


1-1-2014

# Rates And Modes Of Sequence Evolution In Various Lineages Within Chenopodiaceae

James Andrew Naeger  
*Wayne State University,*

Follow this and additional works at: [http://digitalcommons.wayne.edu/oa\\_theses](http://digitalcommons.wayne.edu/oa_theses)

 Part of the [Evolution Commons](#), [Molecular Biology Commons](#), and the [Systems Biology Commons](#)

---

## Recommended Citation

Naeger, James Andrew, "Rates And Modes Of Sequence Evolution In Various Lineages Within Chenopodiaceae" (2014). *Wayne State University Theses*. Paper 351.

**RATES AND MODES OF SEQUENCE EVOLUTION IN VARIOUS LINEAGES  
WITHIN CHENOPODIACEAE**

by

**JAMES A. NAEGER**

**THESIS**

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

**MASTER OF SCIENCE**

2014

MAJOR: BIOLOGICAL SCIENCES

Approved by:

---

Advisor

Date

© COPYRIGHT BY

JAMES A. NAEGER

2014

All Rights Reserved

## **DEDICATION**

“When things get bad enough, then something happens to correct the course.”  
–Jonas Salk, “Man Evolving”, 1985

**I dedicate this thesis to Alyson, who lifted me out of despondency and helped me to believe in love again.**

## ACKNOWLEDGMENTS

I would like to thank my parents, Jill and Denis Naeger, for their love and support. My brothers Jeffrey, Joshua, and Joseph have always stood by me. Without the constant support of my family I would not have been able to finish this thesis. My awesome cats Andy and Nicki also provided emotional support and cuddles at key moments. Without the love of my girlfriend, Alyson Schramm, and her parents Ken and Jean, I'm not sure I would have finished this thesis with the same level of psychological stability that I am thought to currently possess.

Since he joined our lab, Nicholas West has been the best lab mate anyone could ask for (although he did introduce me to Reddit, an act which has undoubtedly set me back many months). Many other lab mates and colleagues, including Suba Nagendran, Amapola Balancio, and Scott Medler, have enriched my laboratory experiences and provided insight to me. In the Biological Sciences Department at Wayne State, few could have assisted me more than Louise Dezur, Rose Mary Priest, Linda van Thiel, and Krystyn Purvis. I wouldn't have stuck with Wayne State and graduate school if it hadn't been for their collegiality and encouragement.

Last but not least, I want to thank Ed Golenberg. As an advisor and a friend, his contributions to my success are without equal. I will forever be indebted to him for the kind guidance he has given me.

## TABLE OF CONTENTS

DEDICATION .....	ii
ACKNOWLEDGMENTS .....	iii
LIST OF TABLES .....	v
INTRODUCTION .....	1
RESULTS .....	11
Floral organ morphology & reproductive strategy .....	11
Phylogenetic Analyses .....	40
Rates testing & tests of selection .....	73
DISCUSSION .....	99
CONCLUSIONS.....	111
FUTURE WORK.....	115
REFERENCES .....	124
ABSTRACT.....	134
AUTOBIOGRAPHICAL STATEMENT.....	136

## LIST OF TABLES

Table 1. Identity (%) of sequence pairs from the <i>rbcS</i> alignment in Fig. 2.....	16
Table 2. Identity (%) of sequence pairs from the <i>rbcL</i> alignment in Figure 3. ....	18
Table 3. Identity (%) of sequence pairs from the <i>matK</i> alignment in Fig. 4. ....	20
Table 4. Identity (%) of sequence pairs from the <i>UFO</i> alignment in Fig. 5.....	24
Table 5. Identity (%) of sequence pairs from the <i>UFO</i> alignment in Fig. 6.....	25
Table 6. Identity (%) of sequence pairs from the AGAMOUS alignment in Fig. 8.....	31
Table 7. Identity (%) of sequence pairs from the AGAMOUS alignment in Fig. 9.....	31
Table 8. Identity (%) of sequence pairs from the APETALA3 alignment in Fig. 10.....	34
Table 9. Identity (%) of sequence pairs from the APETALA3 alignment in Fig. 11.....	35
Table 10. Identity (%) of sequence pairs from the PISTILLATA alignment in Fig. 12. ....	38
Table 11. Identity (%) of sequence pairs from the PISTILLATA alignment in Figure 13. ....	39
Table 12. Identity (%) of sequence pairs of concatenated sequences in the alignment from Fig. 14. .....	70
Table 13. Tajima's test statistics for <i>rbcS</i> .....	78
Table 14. Matrix for <i>rbcS</i> test of neutral selection. ....	78
Table 15. Matrix for <i>rbcS</i> test of purifying selection.....	79
Table 16. Tajima's test statistics for <i>rbcL</i> .....	80
Table 17. Matrix for <i>rbcL</i> test of neutral selection. ....	80

Table 18. Matrix for <i>rbcL</i> test of purifying selection. ....	81
Table 19. Tajima’s test statistics for <i>matK</i> . ....	81
Table 20. Matrix for the <i>matK</i> test of neutral selection. ....	82
Table 21. Matrix for the <i>matK</i> test of purifying selection. ....	82
Table 22. Tajima’s test statistics for <i>UFO</i> . ....	83
Table 23. Matrix for <i>UFO</i> test of neutral selection. ....	84
Table 24. Matrix for <i>UFO</i> test of purifying selection. ....	85
Table 25. Tajima’s test statistics for <i>AGAMOUS</i> . ....	86
Table 26. Matrix for <i>AG</i> test of neutral selection. ....	87
Table 27. Matrix for <i>AG</i> test of purifying selection. ....	88
Table 28. Tajima’s test statistics for <i>AP3</i> . ....	89
Table 29. Matrix for <i>AP3</i> test of neutral selection. ....	90
Table 30. Matrix for <i>AP3</i> test of purifying selection. ....	91
Table 31. Tajima’s test statistics for <i>PI</i> . ....	92
Table 32. Matrix for <i>PI</i> test of neutral selection. ....	93
Table 33. Matrix for <i>PI</i> test of purifying selection. ....	94
Table 34. Tajima’s test statistics for the ‘super-gene’ alignment, including <i>rbcL</i> . ....	95
Table 35. Tajima’s test statistics for the ‘super-gene’ alignment, not including <i>rbcL</i> . ....	96
Table 36. Tajima’s test statistics for the ‘super-gene’ alignment that included <i>rbcS</i> , <i>matK</i> , and <i>AG</i> . ....	97



Table 37. Tajima’s test statistics for the ‘super-gene’ alignment which included *UFO*, *AP3*, and *PI*. ..... 98

**LIST OF FIGURES**

Figure 1. Floral morphology of species observed in this study. .... 13

Figure 2. Pairwise alignment of *rbcS* sequences. .... 16

Figure 3. Pairwise alignment of *rbcL* sequences. .... 18

Figure 4. Pairwise alignment of *matK* sequences. .... 20

Figure 5. Pairwise alignment of all *UFO* sequences and alleles..... 23

Figure 6. Pairwise alignment of selected *UFO* sequences..... 24

Figure 7. Protein folding estimation for the *SpUFO* protein . .... 26

Figure 8. Pairwise alignment of *AGAMOUS* sequences from Exon 2 to Exon 7. .... 29

Figure 9. Pairwise alignment of *AGAMOUS* coding regions. .... 30

Figure 10. Pairwise alignment of *APETALA3* sequences. .... 34

Figure 11. Pairwise alignment of *APETALA3* coding regions from Exon 2 to Exon 7..... 35

Figure 12. Pairwise alignment of *PISTILLATA* sequences. .... 38

Figure 13. Pairwise alignment of *PISTILLATA* coding regions, including *PI* alleles and *PI*-like sequences sampled. .... 39

Figure 14. Phylogenetic tree constructed from the *rbcS* alignment using Geneious tree builder. .... 41

Figure 15. Phylogenetic tree constructed from the *rbcS* alignment using PhyML. .... 42

Figure 16. Phylogenetic tree constructed from the *rbcL* alignment using Geneious tree builder. .... 44

Figure 17. Phylogenetic tree constructed from the <i>rbcL</i> alignment using PhyML.....	45
Figure 18. Phylogenetic tree constructed from the <i>matK</i> alignment using Geneious tree builder. ....	47
Figure 19. Phylogenetic tree constructed from the <i>matK</i> alignment using PhyML.....	48
Figure 20. Phylogenetic tree constructed from the <i>UFO</i> alignment from Figure 5, using Geneious tree builder. ....	50
Figure 21. Phylogenetic tree constructed from the <i>UFO</i> alignment from Figure 5, using PhyML. ....	51
Figure 22. Phylogenetic tree constructed from the alignment of selected <i>UFO</i> sequences from Figure 6, using Geneious tree builder. ....	52
Figure 23. Phylogenetic tree constructed from the alignment of selected <i>UFO</i> sequences from Figure 6, using PhyML. ....	53
Figure 24. Phylogenetic tree constructed from the <i>AG</i> alignment from Figure 8 (which considered exon and intron data) using Geneious tree builder. ....	55
Figure 25. Phylogenetic tree constructed from the <i>AG</i> alignment from Figure 8 (which considered exon and intron data) using the Geneious PhyML plugin. ....	56
Figure 26. Phylogenetic tree constructed from the <i>AG</i> alignment from Figure 9 (which considered only coding sequence) using Geneious tree builder. ....	57
Figure 27. Phylogenetic tree constructed from the <i>AG</i> alignment from Figure 9 (which considered only coding sequence) using the Geneious PhyML plugin. ....	58
Figure 28. Phylogenetic tree constructed from the <i>AP3</i> alignment from Figure 10 (which included exon and intron data) using Geneious tree builder.....	60

Figure 29. Phylogenetic tree constructed from the <i>AP3</i> alignment from Figure 10 (which included exon and intron data) using the Geneious PhyML plugin.....	61
Figure 30. Phylogenetic tree constructed from the <i>AP3</i> alignment from Figure 11 (which included only coding regions) using Geneious tree builder.....	62
Figure 31. Phylogenetic tree constructed from the <i>AP3</i> alignment from Figure 11 (which included only coding regions) using the Geneious PhyML plugin.....	63
Figure 32. Phylogenetic tree constructed from the <i>PI</i> alignment from Figure 12 (which included exon & intron data), using Geneious tree builder. ....	65
Figure 33. Phylogenetic tree constructed from the <i>PI</i> alignment from Figure 12 (which included exon & intron data) using the Geneious PhyML plugin. ....	66
Figure 34. Phylogenetic tree constructed from the <i>PI</i> alignment from Figure 13 (exon and 3'UTR data only) using Geneious tree builder. ....	67
Figure 35. Phylogenetic tree constructed from the <i>PI</i> alignment from Figure 13 (exon and 3'UTR data only) using the Geneious PhyML plugin. ....	68
Figure 36. Pairwise alignment of concatenated 'super-gene' sequences.....	70
Figure 37. Phylogenetic tree constructed from the 'super-gene' alignment from Figure 14 using Geneious tree builder. ....	71
Figure 38. Phylogenetic tree constructed from the 'super-gene' alignment from Figure 14 using the Geneious PhyML plugin. ....	72
Figure 39. Horizontal cladogram of characters along evolutionary branches in <i>Anserineae</i> . ....	114

## INTRODUCTION

The relationship between the rates of morphological change and sequence evolution is a central question in the study of molecular evolution. Do species that show long-term morphological stasis have reduced mutation rates and lower levels of polymorphisms?[1] Do clades with extensive morphological differentiation show evidence of overall increases in rates of evolution and an accumulation of mutations across the genome or do they simply reflect saltational effects of particular mutations?[2-4] Do indel mutations or chromosomal translocations that are outside of models of nucleotide substitution occur in a clocklike manner and do they affect morphological change?[5, 6]

We do not expect that the answers to these questions will be universal. Rates of sequence divergence must ultimately be limited by basal mutation rates, but these can vary over time and genomic space. Functional constraints can further filter what we can detect over time, thereby causing different domains to vary in their apparent rates of divergence. More broadly speaking, it is long acknowledged that demographic and genetic factors cause deviations from a common driving rate of evolution [7, 8]. From a population perspective, subdivided populations that allow interdeme competition may drive shifts from one evolutionary equilibrium to another, causing apparent rapid transitions in characters following Wright [9]. Our increasing knowledge in evolutionary development supported by developmental genetics demonstrates that single gene mutations can result in substantial morphological divergence echoing thoughts of Goldschmidt [10]. Alternatively, divergent mutations that isolate individuals, such as polyploidy, can cause rapid speciation without concomitant phenotypic divergence. Therefore, to address these effects, comparative studies of genetic divergence should be addressed in groups of related taxa where morphology, reproductive strategies, and population demographics vary and can be tracked.

The genus *Spinacia*, from which cultivated spinach is derived, appears to originate from areas in northern Iran and near the Caspian Sea. According to different sources, the genus includes three species, *Spinacia oleracea*, cultivated spinach, and two wild species, *Spinacia tetrandra* and *Spinacia turkestanica*. The relationship among these taxa is not clear, and the species distinctions are not universally accepted. All three taxa are dioecious with flowers developing as unisexual from initiation. Rarely, monoecious individuals develop, and these have often been used to generate inbred populations in establishing cultivated varieties. The written history of spinach cultivation is sparse, and only goes back to the 6th century, although clearly the plant had been utilized for food prior to that time. Therefore, *Spinacia sp.* serve as an excellent system in which to test rates of sequence evolution as they might be affected by morphological differentiation, mating system divergence, and recent domestication and cultivated variety development.

The genus *Spinacia* belongs to the Family Chenopodiaceae. Molecular phylogenetic studies based on chloroplast data indicate that it clusters with *Chenopodium foliosum*, *Monolepis nuttalliana*, *Chenopodium bonus-henricus*, and *Scleroblitum atriplicinum*. The exact phylogenetic relationships, however, are not consistent when different genes are used, even among chloroplast-encoded genes [11-15]. Additionally, the species within this clade vary greatly in their geographical distribution (Australia, Central Asia, Europe, and North America) and in floral morphologies (variable organ number) and reproductive systems (protogyny, monoecy and dioecy). Consequently, comparison of sequence evolution among these related species may shed light on the relationship of floral and reproductive strategy evolution and the rates and modes of genetic divergence.

In this study, we investigate the rates and modes of sequence evolution of seven genes, two of which are chloroplast encoded, one nuclear encoded but whose protein is functional in the chloroplast, three nuclear genes that are floral organ identity genes, and one that is required for floral development. We test whether reproductive strategy through the development of dioecy, monoecy, and protogyny, is reflected differentially among floral developmental versus non floral developmental genes, and whether recent domestication leaves noticeable traces in the phylogenetic relationship between *Spinacia oleracea* and its wild relatives, *Spinacia tetrandra* and *Spinacia turkestanica*. The resultant phylogenetic predictions strongly and consistently support the sister relationship of *Spinacia sp.* to a *Blitum* clade consisting of *Chenopodium bonus-henricus*, *Chenopodium foliosum*, and *Monolepis nuttalliana*. Relative rates tests indicate a generally faster rate of nucleotide substitutions within *Spinacia*. This difference is persistent when only genes implicated in male unisexual floral development are analyzed. In addition, insertion/deletion (indel) events occur more prominently within the *Spinacia* clade and occur in both coding and intron regions and are consistent with a more dynamic nucleotide sequence evolution in this clade. Indeed, the evidence for purifying selection in *Spinacia* suggests that the increased nucleotide substitution rates are not driving protein evolution, in contrast to clear evidence of protein sequence and structure evolution driven by indels. There is no discernible footprint of domestication on sequence evolution, and indeed, we cannot detect phylogenetic signals that would support separation of the *Spinacia* accessions into three distinct taxa.

## MATERIALS AND METHODS

### *Gene sampling*

We selected seven genes to isolate and analyze their sequence evolution in this study. They sample both chloroplast and nuclear genomes and vary in encoding genes involved in floral development, photosynthesis, and gene maturation. The two chloroplast genes in this study are *rbcL* and *matK* both chosen due to their extensive inclusion in phylogenetic studies [14, 16-25], and hence their large available databases. *rbcL* encodes for the large subunit of the ribulose-1, 5-bisphosphate carboxylase/oxygenase (RuBisCO) complex, which is involved in carbon fixation during photosynthesis. The *rbcL* gene is reported to have relatively low rates of nucleotide substitution but this rate varies widely among species [26, 27]. *matK* is the second chloroplast-encoded gene used in this study. Alongside *rbcL*, the *matK* gene has been recommended as part of a 2-locus DNA barcode for land plants by the Consortium for the Barcode of Life (CBOL) Plant Working Group [28]. It is encoded with the chloroplast *TRNK* intron and is reported to evolve with faster rates of nucleotide substitution relative to *rbcL* [17, 18, 25]. The third gene among the chloroplast/photosynthesis genes selected is *rbcS*. While located and expressed from the nucleus, it encodes the small subunit of the ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) complex and therefore controls carbon fixation functions within the chloroplast. This nucleary-located gene would presumably experience similar selective pressures as the chloroplast-located *rbcL* gene, which encodes the large subunit of the RuBisCO complex.

We selected four additional nuclear genes to study based on their involvement in floral development, and hence their potential role in the evolution of floral divergence in this group of species. These genes are *UFO* (*UNUSUAL FLORAL ORGANS*), *AG* (*AGAMOUS*), *AP3* (*APETALA3*), and *PI* (*PISTILLATA*). The *UFO* gene encodes an F-box protein that is thought to

be required for floral organ development [29-31]. Other F-box proteins act to target proteins for ubiquitination and subsequent degradation, and *UFO* is thought to regulate the products of the B-class MADS genes *AP3* and *PI* [32-37]. The remaining three genes selected are MADS box transcription factors and are known as floral organ identity genes. *AGAMOUS (AG)* is a C-class MADS-box gene with suggested functional roles in both stamen and ovary development [38-40]. Its expression is limited to developing floral primordia [41]. *APETALA3 (AP3)* is a B-class floral organ identity gene required for normal organ number, whorl development [42-45], and has been suggested to play a central role in sex determination in spinach [46, 47]. *PISTILLATA (PI)* is also a B-class floral organ identity gene. Its protein forms heterodimers with *AP3* protein to generate a transcription factor complex [48-50]. *PI*, along with *AP3* is expressed in spinach early in male floral development but not in female floral development [47]. Regulation of *AP3* and *PI* expression appears to be central to sex determination in spinach [47, 51]; hence the evolution of their sequences may be correlated with the evolution of reproductive strategies and structures within the species studied herein.

### ***Taxon sampling & plant growth***

We sampled the closest relatives to cultivated spinach, *Spinacia oleracea*, starting with a published phylogeny based on the *rbcL* sequence [14]. We sampled *Chenopodium foliosum* (syn. *Blitum virgatum*), *Chenopodium album*, *Chenopodium bonus-henricus* (syn. *Blitum bonus-henricus*), and *Monolepis nuttalliana* (syn. *Blitum nuttalliana*) (Tables 1A, 1B). We also included *Spinacia tetrandra* and *Spinacia turkestanica*, the prickly-seeded wild relatives of cultivated spinach. Where possible, downloaded sequences for *Beta vulgaris* and *Mesembryanthemum crystallinum* were obtained and used as more distant outgroups. Seeds were obtained through the USDA National Plant Germplasm System, germinated in Miracle



Grow Potting Mix, and plants were grown in a Conviron Adaptis A1000 growth chamber ([www.conviron.com](http://www.conviron.com)) at periods of 14-h day / 10-h night at a constant 20 deg Celsius.

### ***DNA isolation, sequence amplification and sequencing***

Fresh leaf tissue was ground to powder with liquid nitrogen and genomic DNA was extracted using the Qiagen DNeasy Plant Mini Kit according to the manufacturer's instructions ([qiagen.com](http://qiagen.com)). In order to obtain some sequences it was necessary to perform a 3' RACE. To isolate RNA from flowers, floral tissue was ground to powder with liquid nitrogen and whole RNA was extracted using the Qiagen RNeasy Plant Mini Kit and the Direct-zol™ RNA MiniPrep with Trizol® ([zymoresearch.com](http://zymoresearch.com)). The RNAs were analyzed using a Nanodrop spectrophotometer ([nanodrop.com](http://nanodrop.com)). cDNA synthesis was performed using the oligo(dT) primer AP Short (Appendix 'A'), AMV Reverse Transcriptase ([promega.com](http://promega.com)), and recombinant RNasin® according to the manufacturer's instructions.

Two degenerate primers were used to amplify the near-full length *rbcS* sequence, based on the known *rbcS* sequences of *Spinacia oleracea* and *Glycine max* (soybean, GMU39567). For the *UFO* gene, a previously unpublished sequence obtained from our lab for *Spinacia oleracea UFO* was used as a starting point for primer design for the Chenopods. A nested PCR strategy was used with degenerate primers designed against the known *Spinacia UFO* sequence and the published *UFO* sequences of *Medicago truncatula* (XM\_003638251) and *Arabidopsis thaliana* (NM\_102834). Following sequencing of these products, additional specific primers were designed for each species in order to perform 3' RACE. For the *AGAMOUS* gene, two degenerate primers were used to amplify a region spanning the L region and K box, from exon 2 to exon 5. These primers were designed based on an alignment of known coding sequences from some core eudicots: *Spinacia oleracea* (AY660007), *Brassica juncea* (DQ060334), *Vitis vinifera*

(GU133631), *Vicia sativa* (JF313850), and *Nicotiana benthamiana* (JQ699177). These primers returned a single band for each species, each ranging in size from around 550 to 600 base pairs in length when genomic DNA was amplified. Additionally, more primers were designed to amplify the region spanning Exon 3 to Exon 7. Degenerate reverse primers for Exon 7 were designed using the *AGAMOUS* sequence of *Spinacia* and *Arabidopsis thaliana*. These primers returned band sizes ranging from 900 to 2,000 bp. Larger bands required primer walk sequencing and *C. bonus-henricus* required specific primers to be designed and used to confirm sequencing results. To amplify the *AP3* gene, a previously published sequence from our lab for *Spinacia oleracea* (cv America) *AP3* (AY604514.1) was used as a starting point for primer design for the rest of the Chenopods. The *AP3* sequence for *Silene vulgaris* (HQ113124) was also used for protein-based degenerate primer design. Additional primer-walk sequencing was used to obtain most of the *AP3* sequences in our study due to the length of the introns. The c-terminal domain of the *PI* gene (containing the *PI*-motif) was sampled for all the species of *Spinacia* (male and female). A previously published sequence from our lab for *Spinacia oleracea* (America cultivar) *PI* (AY604515.1) was used as a starting point for primer design for the rest of the Chenopods. Degenerate primers reaching from exon 1 to exon 4 were designed based on a nucleotide alignment of the exon sequence for *S. oleracea* (GQ120477), *Vitis vinifera* (DQ988043), *Hydrangea macrophylla* (AB454441), *Brassica juncea* (DQ060333), and *Passiflora edulis* (JN087005). We used a combination of degenerate primer design, 3' RACE, and primer walk sequencing to obtain the remainder of the sequences (including 3' UTR). All of these primers are given in Supplementary Material, Table 1.

PCR products were purified using the Promega Wizard® SV Gel and PCR Clean-Up System ([promega.com](http://promega.com)) and cloned using the Promega pGEM®-T Easy Vector System.

Plasmids were sequenced by Eton Biosciences ([etonbio.com](http://etonbio.com)) using an Applied Biosystems 3730xl DNA Analyzer. When feasible (known homozygotes and when non-degenerate primers were used to generate an amplicon) PCR products were sent to Eton for cleanup and direct sequencing without cloning.

### ***Sequence Assembly, alignment, motif analysis, and protein-folding estimation***

For each sample, we generated a single consensus sequence by assembling overlapping forward and reverse reads using Geneious (version 6.1.7, created by Biomatters, available from [geneious.com](http://geneious.com)). All reads were initially trimmed of vector sequences using the UniVec database, and trimmed of low-quality flanking regions, which included the amplifying primer sequences. Where disagreements occurred between reads, the sequences were inspected visually and a base was called appropriate to the highest quality read(s) in the chromatograms. Finally, the consensus sequence was analyzed and annotated for putative coding vs. non-coding regions. All sequences have been submitted to GenBank and their accession numbers are listed in the Appendix (Supplemental Tables 2-4).

Pairwise alignment was performed with the MAFFT plugin (version 1.3) for Geneious using auto algorithm selection, PAM-200 (k=2) scoring matrix, and a gap open penalty of 1.53. The alignments were then visually scrutinized for accuracy. For some sequences, it was prudent to conduct two or more separate alignments. For some, it is instructive to include the sequences obtained from all *Spinacia sp.* samples and for others these identical sequences are left out for ease of viewing. When the putative coding sequences are annotated, extracted alignments are shown separately in order to compare the effect of intron variation on sequence alignment. The variation in the size of the introns for the *PI* gene made alignment using MAFFT auto algorithm selection problematic, so for this gene a better alignment was generated using the E-INS-i

algorithm. The E-INS-i algorithm is slow and best used for sequences with multiple conserved domains and long gaps [52].

Areas of sequence conservation seen in the alignments were analyzed for motifs using PLACE: the plant cis-acting regulatory DNA elements database [53]. Nucleotide sequences were translated in Geneious and exported as amino acid sequences. Protein structures were estimated using the Phyre2 program [54] using default and ‘intense’ conditions.

All of the sequences for a given species were concatenated [55] and these ‘super-genes’ were aligned using the MAFFT plugin (version 1.3) for Geneious using the E-INS-i algorithm, PAM-200 (k=2) scoring matrix, and a gap open penalty of 1.53.

### ***Phylogenetic analysis***

Based on the MAFFT alignments, phylogenetic trees were constructed using Geneious Tree Builder. We used the Tamura-Nei genetic distance model, the neighbor-joining tree build method, the bootstrap resampling method with 10,000 replicates, and support threshold set to 85%. We also used the Geneious PhyML plugin which uses a maximum-likelihood method [56], using the Tamura-Nei substitution model, bootstrap branch support method with 10,000 replicates, taking the best of both NNI and SPR topology searches, and all other settings at default. For some sequences that are identical or nearly so, hierarchical relationships could not be reliably estimated. A representative sample was used for the final phylogenetic analysis and the remaining sample sequences were excluded. When an alignment of the genomic sequences included all taxa was used, a phylogenetic tree was also generated using both methods described above in order to assess the effect the intron sequences had on phylogenetic inference. When available, *Beta vulgaris* was used as the outgroup within the Chenopodiaceae and

*Mesembryanthemum crystallinum* as the distant outgroup; otherwise, *Chenopodium album* was used as the ‘close outgroup to the other species under study here.

### ***Rates testing and tests of selection***

Rates of evolution were examined and Z tests for positive, neutral, and purifying selection were performed using MEGA version 5.1 [57]. The MAFFT alignments were exported from Geneious into MEGA format, and genes/domains were selected for the exons/introns within the MEGA program. A relative rate test (Tajima’s) was done for each ingroup species, compared to *Spinacia sp.* with *C. album* as the outgroup taxon. All nucleotide substitutions were considered at all codon positions and noncoding sites, with complete deletion of gaps/missing data in the alignment. We also performed codon-based Z-tests of selection for each test hypothesis (neutrality, positive selection, and purifying selection) for each gene in this study. We used the analytical method of variance estimation, the Nei-Gojobori (proportion) setting and completely deleted any gaps/missing data.

