among the 'Mixed *Chenopodium* subclade' could be explained by incomplete lineage sorting, hybridization, introgression, or several other mechanisms. This analysis is not capable of identifying which mechanisms are at play. However, the polytomy layers include many interspecific taxa sharing identical sequences for the plastid marker used. This sequence similarity extends across ploidy differences as well as across North and South America. The sequences of *C. frigidum* of Tierra del Fuego are identical

to *C. berlandieri* subsp. *bushianum*-2 from central Wisconsin. *Chenopodium* species are capable of hybridizing, even with differences in ploidy (Wilson, 1980), but the genetic similarity seems too mixed and wide-spread. Perhaps humans of the past and present, intentionally or not, have facilitated the spread of *Chenopodium* across the Americas. One of the world experts on *Chenopodium* taxonomy speculated that American *Chenopodium* is probably very young genus, possibly as recent as 10 000 years (Mosyakin, 1995). Humans have been in the Americas for nearly 30 000 years (Greenberg et al., 1986; Guidon and Delibrias, 1986).

Sister to the rest of the 'Predominantly American subclade' is a very weakly-supported subclade consisting of two Eastern Hemisphere species, *C. bryoniifolium* of China and *C. mucronatum* of South Africa, together here called the 'Transoceanic subclade". The inclusion of these Old World species within a clade of otherwise American species suggests a transoceanic introduction from one continent to another. Since the 'Predominantly American subclade', of which the Transoceanic subclade is included, is sister to the Eurasian clade, we speculate that American *Chenopodium* probably derived from a transoceanic colonization originating from the Old World.

Domesticated taxa- This study includes representative taxa for three *Chenopodium* cultigens of the Americas: *C. berlandieri* subsp. *nuttalliae* endemic to the tran-Mexican Volcanic belt in Mesoamerica, *C. quinoa* originally from Peru and Bolivia, but currently grown in many countries, and *C. pallidicaule* of the Peruvian/Bolivian altiplano. Each domesticated crop lineage is isolated within the phylogeny supporting the idea that they were each independently domesticated, and not resulting from an introduction from one region to another. The sampled individuals of the Mesoamerican domesticate, *C. berlandieri* subsp. *nuttalliae*, unite within a well-supported clade, the 'Mesoamerican subclade', that contains only wild species. The most

widely cultivated of the American domesticates, *C. quinoa*, does not unite in a clade unto its own, but shares plastid sequences identical to several wild species. This crop is also the only domesticated lineage that does not form a monophyletic lineage. One set of accessions is grouped with *C. fremontii* and *C. hircinum*, whereas all other accessions are placed in a large polytomy with a number of species. This suggests periodic plastid introgression among wild and domesticated populations throughout the history of this crop. Episods of hybridization and introgression could explain why *C. berlandieri* of North America and *C. hircinum* of South America are morphologically indistinguishable and probably should be considered the same species, except *C. hircinum* often has rounded leaf apices similar to many *C. quinoa* cultivars.

The least commonly cultivated domesticate is *C. pallidicaule*, grown in the Altiplano of Peru and Bolivia. Samples of *C. pallidicaule* unite in a highly-supported clade sister to a wild species that grows in the same region, *C. petiolare*. These results are in agreement with the phylogeny based on nuclear sequence data (Ch.2), which included a different accession of wild *C. petiolare*. The results of these two studies suggest *C. petiolare* is likely the progenitor species from which *C. pallidicaule* was domesticated. A similar relationship is observed in the phylogenies of Fuentes et al. (2012a, 2012b). However the relationships in their analyses are different in that two of their four *C. atrovirens* samples unite in a clade with *C. petiolare*. The other two *C. atrovirens* samples join the 'Mixed *Chenopodium'* clade, which is where the two *C. atrovirens* sample in our study lie. The petiolare/pallidicaule and the 'Mixed *Chenopodium'* clades in this study and Fuentes et al. (2012a, 2012b) are genetically quite distinct, and it is unlikely that *C. atrovirens* is actually represented in both clades. We suspect theses samples may be misidentified, however plastid introgression is another possibility.

Assessment of recent taxonomic revisions- Several researchers have proposed nomenclatural changes within subf. Chenopodioideae, the subfamily that was formerly the family Chenopodiaceae (Mosyakin 2013; Fuentes et al. 2012a, 2012b; Mosyakin and Clemants, 2002). Our current results support some of these proposed changes, but not all.

We fully agree with the proposed nomenclatural changes and recognize the genera, *Blitum*, *Teloxys*, *Disphania*, *Lipandra*, *Oxybasis*, and *Chenopodiastrum*.

The *Chenopodium carolsulum* sample included in this study has not been included in previous analyses and should be transferred to *Chenopodiastrum*.

The Fuentes et al. (2012a) study proposed subsuming the Australian genera *Rhagodia* and *Einadia* into *Chenopodium* based on their inclusion within the highly supported core *Chenopodium* clade. Despite *Einadia* not being included in this analyses, the lack of resolution for the Australian clade with respect to the core-*Chenopodium* clade, as seen in this study as well as Fuentes et al. (2012a), does not warrant nomenclatural changes at this time. Additional studies are required to determine if Australian taxa should be subsumed into *Chenopodium*, maintained, or designated their own genus or genera.

Future directions- Molecular studies have significantly and positively impacted the taxonomy of *Chenopodium*. In light of recent taxonomic revisions, especially Fuentes et al. (2012b), the taxonomic confusion surrounding this genus is largely alleviated. With this new, substantially clearer delineation of the genus, researchers can focus their attention on the many species level taxonomic problems that remain.

Plastid sequence data does not seem to have the variation necessary to resolve all species relationships within the core of *Chenopodium*. So despite the considerable refinement and

reduction in *Chenopodium* taxa, the genus is still challenging because species are difficult to distinguish, both morphologically and genetically, they hybridize readily, vary in ploidy from 2*x*-8*x*, and often demonstrate graded variation making species identification difficult, if not impossible. This likely reflects the fact that *Chenopodium* is a relatively young lineage, probably less than five million years old (estimated from Kadereit et al., 2005), yet the genus has taken advantage of its saline tolerance and xeric habits and prolific weedy nature, to colonize the globe. We may hope that through the use of additional molecular markers we can ultimately clarify some of the outstanding systematic problems, such as the association between Australian taxa and *Chenopodium*, the biogeographic history of Old World-New World dispersal, and the history of polyploidy.

REFERENCES

Aellen, P. and T. Just. 1943. Key and Synopsis of the American Species of the Genus Chenopodium L. *American Midland Naturalist* 30: 47-76.

Clegg, M., B. Gaut, G. Learn, Jr., and B. Morton. 1994. Rates and patterns of chloroplast DNA evolution. *Proceedings of the National Academy of Sciences USA* 91: 6795-6801.

Clemants, S. and S. Mosyakin. 2003. *Chenopodium*. In Chenopodiaceae. Flora of North America North of Mexico. Oxford University Press. Volume 4: 275-299.

Fritz, G. and B. D. Smith. 1988. Old collections and new technology: documenting the domestication of *Chenopodium* in Eastern North America. *Midcontinental Journal of Archaeology* 13: 3-27.

Fuentes-Bazan, S., G. Mansion, T. Borsch. 2012a. Towards a species level tree of the globally diverse genus *Chenopodium* (Chenopodiaceae). *Molecular Phylogenetics and Evolution* 62: 359–374.

Fuentes-Bazan, S., P. Uotila, and T. Borsch. 2012b. A novel phylogeny-based generic classification for *Chenopodium* sensu lato, and a tribal rearrangement of Chenopodioideae (Chenopodiaceae). *Willdenowia* 42: 5-24.

Greenberg, J., C.G. Turner II, and S.L. Zegura. 1986. The settlement of the Americas: a comparison of the linguistic, dental and genetic evidence. *Current Anthropology* 4: 477-497.

Guidon, N. and G. Delibrias. 1986. Carbon-14 dates point to man in the Americas 32,000 years ago. *Nature* 321: 769-771.

Hohmann, S., J. Kadereit, and G. Kadereit. 2006. Understanding Mediterranean-Californian disjunctions: molecular evidence from Chenopodiaceae-Betoideae. *Taxon* 55: 67–78.

Kadereit, G., T. Borsch, K. Weising, and H. Freitag. 2003. Phylogeny of Amaranthaceae and Chenopodiaceae and the evolution of C4 photosynthesis. *International Journal of Plant Sciences* 164: 959–986.

Kadereit, G., G. Dietrich, J. Surrey, and F. Helmut. 2005. Origin and age of Australian Chenopodiaceae. *Organisms, Diversity & Evolution* 5: 59–80.

Kadereit, G., E. Mavrodiev, E. Zacharias, and A. Sukhorukov. 2010. Molecular phylogeny of Atripliceae (Chenopodioideae, Chenopodiaceae): implications for systematics, biogeography, flower and fruit evolution, and the origin of C4 photosynthesis. *American Journal of Botany* 97: 1664-1687.

Linnaeus, C. 1753a. *Chenopodium*, *In* Species Plantarum 1: 218-222.

Linnaeus, C. 1753b. *Atriplex*, *In* Species Plantarum 2: 1052-1054.

Moquin-Tandon, A. 1840. *Chenopodearum monographica enumeratio*. Parisiis, apud P.J. Loss.

Mosyakin, S. 1995. New taxa of *Corispermum* L. (Chenopodiaceae), with preliminary comments on the taxonomy of the genus in North America. *Novon* 5: 340-353.

Mosyakin, S. 2013. New nomenclatural combinations in *Blitum*, *Oxybasis*, *Chenopodiastrum*, and *Lipandra* (Chenopodiaceae). *Phytoneuron* 56: 1–8.

Mosyakin, S. and S. Clemants. 2002. New nomenclatural combinations in Dysphania R. Br. (Chenopodiaceae): Taxa occurring in North America. *Ukrainian Botanical Journal* 59: 380-385.

Mosyakin, S. and S. Clemants. 1996. New Infrageneric Taxa and Combinations in *Chenopodium* L. (Chenopodiaceae). *Novon* 6: 398-403.

Pillon, Y., J. Johansen, T. Sakishima, S. Chamala, W. Barbazuk, E. Roalson, D. Price, and E. Stacy. 2013. Potential use of low-copy nuclear genes in DNA barcoding: a comparison with plastid genes in two Hawaiian plant radiations. *BMC Evolutionary Biology* 13: 35.

Seward, A. C. 1931. *Fifth International Botanical Congress*. CUP Archive.

Shaw, J., E. Lickey, J. Beck, S. Farmer, W. Liu, J. Miller, K. Siripun, C. Winder, E. Schilling, and R. Small. 2005. The Tortoise and the Hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany* 92: 142–166.

Shaw, J., E. Lickey, E. Schilling, and R. Small. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in Angiosperms: the tortoise and the hare III. *American Journal of Botany* 94: 275-288.

Smith, B. D. and R. A. Yarnell. 2009. Initial formation of an indigenous crop complex in eastern North America at 3800 B.P. *Proceedings of the National Academy of Science* 106: 6561–6566.

Turner, B., J. Munzinger, S. Duangjai, E. Temsch, R. Stockenhuber, M. Barfuss, M. Chase, and R. Samuel. 2013. Molecular phylogenetics of New Caledonian Diospyros (Ebenaceae) using plastid and nuclear markers. *Molecular Phylogenetics and Evolution* 69: 740-763.

Wahl, H. 1952-1953 (issued 1954). A preliminary study of the genus *Chenopodium* in North. America. *Bartonia* 27: 1–46.

Wilson, H. D. 1980. Artificial hybridization among species of *Chenopodium* sect. *Chenopodium*. *Systematic Botany* 5: 253-263.

Zhu, G., S. L. Mosyakin, and S. E. Clemants. 2003. *Chenopodiaceae*. Pp. 351–414, *in* Wu Zhengyi and P.H. Raven (eds.), Flora of China. Volume 5. Science Press, Beijing & Missouri Botanical Garden Press, St. Louis.

Table 1. Complete list of taxa represented in the plastid phylogeny using the *trnQ-rsp16* intergenic marker. All samples are in the Amaranthaceae, except *Microtea debilis* which is in Phytolaccaceae. Samples in the subfamily Chenopodioides marked with an asterisk were previously included in *Chenopodium*. The taxa are listed in order of appearance in the phylogenies presented in Figs. 1 and 2.

Taxa	Provenience	Source
Microtea debilis Sw.	Dominica	NY Herb, Hill 1994
Beta trigyna Waldst.	Armenia	USDA, PI264352
Guilleminea densa (Humb. & Bonpl. ex Schult.) Moq.	USA, Arizona	USDA, A30706
Amaranthus blitoides S. Watson	USA, New Mexico	UDSA, Ames27956
Amaranthus caudatus L.	Peru	USDA, Ames5388
Amaranthus quitensis Kunth.	Brazil	USDA, PI652430
Salsola L. sp.	USA, Wisconsin	Walsh, W0079
Spinacia oleracea L.	Afghanistan	USDA, Ames23662
Spinacia turkestanica Iljin	Turkmenistan	USDA, Ames26104
Blitum bonus-henricus (L.) Rchb.	Germany	USDA, Ames23109
Blitum nuttallianum Schult.	USA, Nebraska	USDA, Ames30540
Blitum californicum S.Wats. - 1	USA, California	Jellen, BYU501
Blitum californicum S.Wats. - 2	USA, California	WIS herb, Sauer 1970
Blitum capitatum L. - 1	USA, Michigan	USDA, Ames21982
Blitum capitatum L. - 2	USA, Colorado	USDA, PI664496
Blitum capitatum L. - 3	Armenia	USDA, PI658753
Blitum virgatum L.	Armenia	USDA, Ames26923
Teloxys aristata (L.) Moq. - 1	unknown	B and T World Seeds, 30094
Teloxys aristata (L.) Moq. - 2	Mongolia	USDA, PI650750
Dysphania glomulifera Paul G. Wilson	Australia	Kew MSB, 0086550
Dysphania cristata (F. Muell.) Mosyakin & Clemants	Australia	KewMel0076685
Dysphania pumilio (R. Br.) Mosyakin & Clemants	USA, Wisconsin	WIS herb, Nee 1974
Dysphania carinata (R. Br.) Mosyakin & Clemants - 1	USA, Hawaii	BRY Herb, Thorn 1993
Dysphania carinata (R. Br.) Mosyakin & Clemants - 2	Bolivia	MO Herb, Nee 1998, #04837540
Dysphania schraderiana (Schult.) Mosyakin & Clemants - 1	Hungary	Kew MSB, 0236528
Dysphania botrys (L.) Mosyakin & Clemants - 1	unknown	R Jellen, 01/2011
Dysphania botrys (L.) Mosyakin & Clemants - 2	unknown	Walsh, BRY herb (#5)
Dysphania schraderiana (Schult.) Mosyakin & Clemants - 2	unknown	Kew MSB, 0000228
Dysphania graveolens (Willd.) Mosyakin & Clemants - 1	USA, Arizona	BRY Herb, Higgins 2002
Dysphania graveolens (Willd.) Mosyakin & Clemants - 2	USA, Arizona	NY herb, Atwood 1991 (labeled C. incisium)
Dysphania ambrosioides (L.) Mosyakin & Clemants - 1	Mexico	Xalapa market-Epizote de Arbol
Dysphania burkartii (Aellen) Mosyakin & Clemants - 1	Uruguay	NY herb, Schinini 1995
Dysphania burkartii (Aellen) Mosyakin & Clemants - 2	Uruguay	NY herb, Dematteis & Schinini 2005
Dysphania chilensis (Schrad.) Mosyakin & Clemants	Bolivia	NY herb, Solomon & Nee 1987
Dysphania ambrosioides (L.) Mosyakin & Clemants - 2	USA, New Mexico	USDA, PI604781
Lipandra polysperma (L.) S. Fuentes, Uotila & Borsch - 1	United Kingdom	Kew MSB, 0064831
Lipandra polysperma (L.) S. Fuentes, Uotila & Borsch - 2	unknown	Jellen, BYU915
Lipandra polysperma (L.) S. Fuentes, Uotila & Borsch - 3	Denmark	WIS Herb,Crompton &Roze 1973
Oxybasis urbica (L.) S. Fuentes, Uotila & Borsch	unknown	Jellen, BYU918
Oxybasis chenopodioides (L.) S. Fuentes, Uotila & Borsch - 1	USA, Utah	NY herb, Goodrich 2009, #01156632
Oxybasis chenopodioides (L.) S. Fuentes, Uotila & Borsch - 2	USA, Nevada	NY herb, Trehm 2008, #01155032
Oxybasis glauca (L.) S. Fuentes, Uotila & Borsch - 1	Belgium	Kew MSB, 0174916
Oxybasis glauca (L.) S. Fuentes, Uotila & Borsch - 2	unknown	Walsh, BRY herb (#14)
C. carnosulum Moq.	Bolivia	NY herb, Beck 1995
Oxybasis rubra (L.) S. Fuentes, Uotila & Borsch - 1	USA, Minnesota	WIS Herb, Thomson 1977
Oxybasis rubra (L.) S. Fuentes, Uotila & Borsch - 2	Poland	USDA, PI658747
Oxybasis rubra (L.) S. Fuentes, Uotila & Borsch - 3	Belgium	Kew MSB, 0174927
Chenopodiastrum hybridum (L.) S.Fuentes, Uotila & Borsch - 1	Russia	NY herb, 2010
Chenopodiastrum simplex (Torr.) S.Fuentes, Uotila & Borsch	USA, Wisconsin	Walsh, W0017A
Chenopodiastrum hybridum (L.) S.Fuentes, Uotila & Borsch - 2	Russia	NY herb, Babepyxa 1986
Chenopodiastrum murale (L.) S. Fuentes, Uotila & Borsch - 1	Peru	NY Herb, Edwin 1966
Chenopodiastrum murale (L.) S. Fuentes, Uotila & Borsch - 2	Malta	Malta Seeds Ltd.
Chenopodiastrum murale (L.) S. Fuentes, Uotila & Borsch - 3	Argentina	MO Herb, Robles 2004, #6086417
Chenopodiastrum murale (L.) S. Fuentes, Uotila & Borsch - 4	Malta	Malta Seeds Ltd.
Chenopodiastrum murale (L.) S. Fuentes, Uotila & Borsch - 5	Portugal	USDA, PI614895
Chenopodiastrum murale (L.) S. Fuentes, Uotila & Borsch - 6	USA, California	Jellen, BYU939