## RESULTS

### Floral organ morphology & reproductive strategy

The species sampled in this study show considerable variation in their floral morphologies. The species vary in their organ number, timing of organ maturation, and diversity of floral types presented on individual plants. This variation has obvious effects on reproductive strategy.

*Chenopodium album* has hermaphroditic flowers (Figure 1a, 1b). The flower is radially symmetrical and contains three concentric whorls of organs. In the first or outermost whorl, there are five sepals, which appear to be fused but distinct prior to dehiscence. The sepals are all attached below the gynoecium (hypogynous) and are covered with mealy structures (farinose). There are no petals, petaloid organs or primordia in the position of a second whorl. The five stamens form opposite and interior to each sepal. There is a single central pistil with a single ovule and two distinct stigmas. The gynoecia of *C. album* have two carpels that are fused at the base (syncarpous), and the flowers appear in condensed cymes (cymules) in the leaf axils [58, 59].

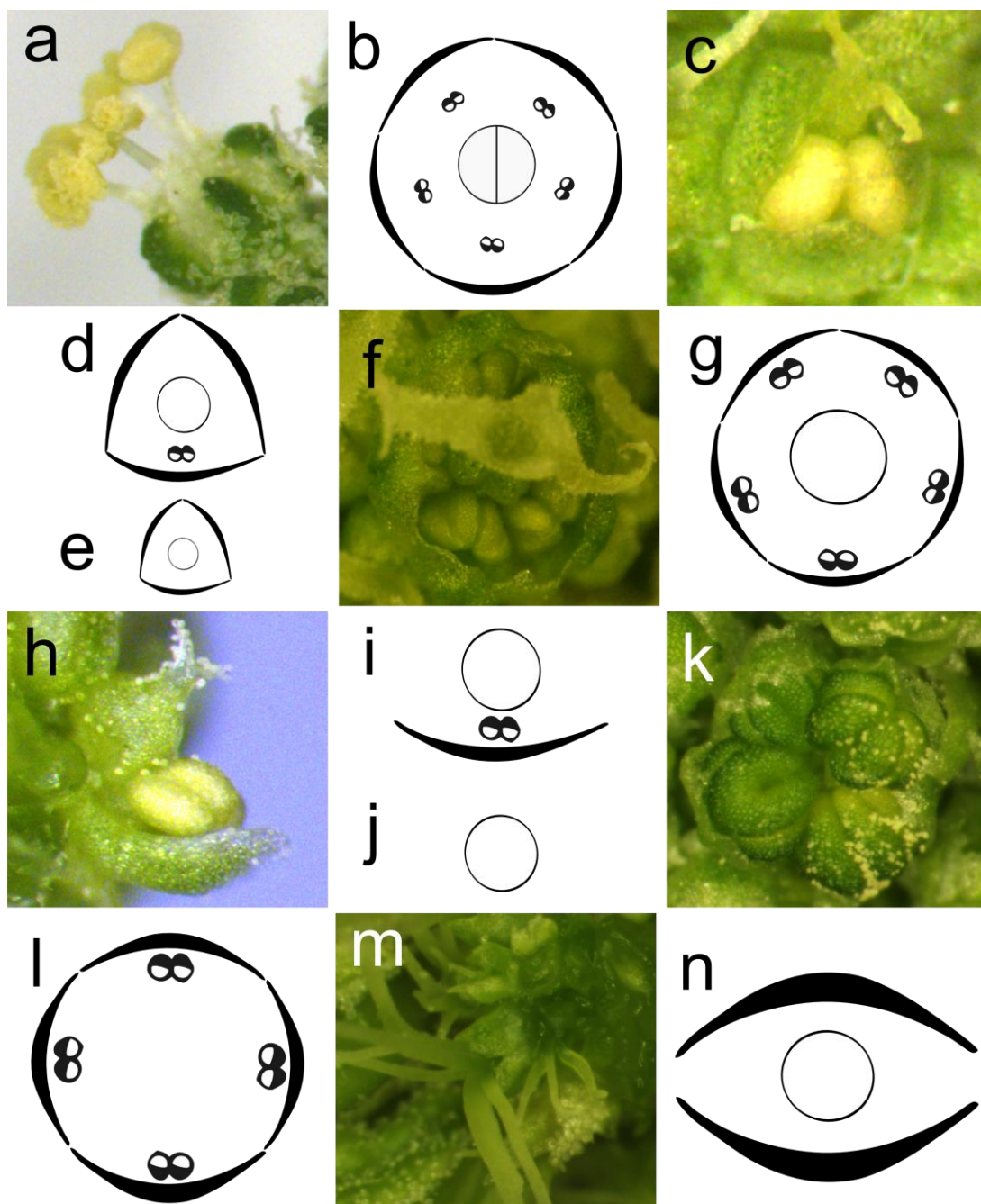
*Chenopodium foliosum* (syn. *Blitum virgatum*) [12] appears to have hermaphroditic flowers as well as unisexual female flowers (gynomonoecy) within a single inflorescence (Figure 1c, 1d). The flowers have 3 non-farinose sepals in the outermost whorl attached below the gynoecium (hypogynous). Each hermaphroditic flower has a single stamen carrying a four-lobed anther, and a single pistil with two stigmas. Female flowers have a single central pistil and no evidence of stamen. Hermaphroditic flowers were dissected and confirmed to be hypogynous.

*Chenopodium bonus-henricus* (syn. *Blitum bonus-henricus*) [12] has hermaphroditic flowers with a single central pistil (Figure 1f, 1g), which is presented above the sepals prior to

the exposure or dehiscence of the stamens (sequential hermaphroditism, protogynous). In the first or outermost whorl there are five distinct sepals, which are all attached below the gynoecium (hypogynous), which do not appear to be farinose. Stamens form opposite and interior to each sepal and each stamen has four-lobed anthers. There is a single central pistil with two distinctive stigmas, which protrude from underneath the sepals before the stamens are fully developed.

*Monolepis nuttalliana* (syn. *Blitum nuttalliana*) [12] has both hermaphroditic flowers as well as unisexual female flowers (gynomonoecy) within a single inflorescence (Figure 1h, 1i, 1j). Each hermaphroditic flower has a single sepal, a single stamen with a four-lobed anther, and a single pistil carrying two distinct stigmas. The sepal is attached below the gynoecium (hypogynous) and is not farinose.

*Spinacia oleracea* is a Type II dioecious species, where unisexual flowers develop without organ abortion [51]. Male flowers have four sepals (Figure 1k, 1l), which develop from sequential appearance of two pairs of sepal primordia. There are four stamens that develop opposite the four sepals. The central (fourth whorl) region shows no evidence of organ primordia formation [60]. Female flowers have two sepals (Figure 1m, 1n). As with the previously described species, there are no primordia for petals. There are also no stamen primordia in the third whorl. The fourth whorl has a single primordium around which the gynoecial girdle forms and develops into the ovary wall. The central region develops into a single ovule [60]. Rarely, flowers of the opposite sex may form on otherwise male or female plants, resulting in sporadic monoecy.



**Figure 1. Floral morphology of species observed in this study.**

*Chenopodium album* flower lateral view (a) and diagram (b). *Chenopodium foliosum* (syn. *Blitum virgatum*) hermaphroditic flower superior view (c) and diagram (d), female flower diagram (e). *Chenopodium bonus-henricus* (syn. *Blitum bonus-henricus*) flower superior view (f) and diagram (g). *Monolepis nuttalliana* (syn. *Blitum nuttalliana*) hermaphroditic flower lateral view (h) and diagram (i), female flower diagram (j). *Spinacia oleracea* male flower, superior view (k) and diagram (l) and *Spinacia oleracea* female flower, lateral view (m) and diagram (n).



## Gene Sequence Structure & Evolution

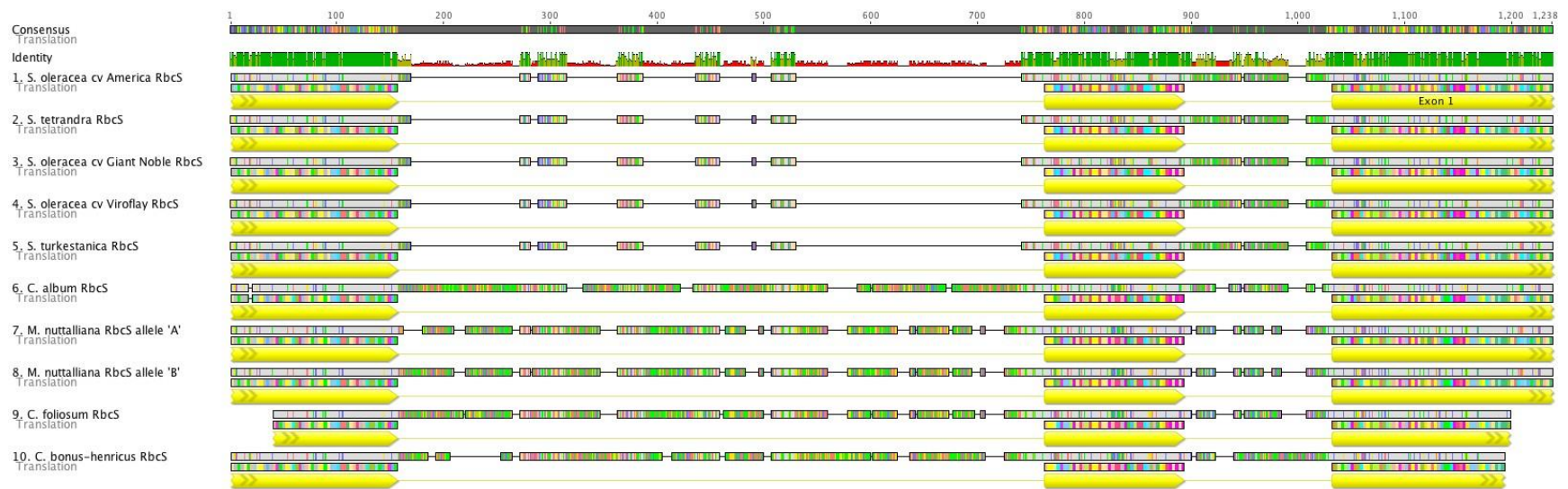
### *rbcS*

The *rbcS* gene sequences obtained are listed in Supplemental Table 2. As seen in Figure 2, this gene has three exons and 2 introns. Primer sequences were trimmed from the final alignment, so most sequences have partially missing data for the beginning of exon 1 and the end of exon 3. However, the data generated does not indicate any insertion/deletion events in the exon regions. For all species, exon 1 is slightly larger than 156 bp in length, exon 2 is exactly 132 bp in length, and exon 3 is slightly larger than 207 bp in length. Intron 1 is where we find most of the variation due to insertions/deletions; intron 1 for *C. album* is 548 bp in length whereas for *S. oleracea cv America* it is just 150 bp. Intron 2 lengths are more similar; for *C. album* is 101 bp in length whereas for *S. oleracea cv America* it is 119 bp.

There are four regions of conservation seen in the first *rbcS* intron and these were analyzed for motifs using the PLACE database [53]. The first conserved region has similarity with many 'GATA motifs, as well as GT-1 binding sites and 'I-box' motifs. The GATA box is required for high-level, light regulated, and tissue specific expression [61]. The consensus GT-1 binding site is found in many light regulated genes and in *rbcS* from many species. The I-box motif is found upstream of many light-regulated genes [62]. We also see a consensus GT-1 binding site at the second conserved region in spinach, and a GATA box in this region in *C. album*. Another GATA box is seen in the third conserved region in the spinach sequence. The fourth conserved region has a GATA box and a consensus GT-1 binding site.

One species, *M. nuttalliana*, returned two sequences for this gene. One of the two sequences differed by a 17-bp AT-rich insertion in the first intron, by a second-position non-synonymous substitution in exon 3, and by a third-position synonymous substitution in exon 3.

The *rbcS* sequences (including both exon and intron regions) for each of the spinach species and cultivars are 99.6% identical (Table 1). As expected, there were no insertion/deletion events in the coding sequences obtained for this gene when comparing all species. There is a high degree of conservation in the exon sequences for this gene but a high amount of intron variation between species (both sequence length and identity). Within the exon regions, sequence identity ranged from 84.0% between *C. album* and *S. oleracea cv America* to 96.9% between *C. foliosum* and *M. nuttalliana*. Intron sequences varied considerably among all the species, both in size and sequence. For example, intron 1 is much smaller in *Spinacia sp.* than in the other Chenopods. The intron regions of the alignments must be considered with some caution due to extensive sequence variation. Among all species there was a range of sequence identities (when considering both exons and introns), from 48.2% identity between *Spinacia sp.* and *C. album* to 88.6% identity between *M. nuttalliana* and *C. foliosum*. Given the high exon sequence identities, this range of values is mainly driven by the intron variation.



**Figure 2. Pairwise alignment of *rbcS* sequences.**

Alignment was performed with the MAFFT plugin (version 1.3) for Geneious using auto algorithm selection, PAM-200 scoring matrix ( $k=2$ ), and a gap open penalty of 1.53. Exons are annotated with yellow arrows indicating directionality, while connecting lines represent introns. Percentage sequence identity is given, from green (100% identity) to black (0%).

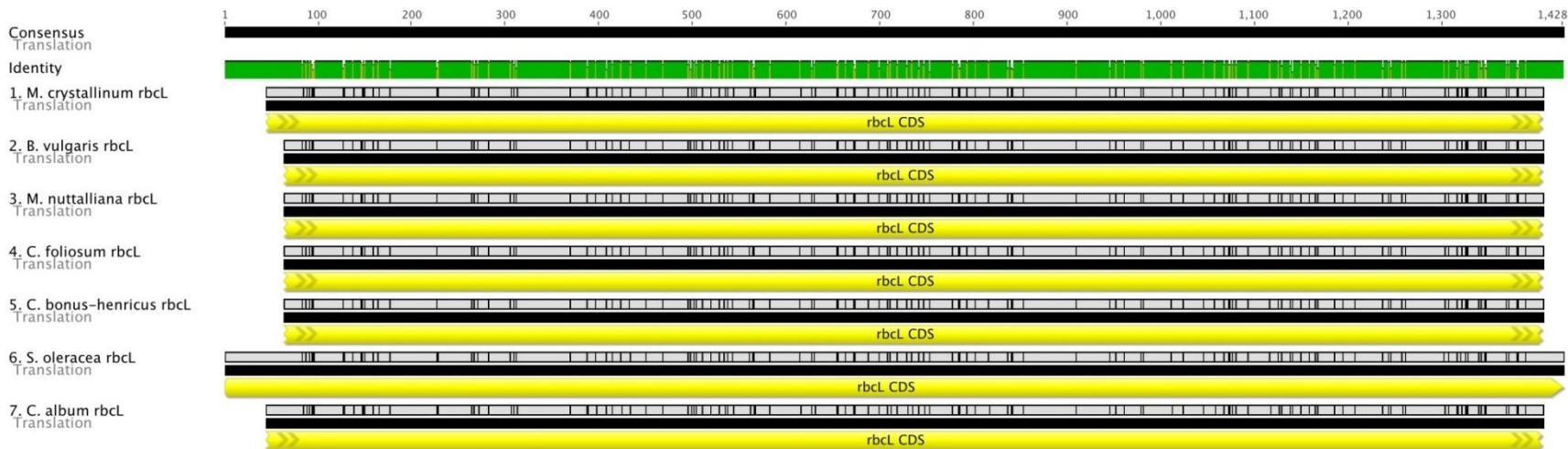
	<i>S. oleracea</i> cv America	<i>S. tetrandra</i>	<i>S. oleracea</i> cv Giant Noble	<i>S. oleracea</i> cv Viroflay	<i>S. turkestanica</i>	<i>C. album</i>	<i>M. nuttalliana</i> RbcS allele 'A'	<i>M. nuttalliana</i> RbcS allele 'B'	<i>C. foliosum</i>	<i>C. bonus-henicus</i>
<i>S. oleracea</i> cv America RbcS	-	-	-	-	-	<b>84</b>	<b>85.1</b>	-	<b>85.6</b>	<b>85.8</b>
<i>S. tetrandra</i> RbcS	99.6	-	-	-	-	-	-	-	-	-
<i>S. oleracea</i> cv Giant Noble RbcS	99.7	99.9	-	-	-	-	-	-	-	-
<i>S. oleracea</i> cv Viroflay RbcS	99.6	99.7	99.9	-	-	-	-	-	-	-
<i>S. turkestanica</i> RbcS	99.6	99.7	99.9	99.7	-	-	-	-	-	-
<i>C. album</i> RbcS	48.3	48.4	48.5	48.4	48.4	-	<b>87.3</b>	-	<b>87.3</b>	<b>87.6</b>
<i>M. nuttalliana</i> RbcS allele 'A'	49	49.1	49.2	49.2	49.1	60.4	-	-	<b>96.9</b>	<b>96</b>
<i>M. nuttalliana</i> RbcS allele 'B'	49	49.1	49.1	49.1	49.1	61.1	98.1	-	-	-
<i>C. foliosum</i> RbcS	46.1	46.1	46.1	46.2	46.1	60.7	86.9	88.6	-	<b>96.4</b>
<i>C. bonus-henicus</i> RbcS	48.7	48.9	48.9	48.8	48.8	58.3	77	78.6	78.8	-

**Table 1. Identity (%) of sequence pairs from the *rbcS* alignment in Fig. 2.**

Percentages above the diagonal (in bold) represent identity when only considering exon sequences from select species.

***rbcL***

We analyzed the published *rbcL* sequences listed in Supplemental Table 2. As seen in the alignment shown in Figure 3, this chloroplast gene contains no introns and the data does not indicate the presence of any insertion/deletion events. For all species, the gene appears to be approximately 1.4 kb in length. This gene has an extremely high degree of conservation observed, from 93.3% identity between *Spinacia sp.* and *M. crystallinum* to 99.2% identity between *C. bonus-henricus* and *C. foliosum* (Table 2). The only variations seen between species for this gene are in the form of single nucleotide substitutions. Of the 1428 sites compared, only 91 sites varied among the species.



**Figure 3. Pairwise alignment of *rbcL* sequences.**

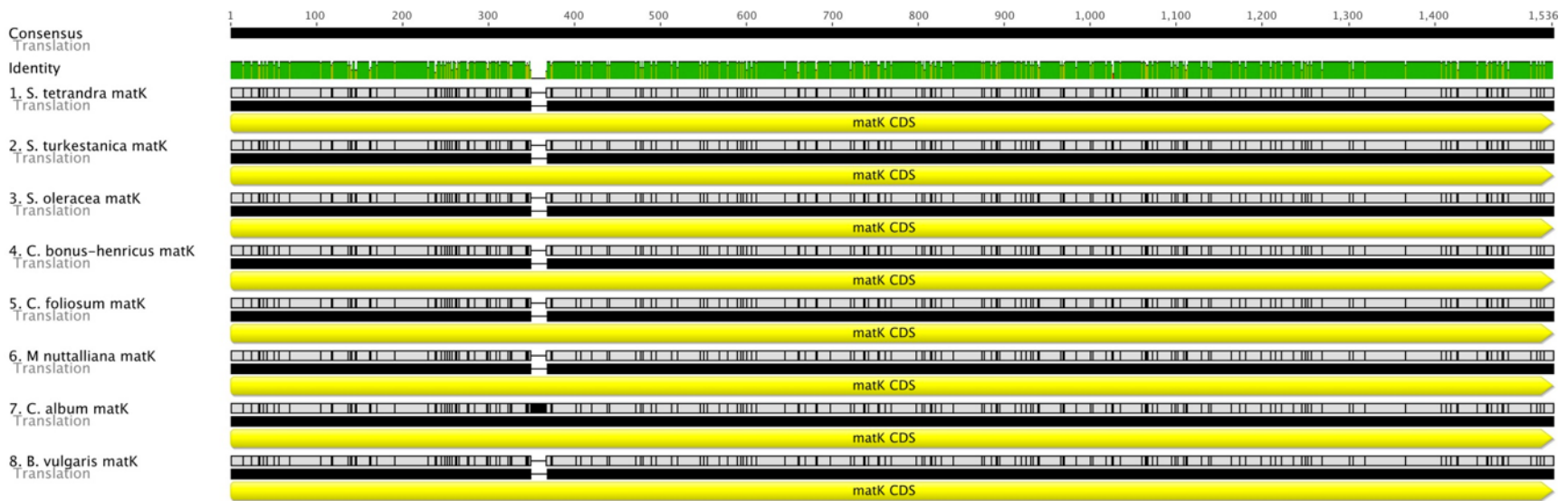
Alignment was performed with the MAFFT plugin (version 1.3) for Geneious using auto algorithm selection, PAM-200 scoring matrix (k=2), and a gap open penalty of 1.53. Exons are annotated with yellow arrows. Percentage sequence identity is given from green (100% identity) to black (0%).

	M. crystallinum	B. vulgaris	M. nuttalliana	C. foliosum	C. bonus-henricus	S. oleracea
B. vulgaris <i>rbcL</i>	94.7		-	-	-	-
M. nuttalliana <i>rbcL</i>	94.2	97.7		-	-	-
C. foliosum <i>rbcL</i>	94.4	97.6	98.4		-	-
C. bonus-henricus <i>rbcL</i>	95	98.1	98.7	99.2		-
S. oleracea <i>rbcL</i>	93.3	96.8	98.2	97.3	97.5	
C. album <i>rbcL</i>	93.8	96.3	96.7	97.8	97.7	96

**Table 2. Identity (%) of sequence pairs from the *rbcL* alignment in Figure 3.**

***matK***

We analyzed the published *matK* sequences listed in Supplemental Table 2. Only a partial sequence was available for *M. crystallinum* so we opted to use *B. vulgaris* as the only outgroup for the final analysis of this gene. As seen in the alignment in Figure 4, this chloroplast gene has no introns. The alignment indicates the presence of a single 18-bp insertion event seen only in *C. album*, 349-bp downstream of the start codon. For all species, the gene is approximately 1.5 kB in length. The *matK* sequences for each of the spinach species are 100% identical (Table 3). Among all the other sequences there is a range of sequence identities, from 91.8% identity between *B. vulgaris* and *C. album* to 99.1% identity between *C. foliosum* and *C. bonus-henricus*.



**Figure 4. Pairwise alignment of *matK* sequences.**

Alignment was performed with the MAFFT plugin (version 1.3) for Geneious using auto algorithm selection, PAM-200 scoring matrix (k=2), and a gap open penalty of 1.53. Exons are annotated with yellow arrows. Percentage sequence identity is given from green (100% identity) to black (0%).

20

	S. tetrandra	S. turkestanica	S. oleracea	C. bonus-henricus	C. foliosum	M nuttalliana	C. album
S. turkestanica matK	100						
S. oleracea matK	100	100					
C. bonus-henricus matK	97.7	97.7	97.7				
C. foliosum matK	97.2	97.2	97.2	99.1			
M nuttalliana matK	96.7	96.7	96.7	98.6	98.8		
C. album matK	92.6	92.6	92.6	93.6	93.2	92.6	
B. vulgaris matK	93.1	93.1	93.1	94	93.5	93.1	91.8

**Table 3. Identity (%) of sequence pairs from the *matK* alignment in Fig. 4.**

## ***UFO***

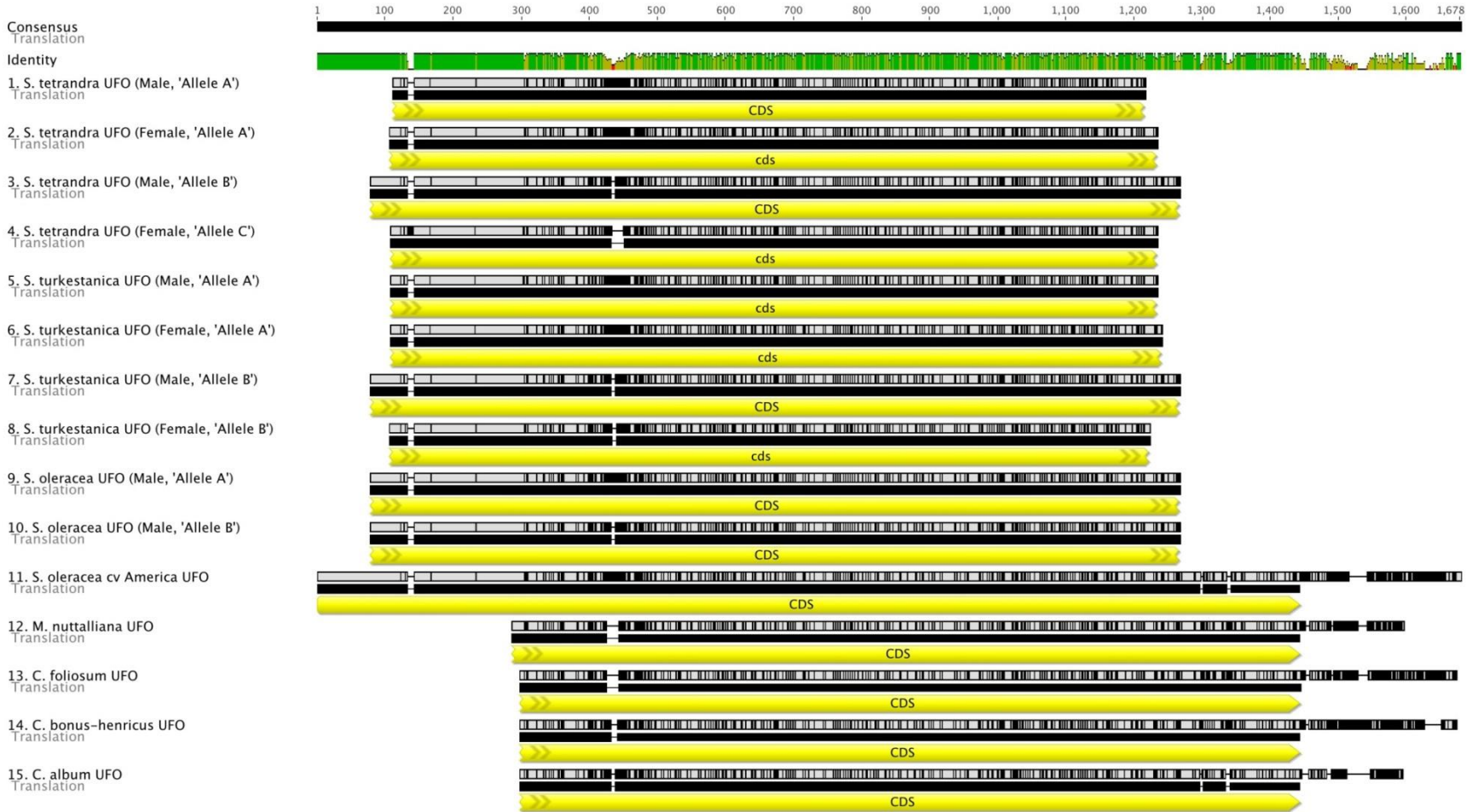
The *UFO* genes sampled are listed in Supplemental Table 2. All of the sequences analyzed for this gene were generated in our laboratory using primers listed in the Supplementary Material, Table 1. We generated two alignments that will be used for the analysis of this gene: one alignment which comprises all the *UFO* sequences generated in our study and another that only includes the four most complete *UFO* sequences (which include 3' end and UTR data). As seen in Figure 5, we identified a number of *UFO* alleles within *Spinacia* populations but could not correlate these with any patterns of sex or species. Thus, an alignment using just the most complete sequences (using data from 3'RACEs) was generated using a single representative sequence for *Spinacia sp.* (Figure 6). This analysis includes data at the 3' end of the gene and in the UTR (data that comprises gaps/missing data that would otherwise not have been used in the phylogenetic analysis of the former alignment).