Chenopodiastrum murale (L.) S. Fuentes, Uotila & Borsch - 7	USA, California	USDA, Ames26140
Atriplex spongiosa F. Muell	Australia	USDA, PI330668
Atriplex leptocarpa F. Muell.	Australia	USDA, PI342565
Atriplex lentiformis (Torr.) S. Watson	South Africa	USDA, PI409120
Atriplex hortensis L.	France	USDA, PI420154
Atriplex patula L. - 1	USA, South Dakota	Walsh, BL09-1
Atriplex patula L. - 2	USA	USDA, W6_30007
Atriplex patula L. - 3	USA, Wisconsin	Walsh, W0032
Atriplex patula L. - 4	USA, Wisconsin	Walsh, W0028
C. vulvaria L. - 1	unknown	Jellen, BYU919
C. album L. - 1	USA, Wisconsin	Walsh, W0078
C. berlandieri Moq. - 1	USA, South Dakota	Walsh, BL14-1
Rhagodia baccata (Labill.) Moq. - 1	Australia	WIS herb, Sauer 1963, #3525
Rhagodia baccata (Labill.) Moq. - 2	Australia	WIS herb, Sauer 1963, #3448
Rhagodia baccata (Labill.) Moq. - 3	Australia	WIS herb, Sauer 1963, #3584
Rhagodia parabolica R.Br.	Australia	WIS herb, Sauer 1963, #3452
C. nitrariaceum (F.Muell.) F.Muell. ex Benth.	Australia	BRY Herb, Stutz 1981
C. auricomum Lindl.	Australia	Kew MSB, 0086701
C. gracilispicum H.W.Kung	China	MO herb, Liu 1998
C. vulvaria L. - 2	Portugal	USDA, PI614896
C. vulvaria L. - 3	Malta	Malta Seeds Ltd.
C. ficifolium Sm. - 1	USA, Wisconsin	Walsh, W0026A
C. ficifolium Sm. - 2	unknown	Jellen, BYU943
C. ficifolium Sm. - 3	Mexico	WIS Herb, Villegas 1963
C. ficifolium Sm. - 4	Switzerland	USDA, Ames25246
C. ficifolium Sm. - 5	USA, Florida	NY Herb, Correll 1981
C. viride L.	unknown	Kew MSB, 0000169
C. oahuense Aellen	USA, Hawaii	Jellen, BYU651
C. iljinii Golosk. - 1	China	MO herb, Liu 1993, #04545929
C. iljinii Golosk. - 2	China	MO herb, Liu 1993, #04769643
C. album L. - 2	Austria	NY Herb, Barta 2004 (C. opulifolium)
C. album L. - 3	USA, Minnesota	WIS herb, Smith 1995
C. album L. - 4	USA, Wisconsin	Walsh, W0013B
C. album L. - 5	USA, Wisconsin	Walsh, W0077
C. album L. - 6	USA, Illinois	NY Herb, Hill 2001
C. album L. - 7	USA, Mississippi	NY Herb, Bryson 2010
C. album L. - 8	USA, Texas	NY Herb, Gutierrez 2007
C. album L. - 9	India	USDA, PI658735
C. album L. - 10	USA, Pennsylvania	Walsh, NY#2
C. album L. - 11	USA, Pennsylvania	NCBI, JN646817
C. album L. - 12	USA, West Virginia	WIS Herb, Utech & Ohara 1982
C. strictum Roth - 1	Germany	Ames23893
C. strictum Roth - 2	USA, California	Jellen, BYU436
C. strictum Roth - 3	USA, Utah	USDA, Ames29770
C. strictum Roth - 4	USA, Illinois	NY Herb, Wahl 1966 (C. opulifolium)
C. missouriense Aellen - 1	USA, Pennsylvania	NY Herb, Berkheimer 1959
C. missouriense Aellen - 2	USA, Missouri	NY Herb, Torin 1966
C. giganteum D. Don - 1	USA, Oklahoma	USDA, PI596371
C. giganteum D. Don - 2	USA, South Carolina	USDA, NSL86650
C. suecicum J. Murr	unknown	Jellen, BYU945
C. formosanum Koidz. - 1	Taiwan	USDA, PI433379
C. formosanum Koidz. - 2	Taiwan	USDA, PI433378
C. bryoniifolium Bunge	China	MO herb, Yellow Plateau Team 1984, #04751988
C. mucronatum Thunb.	South Africa	MO herb, Skead 1981, #5657692
C. petiolare Kunth - 1	Chile	Jellen, BYU562
C. pallidicaule Aellen - 1	Peru	USDA, PI510530
C. pallidicaule Aellen - 2	Peru	USDA, PI510528

C. pallidicaule Aellen - 3	Peru	USDA, PI510525
C. pallidicaule Aellen - 4	Peru	USDA, PI510529
C. pallidicaule Aellen - 5	Bolivia	USDA, PI478406
C. pallidicaule Aellen - 6	Bolivia	USDA, Ames13223
C. atrovirens Rydb. - 1	USA, Arizona	USDA, Ames29812
C. cycloides A. Nelson -1	USA, Texas	NY Herb, Worthington 1988
C. cycloides A. Nelson -2	USA, New Mexico	NY Herb, Higgins 1987
C. desiccatum A. Nelson - 1	USA, Nevada	USDA, Ames29780
C. desiccatum A. Nelson - 2	USA, Nevada	Jellen, BYU834
C. frigidum Phil.	Argentina	MO Herb, Apochian 1995, #5978365
C. hians Standl. - 1	USA, New Mexico	USDA, Ames29788
C. neomexicanum Standl. - 1	USA, New Mexico	USDA, Ames29805
C. nevadense Standl. - 1	USA, Nevada	Jellen, BYU816
C. pratericola Rydb. - 1	USA, Arizona	USDA, Ames28056
C. pallescens Standl. - 1	USA, Kansas	NY Herb, Morse 2006
C. subglabrum (S. Watson) A. Nelson - 1	USA, Kansas	NY Herb, Freemen 2006
C. pratericola Rydb. - 2	USA, Colorado	NY Herb, Holmgren 1999
C. leptophyllum (Moq.) Nutt. ex S. Watson - 1	USA, Utah	BRY Herb, Brotherson 1965
C. leptophyllum (Moq.) Nutt. ex S. Watson - 2	USA, Nevada	BYU Herb, Morefield 1986
C. leptophyllum (Moq.) Nutt. ex S. Watson - 3	USA, Arizona	USDA, Ames29799
C. watsonii A. Nelson - 1	USA, New Mexico	USDA, Ames29813
C. watsonii A. Nelson - 2	USA, Arizona	USDA, Ames29815
C. watsonii A. Nelson - 3	USA, Arizona	Jellen, BYU873
C. subglabrum (S. Watson) A. Nelson - 2	USA, Montana	NY Herb, Lackschewitz 1978, #00990899
C. subglabrum (S. Watson) A. Nelson - 3	USA, Wyoming	NY Herb, Dorn 1993
C. standleyanum Aellen - 1	USA, Wisconsin	Walsh, W0012A
C. standleyanum Aellen - 2	USA, Wisconsin	Walsh, W0017C
C. standleyanum Aellen - 3	USA, Iowa	USDA, PI658755
C. standleyanum Aellen - 4	USA, Indiana	F herb, MacBride 1926
C. standleyanum Aellen - 5	USA, Iowa	USDA, Ames30415
C. berlandieri Moq. - 2	USA, South Carolina	NCBI, JN646819
C. berlandieri Moq. - 3	USA, Ohio	NCBI, JN646819
C. berlandieri Moq. - 4	USA, Missouri	NCBI, JN646821
C. berlandieri Moq. - 5	USA, Pennsylvania	NCBI, JN646829
C. berlandieri Moq. - 6	USA, Maryland	NCBI, JN646830
C. berlandieri Moq. - 7	USA, Ohio	NCBI, JN646831
C. berlandieri Moq. - 8	USA, Missouri	NCBI, JN646832
C. berlandieri Moq. - 9	USA, Missouri	NCBI, JN646833
C. berlandieri Moq. - 10	USA, Missouri	NCBI, JN646834
C. berlandieri Moq. - 11	USA, Alabama	NCBI, JN646835
C. berlandieri Moq. - 12	USA, Pennsylvania	NCBI, JN646836
C. berlandieri Moq. - 13	USA, Maryland	NCBI, JN646837
C. berlandieri Moq. - 14	USA, Iowa	NCBI, JN646841
C. berlandieri Moq. - 15	USA, Arkansas	NCBI, JN646843
C. berlandieri Moq. - 16	USA, Maryland	NCBI, JN646844
C. berlandieri Moq. - 17	USA, Missouri	NCBI, JN646845
C. berlandieri Moq. - 18	USA, Ohio	NCBI, JN646846
C. berlandieri Moq. - 19	USA, South Carolina	NCBI, JN646847
C. berlandieri Moq. - 20	USA, Utah	USDA, PI612858
C. berlandieri Moq. - 21	USA, Iowa	USDA, PI595316
C. berlandieri var. bushianum (Aellen) Cronquist - 1	USA, Wisconsin	WIS Herb, Ziegler & Leykom 1975
C. berlandieri var. bushianum (Aellen) Cronquist - 2	USA, Wisconsin	WIS Herb, Nee 1979
C. berlandieri var. bushianum (Aellen) Cronquist - 3	USA, New Hampshire	NY Herb, Boufford et al. 1985
C. berlandieri var. bushianum (Aellen) Cronquist - 4	USA, Illinois	NY Herb, Hill 2006
C. berlandieri var. bushianum (Aellen) Cronquist - 5	USA, Louisiana	NY Herb, Thomas 1983
C. berlandieri var. bushianum (Aellen) Cronquist - 6	USA, Ohio	OS Herb, Whitkus 1988
C. berlandieri var. sinuatum (Murr) Wahl - 1	USA, Texas	TEX herb, Carr 2007

C. berlandieri var. sinuatum (Murr) Wahl - 2	USA, Arizona	Jellen, BYU575
C. berlandieri var. sinuatum (Murr) Wahl - 3	USA, New Mexico	NCBI, JN646850
C. berlandieri var. berlandieri - 1	USA, Texas	TEX herb, Lott 2005
C. berlandieri var. boscianum (Moq.) Wahl - 1	USA, Missouri	WIS Herb, Palmer 1956
C. berlandieri var. boscianum (Moq.) Wahl - 2	USA, Louisiana	USDA, Ames29307
C. berlandieri var. macrocalycium (Aellen) Cronquist	USA, Maine	USDA, Ames29207
C. berlandieri var. zschackei (Murr) Murr ex Graebn. - 1	USA, Kentucky	Walsh, W0022
C. berlandieri var. zschackei (Murr) Murr ex Graebn. - 2	USA, California	Jellen, BYU422
C. berlandieri var. zschackei (Murr) Murr ex Graebn. - 3	USA, Ohio	OS Herb, Riley 2011
C. berlandieri var. zschackei (Murr) Murr ex Graebn. - 4	Mexico	TEX herb, Henrickson 2001
Chia Roja (unknown sp.)	Mexico	Jellen, BYU567
C. quinoa Willd. - 1	Chile	USDA, PI614885
C. fremontii S. Watson - 1	USA, Utah	Ames29811
C. incanum (S. Watson) A. Heller - 1	USA, Arizona	Jellen, BYU853
C. neomexicanum Standl. - 2	USA, Arizona	USDA, Ames29803
C. watsonii A. Nelson - 4	USA, New Mexico	USDA, Ames29814
C. fremontii S. Watson - 5	USA, Arizona	Jellen, BYU547
C. atrovirens Rydb. - 2	USA, Arizona	Jellen, BYU547
C. fremontii S. Watson - 2	USA, Utah	Walsh, W0064#1
C. berlandieri var. berlandieri - 2		NY Herb,
C. berlandieri var. sinuatum (Murr) Wahl - 4	USA, Arizona	USDA, Ames29767
C. berlandieri var. sinuatum (Murr) Wahl - 5		TEX herb,
C. berlandieri var. berlandieri - 3		TEX herb,
C. berlandieri var. zschackei (Murr) Murr ex Graebn. - 5		TEX herb,
C. hircinum Schrad. - 1	Bolivia	NY Herb, Nee 1991
C. quinoa Willd. - 3	Chile	USDA, PI614880
C. quinoa Willd. - 4	USA, New Mexico	USDA, Ames13735
C. quinoa Willd. - 5	Bolivia	USDA, PI614931
C. quinoa Willd. - 6	Bolivia	USDA, PI614901
C. quinoa Willd. - 7	Bolivia	USDA, PI478418
C. quinoa Willd. - 8	Bolivia	USDA, PI614912
C. quinoa Willd. - 9	Argentina	USDA, PI587173
C. petiolare Kunth - 2	Ecuador	Kew MSB, 0017583
C. hircinum Schrad. - 2	Chile	Jellen, BYU565
C. quinoa Willd. - 10	Peru	USDA, PI510543
C. quinoa Willd. - 11	Peru	USDA, PI510547
C. quinoa Willd. - 12	Peru	USDA, PI510549
C. hians Standl. - 2	USA, Utah	BRY Herb, Holmgren 1992
C. nevadense Standl. - 2	USA, Nevada	Jellen, BYU819
C. pallescens Standl. - 2	USA, Illinois	NY Herb, Lammers 1997
C. pratericola Rydb. - 3	USA, Idaho	NY Herb, Riley 1992
C. incanum (S. Watson) A. Heller - 2	USA, New Mexico	USDA, Ames29792
C. incanum (S. Watson) A. Heller - 3	USA, Wyoming	Jellen, BYU861
C. neomexicanum Standl. - 3	USA, Arizona	USDA, Ames29804
C. neomexicanum Standl. - 4	USA, Utah	USDA, Ames28048
C. fremontii S. Watson - 3	USA, Utah	Walsh, W0061#1
C. fremontii S. Watson - 4	USA, Colorado	Walsh, W0066
C. berlandieri var. sinuatum (Murr) Wahl - 6	USA, Arizona	Jellen, BYU865
C. berlandieri var. sinuatum (Murr) Wahl - 7	USA, Arizona	NCBI, JN646840
C. berlandieri Moq. - 22	USA, Arizona	Jellen, BYU874
C. berlandieri var. zschackei (Murr) Murr ex Graebn. - 6	USA, Wyoming	Jellen, BYU862
C. berlandieri var. zschackei (Murr) Murr ex Graebn. - 7	USA, Colorado	NCBI, JN646839
C. berlandieri var. zschackei (Murr) Murr ex Graebn. - 8	USA, Idaho	NCBI, JN646838
C. berlandieri var. zschackei (Murr) Murr ex Graebn. - 9	USA, New Mexico	Jellen, BYU878
C. berlandieri var. zschackei (Murr) Murr ex Graebn. - 10	USA, Utah	NCBI, JN646827
C. berlandieri subsp. nuttalliae (Saff.) H. D. Wilson & Heiser - 1	Mexico	USDA, PI568155
C. berlandieri subsp. nuttalliae (Saff.) H. D. Wilson & Heiser - 2	Mexico	Jellen, BYU668

C. berlandieri subsp. nuttalliae (Saff.) H. D. Wilson & Heiser - 3	Mexico	USDA, PI568156
C. berlandieri subsp. nuttalliae (Saff.) H. D. Wilson & Heiser - 4	Mexico	Walsh, Xalapa market
C. berlandieri subsp. nuttalliae (Saff.) H. D. Wilson & Heiser - 5	Mexico	NCBI, JN646824
C. berlandieri subsp. nuttalliae (Saff.) H. D. Wilson & Heiser - 6	Mexico	USDA, PI476820 (called C. quinoa)
C. berlandieri subsp. nuttalliae (Saff.) H. D. Wilson & Heiser - 7	Mexico	USDA, PI433231
C. berlandieri subsp. nuttalliae (Saff.) H. D. Wilson & Heiser - 8	Mexico	Walsh, Puebla
C. berlandieri subsp. nuttalliae (Saff.) H. D. Wilson & Heiser - 9	Mexico	USDA, PI433230

Figure 1. A cladogram of the congruent Maximum Likelihood and Bayesian analyses. The numbers reported above the lines are likelihood bootstrap percentages. Numbers below lines are Bayesian posterior probability scores. Likelihood bootstrap values were only reported if 50% or higher. Bayesian posterior probabilities were reported if 0.50 or higher. Dashes denote support values below the designated cutoff. Nodes were collapsed when no support values were higher than the cutoff, visually simplifying phylogenetic associations. Clades within the Chenopodioideae subfamily are congruent with Fuentes et al. (2012a, 2012b), except the Fuentes et al. (2012a) study reported *Dysphania-Teloxys* as sister to *Spinacia-Blitum* and the core-*Chenopodium* clade.

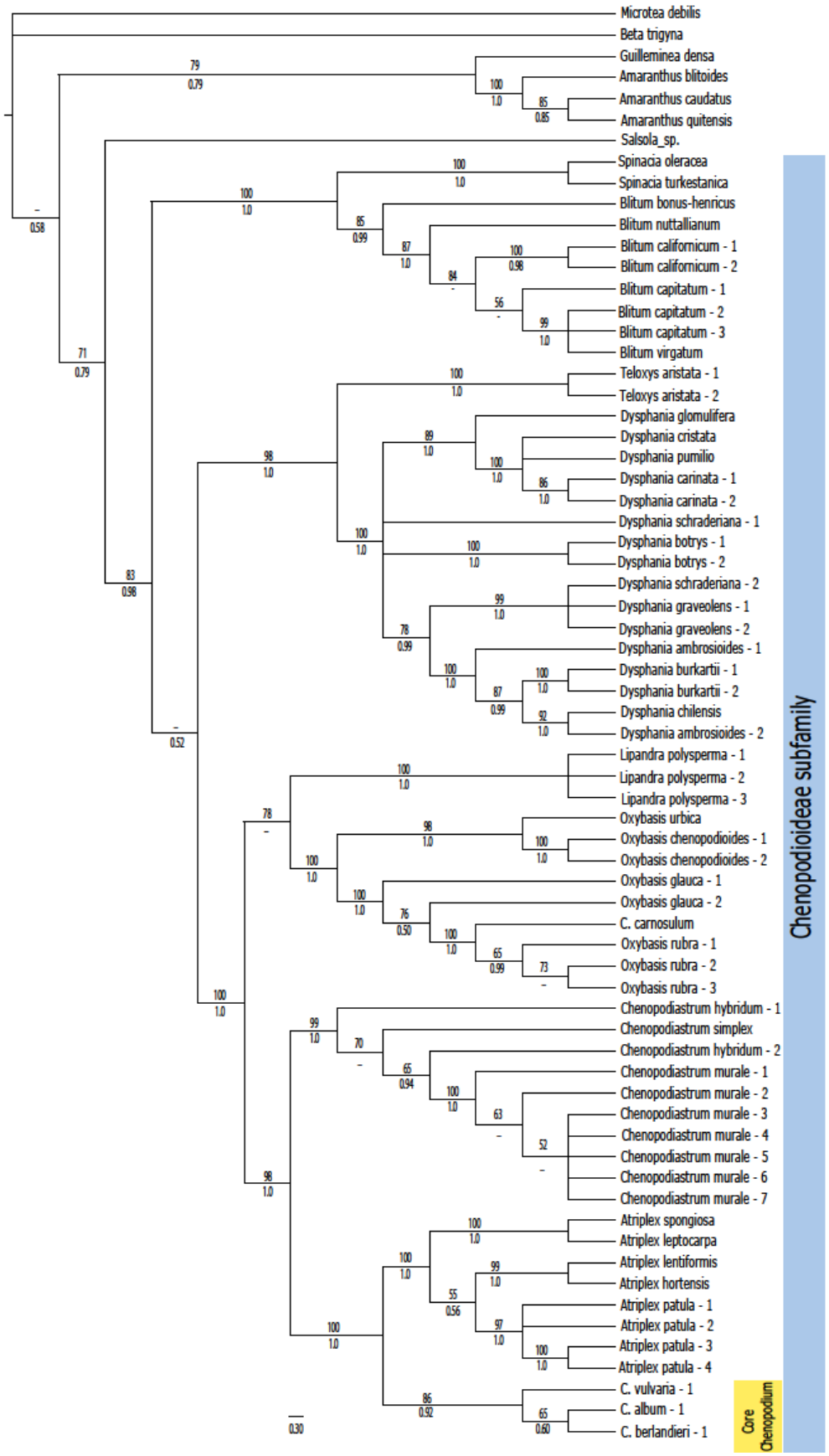

Microtea debilis
Beta trigyna
Guilleminea densa
Amaranthus blitoides
Amaranthus caudatus
Amaranthus quitensis
Salsola_sp.
Spinacia oleracea
Spinacia turkestanica
Blitum bonus-henricus
Blitum nuttallianum
Blitum californicum - 1
Blitum californicum - 2
Blitum capitatum - 1
Blitum capitatum - 2
Blitum capitatum - 3
Blitum virgatum
Teloxys aristata - 1
Teloxys aristata - 2
Dysphania glomulifera
Dysphania cristata
Dysphania pumilio
Dysphania carinata - 1
Dysphania carinata - 2
Dysphania schraderiana - 1
Dysphania botrys - 1
Dysphania botrys - 2
Dysphania schraderiana - 2
Dysphania graveolens - 1
Dysphania graveolens - 2
Dysphania ambrosioides - 1
Dysphania burkartii - 1
Dysphania burkartii - 2
Dysphania chilensis
Dysphania ambrosioides - 2
Lipandra polysperma - 1
Lipandra polysperma - 2
Lipandra polysperma - 3
Oxybasis urbica
Oxybasis chenopodioides - 1
Oxybasis chenopodioides - 2
Oxybasis glauca - 1
Oxybasis glauca - 2
C. carnosulum
Oxybasis rubra - 1
Oxybasis rubra - 2
Oxybasis rubra - 3
Chenopodiastrum hybridum - 1
Chenopodiastrum simplex
Chenopodiastrum hybridum - 2
Chenopodiastrum murale - 1
Chenopodiastrum murale - 2
Chenopodiastrum murale - 3
Chenopodiastrum murale - 4
Chenopodiastrum murale - 5
Chenopodiastrum murale - 6
Chenopodiastrum murale - 7
Atriplex spongiosa
Atriplex leptocarpa
Atriplex lentiformis
Atriplex hortensis
Atriplex patula - 1
Atriplex patula - 2
Atriplex patula - 3
Atriplex patula - 4
C. vulvaria - 1
C. album - 1
C. berlandieri - 1
Chenopodioideae subfamily
Core Chenopodium
0.30

Figure 2. A cladogram of the Core *Chenopodium* species. Maximum Likelihood and Bayesian analyses results were congruent. The numbers reported above the lines are likelihood bootstrap percentages. Numbers below lines are Bayesian posterior probability scores. Likelihood bootstrap values were only reported if 50% or higher. Bayesian posterior probabilities were reported if 0.50 or higher. Dashes denote support values below the designated cutoff. Nodes were collapsed when no support values were higher than the cutoff, visually simplifying phylogenetic associations. Figure 2-A shows the relationship among species and clades near the base of the core Chenopodium clade, while Figure 2-B shows the Mixed Chenopodium clade.

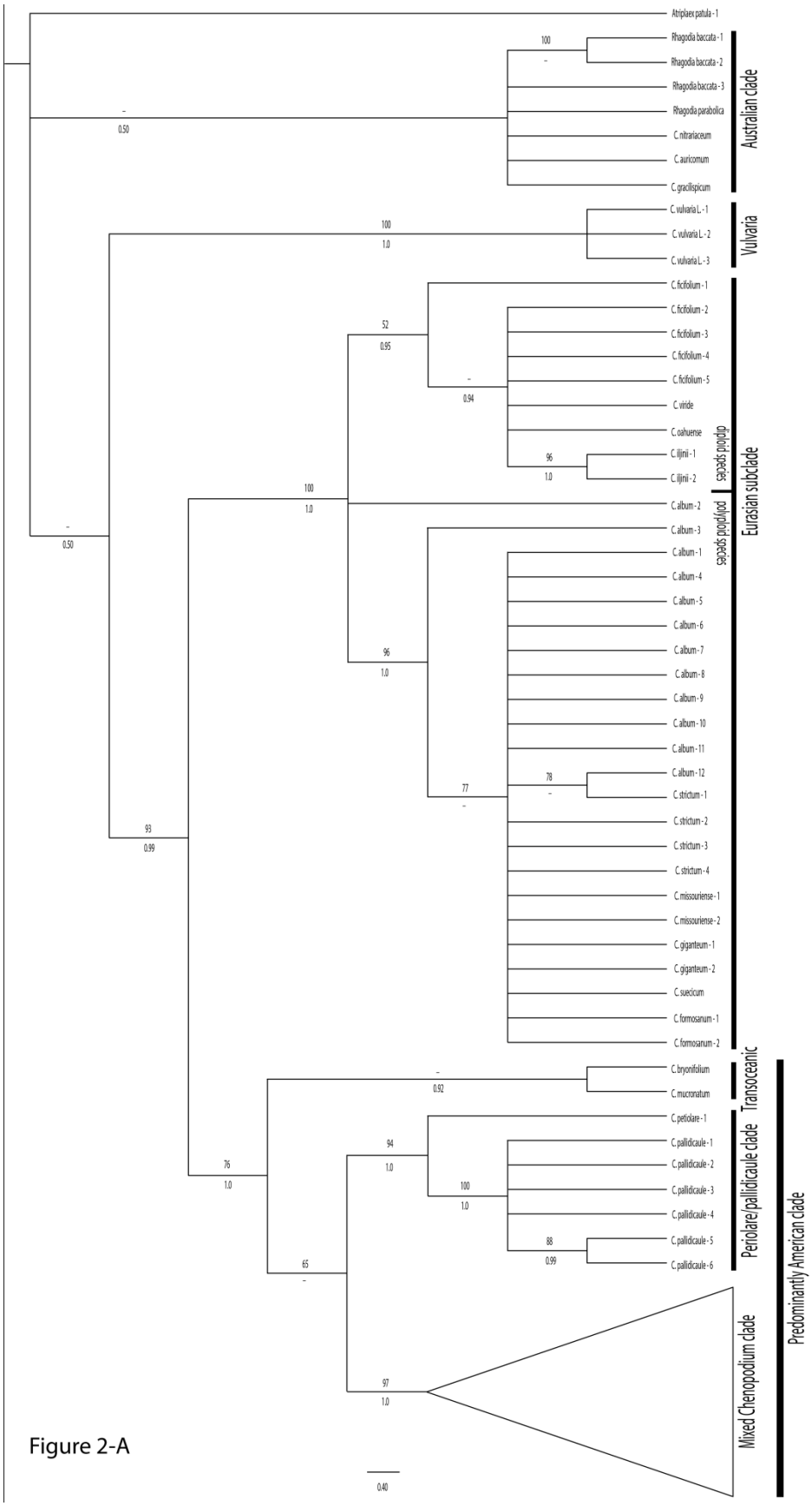
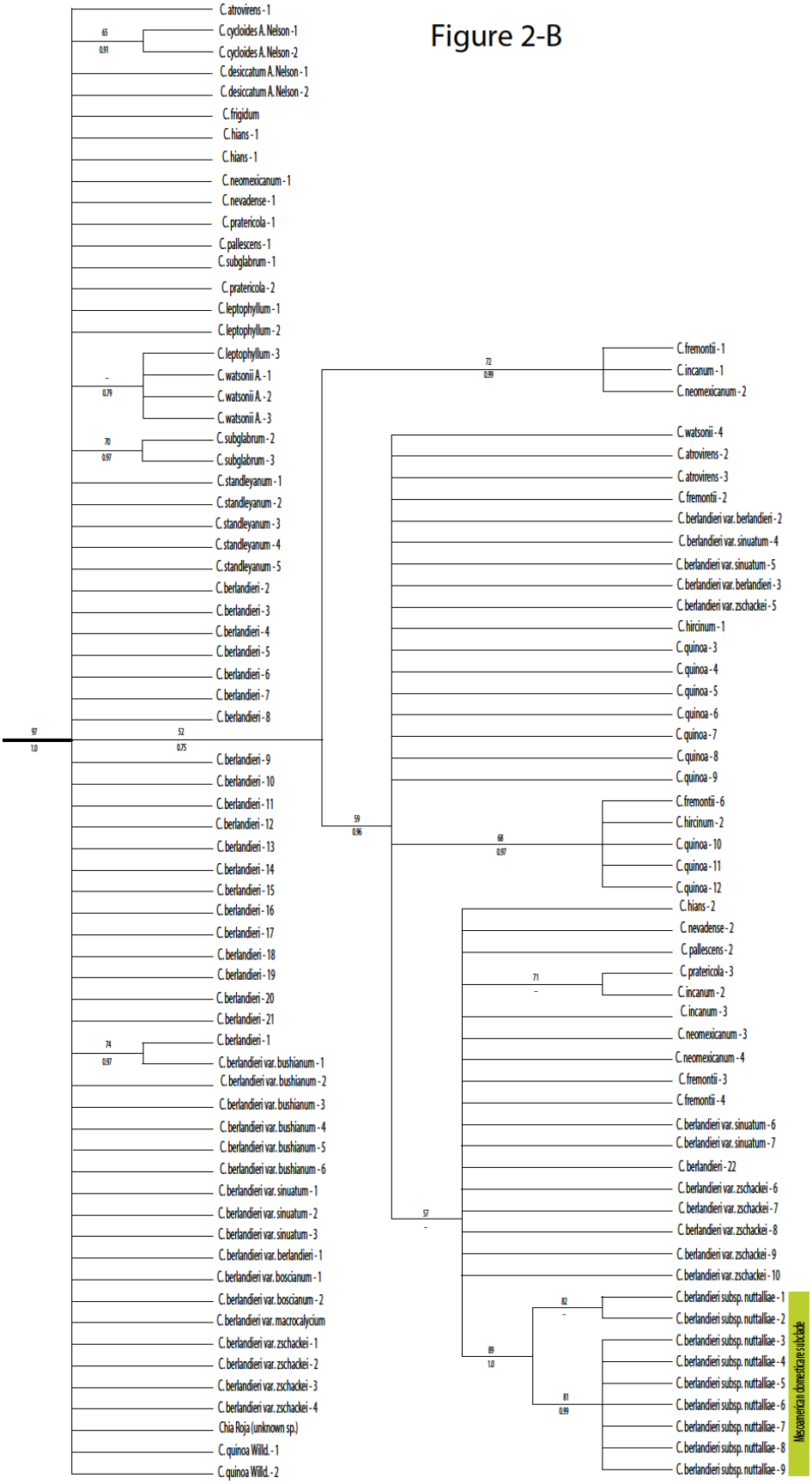

Atriplex patula - 1
Rhagodia baccata - 1
Rhagodia baccata - 2
Rhagodia baccata - 3
Rhagodia parabolica
C. nitrariaceum
C. auricomum
C. gracilispicum
Australian clade
C. vulvaria L. - 1
C. vulvaria L. - 2
C. vulvaria L. - 3
Vulvaria
C. ficifolium - 1
C. ficifolium - 2
C. ficifolium - 3
C. ficifolium - 4
C. ficifolium - 5
C. viride
C. oahuense
C. iljinii - 1
C. iljinii - 2
sepals polyp.
C. album - 2
sepals poly/vod
C. album - 3
C. album - 1
C. album - 4
C. album - 5
C. album - 6
C. album - 7
C. album - 8
C. album - 9
C. album - 10
C. album - 11
C. album - 12
C. strictum - 1
C. strictum - 2
C. strictum - 3
C. strictum - 4
C. missouriense - 1
C. missouriense - 2
C. giganteum - 1
C. giganteum - 2
C. suecicum
C. formosanum - 1
C. formosanum - 2
Eurasian subclade
C. bryoniifolium
C. mucronatum
Transoceanic
C. petiolare - 1
C. pallidicaule - 1
C. pallidicaule - 2
C. pallidicaule - 3
C. pallidicaule - 4
C. pallidicaule - 5
C. pallidicaule - 6
Petiolare/pallidicaule clade
Mixed Chenopodium clade
Predominantly American clade
0.40

Figure 2-A

Figure 2-B


C. atrovirens - 1
C. cycloides A. Nelson -1
C. cycloides A. Nelson -2
C. desiccatum A. Nelson - 1
C. desiccatum A. Nelson - 2
C. frigidum
C. hians - 1
C. hians - 1
C. neomexicanum - 1
C. nevadense - 1
C. pratericola - 1
C. pallescens - 1
C. subglabrum - 1
C. pratericola - 2
C. leptophyllum - 1
C. leptophyllum - 2
C. leptophyllum - 3
C. watsonii A. - 1
C. watsonii A. - 2
C. watsonii A. - 3
C. subglabrum - 2
C. subglabrum - 3
C. standleyanum - 1
C. standleyanum - 2
C. standleyanum - 3
C. standleyanum - 4
C. standleyanum - 5
C. berlandieri - 2
C. berlandieri - 3
C. berlandieri - 4
C. berlandieri - 5
C. berlandieri - 6
C. berlandieri - 7
C. berlandieri - 8
C. berlandieri - 9
C. berlandieri - 10
C. berlandieri - 11
C. berlandieri - 12
C. berlandieri - 13
C. berlandieri - 14
C. berlandieri - 15
C. berlandieri - 16
C. berlandieri - 17
C. berlandieri - 18
C. berlandieri - 19
C. berlandieri - 20
C. berlandieri - 21
C. berlandieri - 1
C. berlandieri var. bushianum - 1
C. berlandieri var. bushianum - 2
C. berlandieri var. bushianum - 3
C. berlandieri var. bushianum - 4
C. berlandieri var. bushianum - 5
C. berlandieri var. bushianum - 6
C. berlandieri var. sinuatum - 1
C. berlandieri var. sinuatum - 2
C. berlandieri var. sinuatum - 3
C. berlandieri var. berlandieri - 1
C. berlandieri var. boscianum - 1
C. berlandieri var. boscianum - 2
C. berlandieri var. macrocalycium
C. berlandieri var. zschackei - 1
C. berlandieri var. zschackei - 2
C. berlandieri var. zschackei - 3
C. berlandieri var. zschackei - 4
Chia Roja (unknown sp.)
C. quinoa Willd. - 1
C. quinoa Willd. - 2
C. fremontii - 1
C. incanum - 1
C. neomexicanum - 2
C. watsonii - 4
C. atrovirens - 2
C. atrovirens - 3
C. fremontii - 2
C. berlandieri var. berlandieri - 2
C. berlandieri var. sinuatum - 4
C. berlandieri var. sinuatum - 5
C. berlandieri var. berlandieri - 3
C. berlandieri var. zschackei - 5
C. hircinum - 1
C. quinoa - 3
C. quinoa - 4
C. quinoa - 5
C. quinoa - 6
C. quinoa - 7
C. quinoa - 8
C. quinoa - 9
C. fremontii - 6
C. hircinum - 2
C. quinoa - 10
C. quinoa - 11
C. quinoa - 12
C. hians - 2
C. nevadense - 2
C. pallescens - 2
C. pratericola - 3
C. incanum - 2
C. incanum - 3
C. neomexicanum - 3
C. neomexicanum - 4
C. fremontii - 3
C. fremontii - 4
C. berlandieri var. sinuatum - 6
C. berlandieri var. sinuatum - 7
C. berlandieri - 22
C. berlandieri var. zschackei - 6
C. berlandieri var. zschackei - 7
C. berlandieri var. zschackei - 8
C. berlandieri var. zschackei - 9
C. berlandieri var. zschackei - 10
C. berlandieri subsp. nuttalliae - 1
C. berlandieri subsp. nuttalliae - 2
C. berlandieri subsp. nuttalliae - 3
C. berlandieri subsp. nuttalliae - 4
C. berlandieri subsp. nuttalliae - 5
C. berlandieri subsp. nuttalliae - 6
C. berlandieri subsp. nuttalliae - 7
C. berlandieri subsp. nuttalliae - 8
C. berlandieri subsp. nuttalliae - 9
Mesoamerican domesticates clade

CHAPTER 2

HOMEOLOG CHARACTERIZATION FOR TWO POLYPLOID LINEAGES IN *CHENOPODIUM* (AMARANTHACEAE) USING INTRONIC SEQUENCE DATA FROM THE SINGLE-COPY NUCLEAR GENE *SALT OVERLY SENSITIVE 1* (*SOS1*)

ABSTRACT

Premise of the study- Chenopodium (Amaranthaceae) is a globally distributed genus composed of approximately 65-100 species. Several polyploid species within the genus are considered noxious agricultural weeds, yet very little research has addressed their evolutionary origin. We use a single/low-copy nuclear marker to characterize allotetraploid and allohexaploid species to help identify their genome donors.

Methods- We used DNA sequences of the single/low-copy nuclear marker, *Salt Overly Sensitive 1* (*SOS1*) to build a phylogeny for *Chenopodium*. Diploid species were sequenced directly, whereas homeologous sequences of polyploid genomes were separated by plasmid mediated cloning. Data were analyzed using maximum likelihood and Bayesian inference.

Key results- Sequences of polyploid species were found in four clades, designated A-D. Two distinct polyploid lineages were identified: one composed of American tetraploid species with A and B class homeologs, and a second composed of Eastern Hemisphere hexaploid species with B, C, and D class homeologs. We infer that the two polyploid lineages were independently derived, presumably each originating once. The American diploid, *C. standleyanum*, was identified as the closest living relative of the A genome donor. The east Asian diploid species, *C. bryoniifolium* allies with American diploid species supporting an Old World origin for American *Chenopodium*.

Conclusions- By characterizing homeologs, we have a better understanding of polyploid genesis for the genetically distinct American tetraploids and Eurasian hexaploids. The identification of *C. standleyanum* as one of the likely genome donors for the domesticated tetraploid, *C. quinoa*, is of potential importance for quinoa breeding.

Key Words: *Chenopodium quinoa*; domestication; migration; *Chenopodium pallidicaule*; *Chenopodium berlandieri* subsp. *nuttalliae*; progenitor

INTRODUCTION

Chenopodium (Amaranthaceae), is a little studied yet widespread, economically and ecologically important genus. *Chenopodium* includes both diploid and polyploid species, with several species adapted to extreme xeric and saline environments. It was domesticated as many as five times and yet is typically considered a serious agricultural weed. A better understanding of *Chenopodium* would contribute to our knowledge of domesticated plants and their evolution from wild progenitors, suggest new potential food crops, and illuminate the origins of polyploid species. Unfortunately, *Chenopodium* is a taxonomically challenging group that has defied taxonomic consensus.

Chenopodium researchers have struggled with species delimitation, generic boundaries, and taxonomic categories for over two centuries (Linnaeus, 1753; Moquin-Tandon, 1840; Standley, 1917; Aellen, 1929; Aellen and Just, 1943; Wahl, 1954; Crawford, 1975; Mosyakin and Clemants, 1996; Clemants and Mosyakin, 2003; Mosyakin, 2013). In the taxonomic literature there is not even consensus on the boundaries of the genus: there are many morphological characters shared by *Chenopodium* species, but these characters display a range of variability and often overlap with other closely related genera. Rodman (1990) conducted the first cladistic analyses of *Chenopodium* in the 1990s based on morphological traits. Subsequent researchers began constructing molecular phylogenies, attempting to avoid many of the problems associated with morphological characters within *Chenopodium* (Manhart and Rettig, 1994; Downie and Palmer, 1994). Collectively the morphological and molecular work contributed to the recognition that the family Chenopodiaceae is a large clade nested within Amaranthaceae. Therefore Chenopodiaceae has been subsumed within Amaranthaceae, which would be paraphyletic if Chenopodiaceae were preserved as a separate family.

Several phylogenies based on plastid and ITS sequencing were published in the early to mid 2000s, each including a relatively small number of *Chenopodium* species (Kadereit et al., 2003, 2005; Müller and Borsch, 2005; Hohmann et al., 2006). These studies demonstrated that *Chenopodium* was polyphyletic, and gradually *Chenopodium* species were moved to different genera to better reflect natural groupings. For example, in 2002, 10 former *Chenopodium* species were reclassified to *Dysphania*, nearly doubling the species in that genus (Mosyakin and Clemants, 2002). More recently, three publications made sweeping changes to *Chenopodium*. Fuentes-Bazan et al. (2012a, 2012b) generated phylogenies constructed from plastid sequence data that include the greatest number of *Chenopodium* species sampled to date. Mosyakin (2013) published proposals for nomenclatural changes based on the plastid phylogenies. In summary, all former *Chenopodium* species more distantly related to *C. album*, the type species, than to *Atriplex*, were removed from *Chenopodium* in an effort to make the genus monophyletic.

Chenopodium is gradually being redefined so as to make it a monophyletic genus. In many phylogenetic analyses the currently recognized species of *Chenopodium* s. str. form a clade that is sister to *Atriplex*. However, the most recent phylogenies (Fuentes-Bazan et al. 2012a, 2012b) show that the Australian genera, *Rhagodia* and *Einadia*, are unresolved with respect to *Chenopodium*. The uncertain relationship of these genera leaves open the possibility that despite recent revisions *Chenopodium* may still be non-monophyletic.

Although recent revisions to *Chenopodium* have delimited the genus to more closely reflect a natural grouping, these revisions have also concentrated the genus down to many of the most difficult species to resolve among using the sampled plastid sequence data. The genus reflects a relatively recent divergence, probably less than five million years old based on date estimates (Kadereit et al., 2005) in conjunction with phylogenies with greater sampling (Fuentes-

Bazan et al., 2012b). Plastid genomes are generally believed to evolve slower than nuclear genomic DNA (Clegg et al., 1994). Not surprisingly, plastid sequence data doesn't provide the variation necessary to resolve the relationships within *Chenopodium*.