As seen in Figure 5, this gene has no introns and appears to be slightly larger than 1.4 KB for all species. There is a highly variable region 414 bp downstream of the start codon in *Spinacia sp.*, where we see a unique sequence in every species under analysis here. There are also two small in-frame insertion events seen only in the *Blitum* group, approximately 100 bp upstream of the stop codon. The *UFO* sequences for the spinach species and cultivars are between 94.5% and 100% identical to each other (Table 4). Among the other species we see a range of sequence identities, from 77.1% identity between *C. album* and *C. bonus-henricus* to 97.6% identity between *C. foliosum* and *M. nuttalliana* (Table 5).

The alignment of the Chenopod *UFO* sequences indicates that there are two highly conserved regions separated by a hypervariable region. The first conserved region of the translated proteins ends with KTPNT at residue 130 in the spinach sequence. Protein folding

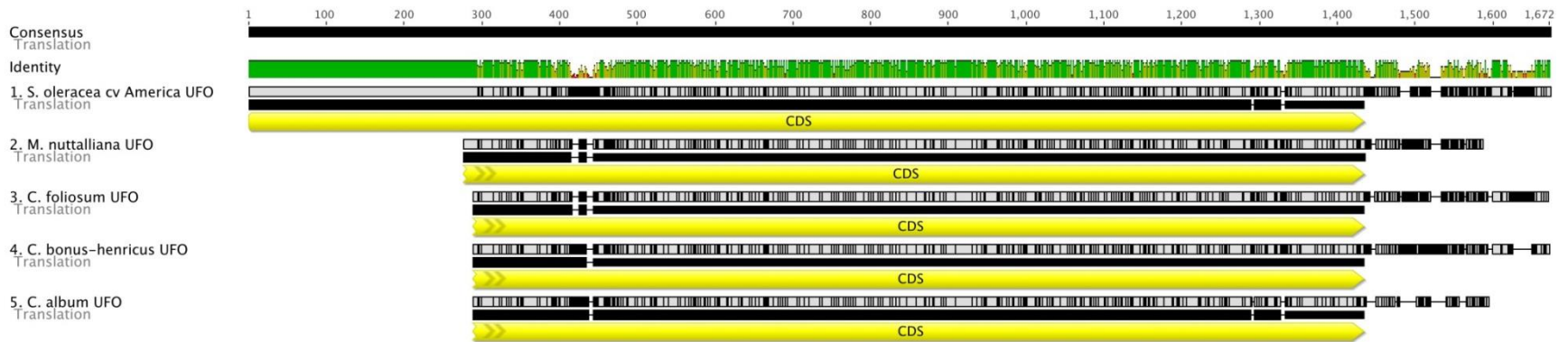


estimations [54] identify this as an F-box domain with 98.8% confidence (Figure 7). F-box domain proteins generally interact with SCF protein complexes to coordinate ubiquitination of targets for degradation in the proteasome [63]. Immediately prior to the variable region, all species have a conserved amino acid sequence ending with KTPNT. This is followed by a moderately variable five amino-acid sequence (SIVYR in both *C. bonus-henricus* and *C. foliosum*, STIYR in *C. album*, SIINR in *M. nuttalliana*, and TNIYR in *Spinacia sp.*) and then an arginine-rich sequence found in all sequences. Nested within this region is the conserved sequence HLNKG. Of particular note, all sequences have a hexanucleotide sequence GAAGGA encoding glycine and glutamic acid with the exception of *C. album*, which has a similar CAAGGA sequence. This hexanucleotide sequence becomes repeated in *Spinacia sp.* generating polymorphisms of 5 to 7 repeat units. Following this repeat region, the sequences diverge until reaching the conserved amino acid sequence beginning with S/T S/T RAYL, which defines the beginning of the second conserved amino acid region. Phyre2 protein-folding analysis identifies this region as a likely beta-propeller of a btb-kelch protein with 99.4% confidence. Therefore, it can be inferred that the hypervariable region in the *UFO* gene bridges two conserved functional protein encoding domains.



**Figure 5. Pairwise alignment of all *UFO* sequences and alleles.**

Alignment was performed with the MAFFT plugin (version 1.3) for Geneious using auto algorithm selection, PAM-200 scoring matrix ( $k=2$ ), and a gap open penalty of 1.53. Exons are annotated with yellow arrows. Percentage sequence identity is given, from green (100% identity) to black (0%).



**Figure 6. Pairwise alignment of selected *UFO* sequences.**

Alignment was performed with the MAFFT plugin (version 1.3) for Geneious using auto algorithm selection, PAM-200 scoring matrix ( $k=2$ ), and a gap open penalty of 1.53. Exons are annotated with yellow arrows.

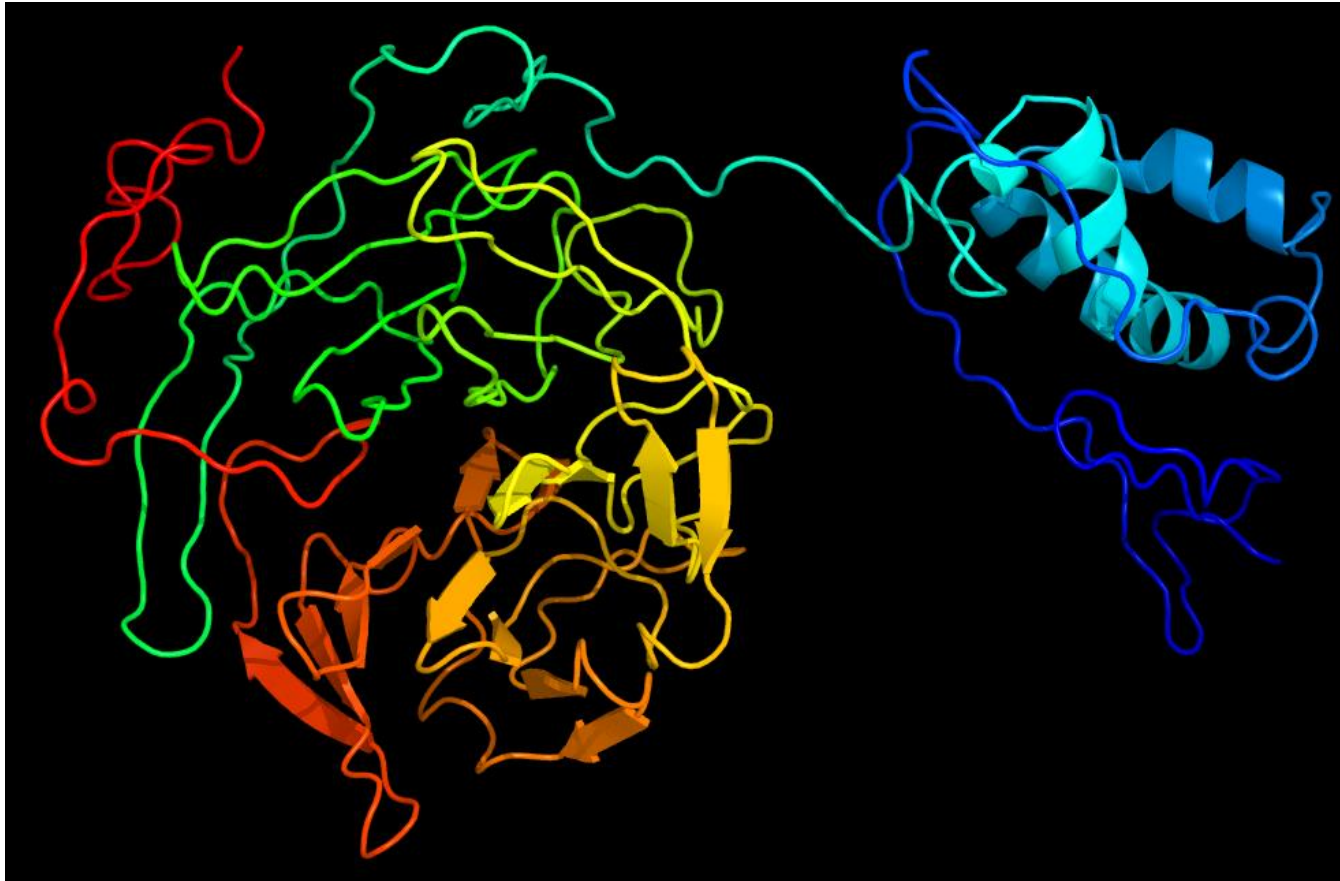
24

	SpTt (Male, 'A')	SpTt (Female, 'A')	SpTt (Male, 'B')	SpTt (Female, 'C')	SpTk (Male, 'A')	SpTk (Female, 'A')	SpTk (Male, 'B')	SpTk (Female, 'B')	SpOI (Male, 'A')	SpOI (Male, 'B')	SpOI cv America	Mnut	Cf	Cbh
<i>S. tetrandra</i> UFO (Female, 'Allele A')	100.0													
<i>S. tetrandra</i> UFO (Male, 'Allele B')	99.5	99.5												
<i>S. tetrandra</i> UFO (Female, 'Allele C')	94.5	94.6	95.1											
<i>S. turkestanica</i> UFO (Male, 'Allele A')	100.0	100.0	99.5	94.6										
<i>S. turkestanica</i> UFO (Female, 'Allele A')	100.0	100.0	99.5	94.6	100.0									
<i>S. turkestanica</i> UFO (Male, 'Allele B')	99.5	99.5	100.0	95.1	99.5	99.5								
<i>S. turkestanica</i> UFO (Female, 'Allele B')	99.3	99.3	99.8	95.0	99.3	99.3	99.8							
<i>S. oleracea</i> UFO (Male, 'Allele A')	100.0	100.0	99.5	94.6	100.0	100.0	99.5	99.3						
<i>S. oleracea</i> UFO (Male, 'Allele B')	99.5	99.5	100.0	95.1	99.5	99.5	100.0	99.8	99.5					
<i>S. oleracea</i> cv America UFO	99.4	99.3	98.7	94.2	99.3	99.3	98.7	98.6	99.2	98.7				
<i>M. nuttalliana</i> UFO	83.9	83.9	84.4	85.1	83.9	83.9	84.4	84.5	83.9	84.4	79.4			
<i>C. foliosum</i> UFO	83.3	83.3	83.9	84.9	83.3	83.3	83.9	83.9	83.3	83.9	76.6	97.5		
<i>C. bonus-henricus</i> UFO	83.8	83.8	84.2	84.9	83.8	83.8	84.2	84.4	83.7	84.2	75.7	92.4	91.4	
<i>C. album</i> UFO	80.7	80.9	81.5	81.8	80.9	80.9	81.5	81.4	81.0	81.5	76.3	78.3	78.2	77.5

**Table 4. Identity (%) of sequence pairs from the *UFO* alignment in Fig. 5.**

	<i>S. oleracea</i> cv America UFO	<i>M. nuttalliana</i> UFO	<i>C. foliosum</i> UFO	<i>C. bonus-henricus</i> UFO
<i>M. nuttalliana</i> UFO	81			
<i>C. foliosum</i> UFO	80	97.6		
<i>C. bonus-henricus</i> UFO	77.4	92.5	91.4	
<i>C. album</i> UFO	76.9	78.7	78.2	77.1

**Table 5. Identity (%) of sequence pairs from the *UFO* alignment in Fig. 6.**



**Figure 7. Protein folding estimation for the *SpUFO* protein .**

The amino acid sequence from *Spinacia oleracea cv America* was used for protein-folding estimation using Phyre2 under intense conditions. Image colored by rainbow N- to C-terminus.

**AG**

The *AG* sequences obtained in this work are listed in Supplemental Table 2. The *AG* gene has a conserved MADS box gene structure [64] consisting of a MADS box encoding region (M) followed by a short 5' sequence, followed by the linker region (I), the keratin-like region (K), and the carboxyl domain encoding region (C). For this gene two degenerate primers were used to amplify a region spanning the L region and K box, from Exon 2 to Exon 5. These primers were designed based on an alignment of known coding sequences from some core eudicots: *Spinacia oleracea* (AY660007), *Brassica juncea* (DQ060334), *Vitis vinifera* (GU133631), *Vicia sativa* (JF313850), and *Nicotiana benthamiana* (JQ699177). These primers returned a single band for each species, each ranging in size from around 550 to 600 base pairs in length when genomic DNA was amplified. More primers were designed to amplify the region spanning Exon 3 to the end of Exon 7. Degenerate reverse primers for Exon 7 were designed using the *AGAMOUS* sequence of *Spinacia* and *Arabidopsis thaliana*. These primers returned band sizes ranging from 900 to 2,000 bp. For *C. bonus-henricus* the sequencing results indicated the presence of two sequences of the *AGAMOUS* gene, which necessitated multiple rounds of cloning to separate the sequences.

This gene has seven exons and six introns. There is major variation in intron sequence length and identity for this gene as seen in Figure 8. The most striking differences are major deletion events seen in intron five of the *AG* sequences from the *Spinacia* species. There is only minor variation in exon sequence and length. The *AG* sequences for each of the spinach species are between 99.3 and 100% identical (Table 6), whereas among all species there is a range of sequence identities, from 30.2% between *S. oleracea* and *C. album* to 95.2% between *C.*

*foliosum* and *M. nuttalliana*. We also generated an alignment using only the putative coding sequences (Figure 9), which showed a range of identities between 89.5% and 100% (Table 7).

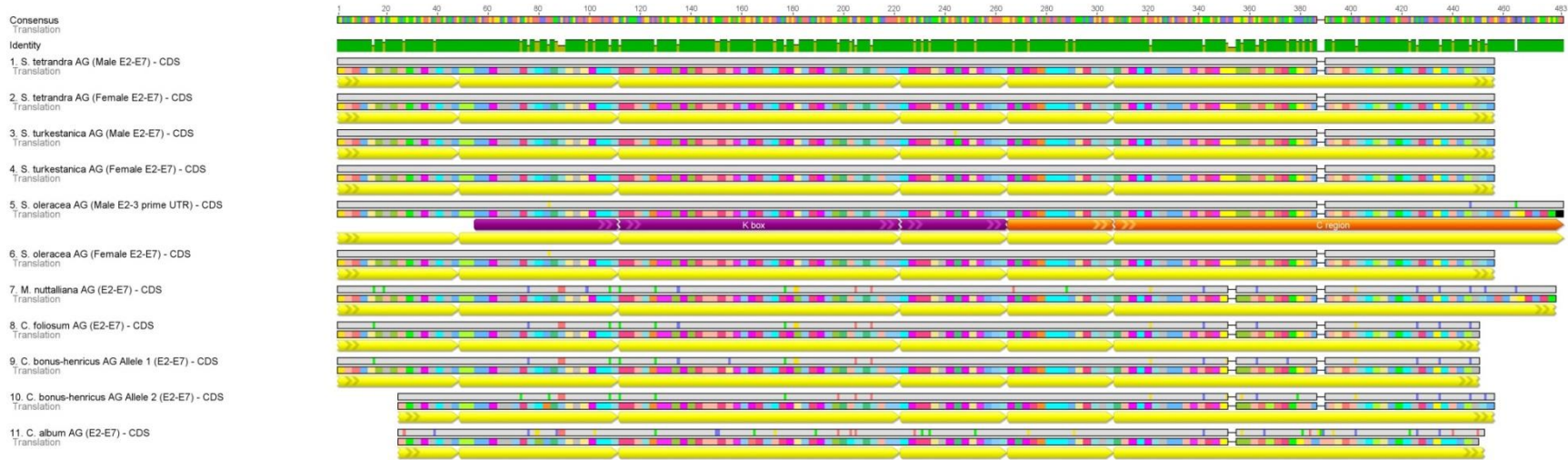
While there is a high degree of conservation seen in the coding portion of these sequences at the amino acid level, we also see intron variation in both sequence and size of the introns. Intron five is 1,526 bp long in *C. album*, but only 228 bp in *S. oleracea cv America*. There is a notable region of conservation in intron 5 and this was analyzed using the PLACE motif database. This intron contains a number of CCAAT-box binding complexes, which are known to be involved in regulation of flowering in *Arabidopsis* [65]. Interestingly, although these ‘islands’ of conservation are seen in *Spinacia sp.*, we see only one canonical CCAAT-box in intron five when we see two in *C. album* and the other species in the *Blitum* group. *C. album* has a unique 3-nucleotide insertion in exon 7. Similarly, there is an independent 3-nucleotide insertion in exon 7 in the *Spinacia* species (Figure 8).



**Figure 8. Pairwise alignment of AGAMOUS sequences from Exon 2 to Exon 7.**

Alignment was performed with the MAFFT plugin (version 1.3) for Geneious using auto algorithm selection, PAM-200 scoring matrix ( $k=2$ ), and a gap open penalty of 1.53. Exons are annotated with yellow arrows. The K-box motif and c-terminal region of *S. oleracea* AG are annotated with purple and orange arrows, respectively.





**Figure 9. Pairwise alignment of *AGAMOUS* coding regions.**

Alignment was performed with the MAFFT plugin (version 1.3) for Geneious using auto algorithm selection, PAM-200 scoring matrix (k=2), and a gap open penalty of 1.53. Shown is putative exon data from exon 2 to exon 7; exons are annotated with yellow arrows. The K-box motif and c-terminal region of *S. oleracea* AG are annotated with purple and orange arrows, respectively.

	SpTt AG (Male)	SpTt AG (Female)	SpTk AG (Male)	SpTk AG (Female)	SpOI AG (Male)	SpOI AG (Female)	Mnut AG	Cf AG	Cbh AG Allele 1	Cbh AG Allele 2
<i>S. tetrandra</i> AG (Female E2-E7)	99.7									
<i>S. turkestanica</i> AG (Male E2-E7)	99.9	99.6								
<i>S. turkestanica</i> AG (Female E2-E7)	99.7	100	99.6							
<i>S. oleracea</i> AG (Male E2-3 prime UTR)	99.3	99.6	99.3	99.6						
<i>S. oleracea</i> AG (Female E2-E7)	99.4	99.7	99.3	99.7	99.7					
<i>M. nuttalliana</i> AG (E2-E7)	41	41.1	40.9	41.1	41.5	40.9				
<i>C. foliosum</i> AG (E2-E7)	41.1	41.2	41.1	41.2	41.1	41.1	95.2			
<i>C. bonus-henricus</i> AG Allele 1 (E2-E7)	40.3	40.4	40.2	40.4	40.3	40.3	93.3	95.1		
<i>C. bonus-henricus</i> AG Allele 2 (E2-E7)	40.9	41	40.9	41	41.1	41	88.6	90.4	89.4	
<i>C. album</i> AG (E2-E7)	30.2	30.3	30.2	30.3	30.2	30.2	48	48.9	49.1	49

**Table 6. Identity (%) of sequence pairs from the AGAMOUS alignment in Fig. 8.**  
This alignment included exon and intron data.

	SpTt (Male)	SpTt (Female)	SpTk (Male)	SpTk (Female)	SpOI (Male)	SpOI (Female)	Mnut	Cf	Cbh Allele 1	Cbh Allele 2
<i>S. tetrandra</i> AG (Female)	100.0									
<i>S. turkestanica</i> AG (Male)	99.8	99.8								
<i>S. turkestanica</i> AG (Female)	100.0	100.0	99.8							
<i>S. oleracea</i> AG (Male)	99.6	99.6	99.3	99.6						
<i>S. oleracea</i> AG (Female)	99.8	99.8	99.6	99.8	99.8					
<i>M. nuttalliana</i> AG	93.6	93.6	93.4	93.6	93.7	93.4				
<i>C. foliosum</i> AG	94.6	94.6	94.4	94.6	94.6	94.4	99.1			
<i>C. bonus-henricus</i> AG Allele 1	94.0	94.0	93.7	94.0	94.0	93.7	98.4	99.3		
<i>C. bonus-henricus</i> AG Allele 2	93.9	93.9	93.7	93.9	94.2	93.9	96.9	97.6	97.4	
<i>C. album</i> AG	90.0	90.0	89.7	90.0	89.5	89.7	89.6	90.3	89.6	90.4

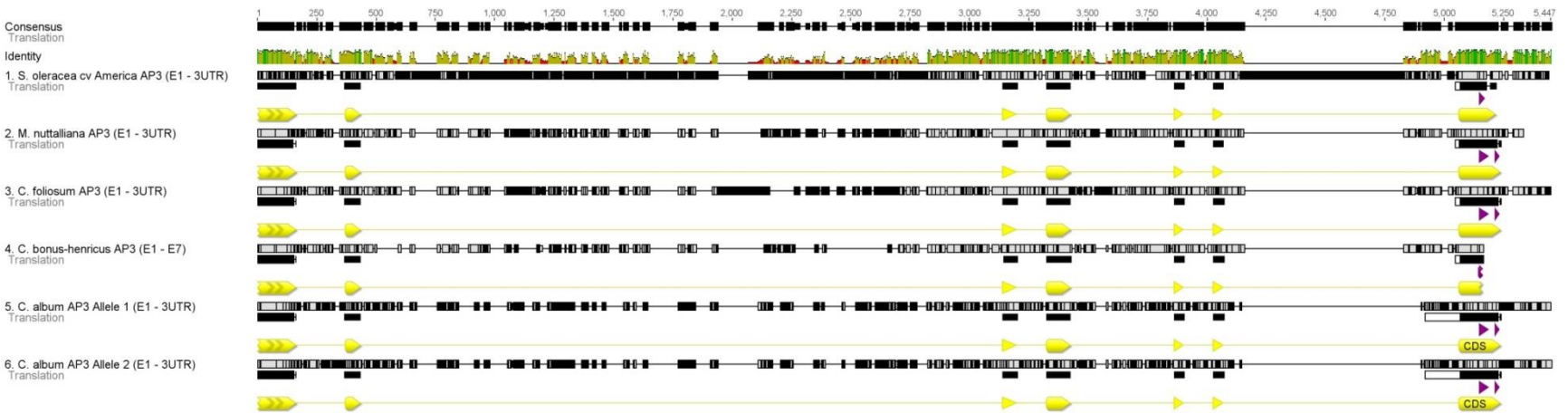
**Table 7. Identity (%) of sequence pairs from the AGAMOUS alignment in Fig. 9.**  
This alignment considered only data from putative coding regions.

### **AP3**

The *AP3* sequences obtained in this work are listed in Supplemental Table 2. Our results indicate a variation in length of the *AP3* gene sequences obtained, ranging from 4,892 bp for *Spinacia sp.* to 3,443 bp in length for *C. foliosum*. For *C. album*, the sequencing results indicated the presence of at least two alleles of the *AP3* gene, which necessitated multiple rounds of cloning to separate the sequences. These alleles differ only slightly in intron 1 and are 98.7% identical (Table 8). For the purposes of our analysis, we use *C. album* ‘Allele 1’ as the outgroup for *AP3*. We could not successfully retrieve the 3’ sequence for *C. bonus-henricus AP3*.

Sequences were aligned and the intron/exon structures were inferred by comparing to other published sequences or to data from 3’ RACE sequences. The *AP3* gene has 7 exons and 6 introns in the species examined here, as seen in Figure 10. There is substantial variation in intron size as well as sequence identity (Table 8) compared to the amount of variation seen in some of the other genes in this study, due to large insertion/deletion events seen in the introns of *AP3*. For example, the length of intron 2 in *Spinacia oleracea cv America* is 2,522 bp, however in *C. bonus-henricus* it is just 1,201 bp in length. In most of the *Blitum* group, intron 6 is approximately 277 bp in length, whereas in *Spinacia* the intron is 943 bp in length. There is also more variation between species in the coding sequence of this gene than we see in some of the other genes in this study. To quantify this variation we generated an alignment using only the putative exon sequence (Figure 11), which showed a range of identities from 77.5% between *Spinacia sp.* and *C. album* ‘Allele 1’ to 95.5% between *C. foliosum* and *C. bonus-henricus* (Table 9). Of particular interest is a region in exon 7, which is where the diagnostic *PI/AP3* motifs can be found [47]. In *Spinacia sp.* we can see a frameshift deletion mutation that has been previously described, but we note here that the closest relatives to spinach do not have that same

mutation (Supplemental Figure 3). *C. album*, *C. foliosum*, and *M. nuttalliana* all have a version of the euAP3 motif as well as the *PI* motif. The sequence of this region in *C. bonus-henricus* is undetermined.



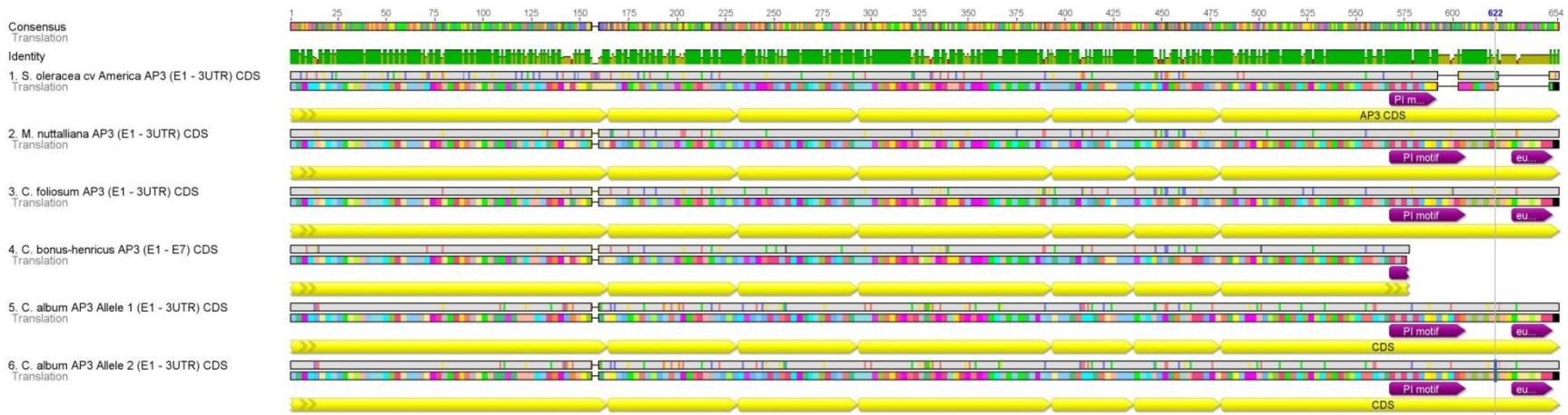
**Figure 10. Pairwise alignment of *APETALA3* sequences.**

Alignment was performed with the MAFFT plugin (version 1.3) for Geneious using auto algorithm selection, PAM-200 scoring matrix ( $k=2$ ), and a gap open penalty of 1.53. Exons are annotated with yellow arrows and *PI* motifs with purple arrows.