There are many allopolyploid species within *Chenopodium* for which neither maternally inherited plastid sequence data nor ITS sequence data, in which concerted evolution has taken place, are adequate for study (Fuentes-Bazan et al., 2012a, 2012b, and BMW personal observation). These genetic regions only reflect a single evolutionary history where as allopolyploid species are created by the joining of two or more distinct genomes (called homeologs), each reflecting a unique history. Only single to low copy nuclear genes have the capacity to reveal the evolutionary history of each donor genome existing within a polyploid, but to our knowledge, no such research has been published to date in *Chenopodium*.

Very little is known about *Chenopodium* polyploids. In the Americas there are three native tetraploid species, *C. berlandieri*, *C. hircinum*, and *C. quinoa*, which are so morphologically similar they should probably be considered the same species.

Eurasian polyploid species are more variable genetically, as evidenced by published chromosome counts for *C. album* including diploid, tetraploid, and hexaploid specimens, and morphologically. Diploid *C. album* appears to be restricted to central Asia, however it is unclear whether this is the ancestral genome-donor population giving rise to polyploid *C. album*. *Chenopodium album* is also morphologically quite variable, consisting of many subspecies, each formerly their own species. Currently *C. album* from Europe is morphologically distinct from Asia, which in turn is morphologically distinct from *C. album* naturalized to North America; *C. album* subsp. *missouriense*. In Wisconsin mixed populations of the European and naturalized *C.*

album are common (BMW, personal observation). Despite immediately adjacent proximity, no morphological intermediates of the two distinct subspecies have thus far been observed. This suggests there is some mechanism at play maintaining these suites of traits, potential a reproductive barrier. As such, the sample of *C. album* subsp. *missouriense* in this study is referred to as *C. missouriense*.

We are only aware of one hypothesis that addresses the potential genome donor of a polyploid lineage. Gandarillas (1968, 1974) proposed that the diploid *C. pallidicaule* was a genome donor for the tetraploid *C. quinoa*. Since both species are domesticated, the mechanism giving rise to *C. quinoa* would be somewhat similar to the origin of bread wheat (hexaploid), where one of the genome donors was itself a domesticated wheat (tetraploid; Hancock, 2004).

Of the five domesticated *Chenopodium* lineages, the only diploid species is *C. pallidicaule*, the most geographically restricted and least utilized crop. The only Asian cultigen, *C. giganteum*, is hexaploid. Of the American polyploid cultigens, *C. quinoa* of South America, and *C. berlandieri* subsp. *nuttalliae* of Mesoamerica are tetraploid. The remaining American cultigen, *C. berlandieri* subsp. *jonesianum* of the eastern United States, has been extinct for at least 250 years, but is presumed to have been tetraploid, the same as modern *C. berlandieri*.

We built a phylogeny using DNA sequences of the low-copy nuclear locus *Salt Overly Sensitive 1 (SOS1)* to address two primary goals: 1) to resolve species relationships within *Chenopodium* that cannot be resolved using plastid sequencing; and 2) to gain a better understanding of the origins of the genomes possessed by polyploid *Chenopodium* species.

MATERIALS AND METHODS

Taxon sampling - In total, 34 *Chenopodium* and 4 outgroup samples were included in this study (Table 1), producing 55 sequences, including polyploid homeologs. The ingroup sampling consisted of 14 diploid and 8 polyploid species. Samples were obtained from various sources, including collections from wild populations, the USDA National Plant Germplasm System, Kew's Millennium seed bank, Missouri Botanical Garden Herbarium, and collaborators (Table 1). Samples obtained as seeds were grown to maturity in a greenhouse, either at Brigham Young University (Provo, Utah, USA) or University of Wisconsin-Madison (Madison, Wisconsin, USA).

DNA extraction - DNA extractions used either fresh leaf material or leaf material stored at -80C until needed for DNA extraction. Leaves were harvested either from greenhouse-grown specimens or directly from wild populations. The only exceptions are the *C. bryoniifolium* sample that was harvested from an herbarium specimen, and the silica gel dried *C. berlandieri* var. *zschackei* sample. Total genomic DNA was extracted from approximately 0.03-0.1 g of frozen leaf tissue or a single small leaf from the *C. bryoniifolium* herbarium specimen. DNA extractions used the protocol of either Todd and Vodkin (1996), or a modified protocol of Alexander *et al.* (2006), using silica membrane columns (Epoch Life Science, Sugar Land, Texas, USA).

Gene selection, target amplification, and sequencing - Primers were independently developed at Brigham Young University and the University of Wisconsin - Madison, based on GenBank accession EU024570.1 (*Chenopodium quinoa* cv. 'Real', *Salt Overly Sensitive 1* locus, *SOS1*).

Primers were designed from flanking exon sequences for introns 16 and 17, each approximately 800 bp in length (Table 2).

Target amplification **-** PCR reactions were performed using either of two methods: 1) 25 µl reactions (1 µl DNA, 0.2 µl *Taq* polymerase (5 units/µl, New England BioLabs, Ipswich, Massachusetts, USA), 0.5 µl of each primer pair (10 µM ea.), 1.25 µl BSA (10 mg/ml), 2 µl dNTP (solution contains 200 µM of each nucleotide), 2.5 µl 10x buffer (1.5 mM $MgCl_2$ included) as follows: initial denaturation was at 94°C for 5 m; followed by 34 cycles of 94°C for 30 s, 50-55°C for 1 m, and 72°C for 1.5 m; and 72°C for 7 m (final extension); or 2) Qiagen Hot Start Taq Master Mix (Qiagen, Valencia, California, USA), using the same thermocycler program, except the initial denaturation was 95°C for 15 m. PCR products were visualized in 1.0-1.2% Agarose gels. The amplified PCR products were purified using either a QIAquick PCR Purification Kit (Qiagen, Valencia, California, USA), or using Agencourt Ampure beads (Beckman Coulter Genomics, Morrisville, North Carolina, USA).

Ploidy determination **-** Chromosome counts were determined manually using a standard Feulgen-acetocarmine staining procedure with root-tip meristematic tissue. Seeds were germinated on moist filter paper for 2-4 days at approximately 22° C in the lab and root-tip meristematic cells of seedlings were arrested using N_2O gas at approximately 140 psi in a pressure chamber for 2-3 hrs per Kato (1999). Root tips were immediately fixed in fresh 3:1 (95% ethanol:glacial acetic acid) and refrigerated for up to one week prior to staining.

Homeolog-specific primers were developed and used to amplify unique sequence product indicative of each of the four homeolog classes observed among polyploid species (Table 2). The presence or absence of PCR product indicated which homeologs were present in a given

sample. Because most of the polyploid taxa appear to be allopolyploids, the homeolog-specific screen, in conjunction with the chromosome counts, allowed the determination of exactly which homeologs were present.

Cloning **-** Diploid species were either direct sequenced or cloned with 10-15 colonies selected and sequenced. Polyploid taxa were cloned, with 8-24 colonies sequenced at a time, until all expected homeolog sequences were obtained, as determined by chromosome counts and homeolog-specific primer results. The pGEM® -T Easy Vector System was used following manufacturer specifications (Promega, Madison, Wisconsin, USA). Clones were grown overnight on an agar plate containing ampicillin and 100 µl of IPTG/Xgal solution. White colonies were picked and tested for the presence of inserts of desired size.

Either of two methods was used for sequencing plasmid inserts. 1) colonies were picked and directly added to PCR reactions as above in Target Amplification, except using M13 primers. Colony PCR fragments visualized with ethidium bromide staining on a 1% agarose gel to identify fragments of appropriate length. Insert amplifications were purified using Agencourt Ampure beads (Beckman Coulter Genomics, Morrisville, North Carolina, USA) and amplified using M13 primers and ABI Big Dye Terminator as per manufacturer's specifications, as follows: 25 cycles of 96°C for 10 s and 50°C for 2 m. Sequences were generated with an ABI 3730 automated sequencer (Applied Biosystems, Foster City, California, USA) at the University of Wisconsin-Madison Biotechnology Center (UWBC).

Alternatively, 2) plasmid inserts were isolated using the GenElute plasmid Miniprep Kit, DNA (Sigma, St. Louis, Missouri, USA) and then quantified using Nanodrop (ND 1000 Spectrophotometer, Nanodrop Technologies Inc., Montchanin, DE, USA). Three hundred to

four hundred nanograms of plasmid DNA was amplified using Big Dye cycle sequencing. M13 primers were used to PCR-amplify the plasmid DNA. The thermocycler protocol was 25 cycles of 96°C for 10 s followed by 50°C for 6 s, and 60°C for 4 m. The amplified PCR product was purified with Sephadex G-50 protocol (GE Healthcare, Little Chalfont, Buckinghamshire, UK). The purified PCR product was sequenced by using ABI3730xl DNA Analyzer (Applied Biosystems, Foster City, California, USA). Once sequenced, vectors were screened using the NCBI VecScreen and the region of interest was confirmed by blasting each sequence against the NR database in NCBI.

Sequence quality and assembly - Sequences were edited and contigs assembled using Sequencher 4.9 (Gene Codes Corporation, Ann Arbor, Michigan, USA). Contigs were manually aligned using MacClade 4.08a OS X (Maddison and Maddison, 2005). Individual sequences were visually analyzed for PCR recombination by assessing consistency in polymorphic indels and nucleotide characters throughout each sequence read. Contigs were excluded from the analysis when PCR recombination was detected. Only samples with both *SOS1* intron 16 and 17 sequences for each determined homeolog were included in the analysis, with the exception of the diploid *C. bryoniifolium* sample which only amplified intron 17.

Phylogenetic Analyses - The program jModelTest v2.1.3 (Darriba et al., 2012) was used to select the most appropriate model (HKY+G for both maximum likelihood and Bayesian analyses). All tree-files were visualized with the program FigTree v1.3.1 (Rambaut, 2009).

Maximum Likelihood Analyses - ML analyses with 100 bootstrap replicates were performed twice using the CIPRES Science Gateway v3.1 (Miller et al. 2010), using the GARLI

2.0 on XSEDE software (Zwickl, 2006). The two likelihood analyses produced identical topologies and their -ln differed by 0.0878, prompting the bootstrap replicates to be combined.

Bayesian Analysis - Each Bayesian phylogenetic analysis was conducted using the CIPRES Science Gateway v3.1, using MrBayes 3.2.2 on XSEDE (Huelsenbeck and Ronquist, 2001). Each analysis consisted of two runs of 100 million generations, starting from a random tree and sampling every 10 000 generations. The temperature was set to 0.31 for the combined intronic dataset and 0.29 for intron 17 dataset to facilitate chain acceptance rates between 10-12%. The first 25% of trees from each run were discarded as burn-in.

RESULTS

Phylogenetic analyses within genus *Chenopodium*- The sequence length for introns 16 and 17 are 802-1014 and 509-935 bp, respectively. A Partition Homogeneity Test yielded a p-value of 0.14, supporting combining the intron 16 and 17 datasets for the *SOS1* gene tree reconstruction. The combined data set had 523 parsimony-informative substitutions. The results of jModelTest indicated HKY+G as the best model of molecular evolution for both ML and Bayesian analyses.

Many complex indels were identified during alignment, approximately 50 among ingroup species. An exact count is complicated by overlapping and nested indels. Due to the great number of indels and their complexity, a full summary will not be provided, and indels were not used for phylogenetic analysis.

The only Asian diploid sample, *C. bryoniifolium* from China, could only be sequenced for intron 17. Tree topologies based on intron 17 data were visually compared to combined intronic data to identify potential non-congruence caused by the missing intron 16 data for *C. bryoniifolium*. The trees were fully congruent in ML and Bayesian analyses, supporting the inclusion of this sample in the combined dataset despite missing data for intron 16.

The ML and Bayesian analyses for the combined data were fully congruent except for branches that had less than 60% bootstrap support for ML, and Bayesian posterior probabilities less than 0.80 (Fig. 1). The two repetitions of ML analyses produced likelihood scores of -10258.5882 and 10258.5004 using the combined intronic data.

The genus *Chenopodium* is composed of three primary clades, 1) the Vulvaria & Auricomum clade, 2) Clade A dominated by species considered native to the Americas, and 3) a large clade that can be further subdivided into 'Clade B', 'Clade C', and 'Clade D', each consisting

entirely of homeologous sequences from polyploid samples, plus one Eurasian diploid, *C. ficifolium*. All homeologs from polyploids resolve into four well-supported clades, A-D, each with ML bootstrap values of 100%, and posterior probability scores of 1.0.

Two distinct allopolyploid genotypes were identified among the nine polyploid species sampled, conforming to tetraploid and hexaploid groupings. The homeologs of each polyploid species separate into clades A-D. The first of the polyploid lineages is composed of American tetraploids with homeologs represented in both the A and B clades. The second polyploid lineage comprises the Eastern Hemisphere hexaploids, consisting of three distinct genomes, with one homeolog in each of the B, C, and D clades.

Two of the Eastern Hemisphere hexaploid samples yielded only two homeologs, not the three homeologs expected for hexaploids. We only succeeded in obtaining DNA sequences of the B and C clade homeologs for the *C. suecicum*-2 sample, failing to demonstrate the presence of a D clade homeolog. For *C. pedunculare*, we succeeded in obtaining sequences for the C and D clade homeologs, but a B clade homeolog was not detected. We twice conducted homeolog-specific screens on these samples without detecting the presence of the missing homeologs.

Here we briefly describe the major clades that will be discussed further below.

Vulvaria & Auricomum Clade: The clade sister to the rest of *Chenopodium* includes *C. vulvaria* (Eurasian) and *C. auricomum* (Australian).

Clade A: This clade consists of both diploid and tetraploid samples from of New World, except for the East Asian diploid, *C. bryoniifolium*. This clade has poor resolution among the tips, but there are two subclades of note: A well-supported clade uniting the domesticated *C. berlandieri* subsp. *nuttalliae* samples, and a moderate to weakly supported clade (64%

ML bootstrap, 0.80 posterior probability) uniting two South American species, the domesticated *C. pallidicaule* and the wild *C. petiolare*.

Clade B: All taxa represented in this clade are polyploid. This is the only clade that includes both Eastern and Western Hemisphere polyploid lineages. All polyploid taxa, except *C. pedunculare*, are represented in this clade. Samples of *C. ficifolium*, a Eurasian diploid, form a clade sister to clade B. Clade B exhibits the least resolution within the phylogeny, however a moderately-supported subclade (68% ML bootstrap, 0.98 posterior probability) unites the two domesticated *C. berlandieri* subsp. *nuttalliae* samples.

Clades C & D: All taxa included in Clades C and D are homeologs belonging to Eurasian hexaploid species. The *C. suecicum*-2 sample could not be sequenced for the "D" clade homeolog.

DISCUSSION

In this study, the outgroup taxa, *Blitum californicum*, *Blitum capitatum*, and *Oxybasis rubra*, were among the nomenclaturally adjusted taxa proposed in Fuentes-Bazan et al. (2012b). The relationship among the outgroup samples, including *Atriplex*, is congruent with plastid based phylogenetic studies (Fuentes-Bazan et al., 2012a, 2012b; Walsh and Emshwiller unpublished data).

Relationships of homeologous sequences of polyploids- This study includes 10 of the approximately 20-25 total American diploid species, all of which are represented in Clade A. The American tetraploids have homeologs represented in Clade A and B. The sequences of A-class homeologs are very similar to the sequences of American diploid species, while the B-class homeologs group within a much larger clade entirely composed of Eurasian taxa, except for the B-class homeologs present in American tetraploids. This suggests that the American tetraploids formed through the union of Western and Eastern Hemisphere genomes. The very close and well supported association of the diploid *C. standleyanum* as sister to the clade composed strictly of American tetraploids, strongly suggests *C. standleyanum* may be the American diploid genome donor and progenitor to the Clade A homeologs. *Chenopodium quinoa* is a domesticated American tetraploid that has been cultivated in Bolivia and Peru for thousands of years, but has been rapidly gaining in global popularity over the last 10 years. The identification of *C. standleyanum* as one of the potential genome donors for *C. quinoa* has important implications for quinoa breeders, as a potential source for genetic variation.

The Eastern Hemisphere polyploids are as yet poorly understood. Many of these species are very difficult to distinguish from each other, and some species, such as *C. album*, have been

reported with varying ploidy levels including diploid, tetraploid, and hexaploid (Skalinska et al., 1976; Krasnikov and Schaulo, 1990; Baltisberger and Voelger, 2006; Atul et al., 2007). It is unclear whether species such as *C. album* are truly variable in ploidy, or whether the reported differences in ploidy levels may represent taxonomic problems or misidentified species. There are two *C. suecicum* samples included in this study, which were determined to be hexaploid, based on chromosome squashes and sequence data. *Chenopodium suecicum* is typically reported as diploid (Baltisberger and Voelger, 2006), and as with *C. album*, there is the question of whether this is a new chromosome count for *C. suecicum* or an error in identification. However, because these samples were determined by two researchers experienced in *Chenopodium* taxonomy, Helena Štorchová and Bohumil Mandák, at the Institute of Experimental Botany in the Czech Republic, we believe these counts are a new hexaploid record for the species.

The American tetraploid lineage is composed of two genetically distinct homeologs, represented in Clades A and B. The Eastern Hemisphere hexaploid lineage in composed of three distinct homeologs, represented in Clades B, C, and D for most samples. Both polyploid lineages have unique genetic compositions not shared by the other lineage, and presumed formed independently, only coincidentally sharing a Clade B homeolog. Furthermore, since we never obtained a DNA sequence of a Clade B homeolog for a native American diploid species, we infer that an Eastern Hemisphere species colonized the Americas at some time in the past, and subsequently became the genome donor for the American tetraploids.

We had limited access to Eastern Hemisphere species for this study. As such, there is not sufficient sampling of Eurasian species, especially diploid taxa, to identify the diploid genome donor that contributed the B clade homeolog. The only diploid in the overall Eurasian clade is *C. ficifolium* which is so distantly related to both the tetraploid and hexaploid lineages that it is

unlikely to be the genome donor, at least directly. It is, however, the most closely related to the genome donor of the B clade homeologs among the species sampled here. The high degree of sequence similarity among the Clade B homeologs of the American tetraploids and Eurasian hexaploids demonstrates that these two polyploid lineages are very closely related, more closely related to each other than either is to *C. ficifolium*. The simplest explanation is that one or more species not included in this analysis, whether extinct or extant, gave rise independently to both the Eurasian hexaploids and the American tetraploids.

We obtained DNA sequences for the B, C, and D Clade homeologs for most of the Eastern Hemisphere hexaploids. Two of the samples, however, only yielded two instead of three homeologs, *C. pedunculare* and *C. suecicum*-2. We cannot determine if the negative amplification results indicate PCR failure, primer mismatch, or that the taxon in question lost the homeolog in question. These samples may have had a different ploidy origin that did not include this genome (e.g., it might have been autoallopolyploid), however, one of the two *C. suecicum* samples possesses the B, C, and D Clade homeologs. Nonetheless, this does not undermine the association of both these accessions with the other Eastern Hemisphere hexaploids.

To date, very few Asian *Chenopodium* species have been included in phylogenetic analyses. The presence of the Asian diploid, *C. bryoniifolium*, in Clade A with all the American *Chenopodium* included in this study suggests an intercontinental dispersal; however, our sampling was not designed to address the directionality of introduction. *Chenopodium bryoniifolium* is morphologically very similar to *C. atrovirens*, except for subtle differences in leaf shape, i.e., basal lobes of leaves are more pronounced in *C. bryoniifolium*, and inflorescences are less densely clustered. Despite the similarities, the trait differences are sufficient to confidently identify this specimen as *C. bryoniifolium* and not a recent introduction

of an American species. The *C. bryoniifolium* nucleotide sequences of *SOS* intron 17 are strikingly similar to American diploid species, identical except for five autapomorphies. There is clearly a close association between American diploid *Chenopodium* and at least this one east-Asian species. American taxa form a clade very similar to Clade A in phylogenetic results based on plastid markers (Fuentes-Bazan et al., 2012a, 2012b; B. Walsh and E. Emshwiller, University of Wisconsin-Madison, unpublished data), however *C. bryoniifolium* is the first non-American species included in this clade.

Generic delimitation **-** Our phylogenetic results suggest that after recent taxonomic reclassifications *Chenopodium* is indeed monophyletic. There is one possible exception, however, in that recent studies based on plastid markers (Fuentes-Bazan et al., 2012a; Kadereit et al., 2010) have demonstrated that Australian taxa tend to form a subclade of mixed genera, unresolved with respect to the *Chenopodium* clade s. str. It is unclear from prior works whether the Australian genera *Rhagodia* and *Einadia* are sister to *Chenopodium* or embedded within the genus, making it paraphyletic. By comparing the relative positions of taxa across the Fuentes-Bazan et al. (2012a) and Kadereit et al. (2010) phylogenies, Australian taxa join together in a well-supported clade, including *C. auricomum*, the only Australian sample included in this study. This suggests that the *C. auricomum* sample can roughly approximate a larger Australian clade including *Rhagodia* and *Einadia*. Based on the relationship of *C. vulvaria* and *C. auricomum* in this study and the well-supported distinct Australian clade observed in other studies, we speculate that if *Rhagodia* and *Einadia* sequences were included in this phylogeny, the results would likely demonstrate that despite the great number of recent taxonomic revisions, *Chenopodium* is still not monophyletic. More studies are needed to determine whether it is appropriate to subsume *Rhagodia* and *Einadia* into *Chenopodium* or remove *C. vulvaria* and/or

C. auricomum from *Chenopodium*. The inclusion or exclusion of Australian taxa in *Chenopodium* will have interesting implications concerning the delimitation of the genus, because some Australian taxa are distinctly *Chenopodium*-like morphologically, but *Rhagodia* and *Einadia* have unique traits not observed elsewhere in *Chenopodium*, such as fruit in the form of fleshy berries.

Overall there is not enough resolution among these nuclear markers to genetically distinguish among most of the taxa in Clades A-D. Within individual clades the sequences are much conserved, but under current sampling it is impossible to say more with confidence. The genetic similarity among the American diploid samples in Clade A is consistent with a rapid radiation following an Old World introduction, but this pattern could also be the result of insufficient data or hybridization, to name a few of the potential confounding factors. It is also possible that species delimitation within *Chenopodium* should be reconsidered, and further study using more variable markers may support decisions to merge some of the more morphologically and genetically similar species. For instance, *C. incanum* and *C. fremontii* are extremely similar morphologically, with nearly every trait overlapping in variability, and they share overlapping distributions in southwestern North America. Similarly, the tetraploid species *C. berlandieri* of North America and *C. hircinum* of South America are nearly indistinguishable morphologically and genetically, typically distinguished from each other by the provenience of the collection rather than inherent characteristics of the plant. Clearly more research in species delimitation is required and future research on both species delimitation and phylogeny may need to employ techniques more typically used for population genetics, such as AFLPs or microsatellites.

Indel patterns were not scored and used in this analysis because they demonstrated varying degrees of utility, and we decided not to include them rather than using some and not

others. The indels tended to group into three categories; 1) *Diagnostic of Clades*: In which the observed indel patterns correlated perfectly with membership in Clades A-D and the Vulvaria & Auricomum Clade. These represent about half of the total indels. Scoring these indels would likely have made little impact on this analysis since the branches these diagnostic indels would support already have the maximum support values. 2) *Sub-clade Specific*: Only two such indel synapomorphies were observed, both in intron 17. One is a 2 bp deletion that is shared by the two samples of *C. petiolare* and the two samples of *C. pallidicaule*. The other is a 10 bp deletion that is shared by both samples of *C. berlandieri* subsp. *nuttalliae* in their A Clade homeolog. 3) Variable: These indels don't seem to conform to any obvious phylogenetic pattern and could represent relatively recent indel genesis, complex nested indels, sequence errors, or patterns requiring more thorough sampling to discern. Several indels are hyper-variable; in intron 17 there is a hyper-variable region of approximately 30 bp that has 17 distinct patterns among the 52 ingroup sequences. The sequence pattern is unlike microsatellites.

One of the relationships distinguished among the American diploids is the moderate to weakly supported clade uniting the South American species, *C. pallidicaule* and *C. petiolare*. *Chenopodium pallidicaule*, common name cañihua or cañawa, is a rustically domesticated crop grown on the Peruvian/Bolivian Altiplano. The wild species, *C. petiolare*, grows along the length of the Andes Mountains from Ecuador to northern Chile. The close evolutionary relationship of the wild *C. petiolare* to a domesticated species, suggests *C. petiolare* may be the closest living descendent of the wild progenitor of *C. pallidicaule*. This possibility is further supported by a unique 2 bp deletion in *SOS1* intron 17, shared by all four samples of *C. petiolare* and *C. pallidicaule*. These species are clearly allied based on sequence and indel data, but the uncharacteristically long branch lengths for the domesticated *C. pallidicaule* samples may reflect

the extent of accumulated genetic change accumulated since this species was genetically isolated and domesticated.

A similar clade unites *C. pallidicaule* and *C. petiolare* in Fuentes et al. (2012a), but with much better support. However, their *C. petiolare* samples unite in a polytomy with two of their four *C. atroviren* samples. Presumably these *C. atrovirens* samples are misidentified since *C. atrovirens* is endemic to western United States. Furthermore, the two *C. atrovirens* samples of Bolivian origin ally with *C. petiolare* (native to Bolivia), while their two North American samples join a polytomy consisting of North American diploid species. The later results are consistent with the plastid sequence data of Walsh and Emshwiller (Chapter 1). The Bolivian specimens are likely *C. petiolare*.

The homeologs of the tetraploid *C. berlandieri* subsp. *nuttalliae* samples unite in a well-supported clade within Clade A, and within a moderately-supported clade (68% ML bootstrap and 0.98 Bayesian posterior probability) for Clade B. These samples represent a domesticated Mesoamerican *Chenopodium* that is often difficult to distinguish from other domesticated species, i.e., *C. quinoa* of Peru/Bolivia and *C. berlandieri* subsp. *jonesianum*, an extinct cultigen of eastern North America. The characters that support these clades demonstrate that *C. berlandieri* subsp. *nuttalliae* is genetically distinct from *C. quinoa*, despite the high degree of morphological similarities due to overlapping variation. This genetic distinction between *C. berlandieri* subsp. *nuttalliae* and *C. quinoa* is further supported by a 10 bp indel unique to *C. berlandieri* subsp. *nuttalliae* in *SOS1* intron 17. Despite these findings, the lack of resolution does not allow us to determine whether these two cultigens were domesticated independently or represent an ancient introduction of domesticated material from one region to another.

With the identification and characterization of two polyploid lineages within *Chenopodium*, it is evident they were formed independently of each other. These data strongly suggest that the diploid *C. standleyanum* is the genome donor of the "A" homeolog in the American tetraploid lineage, however our current sampling was not sufficient to identify the genome donor species for the three homeolog clades of Eastern Hemisphere species. Native American species are by far the best represented taxa in this study and also in the phylogenetic studies of Fuentes-Bazan et al. (2012a, 2012b), which were the largest *Chenopodium* phylogenies to date. However, greater species representation from other continents is required to discern the interesting history and relationships among *Chenopodium* species, polyploid and diploid alike. The inclusion of the Asian species, *C. bryoniifolium* from China, revealed that there was likely a trans-oceanic colonization, but it is unclear whether American *Chenopodium* derived from an east Asian introduction or if *C. bryoniifolium* is derived from an American species. Greater sampling is required to investigate these prehistoric events which allowed *Chenopodium*, a relatively recent genus, to colonize all the habitable continents.

The resolution of DNA sequencing of the loci investigated to date does not appear to be sufficient to distinguish between the closely related species that populate the subclades of *Chenopodium*. The two instances of subclades that were resolved with moderate to good support both included domesticated species, and have implications for the origin of these species, which should be investigated further. 1) This was evident in the clade uniting the South American species, wild *C. petiolare* and domesticated *C. pallidicaule*, suggesting *C. petiolare* may be the progenitor species from which *C. pallidicaule* was domesticated. 2) Each of the two homeologs of the Mesoamerican domesticate *C. berlandieri* subsp. *nuttalliae* formed moderate to well

supported clades, demonstrating that this crop is genetically distinct from its morphological look alike, *C. quinoa*.

REFERENCES

Aellen, P. 1929. Beitrag zur Systematik der *Chenopodium*-Arten Amerikas, vorwiegend auf grund der Samanlung des United States National Museum in Washington, D.C. II. Repertorium Specierum Novarum Regni Vegetabilis. *Feddes Repert* 27: 220-223.

Aellen, P. and T. Just. 1943. Key and synopsis of the American species of the genus *Chenopodium* L. *American Midland Naturalist* 30: 47-76.

Alexander, P., G. Rajanikanth, C. Bacon, and C. Bailey. 2006. Recovery of plant DNA using a reciprocating saw and silica-based columns. *Molecular Ecology Notes* 7: 5-9.

Angiosperm Phylogeny Group. 1998. An ordinal classification for the families of flowering plants. *Annals of the Missouri Botanical Garden* 85: 531–553.

Atul, B., S. Shukla, and D. Ohri. 2007. Genome size variation in some cultivated and wild species of *Chenopodium* (Chenopodiaceae). *Caryologia* 60: 245-250.

Baltisberger, M. and M. Voelger. 2006. Sternbergia sicula. *In* K. Marhold [ed.], IAPT/IOPB chromosome data 1. *Taxon* 55: 444-445.

Clegg, M., B. Gaut, G. Learn, Jr., and B. Morton. 1994. Rates and patterns of chloroplast DNA evolution. *Proceedings of the National Academy of Sciences USA* 91: 6795-6801.

Clemants, S. and S. Mosyakin. 2003. *Chenopodium*. *In* Chenopodiaceae. Flora of North America North of Mexico. Oxford University Press. Volume 4: 275-299.

Crawford, D. 1975. Systematic relationships in the narrow-leaved species of *Chenopodium* of the Western United States. *Brittonia* 27: 279-288.

Darriba, D., G. Taboada, R. Doallo, and D. Posada. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* 9: 772.

Downie, S., and J. Palmer. 1994. A Chloroplast DNA Phylogeny of the Caryophyllales Based on Structural and Inverted Repeat Restriction Site Variation. *Systematic Botany* 19: 236-252.

Fuentes-Bazan, S., G. Mansion, T. Borsch. 2012a. Towards a species level tree of the globally diverse genus *Chenopodium* (Chenopodiaceae). *Molecular Phylogenetics and Evolution* 62: 359–374.

Fuentes-Bazan, S., P. Uotila, and T. Borsch. 2012b. A novel phylogeny-based generic classification for *Chenopodium* sensu lato, and a tribal rearrangement of Chenopodioideae (Chenopodiaceae). *Willdenowia* 42: 5-24.

Gandarillas, H. 1968. Razas de Quinoa. *Boletin Experimental* 34: 5-53.

Gandarillas, H. 1974. Genética y origen de la quinoa. *Boletin Informativo* 9: 1-20.

Hancock, J. 2004. *Plant Evolution and the Origin of Crop Species*. CABI.

Hohmann, S., J. Kadereit, and G. Kadereit. 2006. Understanding Mediterranean-Californian disjunctions: molecular evidence from Chenopodiaceae-Betoideae. *Taxon* 55: 67–78.

Huelsenbeck, J., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17: 754-755.

Kadereit, G., T. Borsch, K. Weising, and H. Freitag. 2003. Phylogeny of Amaranthaceae and Chenopodiaceae and the evolution of C4 photosynthesis. *International Journal of Plant Sciences* 164: 959–986.

Kadereit, G., G. Dietrich, J. Surrey, and F. Helmut. 2005. Origin and age of Australian Chenopodiaceae. *Organisms, Diversity & Evolution* 5: 59–80.

Kadereit, G., E. Mavrodiev, E. Zacharias, and A. Sukhorukov. 2010. Molecular phylogeny of Atripliceae (Chenopodioideae, Chenopodiaceae): implications for systematics, biogeography, flower and fruit evolution, and the origin of C4 photosynthesis. *American Journal of Botany* 97: 1664-1687.

Kato, A. 1999. Air drying method using nitrous oxide for chromosome counting in maize. *Biotechnic and Histochemistry* 74: 160-166.

Krasnikov, A., and D. Schaulo. 1990. Chromosome numbers in representatives of some families of vascular plants in the flora of the Novosibirsk region. II. Bot. Žhurn. (Moscow & Leningrad) 75: 118–120.

Linnaeus, C.. 1753. *Chenopodium*, *In* Species Plantarum 1: 218-222.

Maddison, D. and W. Maddison. 2005. *MacClade*. Sinauer Associates, Sunderland, Massachusetts.

Manhart, J. and J. Rettig. 1994. Gene sequence data. Evolution and Systematics of the Caroyphyllales Symposium, H. Behnke and T. Mabry [eds.], Springer Verlag, Berlin.

Miller, M., W. Pfeiffer, and T. Schwartz. 2010. "Creating the CIPRES Science Gateway for inference of large phylogenetic trees" in Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, LA: 1 - 8.

Moquin-Tandon, A. 1840. *Chenopodearum Monographica Enumeratio*.

Mosyakin, S. 2013. New nomenclatural combinations in *Blitum*, *Oxybasis*, *Chenopodiastrum*, and *Lipandra* (Chenopodiaceae). *Phytoneuron* 56: 1–8.

Mosyakin, S. and S. Clemants. 1996. New Infrageneric Taxa and Combinations in *Chenopodium* L. (Chenopodiaceae). *Novon* 6: 398-403.

Mosyakin, S. and S. Clemants. 2002. New nomenclatural combinations in Dysphania R. Br. (Chenopodiaceae): Taxa occurring in North America. *Ukrainian Botanical Journal* 59: 380-385.

Müller, K. and T. Borsch. 2005. Phylogenetics of Amaranthaceae Based on matK/trnK Sequence Data: Evidence from Parsimony, Likelihood, and Bayesian Analyses. *Annals of the Missouri Botanical Garden* 92: 66-102.

Rambaut, A. 2009. FigTree, version 1.31 [computer program]. Available at http://tree.bio.ed.ac.uk/software/figtree/.

Rodman, J. 1990. Centrospermae revisited, part I. *Taxon* 39: 383-393.

Skalinska M., H. Jankun, H. Wcislo. 1976. Further studies in chromosome numbers of Polish angiosperms. XI. Acta Biologica Cracoviensia, *Series Botanica* 19: 107-148.

Standley, P. 1917. The Chenopodiaceae of the North American Flora. *Bulletin of the Torrey Botanical Club* 44: 411-429.

Swofford, D. 2001. *PAUP** 4.0. Sinauer Associates.

Todd, J., and L. Vodkin. 1996. Duplications that suppress and deletions that restore expression from a chalcone synthase multigene family. *Plant C*ell 8: 687-699.

Wahl, H. 1952-1953 (issued 1954). A preliminary study of the genus *Chenopodium* in North. America. *Bartonia* 27: 1–46.

Zwickl, D. J. (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, The University of Texas at Austin.

Table 1. List of samples included in this analysis. The 'Clade/Homeolog ID' column indicates how many homeologs were detected within a given polyploid sample, types A through D. In the 'Ploidy' column, bold entries indicate samples that were counted via chromosome squash, and italic entries indicate samples that were assayed using homeolog specific primers.

Specimen	ploidy	Clade/homeolog ID	Provenance	SOURCE	Herbaria Deposited
Blitum californicum S.Wats.	**2x**	outgroup	California	Jellen-BYU501	WIS
Blitum capitatum L.	**2x**	outgroup	Colorado	USDA-NSL90209	WIS
Oxybasis glauca (L.) S. Fuentes, Uotila & Borsch	2x	outgroup	unknown	Kew Millennium-0174916	WIS
Atriplex heterosperma Bunge	2x	outgroup	Utah	Jellen-BYU480	
C. vulvaria L. - 1	**2x**	Vulvaria clade	Portugal	USDA-PI614896	USDA-NPGS
C. vulvaria L. - 2	2x	Vulvaria clade	Denmark	Jelen-BYU919	WIS
C. auricomum Lindl.	*2x*	Vulvaria clade	Australia	Kew Millennium-0086701	WIS
C. bryoniifolium Bunge	2x	A	China	MO Herb-1984 #04751988	MO Herb.
C. neomexicanum Standl. - 1	2x	A	Arizona	Jellen-BYU852	Welsh Herbarium, BYU
C. neomexicanum Standl. - 2	2x	A	New Mexico	Jellen-BYU843	Welsh Herbarium, BYU
C. hians Standl.	2x	A	New Mexico	Jellen-BYU842	Welsh Herbarium, BYU
C. watsonii A. Nelson	2x	A	Arizona	Jellen-BYU508	Welsh Herbarium, BYU
C. leptophyllum (Moq.) Nutt. ex S.Wats.	2x	A	Arizona	Jellen-BYU589	Welsh Herbarium, BYU
C. incanum (S. Watson) A. Heller	2x	A	Arizona	Jellen-BYU864	Welsh Herbarium, BYU
C. desiccatum A. Nelson	*2x*	A	Nevada	USDA-Ames 29780	WIS
C. pallidicaule Aellen - 1	**2x**	A	Bolivia	USDA-PI478406	USDA-NPGS
C. pallidicaule Aellen - 2	2x	A	Peru	USDA-PI510530	WIS
C. petiolare Kunth - 1	**2x**	A	Chile	Jellen-BYU562	Welsh Herbarium, BYU
C. petiolare Kunth - 2	**2x**	A	Chile	Jellen-BYU564	Welsh Herbarium, BYU
C. fremontii S. Watson	2x	A	Utah	Jellen-BYU660	Welsh Herbarium, BYU
C. standleyanum Aellen	**2x**	A	Wisconsin	Walsh-W0017C	WIS
C. berlandieri var. boscianum (Moq.) Wahl	4x	A B	Texas	Jellen-BYU937	Welsh Herbarium, BYU
C. berlandieri var. zschackei (Murr) Murr ex Graebn.	*4x*	A B	South Dakota	Walsh-via Milton Haar	WIS
C. quinoa Willd. - 1	**4x**	A B	Peru	CIP-FAO Int'l Nursery	
C. quinoa Willd. - 2	***4x***	A B	Peru	USDA-PI510537	WIS
C. quinoa Willd. - 3	**4x**	A B	Peru	CIP-FAO Int'l Nursery	
C. quinoa Willd. - 4	***4x***	A B	Chile	USDA-PI614880	WIS
C. hircinum Schrad.	***4x***	A B	Chile	Jellen-BYU565	WIS
C. berlandieri subsp. nuttalliae (Saff.) H.D. Wilson & Heiser -1	*4x*	A B	Mexico	Jellen-BYU668	WIS
C. berlandieri subsp. nuttalliae (Saff.) H.D. Wilson & Heiser -2	*4x*	A B	Mexico	USDA-PI433230	WIS
C. ficifolium Sm. - 1	**2x**	Ficifolium clade	Switzerland	USDA-PI 658749	USDA-NPGS
C. ficifolium Sm. - 2	2x	Ficifolium clade	Switzerland	USDA-Ames25246	WIS
C. gigantum D. Don	**6x**	B C D	India	USDA-Ames19046	
C. suecicum J. Murr - 1	**6x**	B C D	Czech Republic	Helena Storchova	Institute of Botany ASCR
C. album L.	*6x*	B C D	Wisconsin	Walsh-W0078	WIS
C. missouriense Aellen	**6x**	B C D	Czech Republic	Helena Storchova	Institute of Botany ASCR
C. suecicum J. Murr - 2	**6x**	B C	Czech Republic	Helena Storchova	Institute of Botany ASCR
C. pedunculare Bertol.	**6x**	C D	Czech Republic	Helena Storchova	Institute of Botany ASCR

Table 2. List of primers and sequences used in this study.