	<i>S. oleracea</i> cv America AP3	Mnut	Cf	Cbh	Ca 'Allele 1'
<i>M. nuttalliana</i> AP3	43.6				
<i>C. foliosum</i> AP3	43.5	88.6			
<i>C. bonus-henricus</i> AP3	40	78.6	75.1		
<i>C. album</i> AP3 'Allele 1'	36.4	52.1	51.9	51.5	
<i>C. album</i> AP3 'Allele 2'	36.2	51.7	51.5	51.1	98.7

**Table 8. Identity (%) of sequence pairs from the *APETALA3* alignment in Fig. 10.**

This alignment included both exon and intron data.



**Figure 11. Pairwise alignment of *APETALA3* coding regions from Exon 2 to Exon 7.**

Alignment was performed with the MAFFT plugin (version 1.3) for Geneious using auto algorithm selection, PAM-200 scoring matrix ( $k=2$ ), and a gap open penalty of 1.53. Exons are annotated with yellow arrows. The K-box motif and c-terminal region of *S. oleracea* *AG* are annotated with purple and orange arrows, respectively.

35

	<i>S. oleracea</i> cv America AP3	Mnut	Cf	Cbh	Ca 'Allele 1'
<i>M. nuttalliana</i> AP3	78.3				
<i>C. foliosum</i> AP3	79.2	95.2			
<i>C. bonus-henricus</i> AP3	85.3	93.7	95.5		
<i>C. album</i> AP3 'Allele 1'	77.5	84	84	84.5	
<i>C. album</i> AP3 'Allele 2'	77.5	84	84	84.5	100

**Table 9. Identity (%) of sequence pairs from the *APETALA3* alignment in Fig. 11.**

This alignment considered only data from putative coding regions.

***PI***

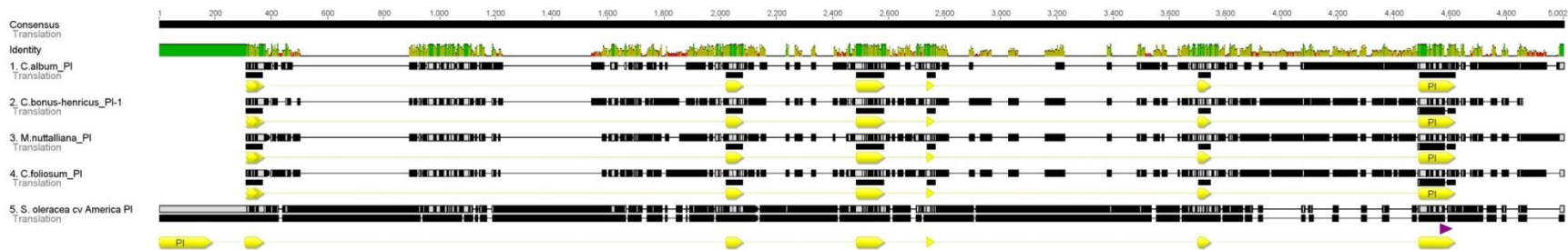
The *PI* sequences obtained and analyzed in this work are listed in Supplemental Table 2. The spinach *PI* gene is approximately 4.1 kB long, with large introns when compared to other genes in this study. There are 7 exons and 6 introns. Exon 7 encodes much of the C-domain and includes the conserved *PI* motif. Previously, *Spinacia oleracea* cv America was reported to have a deletion that removed most of the conserved *PI* motif [47]. To verify this observation, we sampled this domain from gDNA templates in all the putative species of *Spinacia* (both male and female) and found them to be identical (Supplemental Material, Figure 1). For the remainder of our study we used the *Spinacia oleracea* cv America *PI* sequence (GQ120477) as a representative for *Spinacia* sp. in our analysis of *PI* gene evolution in this group of plants.

Intron length and sequence identity varies greatly for *Spinacia* sp., following a similar pattern to what is seen in the comparison of the *AP3* sequences. In *Spinacia* sp., intron 2 is 1,504 bp in length compared to 687 bp for intron 2 of *C. album*. Intron 5 of *Spinacia* is 912 bp in length whereas for *C. album* it is only 261 bp in length. Intron 6 of *Spinacia*, however, is 253 bp in length, while this intron in *C. album* is 563 bp in length. There is no significant variation in intron length or sequence identity within the *Blitum* group, however. When considering exon and intron data, we see a wide range of identities, from 30.1% identity between *Spinacia oleracea* and *C. album* to 90.3% identity between *C. foliosum* and *M. nuttalliana* (Table 10).

Multiple *PI* sequences were found for *C. bonus-henricus* in each of two individuals sampled, and two were found for *C. album*. In our analysis of this gene, it was necessary to determine what was a '*PI*' gene or a *PI*-paralogous gene (or '*PI*-like' gene). This was done through analysis of the C-domain motif. It has previously been shown that *Spinacia* has a truncated version of the *PI*-motif and in our study here we found this same truncated *PI*-motif in

*C. bonus-henricus*, *C. foliosum*, and *M. nuttalliana* (Figure 12). Interestingly, this truncated motif is not seen in *C. album*; instead, we see a canonical *PI*-motif in this sequence, which is considered to be orthologous to the *Spinacia oleracea cv America PI* sequence (GQ120477). The carboxyl-end signatures of the ‘*PI-like*’ genes are not similar to the canonical *PI*-motif or the truncated *PI*-motif seen in the orthologs of *PI* in *C. bonus-henricus*, *C. foliosum*, or *M. nuttalliana*. Thus, these sequences are termed ‘*PI-like*’ and seen as paralogous sequences.





**Figure 12. Pairwise alignment of *PISTILLATA* sequences.**

Alignment was performed with the MAFFT plugin (version 1.3) for Geneious using the E-INS-i algorithm, PAM-200 (k=2) scoring matrix, and a gap open penalty of 1.53. Exons are annotated with yellow arrows and the *SpPI*-motif is indicated with a purple arrow.

	C.album PI	C.bonus-henricus PI-1	M.nuttalliana PI	C.foliosum PI
C.bonus-henricus PI-1	47.8			
M.nuttalliana PI	45.7	81.5		
C.foliosum PI	46.8	84.2	90.3	
S.oleracea cv America PI	30.1	35.8	34.3	35.6

**Table 10. Identity (%) of sequence pairs from the *PISTILLATA* alignment in Fig. 12.**



**Figure 13. Pairwise alignment of *PISTILLATA* coding regions, including *PI* alleles and *PI*-like sequences sampled.**

Alignment was performed with the MAFFT plugin (version 1.3) for Geneious using the E-INS-i algorithm, PAM-200 (k=2) scoring matrix, and a gap open penalty of 1.53. Exons are annotated with yellow arrows and *PI*-motifs are indicated with purple arrows.

39

	C. bonus-henricus PI-1	C. foliosum PI	M. nuttalliana PI	C. bonus-henricus PI-2	S. oleracea cv America PI	C. album PI	C. bonus-henricus PI-like-1
C. foliosum PI (CDS)	96.3						
M. nuttalliana PI (CDS)	95.2	90.9					
C. bonus-henricus PI-2 (CDS)	89.3	80.8	83.9				
S. oleracea cv America PI (CDS)	74.5	68.2	63.3	66.5			
C. album PI (CDS)	73.4	61.5	63.1	65	53.4		
C. bonus-henricus PI-like-1 (CDS)	57.9	55.5	55.1	53.7	52.3	53.5	
C. album PI-like 1 (CDS)	47.9	42.9	46.8	46	38	43.3	59.4

**Table 11. Identity (%) of sequence pairs from the *PISTILLATA* alignment in Figure 13.**

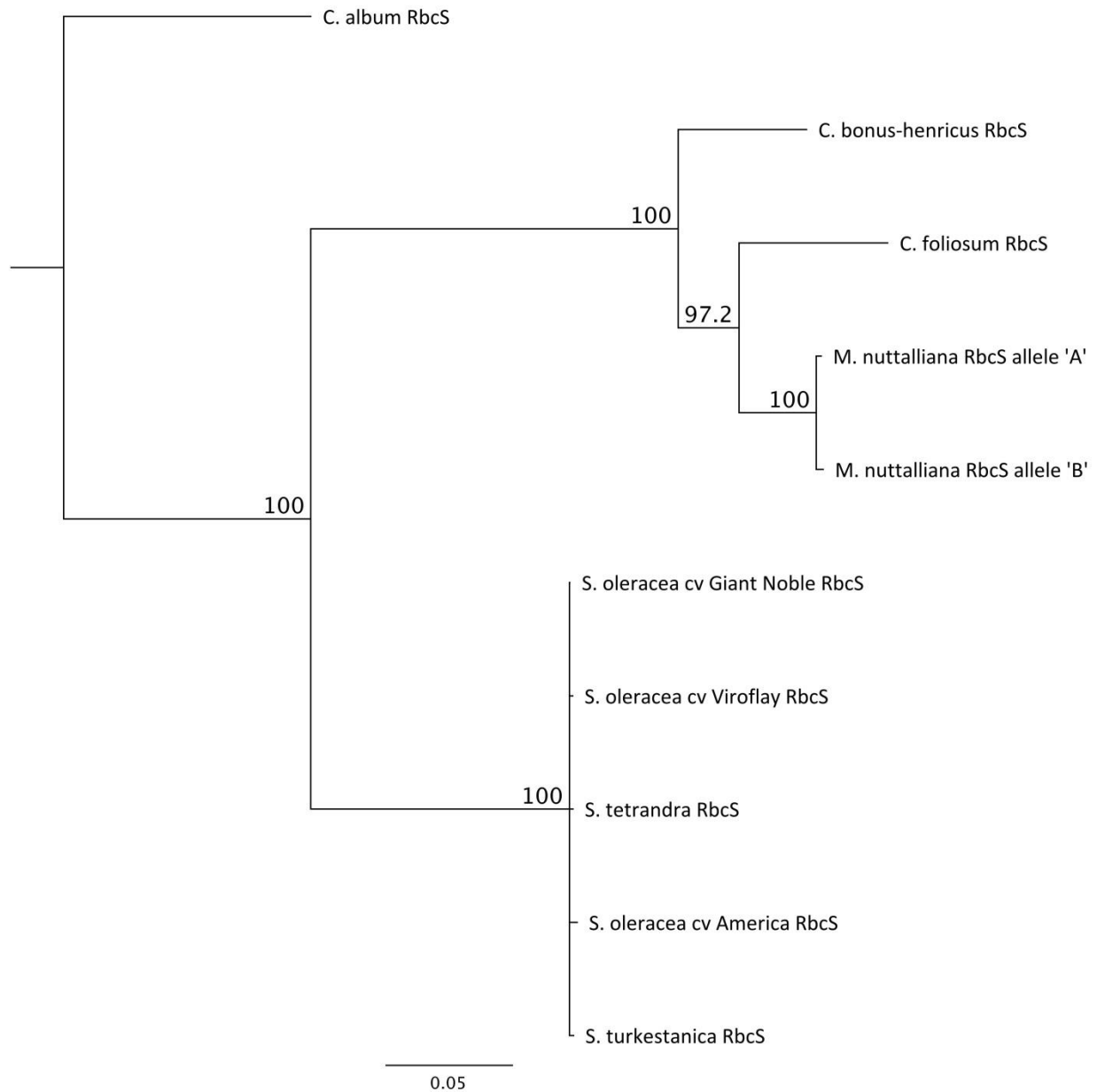
This alignment considered only putative coding data.

## Phylogenetic Analyses

Phylogenetic analysis was performed on each alignment using both the Geneious Tree Builder (which uses a neighbor-joining technique) and the Geneious PhyML plugin (which uses maximum-likelihood). Most trees used *C. album* as the rooted outgroup, although for the *rbcL* and *matK* genes we were able to use far outgroups *M. crystallinum* and *B. vulgaris* as far outgroups. When applicable, we also generated phylogenetic trees from alignments of coding sequences only (e.g. *AG*, *AP3*, and *PI*).

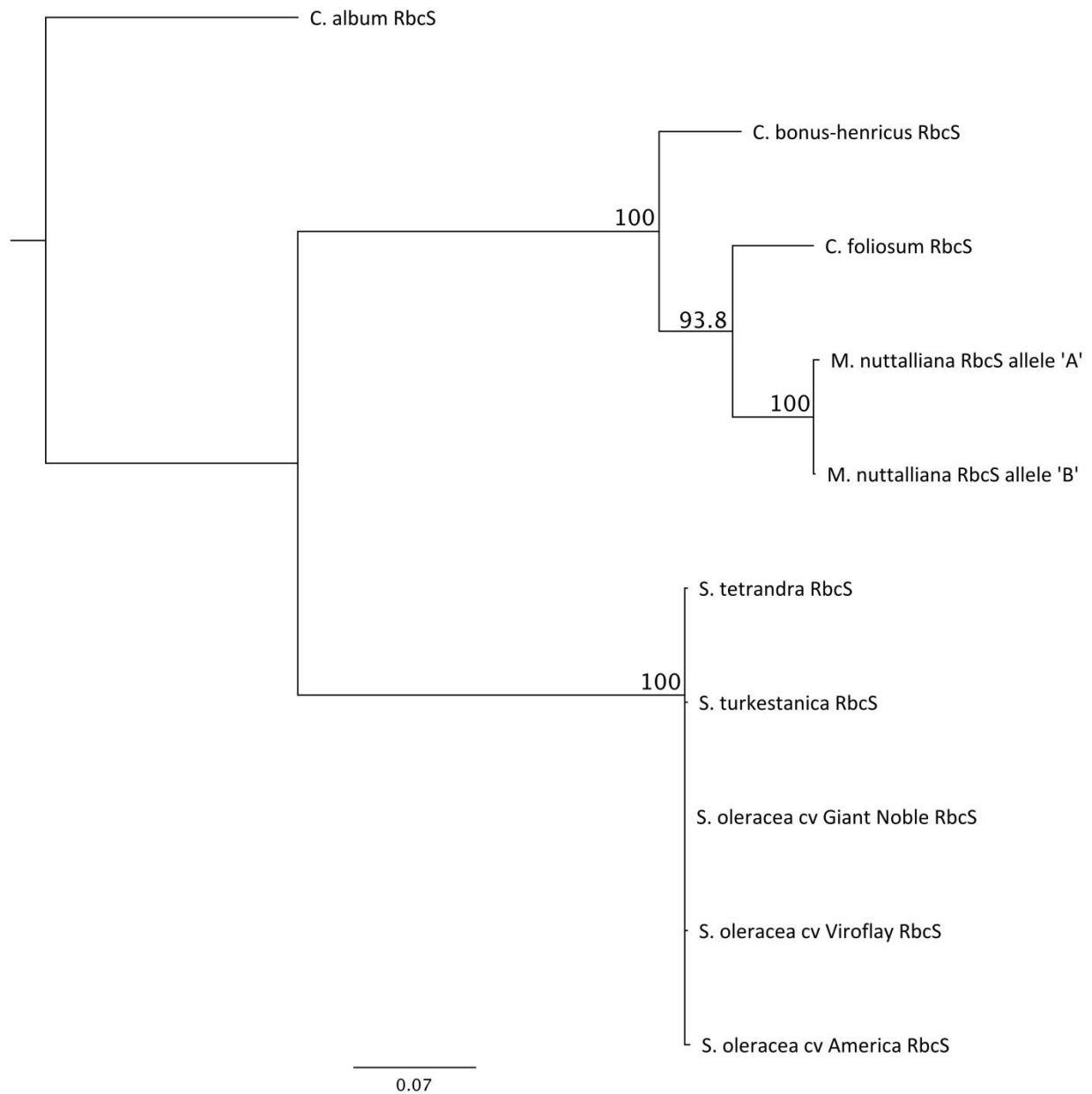
### *rbcS*

Phylogenetic analysis was performed on the *rbcS* alignments using *C. album* as the rooted node, based on its position in the *rbcL* and *matK* trees. These trees both place *C. album* as the near outgroup for the other Chenopods in our study. Trees generated from already published data establish *Beta vulgaris* and *Mesembryanthemum crystallinum* as far outgroups, and cement *C. album*'s role as the close outgroup for our study [11]. There are two major groupings seen in these *rbcS* trees: one being the 'Blitum' group and the other being the *Spinacia* lineage. Within the 'Blitum' group, *Monolepis nuttalliana* and *Chenopodium foliosum* group as sister taxa (98.2% node support in the NJ tree, 93.8% in the ML tree). The two *rbcS* sequences obtained from *M. nuttalliana* cluster, supporting an interpretation of these sequences as alleles (100% and 99.9% node support, respectively). Within the *Spinacia* lineage, we see no hierarchical relationship between species or cultivars that is supported significantly. All nodes were supported at a very high level ( $\geq 78\%$  bootstrap support) with the exception being within the *Spinacia sp.* group. The *Spinacia* sequences are nearly identical, so phylogenetic inference cannot be reliably hypothesized. The same tree topology was generated by both methods when the alignment/tree construction only took into account the extracted exon sequences.



**Figure 14. Phylogenetic tree constructed from the *rbcS* alignment using Geneious tree builder.**

This method used the Tamura-Nei genetic distance model, the neighbor-joining tree build method, bootstrap resampling method with 10,000 replicates, with the support threshold set to 85% and using *C. album* as the rooted outgroup. Topology support (%) among replicates is shown as a branch label.



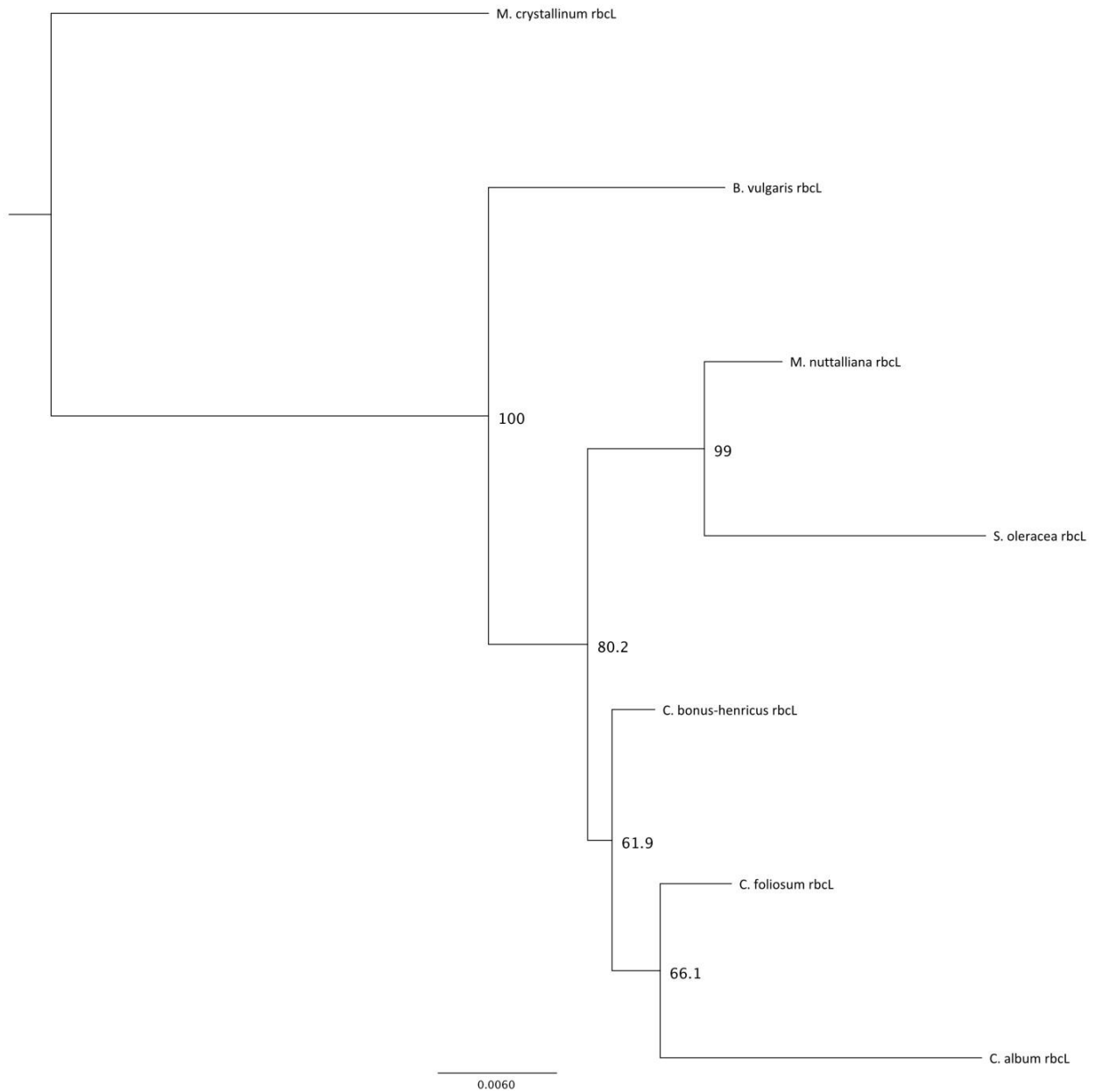
**Figure 15. Phylogenetic tree constructed from the *rbcS* alignment using PhyML.**

This tree was made with the Geneious PhyML plugin which uses a maximum-likelihood method, the Tamura-Nei substitution model, bootstrap branch support method with 10,000 replicates, taking the best of both NNI and SPR topology searches, and all other settings at default. Topology support (%) among replicates is shown as a branch label.

***rbcL***

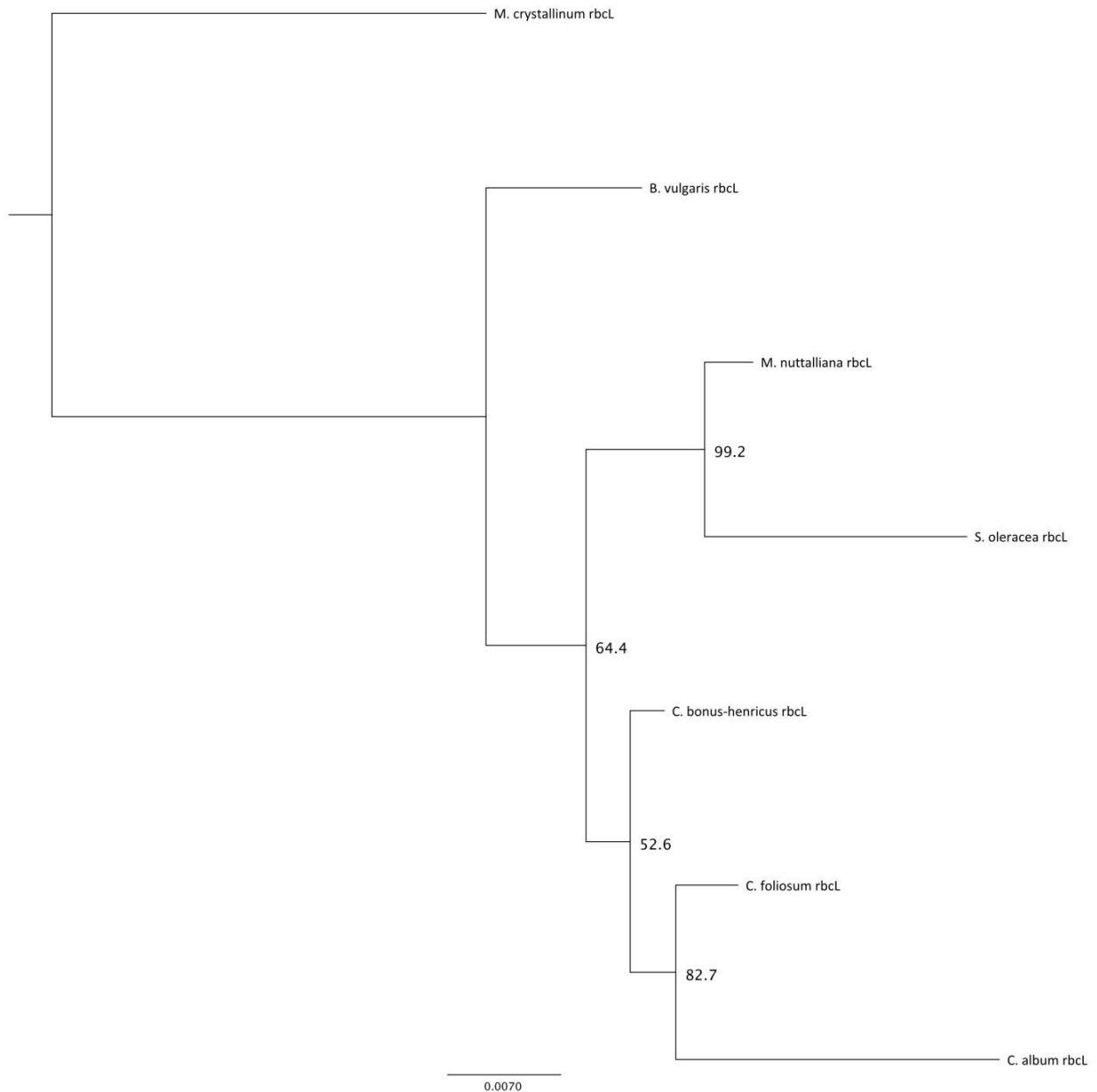
Phylogenetic analysis was performed using *M. crystallinum* as the far outgroup to root the trees. As expected, *Beta vulgaris* is an outgroup to the remaining Chenopods in this study. These remaining species bifurcate into two clades. One clade groups *M. nuttalliana* and *Spinacia* and the second clade groups the remaining three *Chenopodium* species with *C. bonus-henricus* as a sister taxon to the pairing of *C. album* and *C. foliosum*. Bootstrap support for this tree ranges from 100 to 61.9 percent, with the nodes grouping *C. album*, *C. foliosum* and *C. bonus-henricus* receiving the least support. The resulting tree topology from the PhyML analysis is congruent with the neighbor-joining tree. Both trees hypothesize *M. nuttalliana* and *Spinacia* as sister taxa with >99.2% bootstrap support. This result is consistent with the grouping of these species reported by Kadereit, et al. [14]. The placement of *C. album* within a cluster with *C. bonus-henricus* and *C. foliosum* deviates from all other trees in this study and has low support in both tree analyses.

It is worth noting that *C. bonus-henricus*, *C. foliosum*, and *M. nuttalliana* form a monophyletic clade (*Blitum* group) in other gene tree analyses within this study and elsewhere [11]. If the *Blitum* grouping truly reflects actual evolutionary origins, then the differential rates identified in this gene would be a consistent phylogenetic trait. As with the phylogenetic inferences, the low number of unique sequence positions among pairs considered for this *rbcL* analysis makes the analysis of *rbcL* less informative than all others genes examined in this study.



**Figure 16. Phylogenetic tree constructed from the *rbcL* alignment using Geneious tree builder.**

This method used the Tamura-Nei genetic distance model, the neighbor-joining tree build method, bootstrap resampling method with 10,000 replicates, with the support threshold set to 60% and using *M. crystallinum* as the rooted outgroup. Topology support (%) among replicates is shown as a node label.



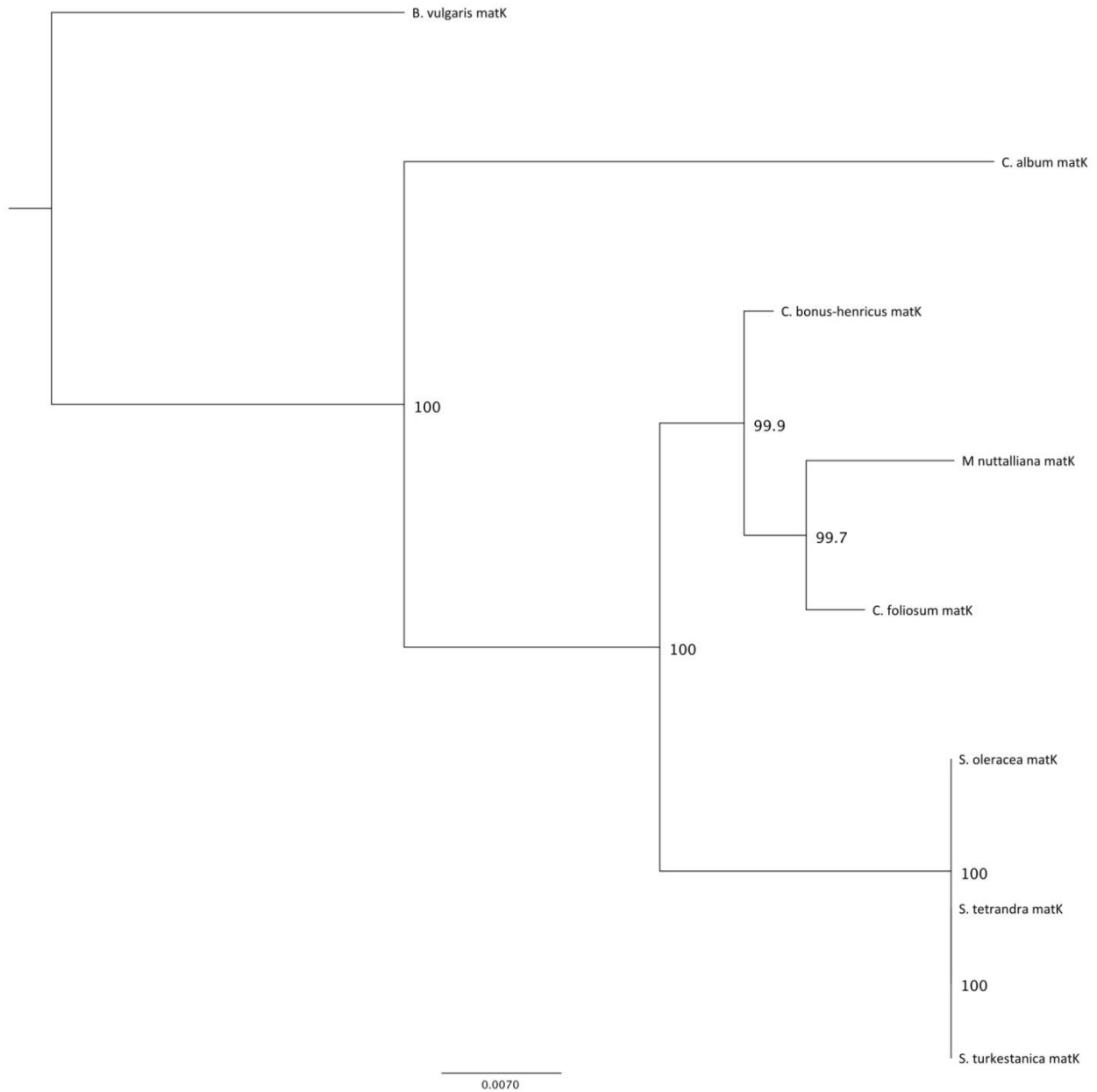
**Figure 17. Phylogenetic tree constructed from the *rbcL* alignment using PhyML.**

This tree was made with the Geneious PhyML plugin which uses a maximum-likelihood method, the Tamura-Nei substitution model, bootstrap branch support method with 10,000 replicates, taking the best of both NNI and SPR topology searches, and all other settings at default. Topology support (%) among replicates is shown as a node label.



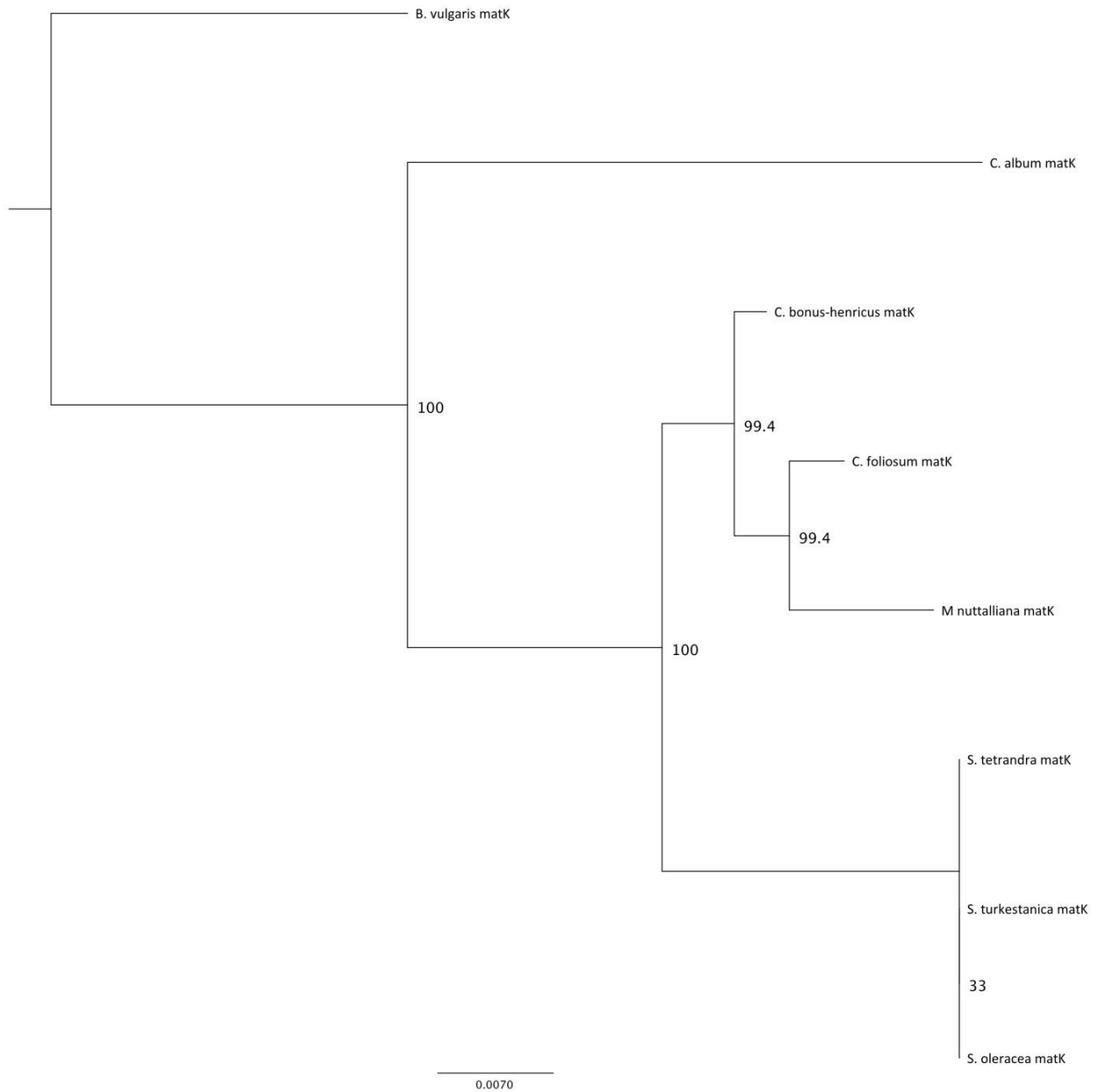
***matK***

Phylogenetic analysis was performed using *Beta vulgaris* to root the trees. The neighbor-joining tree identified *C. album* as sister to all remaining taxa, supporting its position as the near outgroup to the rest (Figure 18). *Spinacia* sp. segregated next, thereby establishing a cluster of *C. bonus-henricus*, *C. foliosum*, and *M. nuttalliana* (the *Blitum* group). All nodes were supported at a very high level ( $\geq 99.7\%$  bootstrap support). The *Spinacia* sequences are 100% identical, so phylogenetic hierarchy cannot be inferred. The tree resulting from the PhyML analysis (Figure 19) is both topologically congruent with the *matK* neighbor-joining tree and with the trees generated from the *rbcS* sequences.



**Figure 18. Phylogenetic tree constructed from the *matK* alignment using Geneious tree builder.**

This method used the Tamura-Nei genetic distance model, the neighbor-joining tree build method, bootstrap resampling method with 10,000 replicates, with the support threshold set to 85% and using *B. vulgaris* as the rooted outgroup. Topology support (%) among replicates is shown as a node label.



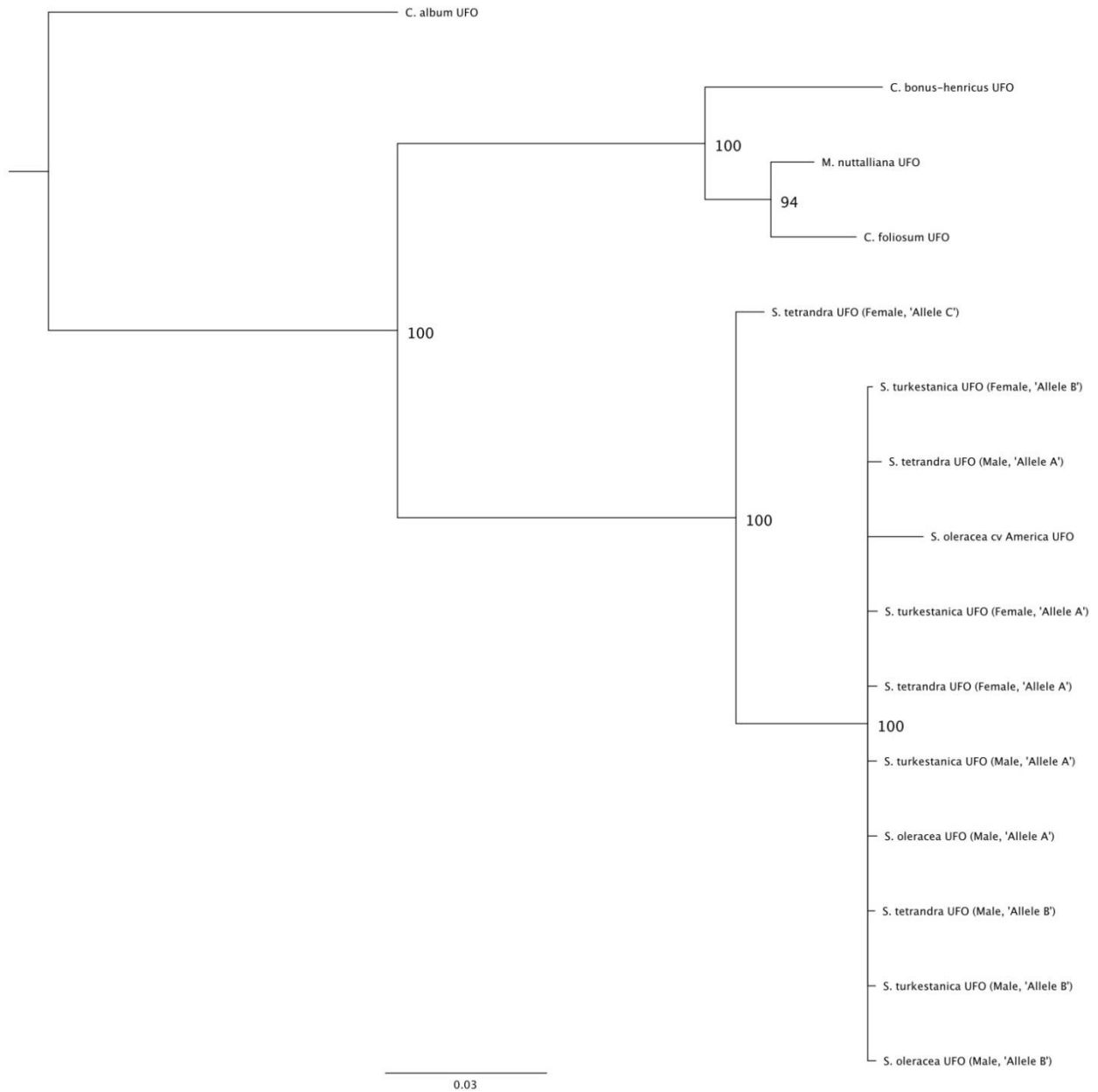
**Figure 19. Phylogenetic tree constructed from the *matK* alignment using PhyML.**

This tree was made with the Geneious PhyML plugin which uses a maximum-likelihood method, the Tamura-Nei substitution model, bootstrap branch support method with 10,000 replicates, taking the best of both NNI and SPR topology searches, and all other settings at default. Topology support (%) among replicates is shown as a node label.

## ***UFO***

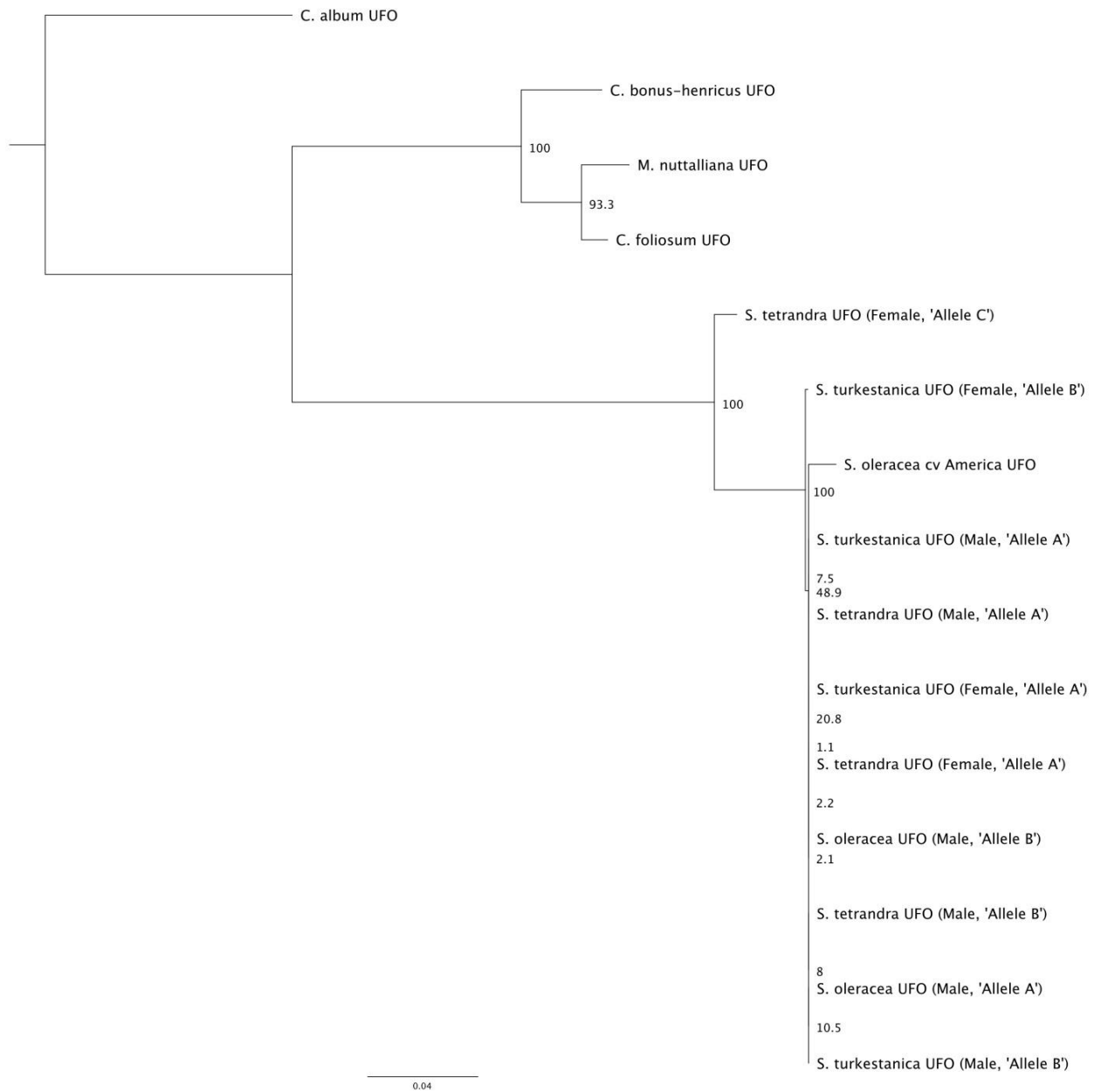
Phylogenetic analysis was performed for the *UFO* gene using *C. album* as the outgroup. We generated both neighbor-joining and maximum likelihood trees for each *UFO* alignment (Figure 5 & Figure 6), for a total of 4 trees for this gene. When considering only the phylogenetic analysis of the limited dataset from Figure 6, both trees have a very high degree of branch node support, in the range of 88.3% to 100%. With *C. album* as the outgroup, both trees have *C. bonus-henricus*, *C. foliosum*, and *M. nuttalliana* (Blitum group) clustering together, with *C. foliosum* and *M. nuttalliana* as sister taxa (Figures 22 & 23). This tree topology closely mirrors that topology seen in other trees in our study (again with the exception being the *rbcL* tree). The *Spinacia* lineage branches off from *C. album* and then the clade of *C. bonus-henricus*, *C. foliosum*, and *M. nuttalliana* branches off together. In the *UFO* trees that included all of the sequences from the various *Spinacia* species and cultivars (Figures 20 & 21), we see a single sequence obtained from *Spinacia tetrandra* that does not cluster with the other *UFO* sequences from *Spinacia*. Here we treat this as an allelic difference (Allele “C”) rather than a gene duplication, as we were unable to obtain this allele from any other individuals of this species. Had this represented a duplicated gene, we would have expected to isolate this sequence in other individuals.

We also used the PhyML plugin for Geneious to analyze the alignments, and the resulting tree topologies matched the neighbor-joining trees. The PhyML tree consisting of the four selected *UFO* sequences (Figure 23) had a lower node support between *C. foliosum* and *M. nuttalliana* (59.7) than the same node had in the equivalent neighbor-joining tree (Figure 22). This node between *C. foliosum* and *M. nuttalliana* is very highly supported in the PhyML tree of the larger *UFO* data set, however (Figure 21).



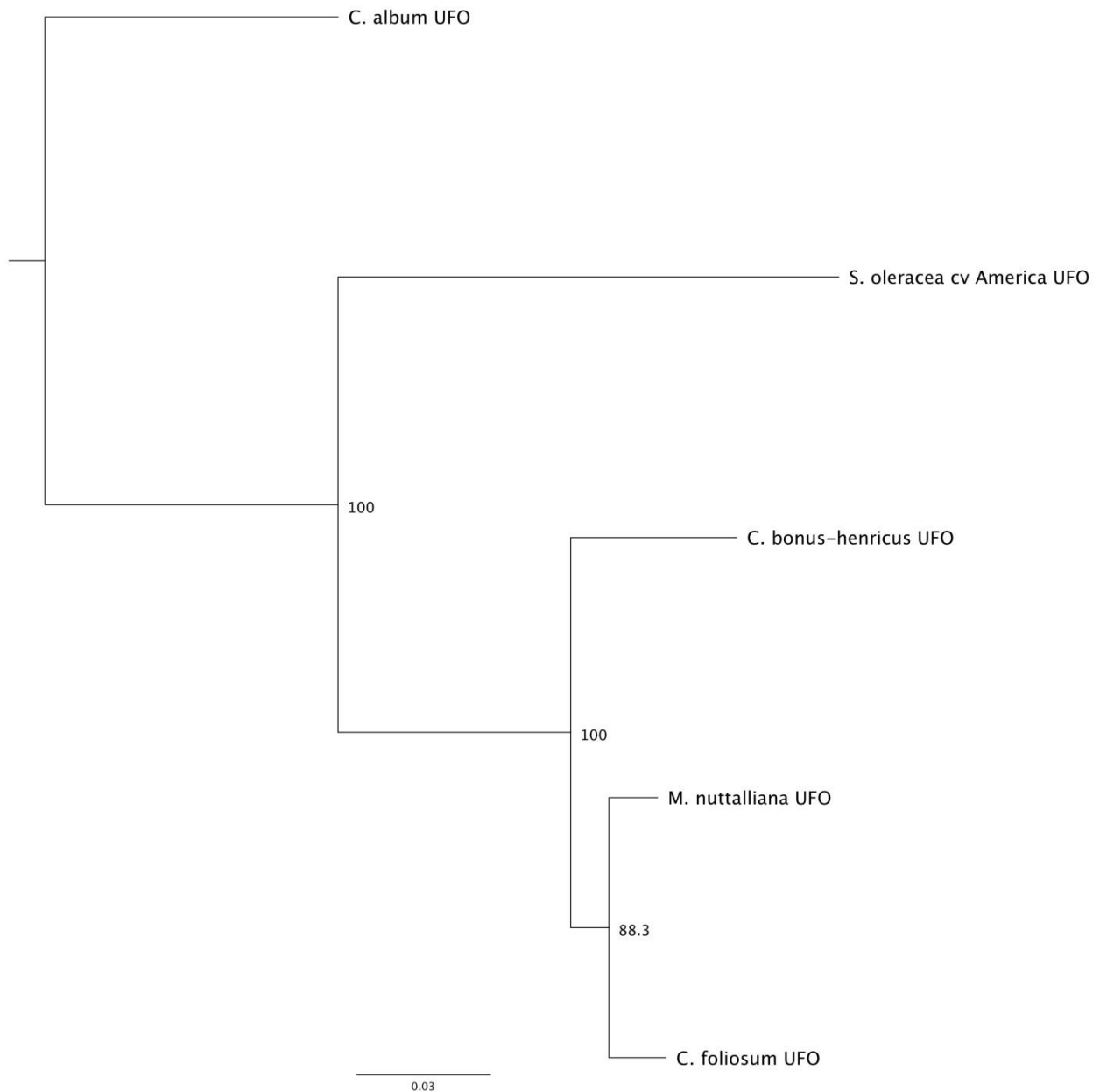
**Figure 20. Phylogenetic tree constructed from the *UFO* alignment from Figure 5, using Geneious tree builder.**

This alignment included all sequences obtained, and the tree construction method used the Tamura-Nei genetic distance model, the neighbor-joining tree build method, bootstrap resampling method with 10,000 replicates, using *C. album* as the rooted outgroup. Topology support (%) among replicates is shown as a node label.



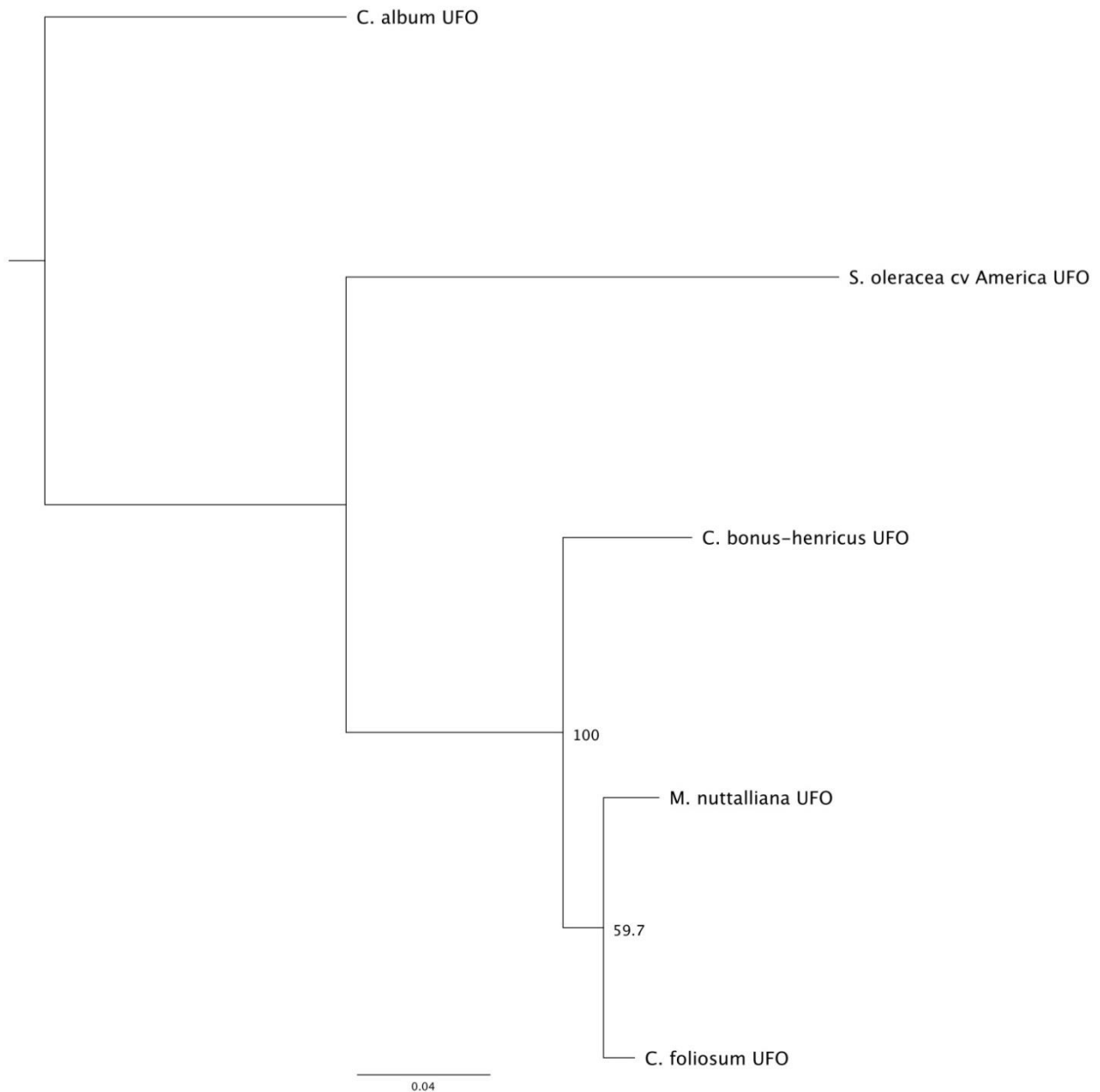
**Figure 21. Phylogenetic tree constructed from the *UFO* alignment from Figure 5, using PhyML.**

This alignment included all sequences obtained, and the tree construction method uses a maximum-likelihood method, the Tamura-Nei substitution model, bootstrap branch support method with 10,000 replicates, taking the best of both NNI and SPR topology searches, and all other settings at default. Topology support (%) among replicates is shown as a node label.