BYU Amplification Primers

i16 forward: TGG CCA GCT GCA TGA TTT C
i16 reverse: TCC TCT CCT TCT GTC TCA CTT TCA GT
i17 forward: TGT CCG CAC TAC CTT TCC TGA
i17 reverse: ACT TGC CTC GTC TTA ACA ACT CGT

UW-Madison Amplification Primers

SOS1i16F2: TGT TAC ATA TGC GCT GCA TTT TTA CG
SOS1i16R: TTT CAG TGA TGA CTG CAG AAG
SOS1i17F: TCC TCA CTA CCT TTC CTG AG
SOS1i17R2: GAA CAG CAT CAT GAA GAT GAA G

Homeolog-Specific Primers

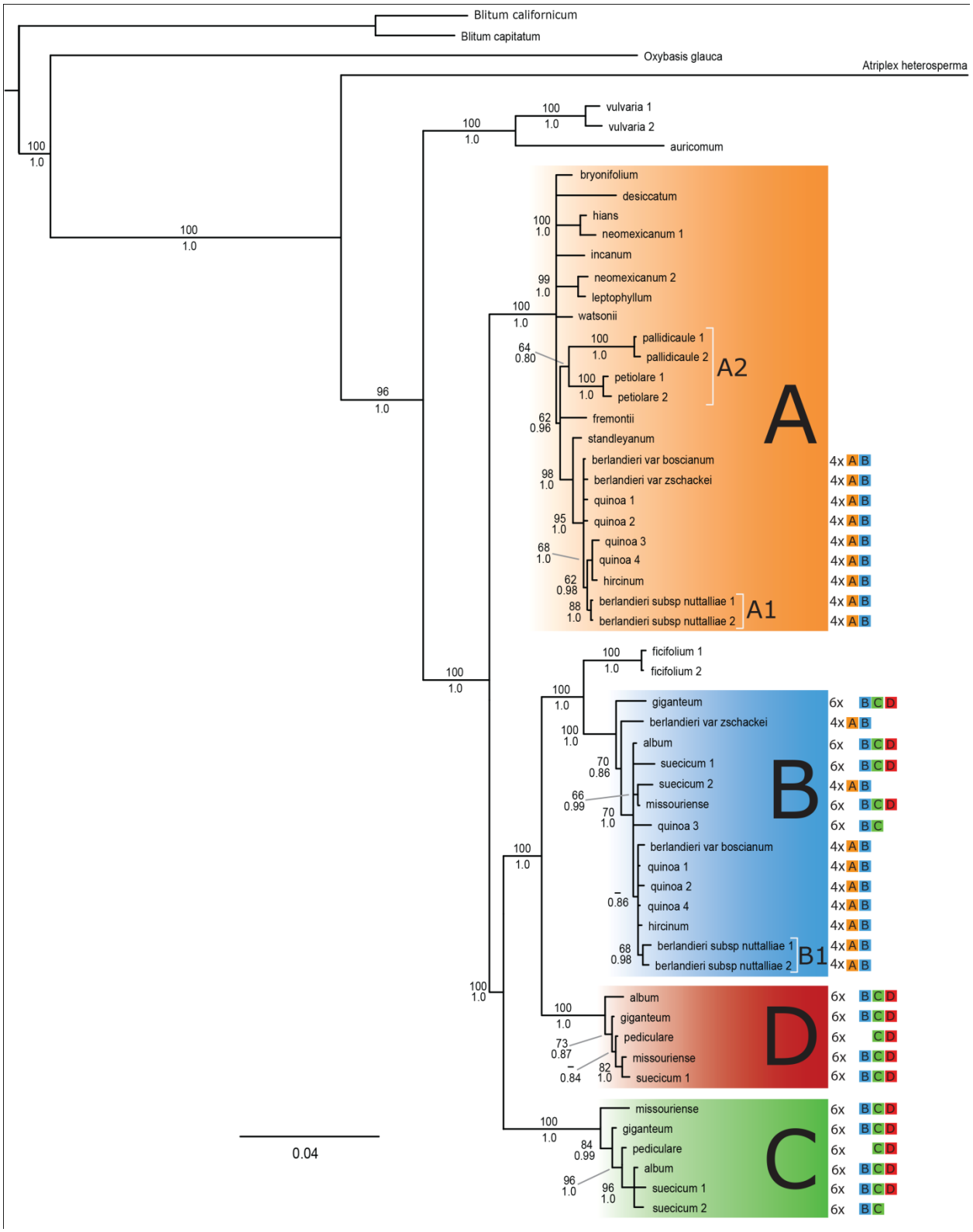
i16A+ & i16R_B- for A specific band
i16C- & i16B+ for B specific band
i16C+ & i16R_B- for C specific band
i16D+ & i16R_B- for D specific band

i16A+: GTA TAG ATA CCT TTC CTC CG
i16R_B-: GTC TTC ATA TCC TTC CCT AA
i16C-: AAK CCA TTC ATA GTT TTC TC
i16B+: GCT AGT CTT CAT ATC CYT AA
i16C+: CTC TTG TTT TCA ATA ACT TG
i16D+: CAT TAT TTA TAG GGA TAT ATA TC

Figure 1. The image shows the Maximum Likelihood phylogenetic construction. The trees produced using Bayesian analyses are fully congruent with the cladogram depicted. Small subclades discussed in the text are indicated by white brackets.

The numbers reported above the lines are likelihood bootstrap percentages. Numbers below lines are Bayesian posterior probability scores. Likelihood bootstrap values were only reported if 60% or higher. Bayesian posterior probabilities were only reported if 0.80 or higher. Dashes denote support values below the designated cutoff. Nodes were collapsed when no support values were higher than the cutoff, visually simplifying phylogenetic associations.

The homeologs of polyploid samples group within four clades designated A-D. Two polyploid lineages were identified: tetraploid species (AB) native to the Americas, and hexaploid species (BCD) native to Eurasia. Along the right margin is a key to aid in identifying the number of homeolog sequences present within a given sample, and in which clades those homeologs will be found. Diploid samples are not marked.


Blitum californicum
Blitum capitatum
Oxybasis glauca
Atriplex heterosperma
vulvaria 1
vulvaria 2
auricomum
bryonifolium
desiccatum
hians
neomexicanum 1
incanum
neomexicanum 2
leptophyllum
watsonii
pallidicaule 1
pallidicaule 2
A2
petiolare 1
petiolare 2
A
fremontii
standleyanum
berlandieri var boscianum
berlandieri var zschackei
quinoa 1
quinoa 2
quinoa 3
quinoa 4
hircinum
berlandieri subsp nuttalliae 1
berlandieri subsp nuttalliae 2
A1
4x A B
ficifolium 1
ficifolium 2
giganteum
berlandieri var zschackei
album
suecicum 1
suecicum 2
missouriense
quinoa 3
B
6x B C D
4x A B
6x B C D
6x B C D
4x A B
6x B C D
6x B C
berlandieri var boscianum
quinoa 1
quinoa 2
quinoa 4
hircinum
berlandieri subsp nuttalliae 1
berlandieri subsp nuttalliae 2
B1
4x A B
album
giganteum
pediculare
missouriense
suecicum 1
D
6x B C D
6x C D
missouriense
giganteum
pediculare
album
suecicum 1
suecicum 2
C
0.04

Chapter 3

USING MOLECULAR MARKERS TO TRACE THE DOMESTICATION HISTORY OF THE EXTINCT *CHENOPODIUM* CULTIGEN

ABSTRACT

Premise of the study- Researchers have speculated on the origin of thin-seed coated, domesticated *Chenopodium* seeds (Amaranthaceae) recovered from Eastern North American (ENA) archaeological sites for nearly 75 years. Three hypotheses were proposed: 1) *Chenopodium* was domesticated in Mesoamerica and subsequently introduced to ENA (Wilson, 1981); 2) *Chenopodium* was domesticated independently in both ENA and Mesoamerica (Smith, 1984); and 3) *Chenopodium* was domesticated in ENA and later introduced to Mesoamerica (Heiser, 1990). A recent study addressed these hypotheses (Kistler and Shapiro, 2010) by sequencing DNA from archaeological seeds using a plastid sequence marker. Their findings supported the second hypotheses. Our analysis includes increased sampling from more sites to confirm whether their results will be corroborated.

Methods- We used DNA sequences of the plastid intergenic *trnQ-rsp16* region and single/low-copy nuclear *Salt Overly Sensitive 1* (*SOS1*) region, each including genetic markers unique to the Mesoamerican domesticate. Archaeological seeds were sequenced and compared to modern sequences.

Key results- Plastid sequences were obtained from 12 archaeological seeds representing 5 sites. Three seeds yielded sequences for SOS1.

Conclusions- Collectively the archaeological seeds have a mix of plastid haplotypes similar to wild populations, suggesting occasional hybridization with wild populations. Approximately 40% of seeds yielded plastid sequences similar to the Mesoamerican domesticate, indicating a single domestication gave rise to the Mesoamerican and ENA cultigens. The three nuclear sequences obtained from archaeological seeds were similar to those of the modern

Mesoamerican domesticate. These results in conjunction with current archaeological evidence support Hypothesis 3: domesticated *Chenopodium* arose in ENA and was subsequently introduced to Mesoamerica.

Key Words: *Chenopodium berlandieri* subsp. *nuttallaie*; *Chenopodium berlandieri* subsp. *jonesianum*; aDNA; ancient DNA; degraded DNA

INTRODUCTION

For over a century researchers have attempted to determine where crops originated. Vavilov (1992) traveled the world in the 1920s and observed that crops did not appear randomly, but grouped within discrete regions. Vavilov characterized many of these regions he described as centers of variation, which correlate strongly with centers of crop domestication. He observed that crop plants most often arose in regions where the greatest amount of morphological variability occured. However, some crops could not be definitively traced to a single center of domestication, contrasting sharply with the long held belief that the creation of domesticated crops was a rare occurrence, with each crop arising once. Many crops became the subject of controversies over single vs. multiple domestications. Many of these controversies are currently still under debate. Over the last 50 years researchers have found that domestication is more common and more complicated than once thought, as suggested by the multiple independent domestications determined for beans, squash, and cotton (Decker, 1985; Gepts and Bliss, 1985; Wendel et al., 1989; Decker-Walters et al., 2002).

There are two regions, corresponding to centers of domestication, where it is very difficult to determine in which location certain domesticated crops originated: Mesoamerica, and the woodlands and river valleys throughout much of the mid-continental United States encompassing the middle portion of the Mississippi River and several of the tributary rivers (Fig. 1), often referred to as eastern North America (ENA).

Mesoamerica is well accepted as a center of domestication dating back to the 1920s with the work of Vavilov (1992). Mesoamerica has given rise to several economically important crops such as maize (*Zea mays* L.) and beans (*Phaseolus vulgaris* L.), as well as more localized

domesticates, such as *Amaranthus* spp. (MacNeish, 1970). These crops are still cultivated in Mexico today.

ENA is a lesser known and more recently recognized center of domestication. Bruce Smith (1985) championed the recognition of this center of independent domestication, which still remains controversial. Vavilov was not aware of ENA as a center of domestication because by the 1920s the indigenous peoples of this region were almost completely displaced, and suffered catastrophic demographic decline due to Old World diseases to which Native Americans had no immunities, and surviving populations had been encouraged or forced to abandon many cultural traditions, including the cultivation of traditional crops (Jones et al. 2011; Giago 2006). However, archaeological evidence throughout this region suggests five domesticated crops potentially originated in ENA, marshelder (*Iva annua*; extinct), little barley (*Hordeum pusillum*; extinct), sunflower (*Helianthus annuus*), squash (*Cucurbita pepo*), and a thin seed-coated Chenopod (*Chenopodium berlandieri* subsp. *jonesianum*; extinct), from hence-forth referred to as 'jonesianum' (Smith, 1987; Hunter, 1992). Subsistence practices relied heavily on these crops for thousands of years until around 800-1000 CE, when an agricultural transition took place and shifted from reliance on traditional crops to maize (Ambrose, 1987. Over the last 1000 years domesticated marshelder, little barley, and 'jonesianum' went extinct. Fortunately, a good amount of archaeological work has been conducted throughout ENA and the traditional crops potentially domesticated in this region are well represented in museum collections.

Of the domesticated crop remains recovered from archaeological sites across ENA, only marshelder and little barley are unique to this region, and the recognition of little barley as truly domesticated is still undecided (Hunter, 1992). As for squash, sunflower, and *Chenopodium*, researchers have reported finding these domesticates in both Mesoamerica and ENA, each crop

sparking debates over their individual origins of domestication. In all three cases the data are not conclusive enough for researchers to reach a definitive conclusion, so the controversies continue. For sunflower, the consensus seems to currently favor an origin in ENA followed by an introduction to Mesoamerica (Harter et al., 2004). Genetic analyses for squash (*Cucurbita pepo*) and domesticated *Chenopodium* suggest they were independently domesticated in both Mesoamerica and ENA (Decker-Walters et al., 2002; Kistler and Shapiro, 2011).

The origin of domesticated *Chenopodium* has proven to be one of the most challenging and interesting of the controversial crops. Large caches of *Chenopodium* seeds were occasionally recovered from archaeological sites in ENA, such as the approximately 10 million seeds excavated from Ash Cave, OH (Smith, 1985), and the 3.5 million seeds from Newbridge, IL (Maina, 1967, misreported as *C. album*). But *Chenopodium* seeds are relatively small (1-2 mm diameter) and often went unnoticed at archaeological excavations until the implementation of soil flotation, which makes possible the recovery of tiny organic objects, such as fish scales, fine bones, small seeds, and snail shell fragments (Struever, 1968). Soil flotation has revealed that *Chenopodium* seeds are present in the majority of archaeological sites dating between 2200-1000 B.P. and are often the most abundant seed type present. Since the 1930s researchers had speculated that *Chenopodium* may have been cultivated throughout ENA. In 1985, Smith using electron microscopy, identified seed coat thickness as the first quantitative character for distinguishing between wild vs. domesticated *Chenopodium*; seed coat thickness greater than 40 microns indicates wild and less than 20 microns indicates domesticated seeds. These measurements of seed coat thickness appear to be applicable to all species of *Chenopodium*.

Once researchers were able to adequately recover and distinguish domesticated *Chenopodium* from wild varieties, comparisons between domesticated *Chenopodium* from ENA

and Mesoamerican was possible. The extant *Chenopodium* domesticate cultivated in Mesoamerica is a relatively localized crop, often called huazontle (*C. berlandieri* subsp. *nuttalliae*), from hence forth referred to as 'nuttalliae' for brevity. 'Nuttalliae' is characterized by a reduction in seed coat thickness, sometimes entirely absent, and typically has tan colored seeds in contrast to the black seeds more typical of wild forms. The dense elongated terminal inflorescences of 'nuttalliae' are traditionally consumed battered and fried, however, the leaves and seeds are also edible, and possibly as nutritious as the closely related quinoa (*C. quinoa*), which are high in protein, essential amino acids, and minerals (Schlick and Bubenheim, 1993). The current range of 'nuttalliae' encompasses the Trans-Mexican Volcanic belt, roughly extending from Guadalajara to Puebla, and including the ancient city Tenochtitlán, currently known as Mexico City. This region of Mexico is notorious for poor preservation of botanical remains, and no domesticated *Chenopodium* remains are known from pre-European contact archaeological sites (McClung de Tapia and Aguilar-Hernández, 2001).

The extinct domesticated *Chenopodium* of ENA, 'jonesianum', was potentially the most important crop during the Woodland period (approximately 2500-1000 B.P.). As stated earlier, the starchy seeds of 'jonesianum' are often the most abundant seed type recovered in botanical assemblages and large caches of seeds have been recovered. The large cache of seeds recovered from Ash Cave, OH (Smith, 1985), were buried in a woven sack, and some of these seeds show signs of germination and insect damage suggesting they were stored viable for future sowing. It was common practice for seeds interred in food caches to be parched or toasted, killing insects, bacteria, fungi, and the seed embryos, greatly enhancing stored food preservation. 'Jonesianum' seeds were certainly being consumed by many people based on their presence in the majority of human coprolites predating 1000 B.P recovered from Kentucky caves and rockshelters. These

seeds were consumed along with other domesticated seeds such as sunflower (Faulkner, 1991; Gremillion, 1996).

Comparisons between the domesticated *Chenopodium* of Mesoamerica and ENA are problematic because typically only the seeds of 'jonesianum' are preserved, greatly limiting morphological comparisons to 'nuttalliae'. Furthermore, 'jonesianum' seeds are often charred, poorly preserved, fragmented, and physically altered by the conditions of preservation, including shrinking and shape alterations caused by the inward collapsing seed coat. There are few archaeological examples of 'nuttalliae', so comparisons involve 1000+ year old 'jonesianum' seeds compared to modern 'nuttalliae' seeds resulting in potentially erroneous conclusions. The primary distinguishing characteristic between the two domesticates is that 'jonesianum' typically has black colored seeds, similar to wild type seed color, and 'nuttaliae' typically has tan seeds. However, tan to brownish-tan 'jonesianum' seeds have been recovered from several Ozark rockshelter archaeological sites (Arkansas) and the Riverton site, IL (Fritz, 1986; Smith and Yarnell, 2009). The tan seeds recovered from Riverton are the oldest known 'jonesianum' seeds recovered within ENA, determined to be 3800 B.P. (Smith and Yarnell, 2009).

Differing interpretations of the morphological and archaeological data has resulted in the proposal of three hypotheses.

Hypothesis #1, a single domestication originating in Mesoamerica (Wilson, 1981): The tan 'jonesianum' seeds recovered from ENA share the same morphological characters indicative of domesticated 'nuttalliae'. The two populations represent a common domestication originating in Mesoamerica and followed by a subsequent introduction to ENA. Fritz (1986) speculated that 'nuttalliae' was introduced into the Ozark region and spread throughout ENA, perhaps

hybridizing with wild taxa to develop the black seed coated 'jonesianum'. This hypothesis is the least compatible with current archaeological data. In ENA, 'jonesianum' has a rich, nearly 4000 year history with the oldest site, Riverton, IL, exhibiting tan seeds similar to 'nuttalliae'. A Mesoamerican introduction is still plausible, but fully domesticated 'nuttalliae' would need to have at least a 4000 year history. Currently, domesticated *Chenopodium* has about a 500 year history from Mesoamerican archaeological sites (McClung de Tapia and Aguilar-Hernández, 2001). Perhaps the ancestral remains of 'nuttalliae' had a much smaller distribution than today and have so far evaded detection.

Hypothesis #2, independent domestications in both Mesoamerica and ENA (Smith, 1984): The peoples throughout ENA were industrious agriculturalists propagating a suite of crops and definitively giving rise to at least one domesticated cultigen (marshelder). 'Jonesianum' represents a cultigen of *Chenopodium* derived independently from 'nuttalliae', but coincidentally domesticated from the same wild progenitor species, *C. berlandieri*. Morphological similarities between 'nuttalliae' and 'jonesianum' are the result of the selective pressures of domestication converging on a similar suite of traits, coincidentally resembling each other. This hypothesis is the most consistent with current archaeological data. As it stands, archaeological data suggest 'nuttalliae' may have been domesticated more recently than 'jonesianum'.

Hypothesis #3, a single domestication originating in ENA: Heiser (1990) stated that fully-domesticated tan 'jonesianum' seeds were introduced to Mesoamerica from ENA. Heiser however did not argue or support his claim, but merely stated this hypothesis in the closing sentence of his publication. This hypothesis is the least accepted of the three, but the fully domesticated, nearly 4000 year old, tan seeds recovered from Riverton, IL, suggest it has merit. Furthermore, the tan and black seeds recovered from ENA could be considered morphological

variation, which is the primary observation Vavilov used to identify centers of domestication in the 1920s. Crops in regions exhibiting more morphological variation than neighboring regions were interpreted as having a longer history of cultivation and the probable location of origin.

Morphological evidence alone is inconclusive and researchers generally agree that additional lines of evidence are needed. The genetic analysis of Kistler and Shapiro (2011) is the only analysis conducted using molecular instead of morphological data, and their results are consistent with hypothesis #2, the independent origins of 'nuttalliae' and 'jonesianum'. DNA sequences from a non-coding plastid region were obtained from three archaeological 'jonesianum' seeds, one each from three sites, and compared with modern 'nuttalliae' and wild *C. berlandieri* sequences. The genetic signatures obtained from 'jonesianum' seeds more closely matched wild *C. berlandieri* than Mesoamerican 'nuttalliae', thus suggesting that both domesticated lineages were independently domesticated. We feel the small sampling of 'jonesianum' (n=3) and limited geographic representation of wild taxa justifies further study of the origins of domesticated *Chenopodium.* In addition, degraded DNA analyses use plastid or mitochondrial sequence data since the abundance of these organelles, hundreds per cell, greatly enhance the potential for successful sequence amplification. However, organelles are generally maternally inherited and only allow an incomplete glimpse of the organism's evolutionary history, especially since *C. berlandieri* are allotetraploid (4x), a hybrid of two genetically distinct species.