**Figure 22. Phylogenetic tree constructed from the alignment of selected *UFO* sequences from Figure 6, using Geneious tree builder.**

This method used the Tamura-Nei genetic distance model, the neighbor-joining tree build method, bootstrap resampling method with 10,000 replicates, with the support threshold set to 75% and using *C. album* as the rooted outgroup. Topology support (%) among replicates is shown as a node label.



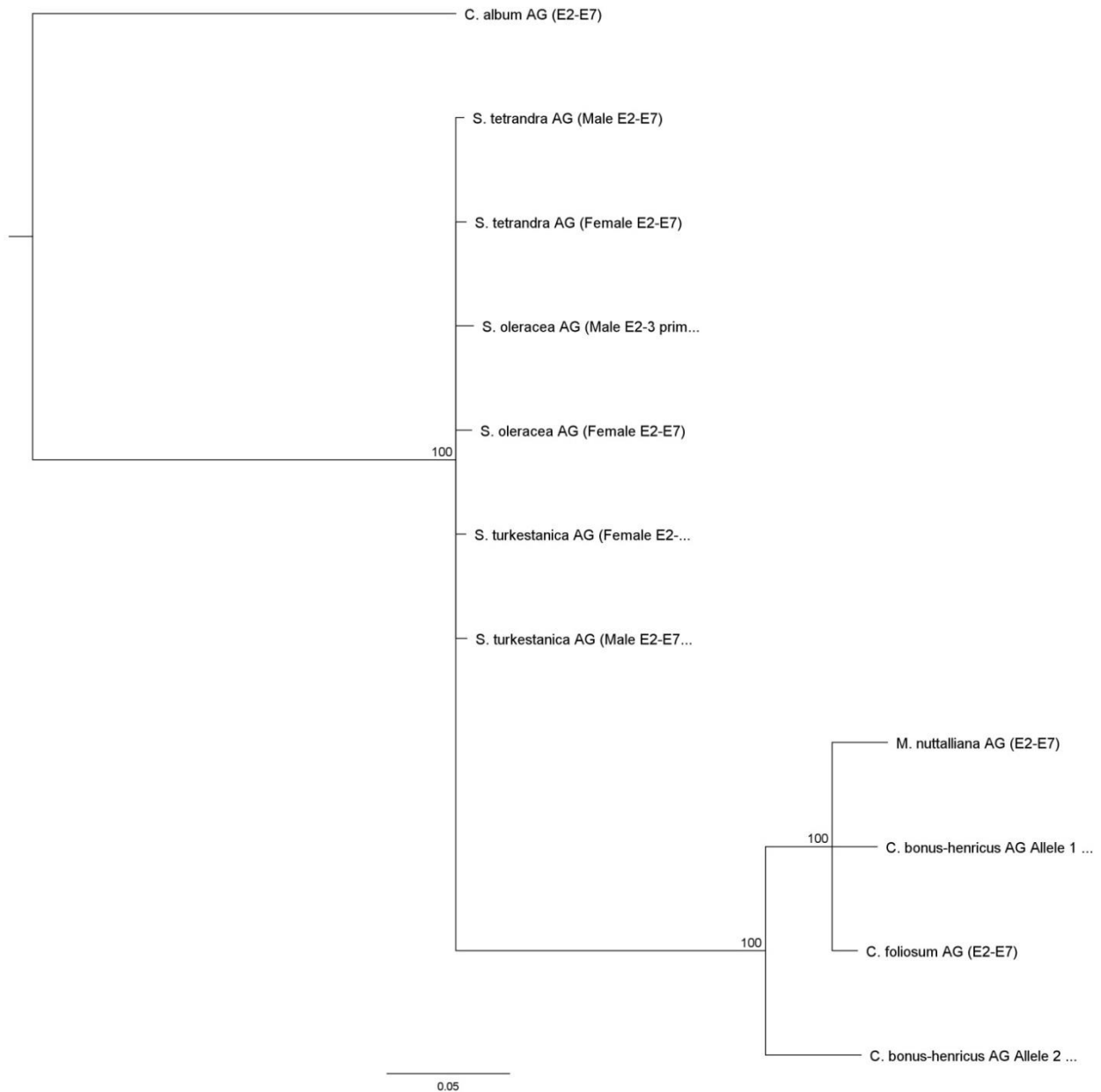
**Figure 23. Phylogenetic tree constructed from the alignment of selected *UFO* sequences from Figure 6, using PhyML.**

The Geneious PhyML plugin uses a maximum-likelihood method, and we used the Tamura-Nei substitution model, bootstrap branch support method with 10,000 replicates, taking the best of both NNI and SPR topology searches, and all other settings at default. Topology support (%) among replicates is shown as a node label.



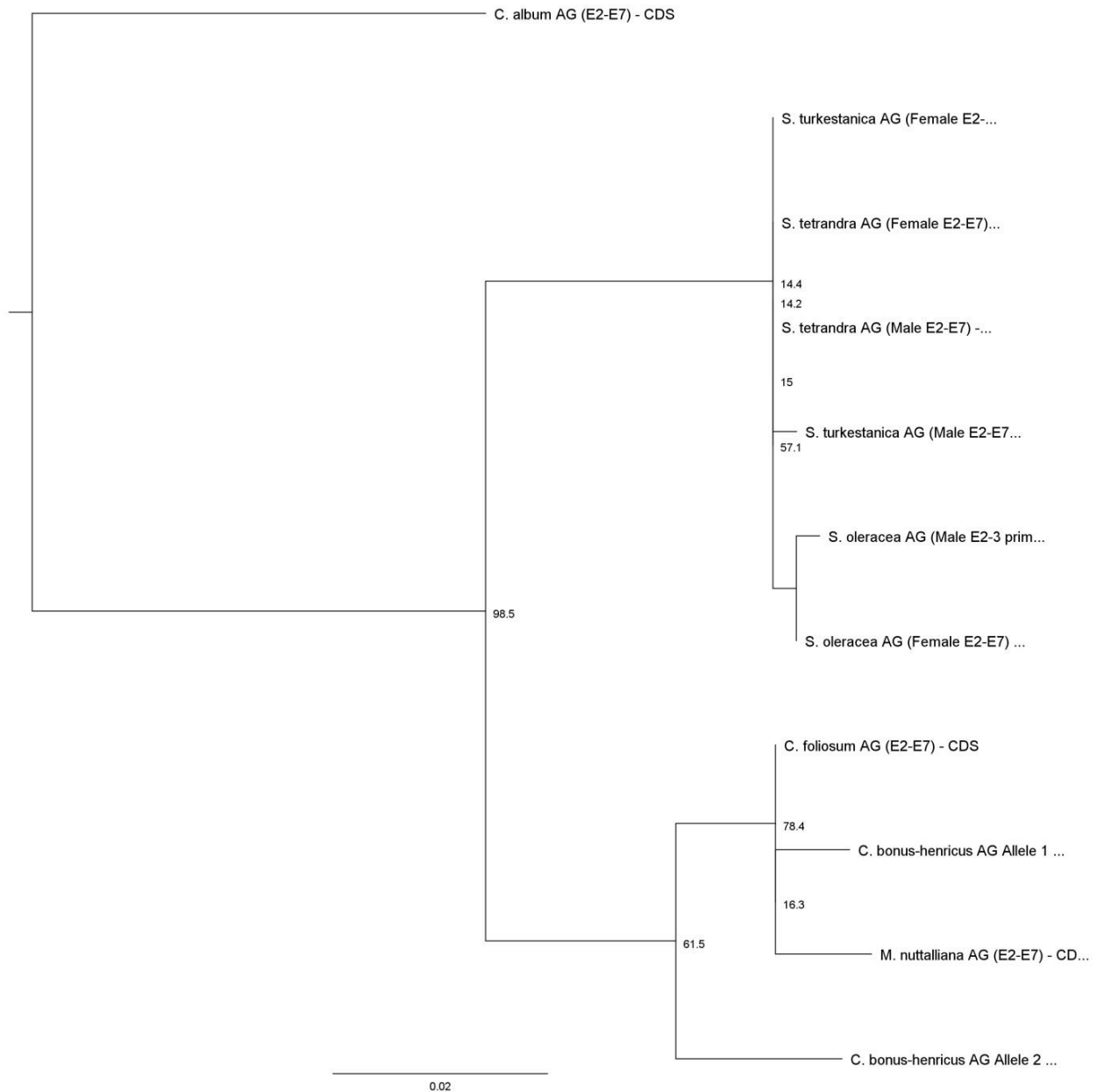
**AG**

The neighbor-joining tree generated using the entire *AG* sequence alignment differs from those generated from the *matK*, *rbcS*, and *UFO* alignments in this study in that the Blitum group branches off of the *Spinacia* group (Figure 24). The major differences seen in the analysis of *AG* sequences between species are in the introns and they are in the form of large insertion/deletion events (data that is not taken into account with the phylogenetic analyses, which treat these regions as gaps or missing data.) Therefore, the alignment of the coding sequences was also analyzed using both the Geneious Tree Builder (neighbor-joining, Figure 26) and the Geneious PhyML plugin (maximum-likelihood, Figure 27) using *C. album* as the rooted node. All the major branch nodes for the CDS tree generated using the Geneious Tree Builder parameters have greater than 90.7% consensus support, with the exception of within the *Spinacia* group (the sequences for which are nearly identical, making phylogenetic inference unreliable). The PhyML trees matched the same basic topology seen with the neighbor-joining tree when only considering the coding sequence (also with the *matK*, *rbcS*, and *UFO* trees previously generated). All trees used *C. album* as the rooted outgroup and grouped *C. bonus-henricus* *AG* ‘Allele 2’ outside the main Blitum *AG* group, supporting this sequence as arising from a gene duplication event and not an allelic difference.



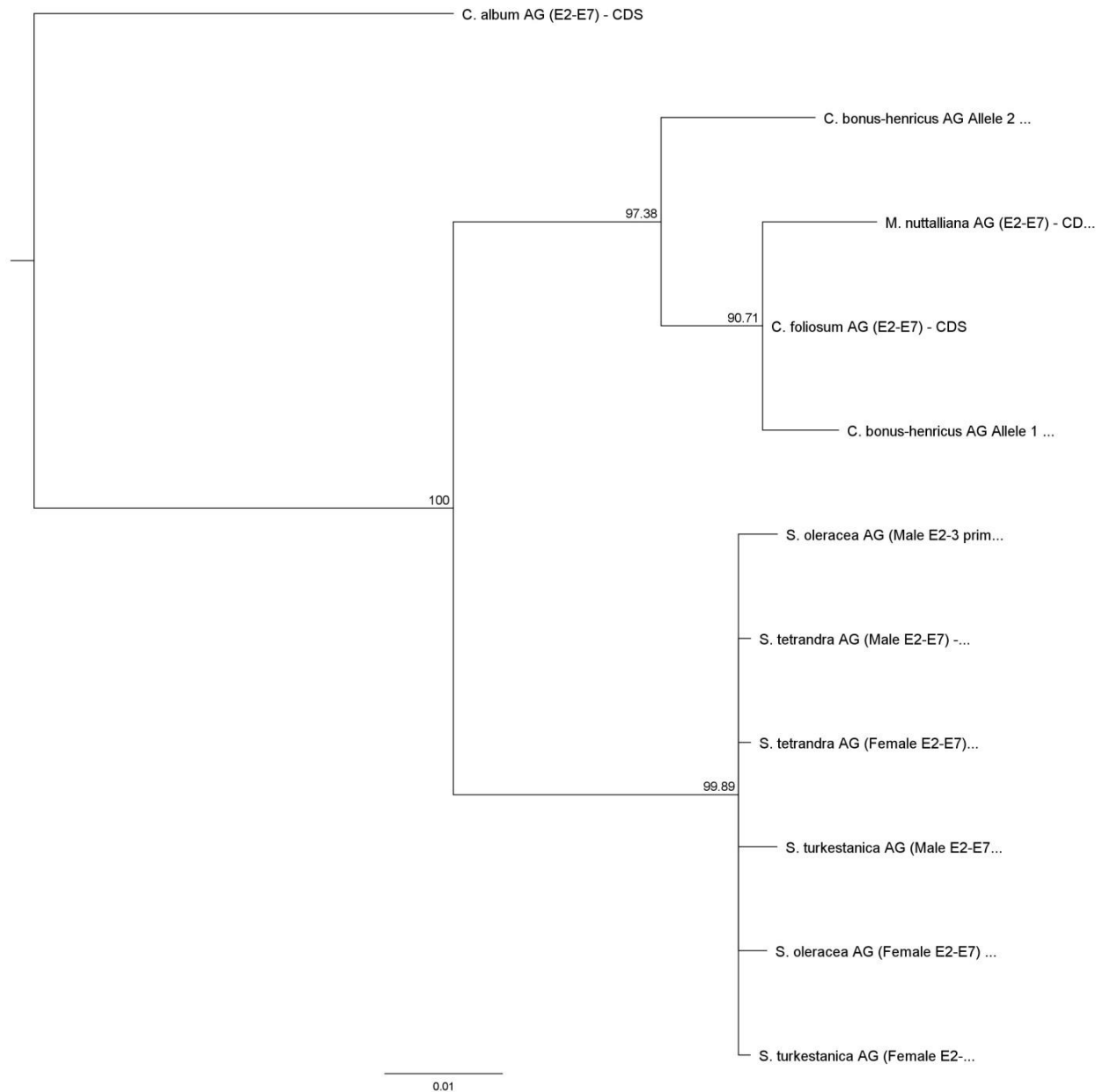
**Figure 24. Phylogenetic tree constructed from the AG alignment from Figure 8 (which considered exon and intron data) using Geneious tree builder.**

This method used the Tamura-Nei genetic distance model, the neighbor-joining tree build method, bootstrap resampling method with 10,000 replicates, with the support threshold set to 85% and using *C. album* as the rooted outgroup. Topology support is 100 % at each supported node, shown as a branch label.



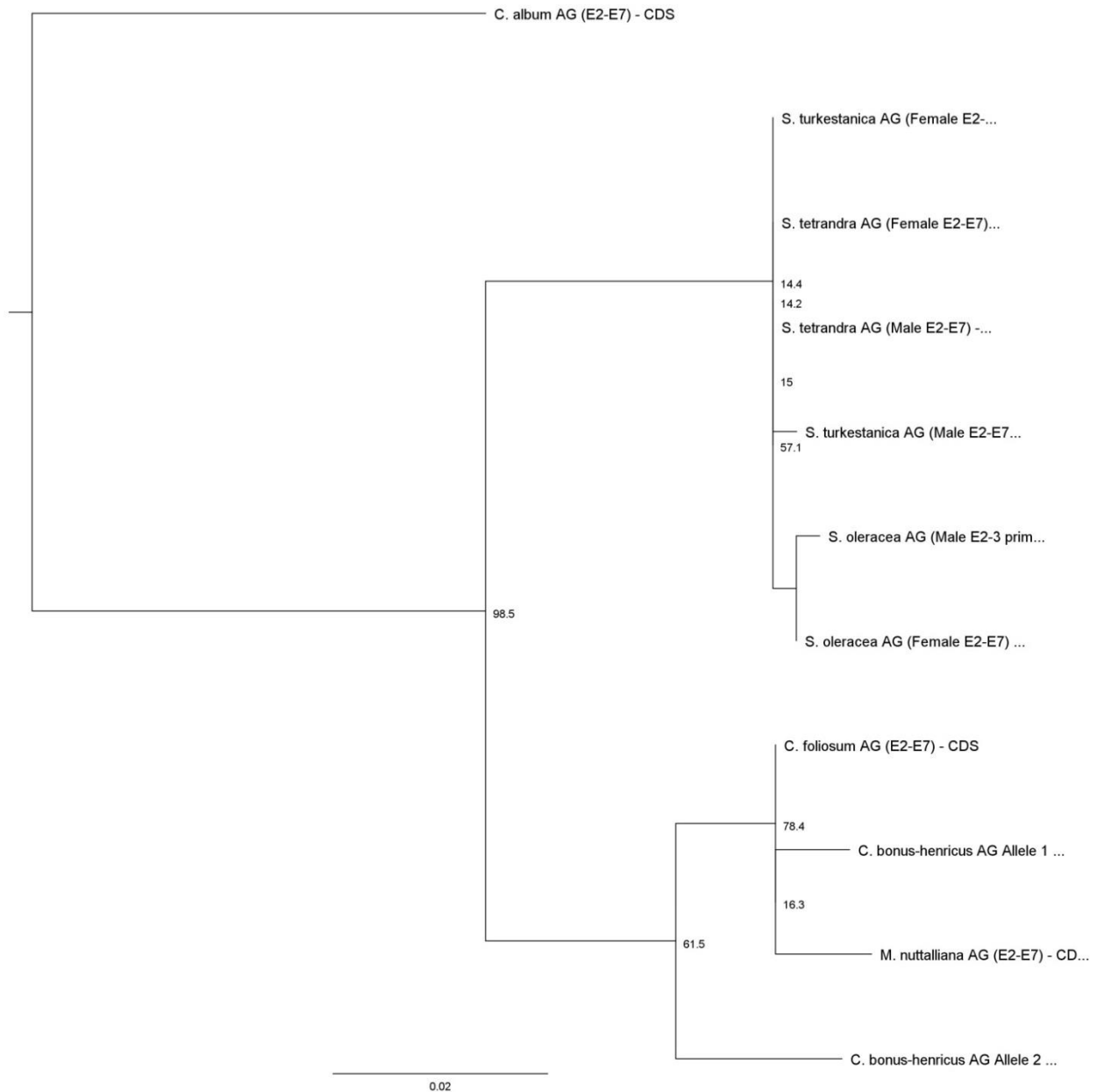
**Figure 25. Phylogenetic tree constructed from the AG alignment from Figure 8 (which considered exon and intron data) using the Geneious PhyML plugin.**

This method uses maximum-likelihood, the Tamura-Nei substitution model, bootstrap branch support method with 1,000 replicates, taking the best of both NNI and SPR topology searches, and all other settings at default. Topology support (%) among replicates is shown as a node label.



**Figure 26. Phylogenetic tree constructed from the AG alignment from Figure 9 (which considered only coding sequence) using Geneious tree builder.**

This method used the Tamura-Nei genetic distance model, the neighbor-joining tree build method, bootstrap resampling method with 10,000 replicates, with the support threshold set to 85% and using *C. album* as the rooted outgroup. Topology support (%) among replicates is shown as a branch label.

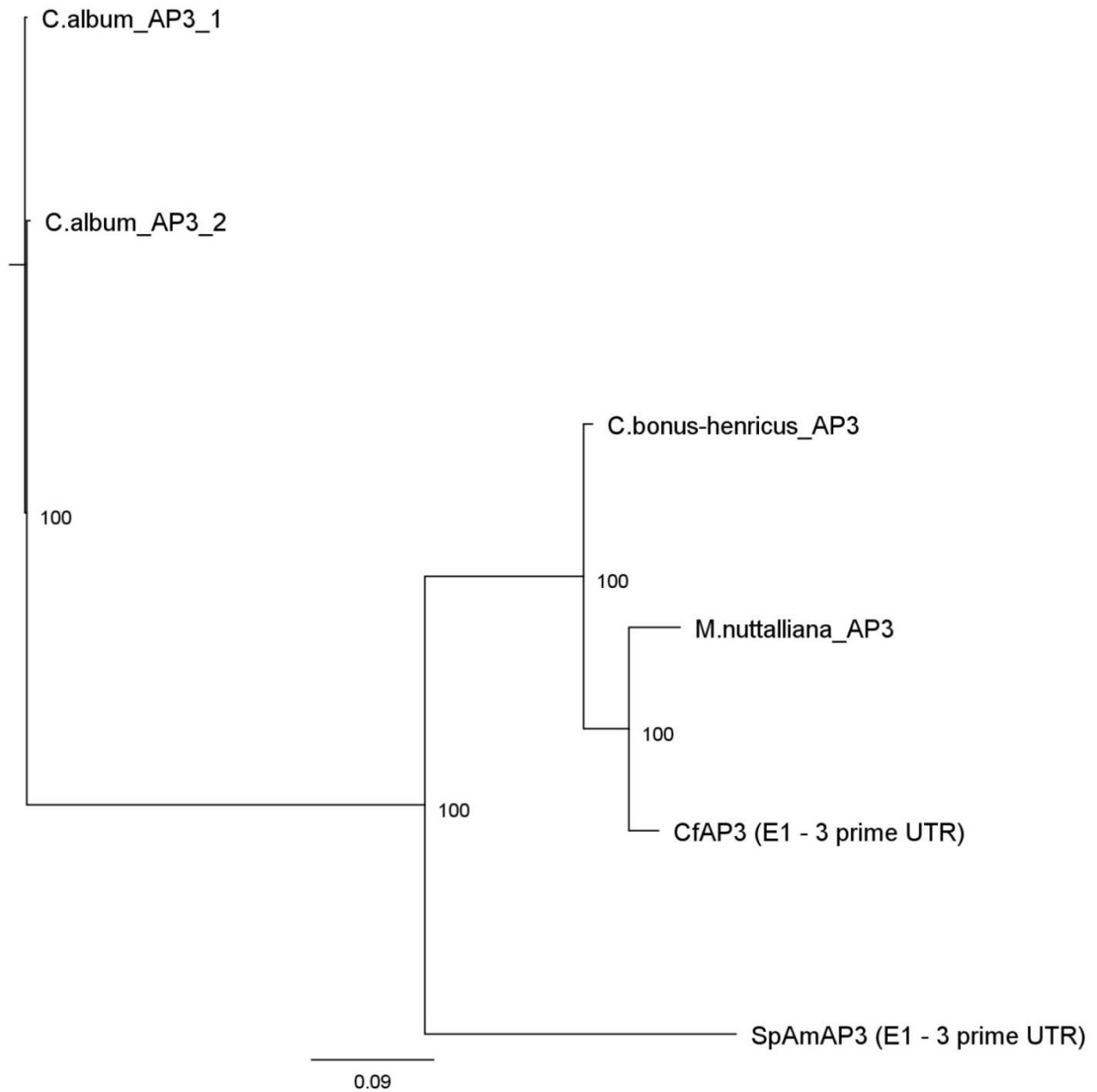


**Figure 27. Phylogenetic tree constructed from the AG alignment from Figure 9 (which considered only coding sequence) using the Geneious PhyML plugin.**

This method uses maximum-likelihood, the Tamura-Nei substitution model, bootstrap branch support method with 1,000 replicates, taking the best of both NNI and SPR topology searches, and all other settings at default. Topology support (%) among replicates is shown as a node label.

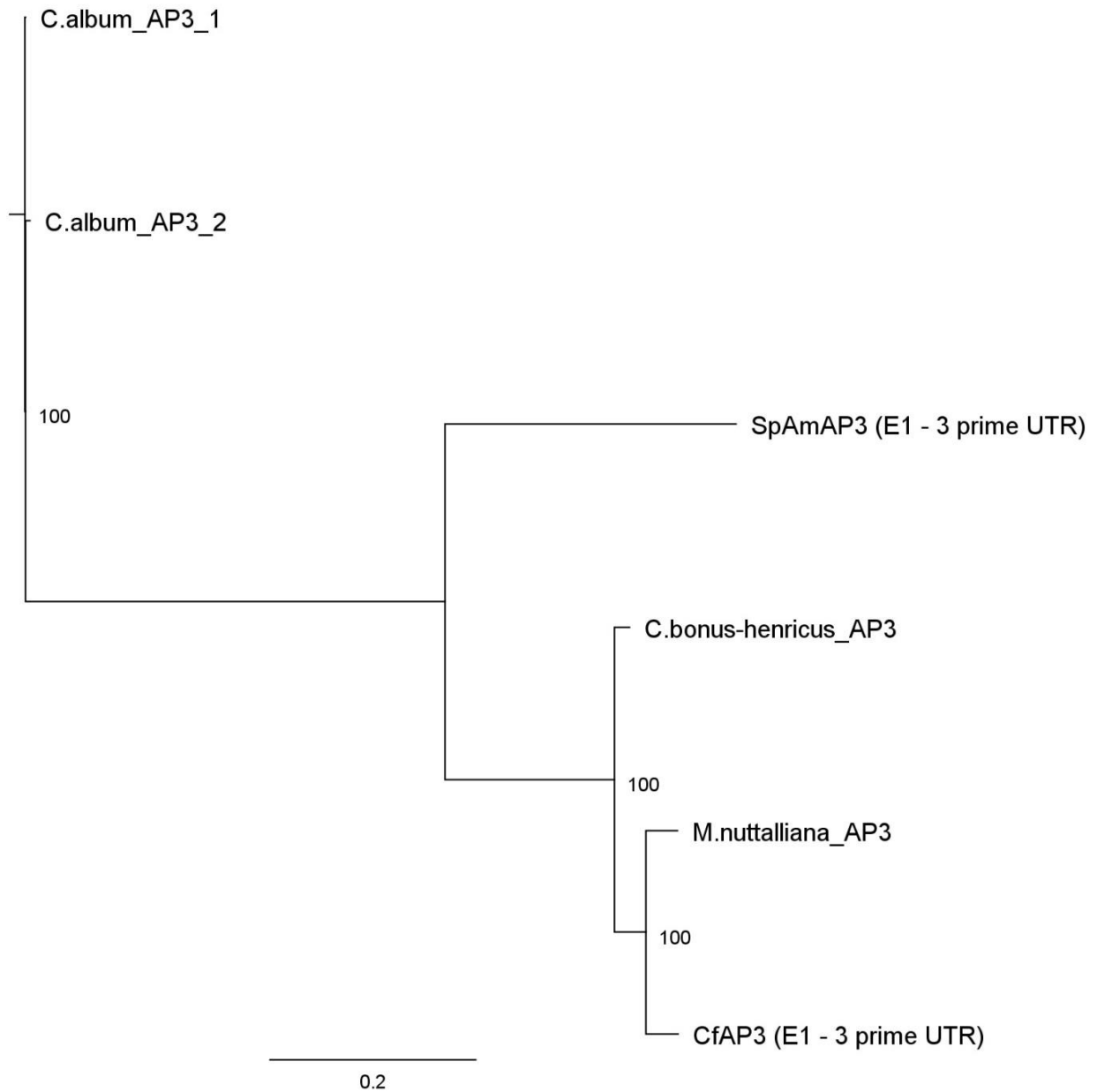
***AP3***

Phylogenetic analysis was performed on the entire *AP3* alignment using both the Geneious Tree Builder (neighbor-joining method, Figure 28) and the Geneious PhyML plugin (maximum-likelihood method, Figure 29) using *C. album AP3* ‘Allele 1’ as the outgroup for rooting. Branch node support for all nodes in both trees is 100%. The *AP3* alignment of the putative exon sequences from Figure 11 was also used for phylogenetic analysis, using the neighbor-joining and maximum-likelihood methods (Figures 30 & 31, respectively). Branch node support for all nodes in these two trees is >86%. All four *AP3* trees generated match the same basic topology as those generated for *matK*, *rbcS*, *UFO*, and *AG*, with *Spinacia sp.* branching from *C. album* first, followed by a splitting of *C. bonus-henricus*, leaving *C. foliosum* and *M. nuttalliana* as sister taxa.