The focus of the present study is to reinvestigate the relationship among the domesticated *Chenopodium* subspecies of Mesoamerica and ENA with the goal to resolve differences among domesticated populations if they indeed were domesticated independently. Our study attempts to test the findings of Kistler and Shapiro (2011) including a larger sampling of archaeological

seeds from a greater number of archaeological sites. Furthermore, the current study attempts to obtain both plastid and single/low copy nuclear sequence data for all samples. Comparisons of genetic markers will be used to determine if 'jonesianum' sequences are most similar to 'nuttalliae' suggesting a single origin/domestication, or if the domesticates are more similar to wild *C. berlandieri* suggesting separate origins and independent domestications.

Summary of the three hypotheses being tested:

Hypothesis #1: 'Nuttalliae' was introduced into ENA, originating from a single domestication event in Mesoamerica (Wilson, 1981).

Hypothesis #2: 'Nuttaliae' and 'jonesianum' were independently domesticated, coincidentally from the same progenitor species; thus two separate origins (Smith, 1983).

Hypothesis #3: 'Jonesianum' was introduced into Mesoamerica, originating from a single domestication event in ENA (Heiser, 1990).

METHODS

Taxon sampling, modern specimens- Sequences of modern samples comprise the *trnQ-rsp16* plastid dataset for Walsh and Emshwiller (unpublished, but will be). This dataset consists of a total of 234 samples, representing 77 species. The *Chenopodium* sampling represents a global distribution, but includes nine 'nuttalliae' accessions and 51 accessions of wild *C. berlandieri*. This study includes representatives of the six varieties of wild *C. berlandieri* as described in the most recent treatment of North American *Chenopodium* (Clemants and Mosyakin, 2004) in an attempt to identify the taxa most closely related to the wild progenitors of 'jonesianum' and 'nuttalliae'. The goal of taxon sampling was to obtain a broad geographic representation across Mexico and the United States.

Taxon sampling, archaeological specimens- Archaeological specimens for 'jonesianum' were obtained representing 7 archaeological site collections (Fig. 1 & Table 1). These collections were on loan from five museums with permission to conduct destructive analyses. This analysis of multiple archaeological collections is intended to represent the diversity of 'jonesianum' spanning time and geography. These seven sites were specifically selected for DNA analyses because they met at least two of the three following criteria: 1) quantity of seeds in each collection was sufficiently large so as not to significantly deplete the collection, 2) seeds exhibit excellent preservation, and 3) these seed collections are well characterized in the literature with radiocarbon dates and morphological studies confirming traits indicative of domestication.

Target region, plastid- Thirteen plastid intergenic spacer regions were sequenced to identify variability among extant samples of two wild *C. berlandieri* and one 'nuttalliae'. The *trnQ-rsp16* intergenic spacer region demonstrated the highest variability observed among plastid sequence

comparisons, with two 'nuttalliae'-unique characters, instead of the zero to one observed for other plastid markers. Any persisting DNA contained within archaeological seeds was assumed to be fragmented due to degradation, so two sets of primers were designed (Table 2), each flanking one of the 'nuttalliae'-unique characters approximately centered within a 100-250 base amplification.

Target region, nuclear- A 13-base indel unique to 'nuttalliae' was identified within the A-version homeolog as well as a 'nuttalliae'-unique character within the 17th intron of *salt overly sensitive 1* gene (*SOS1*), a single-copy nuclear marker. Primers were designed that took advantage of the many indels present in this region to amplify an approximately 130 base segment of just the A-version homeolog, encompassing the diagnostic 'nuttalliae'-unique characters (chapter 2).

DNA extraction, amplification, and sequencing for extant samples- Extant samples were processed at the University of Wisconsin-Madison. The extraction and amplification of genomic DNA from modern taxa was conducted as described in Chapter 2 for *SOS1* intron 17, and Chapter 1 for the plastid *trnQ-rsp16* intergenic spacer.

DNA extraction, amplification, and sequencing for archaeological samples- The extraction of genomic DNA and the setup for PCR reactions for archaeological seeds was conducted in the dedicated degraded DNA laboratory at the University of Illinois at Urbana-Champaign, managed by Dr. Ripan Malhi. In the summer and fall of 2013, three trips were made to the degraded DNA facility where DNA extractions and the setup for amplification reactions were generated. Genomic DNA was extracted using bleached mortar and pestles and a DNeasy Plant Mini Kit (Qiagen, Valencia, California, USA), as per manufacturers specifications. PCR reactions were

performed in 25 µl reactions (2-5 µl template, 0.2 µl *Taq* polymerase (5 units/µl, New England BioLabs, Ipswich, Massachusetts, USA), 0.5 µl of each primer (10 µM ea.), 1.25 µl BSA (10 mg/ml), 2 µl dNTP (reagent consists of 200 µM of each nucleotide), 2.5 µl 10x buffer (1.5 mM $MgCl_2$ included) as follows: initial denaturation was at 96°C for 2 m; followed by 45 cycles of 96°C for 30 s, 52°C for 30 s, and 72°C for 30 s; followed by a final 72°C for 5 m. PCR products were purified using Agencourt Ampure beads (Beckman Coulter Genomics, Morrisville, North Carolina, USA).

As with any study involving degraded DNA, many precautions were taken to prevent and detect potential sources of contaminating DNA. Modern and archaeological samples were processed in separate facilities. The degraded DNA laboratory had not previously been used for *Chenopodium* research. The degraded DNA laboratory enforced strict rules, such as all researchers wore head-to-toe sterile gowns and followed entry protocols to eliminate the introduction of environmental contamination. All incoming objects entering the degraded DNA laboratory were surface treated with DNA AWAY surface decontaminant (Thermo Scientific, Madison, Wisconsin, USA). Furthermore, all disposable lab equipment used at the degraded DNA laboratory was purchased new and sent directly there. The few items transferred from UW-Madison to the degraded laboratory, such as mortars and pestles, were washed and soaked in undiluted Clorox bleach (6% sodium hypochlorite) for 20 minutes before use. In the case of ceramic mortars and pestles, six null DNA extractions (negative controls) were conducted, immediately following the extraction of varying amounts of fresh leaf material. Mortars and pestles were soaked in bleach for varying amounts of time, from 2-10 minutes. No PCR amplifications or sequences could be generated from the null samples.

PCR reactions were setup in the degraded DNA laboratory and sealed tubes were then transferred to Dr. Ripan Malhi's modern sample laboratory, three blocks away, and amplified in a thermocycler. Negative controls were included with all PCR amplifications. Most PCR were frozen and later visualized in Dr. Sara Patterson's laboratory at the University of Wisconsin-Madison; a laboratory in which *Chenopodium* research had not previously taken place.

As a means to identify potential contaminating DNA from PCR product, primers were designed to amplify non-overlapping segments of the *trnQ-rsp16* plastid intergenic spacer region. Each of the two segments amplified one 'nuttalliae'-specific character. Inconsistencies among the sequenced DNA regions does not necessarily indicate contamination, but was implemented to help identify inconsistencies among two contiguous regions. Furthermore, attempts were made to amplify all samples for all regions multiple times to better detect inconsistencies in PCR product which might indicate PCR contamination or the use of incorrect DNA template. Duplicate amplifications were always conducted on different days.

In ancient DNA analyses, when a sample is sufficiently large, two separate extractions are conducted to verify consistency. Archaeological *Chenopodium* seeds are generally quite small, approximately 1 mm diameter. Despite their small size, four seeds from the Cow Ford site, AR, the structurally best preserved seeds, were cut in half with each half extracted separately.

Sequence generation, assembly, and character scoring- Cycle sequencing reactions were generated using purified PCR products and BigDye Terminator as per manufacturer's specifications (Life Technologies, Grand Island, New York, USA). Products were further purified with Agencourt CleanSEQ beads (Beckman Coulter Genomics, Morrisville, North

Carolina, USA) and submitted to either the UIUC Core Sequencing Facility at Urbana-Champaign or UW Biotechnology Center at Madison.

Sequences were edited and contigs generated using Sequencher 4.9 (Gene Codes Corporation, Ann Arbor, MI, USA). Contigs were manually aligned using MacClade 4.08a OS X (Maddison and Maddison, 2005).

The scoring for each region of each sample was originally based on the binary choice, either wild-type or 'nuttalliae'-like, based on the nucleotide observed at the 'nuttalliae'-unique character for each region. However, more variation or haplotypes were observed than encompassed by scoring a single nucleotide. Instead, the scoring methodology was slightly altered to score each haplotype observed in the sequence data. Sample results were then used to compare genetic and physical traits to identify potential patterns in the data.

RESULTS

In total, 12 seeds among five archaeological sites were successfully sequenced (Figure 2). The sequence length for *trnQ-rsp16* segments 1 and 2 and the segment of *SOS1* intron 17, are 138-206, 115-170, and 221-235, respectively. Amplification and sequence quality were considerably weaker/poorer than observed among modern samples.

Character Scoring- Character scoring was not as simple as originally anticipated, based on the sequences of four samples. Instead of just wild-type and 'nuttalliae'-like haplotypes, a third haplotype was observed, genetically identical to sequences of a taxonomically challenging group of Eurasian polyploids, occasionally referred to as the *C. album* complex. Furthermore, three variants of wild-type haplotype were observed, which were also observed among modern sequences once greater sampling was analyzed (Chapter 1).

Archaeological sequence data were scored as one of three distinct categories determined by manual comparisons within a sequence alignment with modern taxa: wild-type, Eurasian, or 'nuttalliae'-like. The 'nuttalliae'-like haplotype is characterized by sequence data identical to those observed in modern 'nuttalliae' samples, based on nine 'nuttalliae' accessions for plastid sequence data and five accessions for *SOS1*. The wild-type haplotypes differs from 'nuttalliae'-like sequences by one or more nucleotide markers, always differing from 'nuttalliae' at the 'nuttalliae'-unique marker. Two versions of the wild-type haplotype were observed in segment 1, differing by 1 bp, and three haplotypes observed in segment 2, differing by 1 bp each. We cannot determine if the variability observed is due to poor sequence quality, or representing actual variation as observed in modern wild populations in Chapter 1. The variable residues tend to be at hotspots of variability among wild samples, however only the wild-type designation of

Segment 1 (Fig. 2) identically matches any sequences represented in Chapter 1. The Eurasian haplotype is quite distinct from wild-type and 'nuttalliae'-like haplotypes, characterized by a 17 base indel and several unique nucleotide characters. This haplotype is very similar to sequences of *C. album*, a currently globally distributed Eurasian invasive particularly adept at thriving within anthropogenic disturbances.

Within plastid sequence data, the wild-type haplotype class was the most frequently occurring sequence pattern, detected in 58% of the archaeological seeds assayed, 7 of the 12 seeds. The 'nuttalliae'-like haplotype was the second most frequently occurring haplotype class, detected at 33% frequency, 4 of 12 seeds. The Eurasian haplotype was detected in a single seed from the Cow Ford site, AR, representing 8% of the detected variability.

In all instances when both plastid segments successfully amplified, the contiguous segments were of the same haplotype determination. Furthermore, since samples could not be successfully divided and extracted independently of each other to confirm consistency, replicates of amplifications were attempted. In all instances of replicate amplifications the haplotype determinations were in agreement.

Sequence contamination- Archaeological seed amplifications for plastid markers began sporadically demonstrating weak positive reactions in negative controls shortly into the second of three trips to the degraded DNA laboratory. Some negative controls generated the 'nuttalliae'-like haplotype while other amplifications produced the wild-type haplotype. The source of the contamination is still unknown, but did not appear to spread to other archaeological seed amplifications. Some adjacent amplification reactions produced varying haplotypes differing from the sequences observed from the failed negative controls, while most reactions did not

amplify at all. As a conservative precaution, we decided to only consider plastid sequence data obtained during the first trip, when no contamination was detected. Unfortunately, sequence data for the Newbridge site, IL, were derived from amplifications generated during the second trip to the degraded DNA laboratory and will not be discussed further other than to report our observations which may be of potential use in future studies: One semi-charred domesticated Newbridge seed was successfully sequenced for *trnQ-rsp16* segment 2 possessing the 'nuttalliae'-like haplotype, and two wild *C. berlandieri* seeds recovered from a different feature than the domesticated seeds were successfully sequenced each possessing a wild-type haplotype in segment 2.

Amplifications for *SOS1* were attempted once during the first trip to the degraded DNA laboratory without successful amplification. During the second and third trip many attempts to amplify *SOS1* were tried. In total three domesticated archaeological samples, spanning two archaeological sites, weakly amplified for *SOS1*. All three sequences demonstrated a 'nuttalliae'-like haplotype. No amplification of negative control reactions were observed for *SOS1* amplifications.

Testing sample consistency by dividing seed samples- To test within sample consistency, four Cow Ford seeds were cut in half and DNA extractions were attempted on each half. Only one seed-half successfully amplified and sequenced, so consistency could not be determined using this method. The Cow Ford seeds were selected for this test because they exhibit the best preservation observed among all archaeological seed assemblages included in this study, and thus were considered the most likely samples to work for this analysis. Since sequence data was successfully generated from the four whole-seed extractions from the Cow Ford collection, but could not be reliably obtained from half-seed extractions, dividing archaeological *Chenopodium*

seeds in half is not a useful method of demonstrating within-sample sequence consistence under the procedure used in this study.

DISCUSSION/CONCLUSION

We were surprised to observe mixed haplotypes among seed assemblages within multiple archaeological sites, as well as the presence of the 'nuttalliae'-like haplotype being detected in multiple sites, including Cloudsplitter, the oldest site included in this study (Fig. 2). With such an unexpected pattern it is easy to suspect error or contamination may have played a role. However, many precautions were implemented to prevent contamination and detect inconsistencies within the data, and we are confident that these sequences accurately represent DNA inherent in the archaeological samples themselves.

Our results differ from the Kistler and Shapiro (2011) analysis of 'jonesianum' in that we observed three haplotype classes mixed within archaeological seed assemblages while Kistler and Shapiro (2011) only observed one, the wild-type haplotype. Our results suggest the wild-type haplotype is the most abundant, present in greater than 50% of analyzed sample. It is possible that the variability was not observed by Kistler and Shapiro (2011) due to limited sampling, as they collected sequence data from three archaeological seeds total, resulting in the chance detection of only the most frequently occurring haplotype. Furthermore, of the three samples total from the Holman rockshelter, AR, two from this study and one from Kistler and Shapiro (2011), only the wild-type haplotype was detected. If this reflects a higher proportion of the wild-type haplotype than observed in other populations, then the odds of Kistler and Shapiro (2011) observing variation would be further reduced.

In retrospect we would expect wild-type haplotypes to be detected within cultivated populations of 'jonesianum'. After all, 'jonesianum' was being cultivated for several thousand years within the natural range of its wild progenitor species, *C. berlandieri* (also the likely

progenitor for 'nuttalliae'). Detection of the Eurasian haplotype is also somewhat expected. Currently, *C. album* is considered one of the most serious and pervasive agricultural weeds in the Midwest. It thrives in nearly any human-made disturbance and likely persisted in close proximity with American Indians, especially once agriculture became the primary means of subsistence.

The only truly unexpected finding is the detection of the 'nuttalliae'-like haplotype within 'jonesianum' populations. This was a surprise to us because we expected to observe and confirm the findings of Kistler and Shapiro (2011). We are as certain as we can be that we sequenced DNA associated with the archaeological seeds and not from a modern source.

'Nuttalliae'-like haplotype and testing hypotheses- Using plastid sequence data the 'nuttalliae'-like haplotype was detected in four archaeological seed assemblages: Cloudsplitter, Ash cave, Cow Ford, and Edens Bluff. The presence of the 'nuttalliae'-like haplotype among cultivated populations of 'jonesianum' spanning over 2000 years and over 600 miles apart suggests this haplotype is associated with and indicative of a cultivated *Chenopodium* lineage. The few *SOS1* sequences generated also detected the 'nuttalliae'-like haplotype among three of the same sites as the plastid data did.

The detection of the 'nuttalliae'-like haplotype among multiple archaeological populations of 'jonesianum' using both plastid and nuclear markers is not consistent with Hypothesis #2: 'nuttalliae' and 'jonesianum' independently domesticated. It is profoundly unlikely for these non-coding genetic markers to have converged upon identical sequences for two independently domesticated lineages. These results support rejecting Hypothesis #2.

The shared 'nuttalliae'-like haplotypes of 'nuttalliae' and 'jonesianum' strongly suggests they derived from a single common ancestor. Unfortunately the sequence data cannot determine whether the Mesoamerican 'nuttalliae' was introduced to ENA, or if 'jonesianum' was introduced to Mesoamerica.

Current archaeological data are most consistent with Hypothesis #3: domesticated 'jonesianum' introduced to 'Mesoamerica'. The archaeological history of domesticated *Chenopodium* is more than 3000 years older in ENA than in Mesoamerica. For Mesoamerica to be the source of the 'nuttalliae'-like haplotype in ENA, 'nuttalliae would need to have been introduced prior to 3450 B.P. to be detected in the Cloudsplitter seed assemblage, and earlier than 3800 B.P. to explain the tan, fully domesticated seeds recovered from the Riverton site, IL. This suggests 'nuttalliae' should have a roughly 4000 year history in Mesoamerica, but so far no domesticated *Chenopodium* remains of this antiquity have been recovered. Although we acknowledge that the absence of archaeological domesticated *Chenopodium* seeds from an area known for poor preservations is not very compelling, many 1000+ year old *Chenopodium* seeds have been recovered from habitation sites within the current range of cultivated 'nuttalliae'. None of the recovered *Chenopodium* seeds show traits indicative of domestication (McClung de Tapia and Aguilar-Hernández, 2001).

Other than the genetic and archaeological data, there isn't any additional information to draw upon. As such, the most parsimonious explanation of the data supports Hypothesis #3: 'jonesianum' was domesticated in ENA and subsequently introduced to Mesoamerica.

Eurasian haplotype- The detection of a Eurasian haplotype among the Cow Ford seed assemblage is not a surprising result. The Eurasian-haplotype is typically associated with *C.*

album, an invasive species of Eurasian origin. This weedy species is an aggressive colonizer currently associated with urban and agricultural settings and seemingly hyper-adapted to human disturbances. Furthermore, *C. album* appears to occasionally hybridize with *C. berlandieri*, conservatively estimated at 5% frequency based on morphological and sequence data (personal observation). These data suggest that *C. album* may have been associated with anthropogenic disturbances many thousands of years ago in both the Americas and in Europe (Stokes and Rowley-Conwy, 2002) and colonized the New World at least 1000 years before the earliest known European contact - Norse sailors between AD 985-1000 (Pálsson and Magnusson, 1965).

Tan colored seeds- Tan seeds have been recovered from many sites in ENA, apparently spanning over three thousand years: 12 Arkansas sites indicated in Fritz (1986), Blood Run site in Iowa (Green and Tolmie, 2004), and Riverton site in Illinois (Smith and Yarnell, 2009). No genetic, geographic, or temporal pattern is evident to account for the tan and dark domesticated *Chenopodium* seeds.

Summary- The findings of this study are not consistent with the findings of Kistler and Shapiro (2011), demonstrating that unique genetic characters in plastid and nuclear markers were detected in both the Mesoamerican and ENA domesticated *Chenopodium*, suggesting a single origin. However, the analysis was not able to determine whether the single origin occurred in Mesoamerica or ENA. The rich 3 800 year archaeological history of domesticated *Chenopodium* in ENA when compared to the relatively recent 500 year history in Mesoamerica, suggests *Chenopodium* was first domesticated in ENA then subsequently introduced to Mesoamerica, as hypothesized by Heiser (1990).

We encourage other researchers to replicate our results, especially with the inclusion of archaeological collections that have not been previously analyzed.

REFERENCES

Alexander, P. J., G. Rajanikanth, C. D. Bacon, and C. D. Bailey. 2006. Recovery of plant DNA using a reciprocating saw and silica-based columns. *Molecular Ecology Notes* 7: 5-9.

Ambrose, S. H. 1987. Chemical and isotopic techniques of diet reconstruction in eastern North America. In *Emergent Horticultural Economies of the Eastern Woodlands*, W. Keegan [ed.], pp. 87-107. Southern Illinois University at Carbondale, Center for Archaeological Investigations.

Charles, T. 2001. The Pumpkin Site: 38GR226, Archaeological Investigation of a Prehistoric Middle Woodland Village in Northern Greenville County, South Carolina". Research Manuscript Series. Book 203. http://scholarcommons.sc.edu/archanth_books/203

Clemants, S. E. and S. L. Mosyakin. 2004. *Chenopodium*. In: Flora of North America Editorial Committee, eds. 1993+. Flora of North America North of Mexico. 16+ vols. New York and Oxford. Vol. 4, pp. 275-299.

Decker, D. S. 1985. Numerical analysis of allozyme variation in *Cucurbita pepo*. *Economic Botany* 42: 4-15.

Decker-Walter, D. S., J. E. Staub, S. M. Chung, E. Nakata, and H. D. Quemada. 2002. Diversity in free-living populations of *Cucurbita pepo* (Cucurbitaceae) as assessed by random amplified polymorphic DNA. *Systematic Botany* 27: 19-28.

Emshwiller, E. 2006. Chapter 8: Genetic data and plant domestication. *In* M. A. Zeder, D. Bradley, E. Emshwiller, and B. Smith [eds.], Documenting Domestication: New Genetic and Archaeological Paradigms, 99-122. University of California Press, USA.

Faulkner, C. T. 1991. Prehistoric Diet and Parasitic Infection in Tennessee: Evidence from the Analysis of Desiccated Human Paleofeces. *American Antiquity* 56: 687-700.

Fritz, G. J. 1986. Prehistoric Ozark Agriculture, the University of Arkansas Rockshelter Collections. Ph.D. dissertation, Department of Anthropology, University of North Carolina at Chapel Hill, USA.

Fritz, G. J. 2000. Levels of native biodiversity in Eastern North America. *In* P. E. Minnis and W. J. Elisens [eds.], Biodiversity in Native America, 223-247. University of Oklahoma Press, Norman, USA.

Gepts, P. and F. A. Bliss. 1985. F1 hybrid weakness in the common bean-Differential geographic origin suggests two gene pools in cultivated bean germplasm. *Journal of Heredity* 76: 447-450.

Giago, T. 2006. *Children left behind: dark legacy of Indian mission boarding schools*. Santa Fe, NM: Clear Light.

Green, W. and C. Tolmie. 2004. Analysis of plant remains from Blood Run. *Plains Anthropologist* 49: 525-542.

Gremillion, K. J. 1996. Early Agricultural Diet in Eastern North America: Evidence from Two Kentucky Rockshelters. *American Antiquity* 61: 520-536.

Harter, A. V., K. A. Gardner, D. Falush, D. L. Lentz, R. A. Bye, and L. H. Rieseberg. 2004. Origin of extant domesticated sunflowers in eastern North America. *Nature* 430: 201-205.

Heiser, C. 1990. New perspectives on the origin and evolution of new world domesticated plants: Summary. *Economic Botany* 44: 111-116.

Hunter, A. A. 1992. Utilization of *Hordeum pusillum* (little barley) in the midwest United States: applying Rindos' co-evolutionary model of domestication. Ph.D. dissertation, University of Missouri-Columbia. Ann Arbor, USA.

Ison, C. R. 2000. Fire on the Edge: Prehistoric Fire Along the Escarpment Zone of the Cumberland Plateau. *In* Proceedings: workshop on fire, people, and the central Hardwood landscape, pp. 36-45. USDA Forest Service General Technical Report.

Jones, V. H. 1936. The vegetal remains of Newt Kash Hollow Shelter. In: Webb, W.S.; Funkhouser, W.D. [eds.] *Rock Shelters in Menifee County, Kentucky*, Reports in Archaeology and Anthropology University of Kentucky, Lexington. 3(4): 147-165.

Jones, J., D. Bosworth, and A. Lonetree. 2011. "American Indian Boarding Schools: An Exploration of Global Ethnic & Cultural Cleansing", Ziibiwing Center of Anishinabe Culture & Lifeways,

Kistler, L. and B. Shapiro. 2011. Ancient DNA confirms a local origin of domesticated chenopod in eastern North America. *Journal of Archaeological Science 38*: 3549-3554.

MacNeish, R. 1970. The Prehistory of the Tehuacan Valley. University of Texas Press, Austin, USA.

Maddison, D. R. and W. P. Maddison. 2005. MacClade. Sinauer Associates, Sunderland, Massachusetts.

Maina, S. L. 1967. Ethnobotany of the Prehistoric Newbridge Site, Lower Illinois River Valley. University of Chicago, Department of Botany, USA.

McClung de Tapia, E. and B. Aguilar-Hernández. 2001. Vegetation and Plant use in Postclassic Otumba. *Ancient Mesoamerica* 12: 113–125.

Schlick, G. and D. L. Bubenheim. 1993. Quinoa: An Emerging "New" Crop with Potential for CELSS. NASA Technical Paper 3422, Ames Research Center, Moffett Field, California, USA.

Smith, B. D. 1984. *Chenopodium* as a prehistoric domesticate in eastern North America: evidence from Russell Cave, Alabama. *Science* 226: 212-257.

Smith, B. D. 1985. *Chenopodium berlandieri* ssp. *jonesianum*: Evidence for a Hopewellian domesticate from Ash Cave, Ohio. *Southeastern Archaeology* 4: 107-133.

Smith, B. D. 1987. The Independent Domestication of Indigenous Seed-Bearing Plants in Eastern North America. *In* W. Keegan [ed.], Emergent Horticultural Economies of the Eastern Woodlands. Southern Illinois University at Carbondale, Center for Archaeological Investigations, *Occasional Paper* 7: 3-47.

Smith, B. D. and C. W. Cowan. 1987. Domesticated *Chenopodium* in Prehistoric Eastern North America: New Accelerator Dates from Eastern Kentucky. *American Antiquity* 52: 355-357.

Smith, B. D. and R. A. Yarnell. 2009. Initial formation of an indigenous crop complex in eastern North America at 3800 B.P. *Proceedings of the National Academy of Sciences* 106: 6561–6566.

Stokes, P. and P. Rowley-Conwy. 2002. Iron Age Cultigen? Experimental return rates for Fat Hen (*Chenopodium album* L.). *Environmental Archaeology* 7: 95-99.

Struever, S. 1968. Flotation Techniques for the Recovery of Small-Scale Archaeological Remains. *American Antiquity* 33: 353-362.

Vavilov, N. I. 1992. Origin and Geography of Cultivated Plants. V. F. Dorofeyev [ed.], Cambridge University Press, Cambridge, USA. (Translated from Russian by Doris Löve.)

Wendel, J. F., P. D. Olson, and J. Stewart,. 1989. Genetic Diversity, Introgression, and Independent Domestication of Old World Cultivated Cottons. *American Journal of Botany* 76: 1795-1806.

Wilson, H. D. 1981. Domesticated *Chenopodium* of the Ozark Bluff Dwellers. *Economic Botany* 35: 233-239.

Wilson, H. D. and C. B. Heiser Jr. 1979. The Origin and Evolutionary Relationships of `Huauzontle' (*Chenopodium nuttalliae* Safford), Domesticated Chenopod of Mexico. *American Journal of Botany* 66: 198-206.

Table 1. DNA sequencing was attempted on domesticated *Chenopodium* seeds recovered from seven archaeological sites. The seeds depicted are from the indicated archaeological sites and to scale relative to each other. For size reference, Cow Ford seeds are approximately 1 mm diameter. The sites are arranged oldest to youngest. This photograph was taken after DNA analyses were completed.

Archaeological Site	Provenience	Site Number	Age Estimate	Lending Museum
Cloudsplitter	Kentucky	15MF36	3450 BP [1]	William S. Webb Museum of Anthropology (KY)
Russell Cave	Alabama	1JA181	1974 BP [2]	Smithsonian National Museum of Natural History (DC)
Edens Bluff	Arkansas	3BE6	1930 BP [3]	The University Museum of University of Arkansas (AR)
Ash Cave	Ohio	33HO1	1720 BP [4]	Peabody Museum of Archaeology and Ethnology Harvard University (MA)
Cow Ford	Arkansas	3BE7	1620 BP [3]	The University Museum of University of Arkansas (AR)
Newbridge	Illinois	11GE456	AD 400-700 [5]	Illinois State Museum (IL)
Holman	Arkansas	3MA34	920 BP [3]	The University Museum of University of Arkansas (AR)

1 - Smith & Cowen 1987 2 - Smith 1984 3 - Fritz 1986 4 - Smith 1985 5 - Personal conumication with Dr. Bonnie Styles excavated the Newbridge site (Illinois State Museum)

Table 2. The primers used to amplify archaeological samples. Each of these three regions of amplification includes one 'nuttalliae'-unique character. The nuclear region also includes a 13 bp indel (presumably a deletion). The two plastid regions are contiguous segments.

Plastid primers

1st segment (Q2)

trnQ_2_A_F	CCC GCT ATT CGG AGG TTC G
trnQ_2_A_R	ACT TGT ATC CGT GTG TSA TTT GAA CAC

2nd segment (Q1)

trnQ_1_AF	ATT CGT TGG AGA TTT AGT CAA GAT GG
trnQ_1_AR	GGA AAA TCA ACK TTC TTT GCC GG

Nuclear primers

SOS1

SOS17_2_AF	GTC AAT GCC ATA TCC AAG CAT TTC
SOS17_2_AR2	TGG GAT ATA TTG TTA TTA AGC CTT TAA

Figure 1. The upper red circle approximately encompasses archaeological sites where *C. berlandieri* subsp. *jonesianum* (jonesianum) has been recovered. There is one site outside of this region, the Pumpkin site (38GR226) located in South Carolina (Charles, 2001). The indicated region is Eastern North America (ENA), and is one of the centers of domestication in North America. The seven archaeological sites from which samples were analyzed are indicated on the map, as well as the estimated age, and the color of the seeds: black, tan, and uncertain due to charring (gray). In addition to the archaeological sites included in this analysis, the Riverton site, IL, the oldest site from which 'jonesianum' has been recovered, has also been included.

The lower red circle encompasses the approximate range of modern *C. berlandieri* subsp. *nuttalliae* (nuttalliae) cultivation. Tan seeded varieties are the most commonly cultivated form in Mesoamerica.

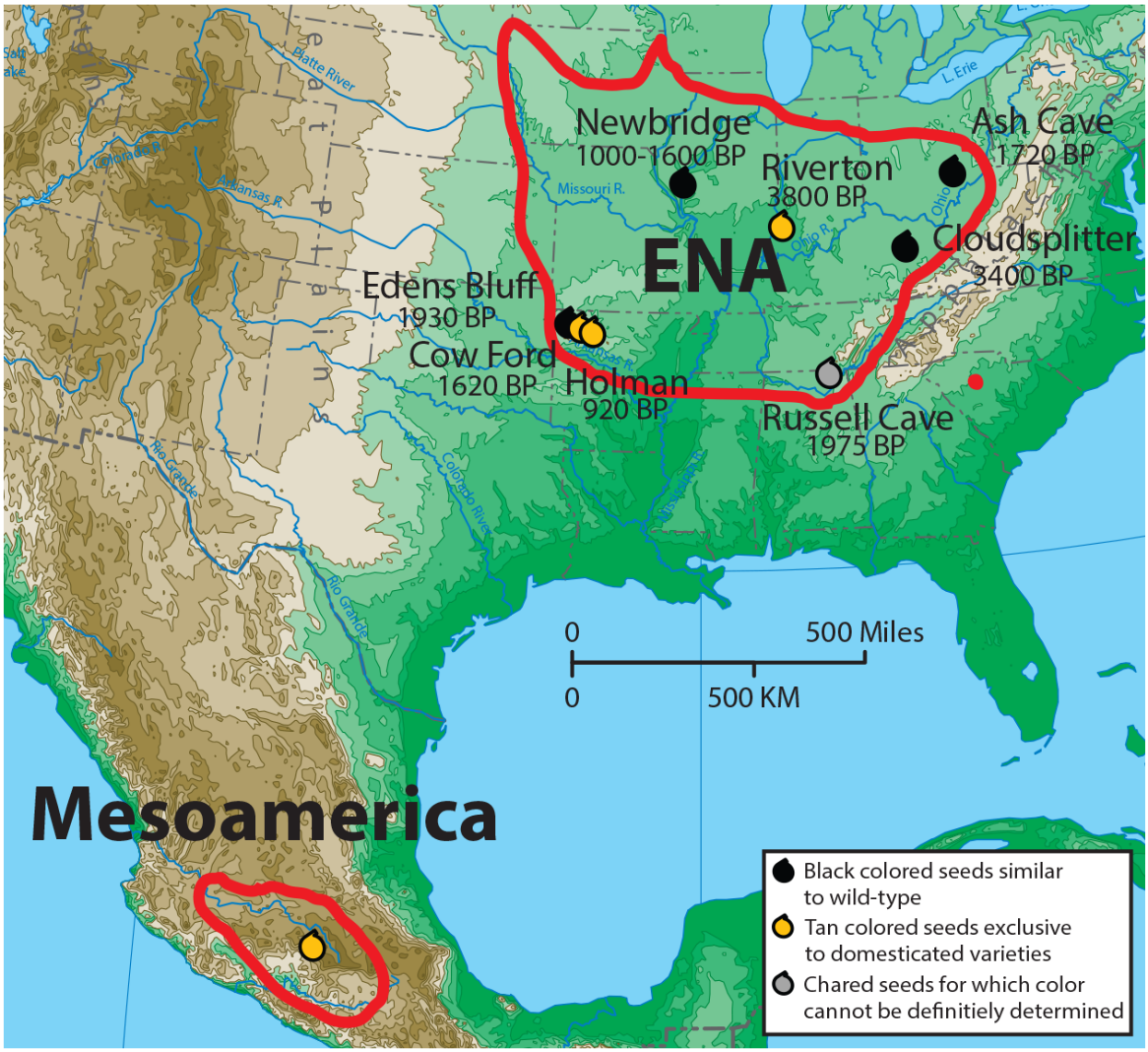
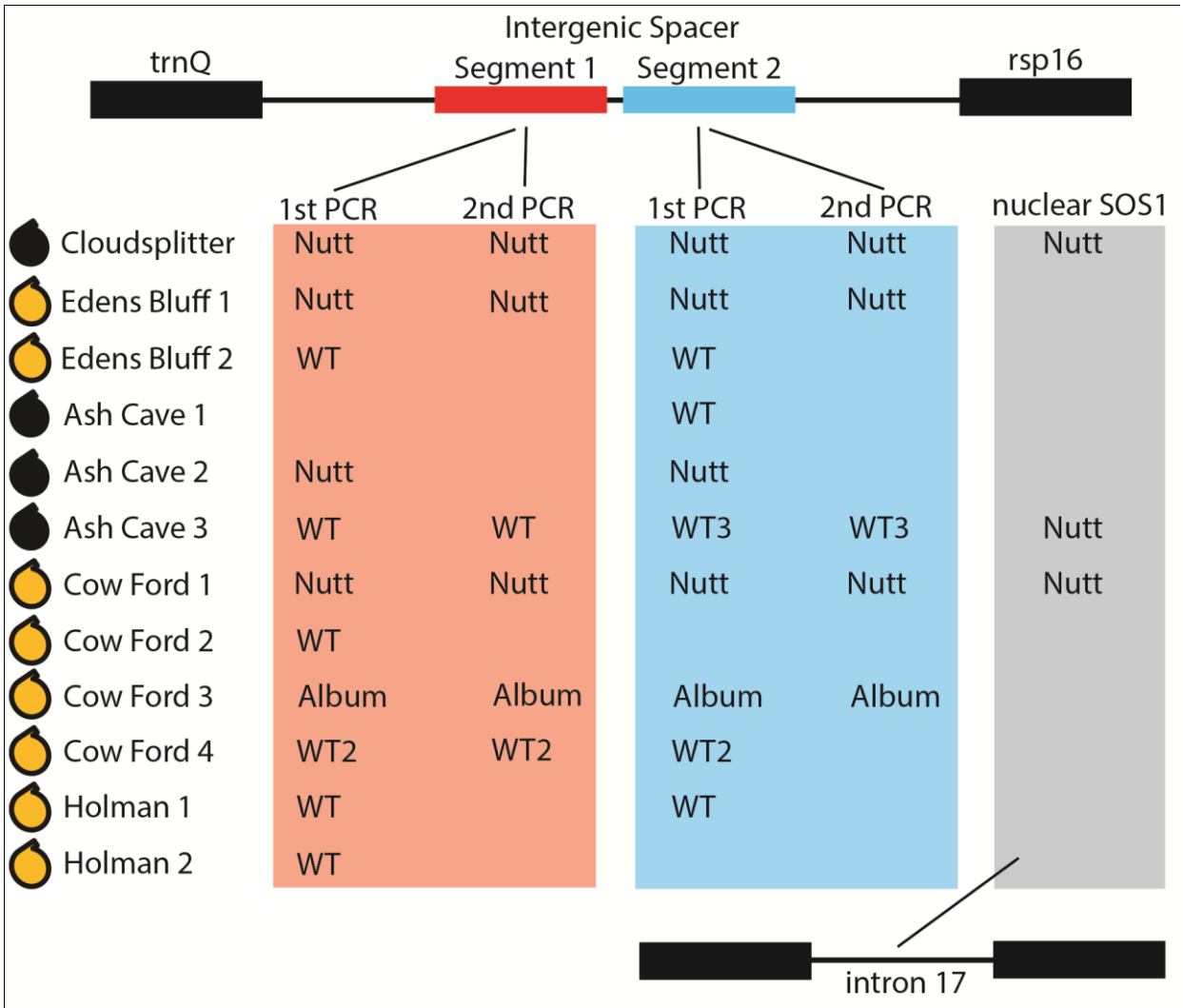

Newbridge
1000-1600 BP
Riverton
3800 BP
Ash Cave
1720 BP
Cloudsplitter
3400 BP
ENA
Edens Bluff
1930 BP
Cow Ford
1620 BP
Holman
920 BP
Russell Cave
1975 BP
Mesoamerica
0
500 Miles
0
500 KM
Black colored seeds similar to wild-type
Tan colored seeds exclusive to domesticated varieties
Chared seeds for which color cannot be definitiely determined

Figure 2. Summary of archaeological seed sequence data. Sequence data was obtained from three regions: two non-overlapping segments of the plastid *trnQ-rsp16* intergenic spacer, and a segment of intron 17 from the single/low-copy nuclear loci, *salt overly sensitive 1* (*SOS1*). Since individual seed samples could not be divided and extracted separately, when possible sequence results were replicated to confirm consistence. In all instances replications were congruent.

Nutt: Indicates 'nuttalliae'-like sequences present in both the plastid and nuclear datasets. The plastid sequences are identical to *C. berlandieri* subsp. *nuttalliae*-1 and 2 in Chapter 1. The three samples successfully sequenced for *SOS1* produced 'nuttalliae'-like sequences.

Album: Indicates a haplotype with a sequence nearly identical to the Eurasian invasive species *C. album*. This and allied species usually demonstrate a high degree of sequence conservation (Chapter 1), so sequence discrepancies are most likely an artifact of poor sequence quality and tend to occur near the ends of the contig.

WT: Indicates the wild-type haplotype for plastid sequences, identical to *C. berlandieri*-1 in Chapter 1. Variants of this sequence were observed, however it is difficult to determine if this reflects actual variation in wild populations, as observed in Chapter 1, or is an artifact of poor sequence quality.



Sites	Seed Attempted	Seeds Sequenced	Summary
Cloudsplitter	6	1	1 Nutt
Edens Bluff	5	2	1 Nutt, 1 WT
Ash Cave	9	3	1 Nutt, 2 WT
Cow Ford	8	4	1 Nutt, 2 WT, 1 Album
Holman	3	2	2 WT

58% wild-type
33% 'nuttalliae'-like
8% eurasian