**Figure 28. Phylogenetic tree constructed from the AP3 alignment from Figure 10 (which included exon and intron data) using Geneious tree builder.**

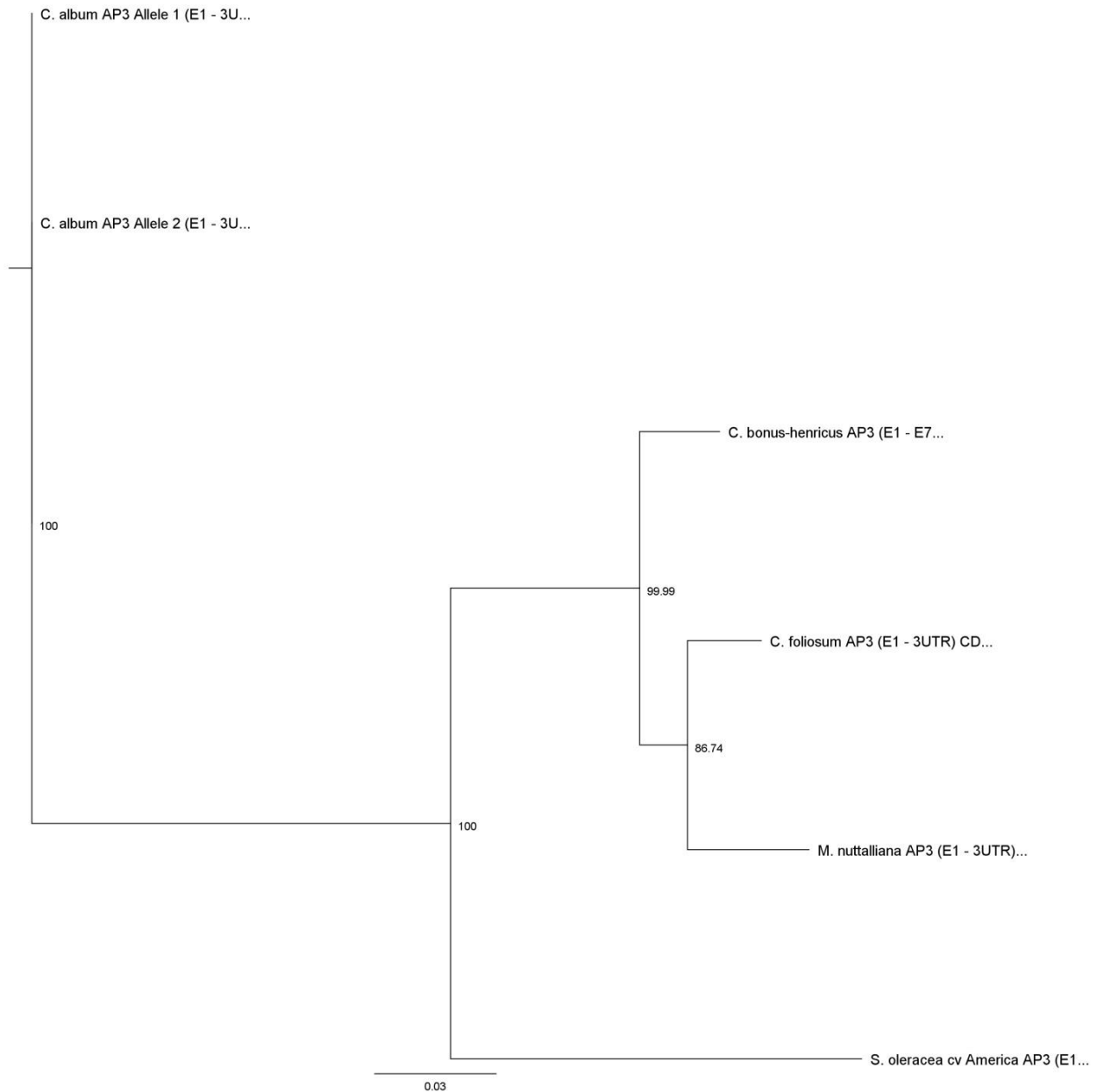
This method used the Tamura-Nei genetic distance model, the neighbor-joining tree build method, bootstrap resampling method with 10,000 replicates, with the support threshold set to 85% and using *C. album* as the rooted outgroup. Topology support (%) among replicates is shown as a node label.



**Figure 29. Phylogenetic tree constructed from the AP3 alignment from Figure 10 (which included exon and intron data) using the Geneious PhyML plugin.**

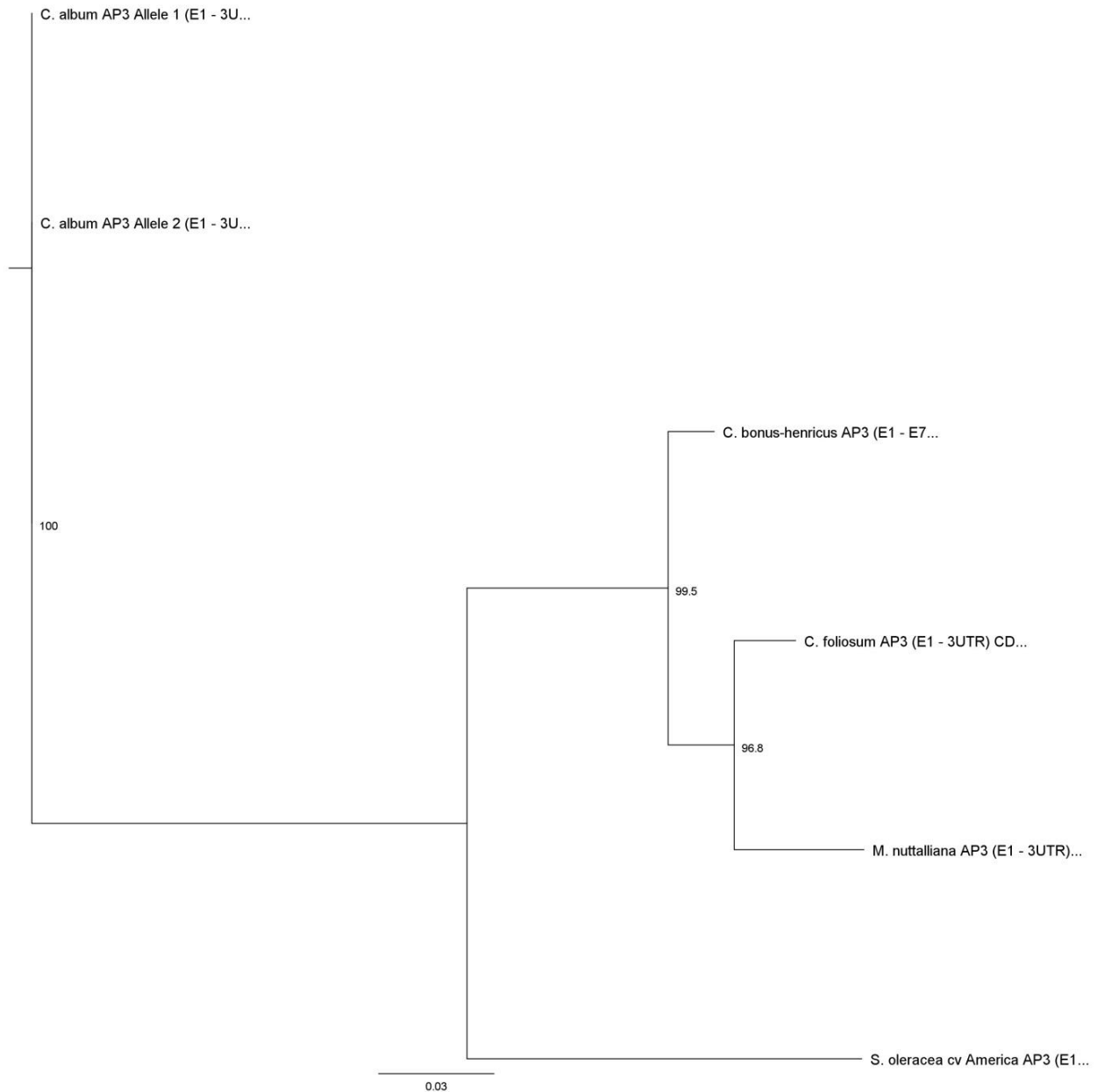
This method uses maximum-likelihood, the Tamura-Nei substitution model, bootstrap branch support method with 1,000 replicates, taking the best of both NNI and SPR topology searches, and all other settings at default, with *C. album* as the rooted outgroup. Topology support (%) among replicates is shown as a node label.





**Figure 30. Phylogenetic tree constructed from the AP3 alignment from Figure 11 (which included only coding regions) using Geneious tree builder.**

This method used the Tamura-Nei genetic distance model, the neighbor-joining tree build method, bootstrap resampling method with 10,000 replicates, with the support threshold set to 85% and using *C. album* as the rooted outgroup. Topology support (%) among replicates is shown as a node label.



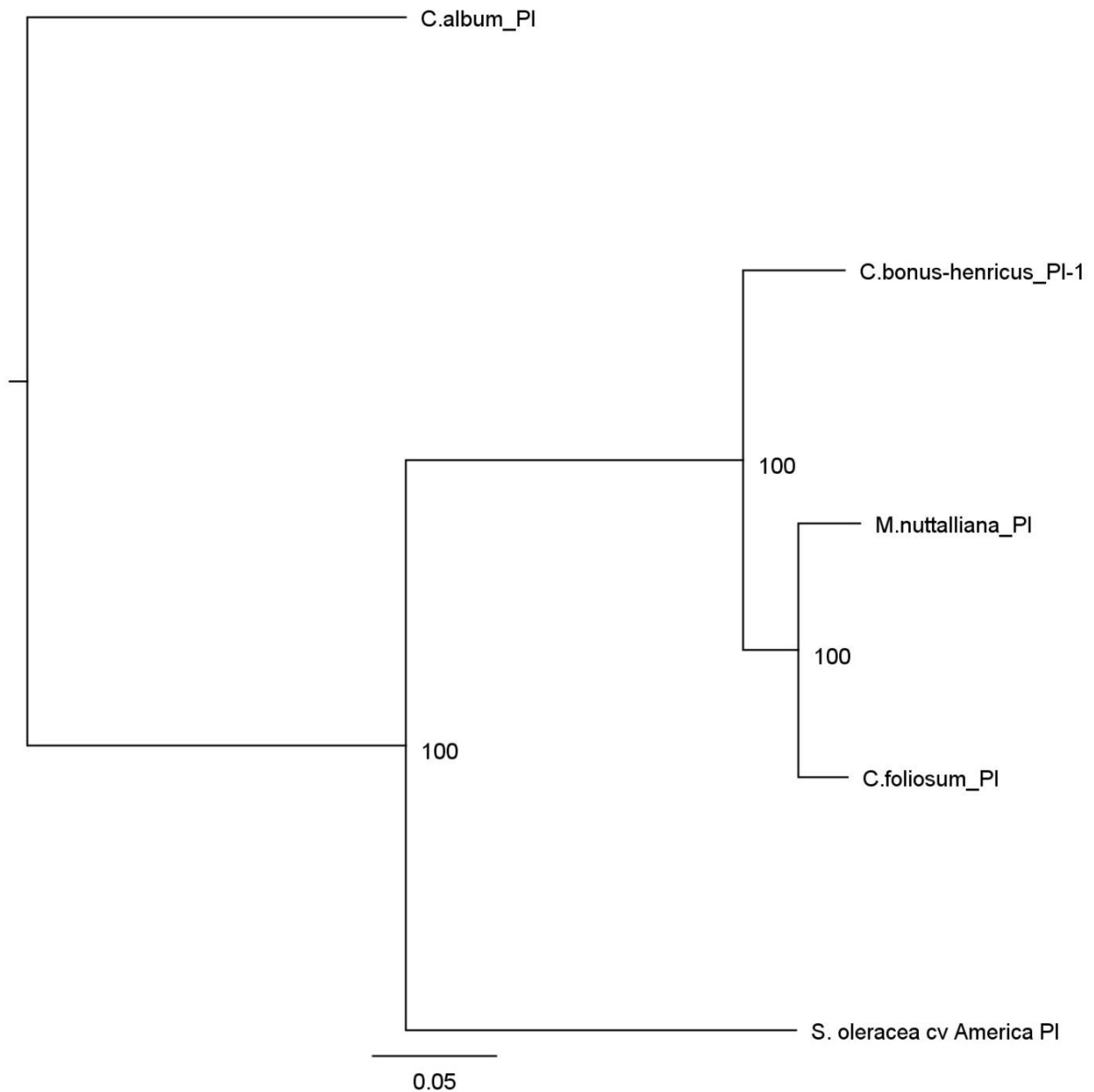
**Figure 31. Phylogenetic tree constructed from the AP3 alignment from Figure 11 (which included only coding regions) using the Geneious PhyML plugin.**

This method uses maximum-likelihood, the Tamura-Nei substitution model, bootstrap branch support method with 1,000 replicates, taking the best of both NNI and SPR topology searches, and all other settings at default, with *C. album* as the rooted outgroup. Topology support (%) among replicates is shown as a node label.

***PI***

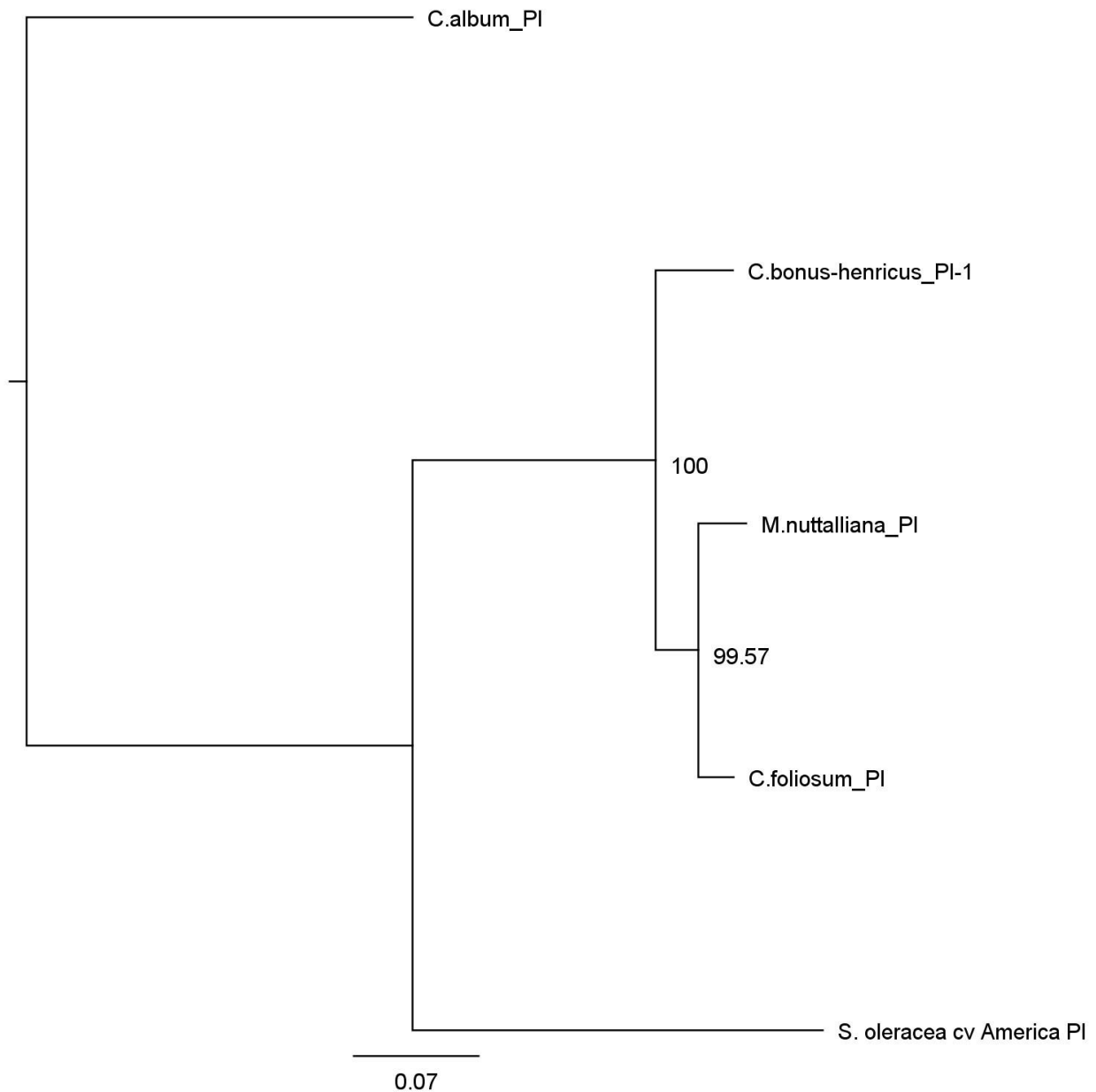
We used the neighbor-joining and PhyML methods to generate trees from the alignments from Figure 12 that included introns (Figures 32 & 33). The tree topologies seen in these analyses match precisely with those generated for *matK*, *rbcS*, *UFO*, *AG*, and *AP3*, with *Spinacia sp.* branching from *C. album* first, followed by a splitting of *C. bonus-henricus*, leaving *C. foliosum* and *M. nuttalliana* as sister taxa. All nodes from these two consensus trees were supported at >99.5% with 10,000 replicates.

For the PISTILLATA gene, we were only able to obtain some of the homologous or allelic sequences from 3' RACE experiments, so the analysis that includes these sequences utilizes exon data only (Figure 13). For the purposes of our homolog analysis, we use the *C. album PI* ortholog as the outgroup. We used the neighbor-joining and PhyML methods to generate trees for the alignment of the exon sequences from Figure 13, including *PI* alleles and *PI-like* sequences (Figures 34 & 35). When using *C. album PI-* as the rooted outgroup, we see the *PI-like* paralogous sequences branching off and grouping together; the remaining orthologous sequences form the same topology shared with the two previously-mentioned *PI* phylogenies (which included only sequences orthologous to the *Spinacia oleracea cv America PI* sequence (GQ120477)). In these trees we see that the *C. bonus-henricus PI* alleles are sister to each other; these two sequences differ mostly in the 3'UTR region, having 89.3% identity with each other (Table 11). This topology suggests a shared ancestral duplication that we did not see in the other species.



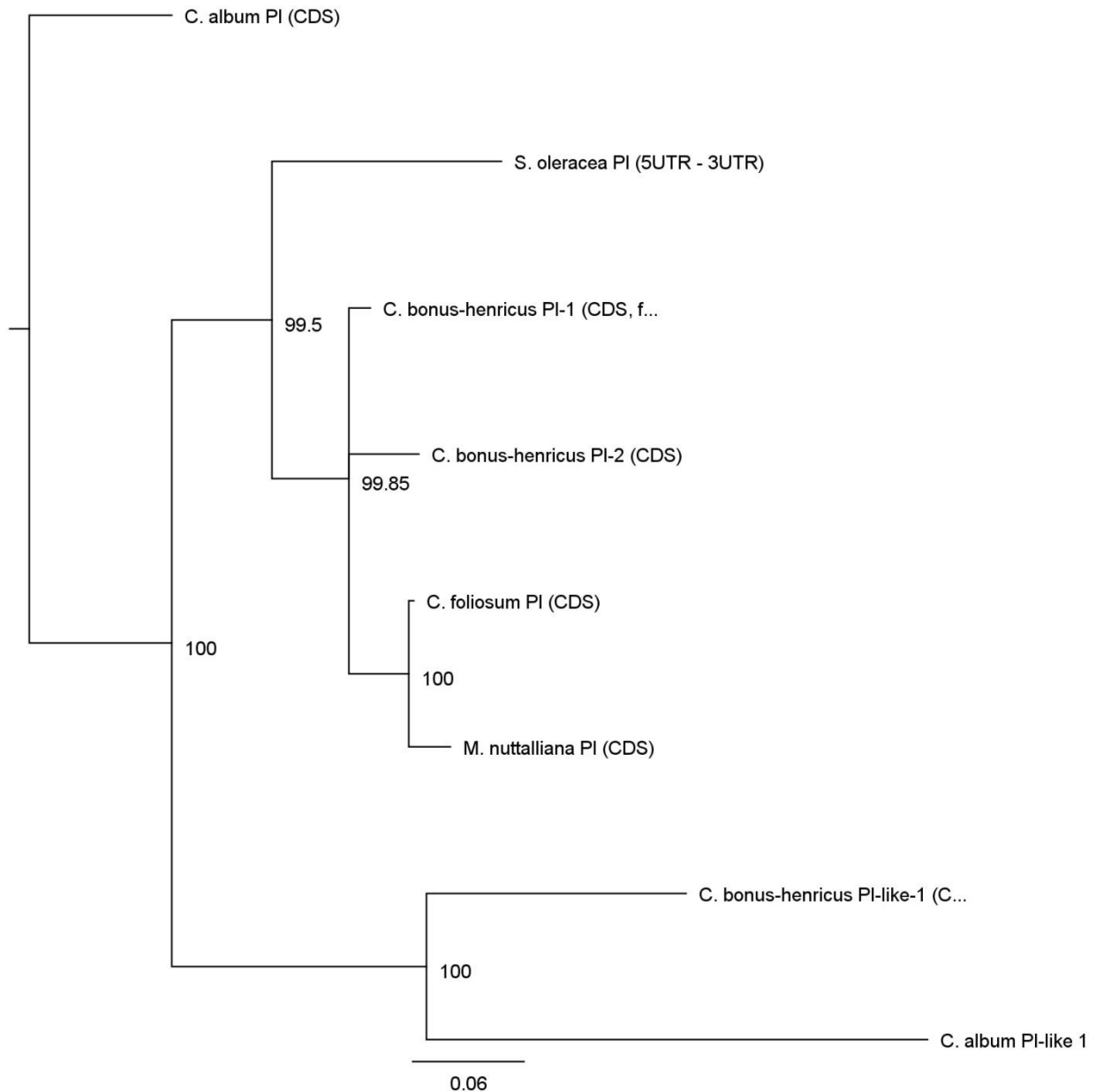
**Figure 32. Phylogenetic tree constructed from the *PI* alignment from Figure 12 (which included exon & intron data), using Geneious tree builder.**

This method used the Tamura-Nei genetic distance model, the neighbor-joining tree build method, bootstrap resampling method with 10,000 replicates, and using *C. album* as the rooted outgroup. Topology support among replicates is 100% at every node.



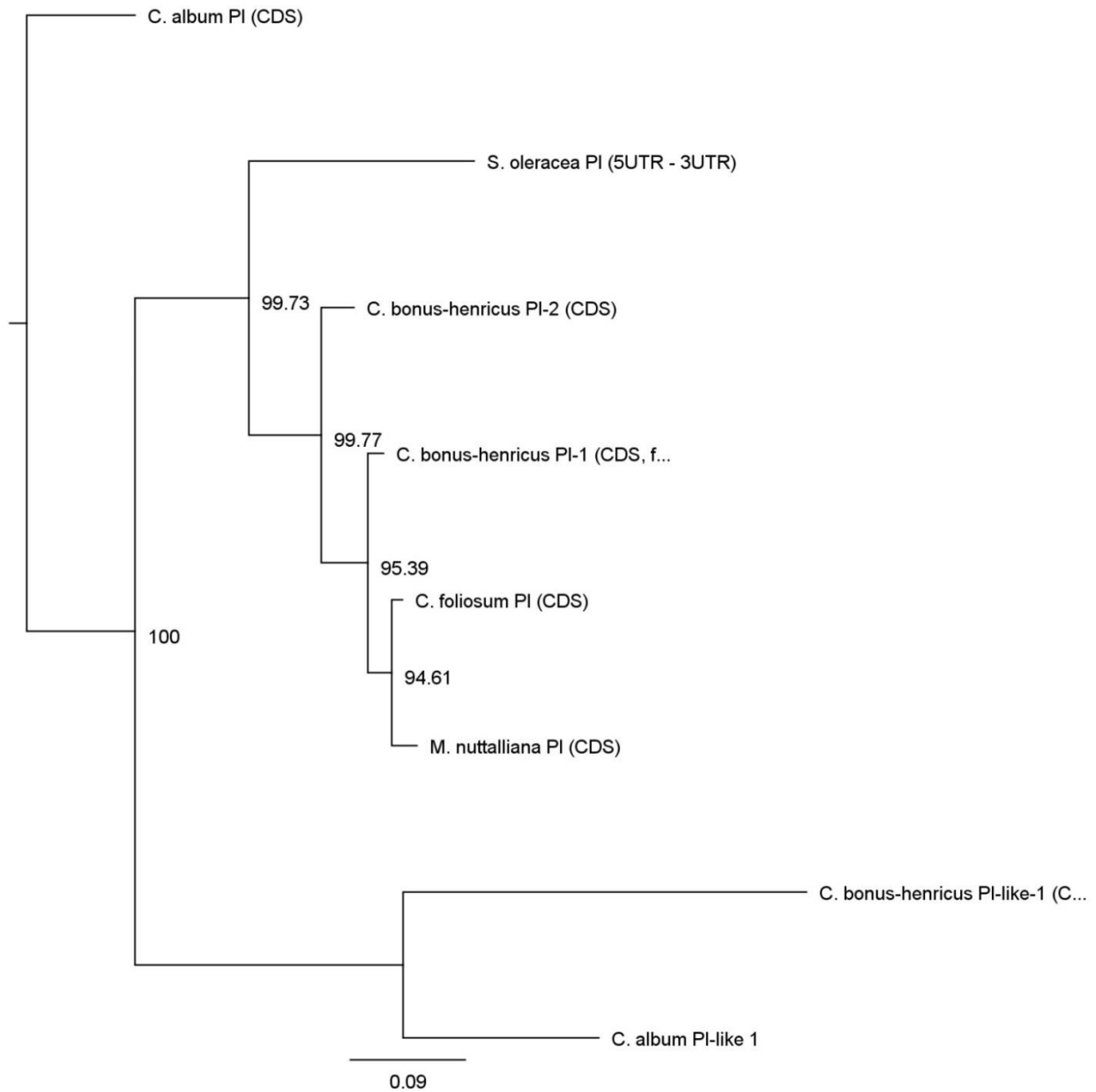
**Figure 33. Phylogenetic tree constructed from the *PI* alignment from Figure 12 (which included exon & intron data) using the Geneious PhyML plugin.**

This method uses maximum-likelihood, the Tamura-Nei substitution model, bootstrap branch support method with 10,000 replicates, taking the best of both NNI and SPR topology searches, and all other settings at default. Topology support (%) among replicates is shown as a node label and *C. album* was used as the rooted outgroup.



**Figure 34. Phylogenetic tree constructed from the *PI* alignment from Figure 13 (exon and 3'UTR data only) using Geneious tree builder.**

This method used the Tamura-Nei genetic distance model, the neighbor-joining tree build method, bootstrap resampling method with 10,000 replicates, with the support threshold set to 85% and using the *C. album* *PI* ortholog as the rooted outgroup. Topology support (%) among replicates is shown as a node label.



**Figure 35. Phylogenetic tree constructed from the *PI* alignment from Figure 13 (exon and 3'UTR data only) using the Geneious PhyML plugin.**

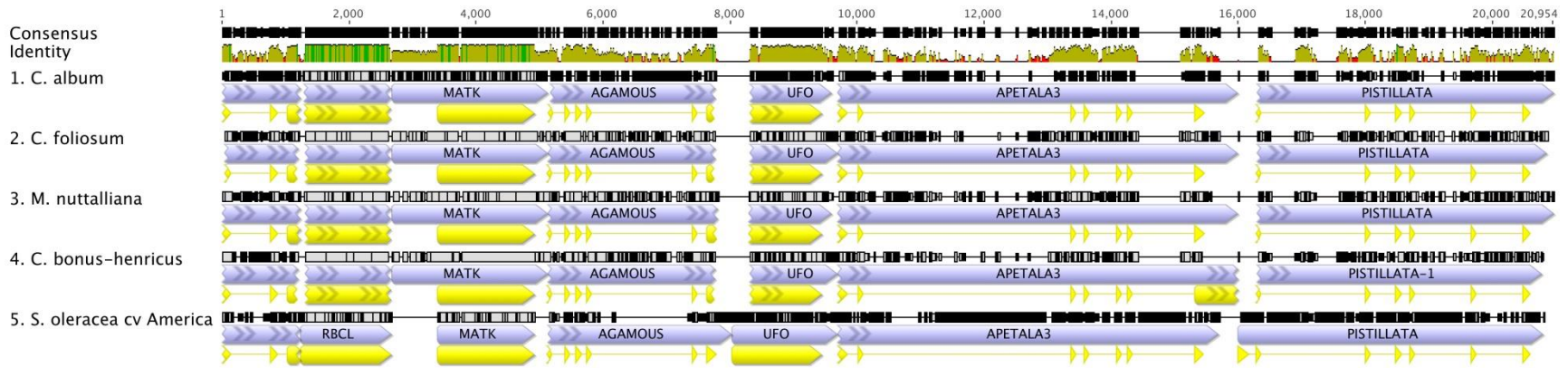
This method uses maximum-likelihood, the Tamura-Nei substitution model, bootstrap branch support method with 10,000 replicates, taking the best of both NNI and SPR topology searches, and all other settings at default. Topology support (%) among replicates is shown as a node label.

### *Concatenated sequences*

All of the sequences for a given species were concatenated [55] and these ‘super-genes’ were aligned using the MAFFT plugin (version 1.3) for Geneious using the E-INS-i algorithm, PAM-200 (k=2) scoring matrix, and a gap open penalty of 1.53. Alignment results indicated the best result when the E-INS-i algorithm was used. This algorithm is slow and best used for sequences with multiple conserved domains and long gaps [52], although visual inspection and manual correction for errors was still used. The resulting alignment is over 20 kB in length (Figure 36). Among species there is a range of sequence identities from 41.3% between *S. oleracea cv America* and *C. album* to 91.5% between *C. foliosum* and *M. nuttalliana* (Table 12).

Phylogenetic analysis was performed on the ‘super-gene’ alignment using both the Geneious Tree Builder (neighbor-joining, Figure 37) and the Geneious PhyML plugin (maximum-likelihood, Figure 38) using *C. album* as the rooted node. Both trees matched the basic topology as those generated for *matK*, *rbcS*, *UFO*, *AG*, *AP3*, and *PI*: with *Spinacia sp.* branching from *C. album* first, followed by a splitting of *C. bonus-henricus*, leaving *C. foliosum* and *M. nuttalliana* as sister taxa.



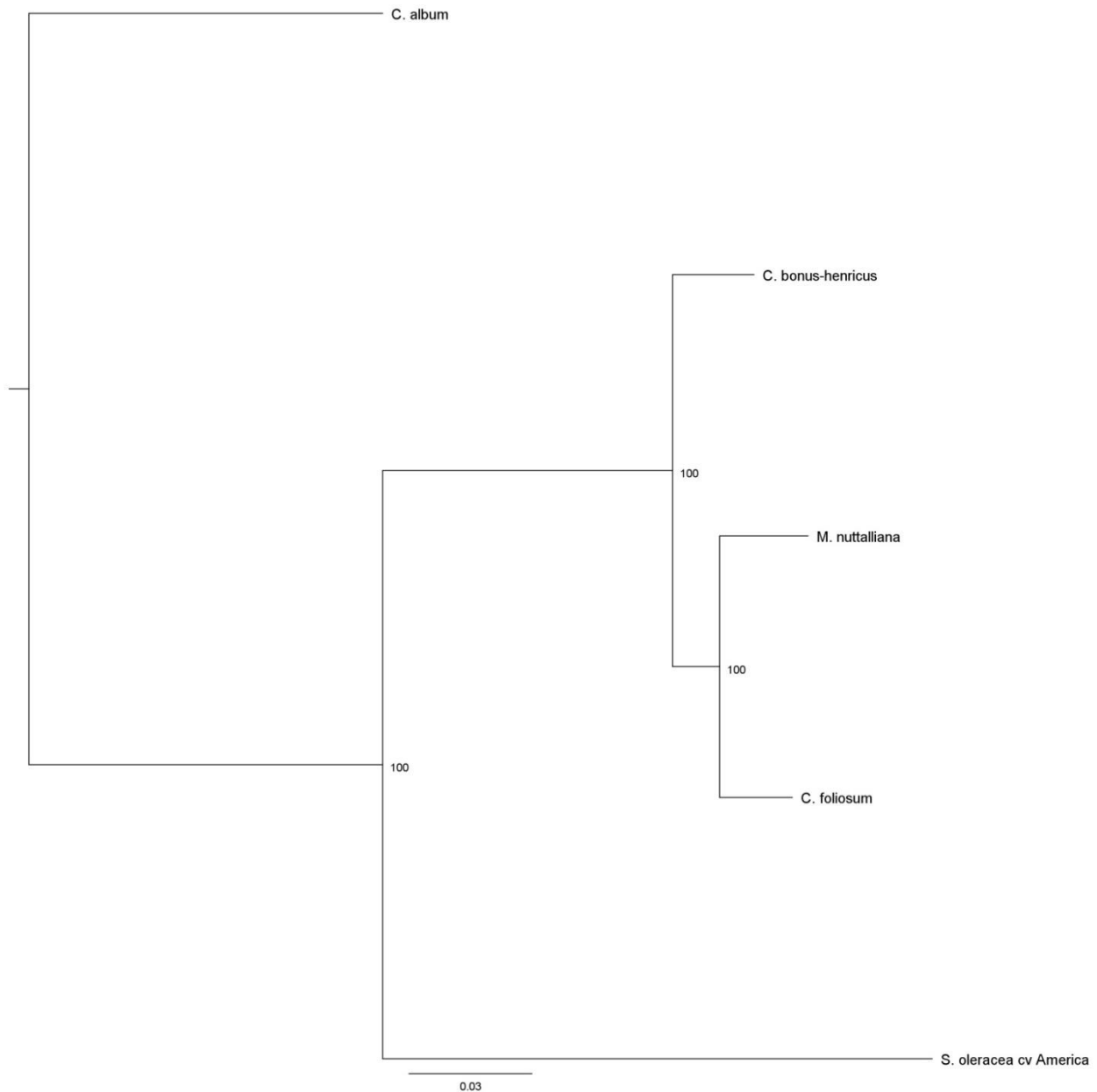


**Figure 36. Pairwise alignment of concatenated ‘super-gene’ sequences.**

Alignment was performed with the MAFFT plugin (version 1.3) for Geneious using the E-INS-i algorithm, PAM-200 scoring matrix (k=2), and a gap open penalty of 1.53. Each individual gene is annotated with a grey arrow and exons are annotated with yellow arrows.

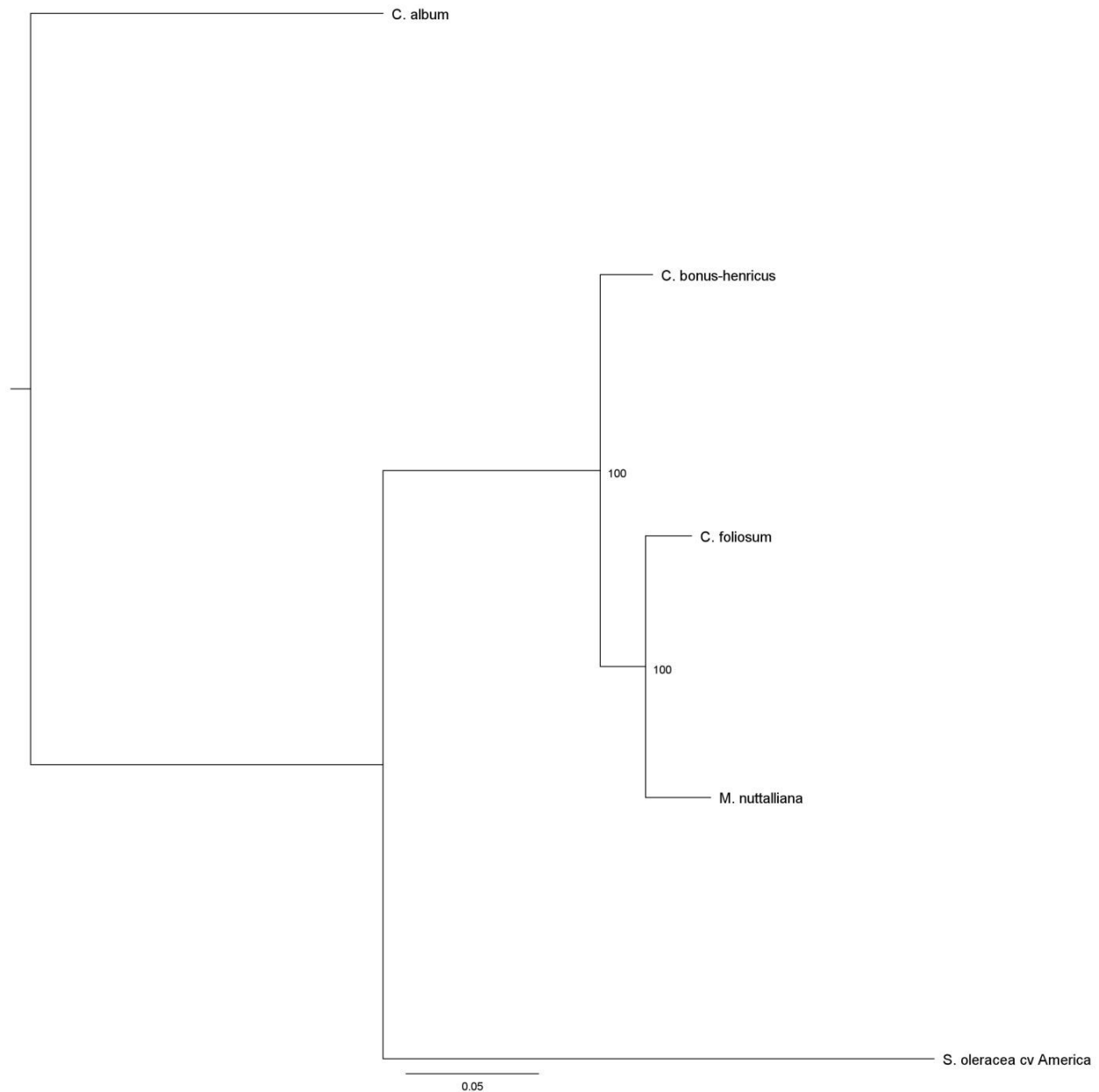
	C. album	C. foliosum	M. nuttalliana	C. bonus-henricus
C. foliosum	61			
M. nuttalliana	60.8	91.5		
C. bonus-henricus	61.3	86.3	85.8	
S. oleracea cv America	41.3	45.7	45.6	44.7

**Table 12. Identity (%) of sequence pairs of concatenated sequences in the alignment from Fig. 14.**



**Figure 37. Phylogenetic tree constructed from the ‘super-gene’ alignment from Figure 14 using Geneious tree builder.**

This method used the Tamura-Nei genetic distance model, the neighbor-joining tree build method, bootstrap resampling method with 10,000 replicates, with the support threshold set to 85% and using *C. album* as the rooted outgroup. Topology support among replicates is 100% and is shown as a node label.



**Figure 38. Phylogenetic tree constructed from the ‘super-gene’ alignment from Figure 14 using the Geneious PhyML plugin.**

This method uses maximum-likelihood, the Tamura-Nei substitution model, bootstrap branch support method with 10,000 replicates, taking the best of both NNI and SPR topology searches and all other settings at default. Topology support among replicates is 100% and is shown as a node label.

### Rates testing & tests of selection

We tested for uniformity of rates of nucleotide substitution in the genes in this work by performing a relative rates test (Tajima's) for each selected ingroup species compared to *Spinacia sp.*, with *C. album* or *B. vulgaris* as the outgroup taxon.

#### *rbcS*

There is no evidence of differential rates of evolution between species pairs for the *rbcS* gene (Table 13). Subsequently, we tested for patterns of selection within the coding regions. For *rbcS* we are able to reject the null hypothesis of strict neutrality between the sequences analyzed (Table 14). There is no evidence for positive selection; however, we do see evidence of purifying selection (Table 15).

#### *rbcL*

Broadly speaking *M. nuttalliana*, *C. bonus-henricus*, and *C. foliosum* do not differ significantly in the rates of nucleotide substitution. Similarly, *Spinacia oleracea* and *C. album* appear to have similar rates (Table 16). For *rbcL* we are able to reject the null hypothesis of strict neutrality between the sequences analyzed, with the exception being between *C. foliosum* and *C. bonus-henricus* (Table 17). As with *rbcS*, there is no support for positive selection, but there is significant evidence of purifying selection in each sequence pair analyzed (Table 18).

#### *matK*

There is no evidence of differential rates of evolution between *Spinacia sp.* and *C. foliosum*, between *Spinacia sp.* and *M. nuttalliana*, and between *C. bonus-henricus* and *C. foliosum* (Table 19). There is evidence of unequal rates of evolution between *Spinacia sp.* and *C. bonus-henricus*, between *C. bonus-henricus* and *M. nuttalliana*, and between *C. foliosum* and *M.*

*nuttalliana*. We also analyzed the *matK* sequences for evidence of selection. For this gene we are able to reject the null hypothesis of strict neutrality between the sequences analyzed, with exception being between *C. bonus-henricus* and *C. foliosum* (Table 20). For all other sequence pairs, we see evidence of purifying selection (Table 21).

### ***UFO***

Within the *Spinacia sp.* group, we did not find any evidence to suggest unequal rates of evolution when using *C. album* as the outgroup. However, rates of evolution are not equal between *Spinacia sp.* and each ingroup taxon (Table 22). Based on the number of unique differences seen in *Spinacia*, the rate of evolution appears to be faster than its closest relatives.

We tested for patterns of selection within the dataset that included all of the *Spinacia UFO* alleles. For *UFO* we are able to reject the null hypothesis of strict neutrality between ingroup species, and between the sequences for *Spinacia sp.* and *UFO* “Allele C” from *Spinacia tetrandra* (Table 23). There is no evidence for positive selection; however, we do see evidence of purifying selection (Table 24). However, these data do not take into account indel regions, which are treated as gaps/missing data. In the hypervariable region, we did not see any association between the number of repeat units and any other SNP pattern in the rest of the gene sequence.

### ***AG***

Rates of nucleotide substitution among the *AG* genes appear to be constant, with the exception being between *Spinacia oleracea cv America* and *C. foliosum* (Table 25). However, this analysis treats gaps in the alignment by deleting them and does not take into account the major sequence differences seen as indels. We found evidence that the sequences are not experiencing neutral selection for all species pairs considered, except for between *M. nuttalliana*

and *C. foliosum*, between *C. foliosum* and *C. bonus-henricus* ‘Allele 1’, and between AG alleles of *C. bonus-henricus* (Table 26). We found evidence for purifying selection between all of those pairs for which the neutral model was rejected (Table 27).

### ***AP3***

For the *AP3* gene, we did find that evolutionary rates were not equal between *C. foliosum* and *S. oleracea*, between *C. bonus-henricus* and *S. oleracea*, between *C. bonus-henricus* and *C. foliosum*, and between *M. nuttalliana* and *C. bonus-henricus* (Table 28). These rate inequalities combined with the branch lengths seen in phylogenetic analyses indicate that the *AP3* gene is evolving at a faster rate in *Spinacia sp.* compared to that rate seen in the *Blitum* group. We also tested for directionality of selection in MEGA, and we are able to reject the null hypothesis of strict neutrality between *Spinacia oleracea* and the other species’ *AP3* sequences (Table 29). There is evidence of purifying selection in *Spinacia oleracea* (Table 30). Among all other species pairs, we cannot reject the null hypothesis of neutral evolution.

### ***PI***

Rates testing was performed on the complete *PI* sequences from Figure 12. Rates of evolution are unequal between *Spinacia sp.* and *C. foliosum*, between *Spinacia sp.* and *C. bonus-henricus*, but rates are equal between *S. oleracea* and *M. nuttalliana* and in the other species pairs examined (Table 31). We are able to reject the null hypothesis of strict neutrality of evolution between *Spinacia oleracea* and the ingroup species, as well as between *C. album* and the ingroup species. We see evidence of purifying selection for these same species pairs (Tables 32 & 33).

### *Concatenated sequences*

The entire concatenated dataset was used for rates analysis. All nucleotide substitutions were considered at all codon positions and noncoding sites, with complete deletion of gaps/missing data in the alignment (Table 34). This analysis was also run on a concatenated dataset that excluded the *rbcL* gene from consideration (Table 35), as its analyses were incongruent with the others and has the potential to generate evolutionary noise [66]. The *rbcL* gene does not provide convincing support and is an outlier in this and other analyses [13], due to its potential non-clocklike rate of non-synonymous substitution [67]. The differences between these two rates tests and those generated on the individual genes in this study indicate that the *rbcL* gene appears to be evolving at a rate inconsistent with other genes in this work and is a potential source of error in the phylogenetic analysis of the species in *Anserineae*.

We subdivided the concatenated dataset into a super-gene consisting of genes thought to be related to the evolution of dioecy (*AP3*, *PI*, and *UFO*) and a super-gene consisting of genes not thought to be related to dioecy (*rbcS*, *matK*, and *AG*). These concatenated sequence alignments were created using the MAFFT plugin (version 1.3) for Geneious and the E-INS-i algorithm, PAM-200 (k=2) scoring matrix, and a gap open penalty of 1.53, although visual inspection and manual correction was still necessary. The alignments were exported into MEGA and annotated for genes and coding domains, then used for rates analysis. The rates testing of genes not thought related to dioecy (Table 36) indicates that *Spinacia sp.* appears to be evolving faster than *C. foliosum* and *C. bonus-henricus*, and that *M. nuttalliana* is also evolving at a slightly faster rate than its' closest relative *C. foliosum*. When we consider just the B-class and *UFO* genes (Table 37), *Spinacia sp.* appears to be evolving faster than the entire '*Blitum*' group, but that *M. nuttalliana* is not evolving at a rate faster than *C. foliosum*, instead evolving at a rate

faster than *C. bonus-henricus*. It is highly likely that these differences in rates analysis that are seen within the '*Blitum*' group reflect a loss of resolution due to the high degree of similarity seen between these species across all genes in this study, and the non-inclusion of indel variation in the rates analyses.



	<i>M. nuttalliana</i> ‘A’	<i>C. foliosum</i>	<i>C. bonus-henricus</i>
<i>C. foliosum</i>	0.00 (1.00000)		
<i>C. bonus-henricus</i>	0.01 (0.90683)	0.36 (0.54722)	
<i>S. oleracea</i>	2.14 (0.14343)	0.94 (0.33280)	2.45 (0.11719)

**Table 13. Tajima’s test statistics for *rbcS*.**

Test statistics from the *rbcS* relative rates tests (Tajima’s) for each species pair with the P-values given in parentheses. A P-value less than 0.05 is often used to reject the null hypothesis of equal rates between lineages. None of the P-values are significant for *rbcS*.

	<i>C. album</i>	<i>M. nuttalliana</i> ‘A’	<i>C. foliosum</i>	<i>C. bonus-henricus</i>
<i>M. nuttalliana</i> ‘A’	-4.941***			
<i>C. foliosum</i>	-4.675***	-2.475*		
<i>C. bonus-henricus</i>	-4.546***	-2.992**	-3.877***	
<i>S. oleracea</i>	-3.722***	-3.904***	-3.667***	-3.744***

**Table 14. Matrix for *rbcS* test of neutral selection.**

Test statistic (dN – dS) is shown for each species pair. Significance is shown at the 5% level (\*), at the 1% level (\*\*), and the 0.1% level (\*\*\*).

	<i>C. album</i>	<i>M. nuttalliana</i> 'A'	<i>C. foliosum</i>	<i>C. bonus-henricus</i>
<i>M. nuttalliana</i> 'A'	4.941***			
<i>C. foliosum</i>	4.675***	2.424**		
<i>C. bonus-henricus</i>	4.546***	3.068***	3.863***	
<i>S. oleracea</i>	3.722***	3.914***	3.684***	3.777***

**Table 15. Matrix for *rbcS* test of purifying selection.**

Test statistic (dS - dN) is shown for each species pair. Significance is shown at the 5% level (\*), at the 1% level (\*\*), and the 0.1% level (\*\*\*).

	<i>M. nuttalliana</i>	<i>C. foliosum</i>	<i>C. bonus-henricus</i>	<i>S. oleracea</i>
<i>C. foliosum</i>	0.05 (0.82726)			
<i>C. bonus-henricus</i>	1.47 (0.22525)	3.60 (0.05778)		
<i>S. oleracea</i>	6.55 (0.01052)*	3.46 (0.06298)	8.76 (0.00308)*	
<i>C. album</i>	8.40 (0.00376)*	10.80 (0.00102)*	19.20 (0.00001)*	0.89 (0.34523)

**Table 16. Tajima’s test statistics for *rbcL*.**

Test statistics from the *rbcL* relative rates tests (Tajima’s) for each species pair with the P-values given in parentheses. A P-value less than 0.05 is often used to reject the null hypothesis of equal rates between lineages & is indicated by an asterisk (\*).

	<i>M. nuttalliana</i>	<i>C. foliosum</i>	<i>C. bonus-henricus</i>	<i>S. oleracea</i>
<i>C. foliosum</i>	-2.367*			
<i>C. bonus-henricus</i>	-2.360 *	-1.764		
<i>S. oleracea</i>	-4.984***	-4.586***	-4.890***	
<i>C. album</i>	-5.226***	-5.013***	-4.678***	-6.172***

**Table 17. Matrix for *rbcL* test of neutral selection.**

Test statistic (dN – dS) is shown for each species pair. Significance is shown at the 5% level (\*), at the 1% level (\*\*), and the 0.1% level (\*\*\*).

	<i>M. nuttalliana</i>	<i>C. foliosum</i>	<i>C. bonus-henricus</i>	<i>S. oleracea</i>
<i>C. foliosum</i>	2.351**			
<i>C. bonus-henricus</i>	2.380**	---		
<i>S. oleracea</i>	4.968***	4.591***	4.891***	
<i>C. album</i>	5.129***	4.914***	4.550***	6.134***

**Table 18. Matrix for *rbcL* test of purifying selection.**

Test statistic (dS - dN) is shown for each species pair. Significance is shown at the 5% level (\*), at the 1% level (\*\*), and the 0.1% level (\*\*\*). Dashes (---) indicate a species pair for which the null hypothesis of strict neutrality between sequences could not be rejected.

	<i>M. nuttalliana</i>	<i>C. foliosum</i>	<i>C. bonus-henricus</i>
<i>C. foliosum</i>	4.76 (0.02905)*		
<i>C. bonus-henricus</i>	10.71 (0.00106)*	3.00 (0.08326)	
<i>S. oleracea</i>	0.00 (1.00000)	1.98 (0.15985)	6.82 (0.00902)*

**Table 19. Tajima's test statistics for *matK*.**

Test statistics from the *matK* relative rates tests (Tajima's) for each species pair with the P-values given in parentheses. A P-value less than 0.05 is often used to reject the null hypothesis of equal rates between lineages & is indicated by an asterisk (\*).

	<i>M. nuttalliana</i>	<i>C. foliosum</i>	<i>C. bonus-henricus</i>	<i>S. oleracea</i>
<i>C. foliosum</i>	-2.494*			
<i>C. bonus-henricus</i>	-2.540*	-1.149		
<i>S. oleracea</i>	-4.289***	-3.214**	-2.588**	
<i>C. album</i>	-4.197***	-3.541***	-3.261***	-3.486***

**Table 20. Matrix for the *matK* test of neutral selection.**

Test statistic (dN – dS) is shown for each species pair. Significance is shown at the 5% level (\*), at the 1% level (\*\*), and the 0.1% level (\*\*\*).

	<i>M. nuttalliana</i>	<i>C. foliosum</i>	<i>C. bonus-henricus</i>	<i>S. oleracea</i>
<i>C. foliosum</i>	2.494**			
<i>C. bonus-henricus</i>	2.540**	---		
<i>S. oleracea</i>	4.289***	3.214***	2.588**	
<i>C. album</i>	4.197***	3.541***	3.261***	3.486***

**Table 21. Matrix for the *matK* test of purifying selection.**

Test statistic (dS - dN) is shown for each species pair. Significance is shown at the 5% level (\*), at the 1% level (\*\*), and the 0.1% level (\*\*\*). Dashes (---) indicate a species pair for which the null hypothesis of strict neutrality between sequences could not be rejected.

	<i>M. nuttalliana</i>	<i>C. foliosum</i>	<i>C. bonus-henricus</i>
<i>C. foliosum</i>	0.00 (1.00000)		
<i>C. bonus-henricus</i>	0.00 (1.00000)	0.02 (0.88150)	
<i>S. oleracea</i>	7.71 (0.00548)*	7.08 (0.00779)*	6.64 (0.00995)*

**Table 22. Tajima's test statistics for *UFO*.**

Test statistics from the *UFO* relative rates tests (Tajima's) for each species pair using *C. album* as the outgroup, with the P-values given in parentheses. A P-value less than 0.05 is often used to reject the null hypothesis of equal rates between lineages & is indicated by an asterisk (\*).

	<i>S. oleracea</i> cv <i>America</i>	<i>S. tetrandra</i> ‘Allele C’	<i>M. nuttalliana</i>	<i>C. foliosum</i>	<i>C. bonus-henricus</i>
<i>S. tetrandra</i> ‘Allele C’	-5.413***				
<i>M. nuttalliana</i>	-8.586***	-8.693***			
<i>C. foliosum</i>	-9.002***	-9.131***	-2.509*		
<i>C. bonus-henricus</i>	-8.565***	-8.983***	-4.052***	-4.953***	
<i>C. album</i>	-10.957***	-11.518***	-9.625***	-10.246***	-9.949***

**Table 23. Matrix for *UFO* test of neutral selection.**

Test statistic (dN – dS) is shown for each species pair. Significance is shown at the 5% level (\*), at the 1% level (\*\*), and the 0.1% level (\*\*\*).

	<i>S. oleracea</i> cv <i>America</i>	<i>S. tetrandra</i> 'Allele C'	<i>M. nuttalliana</i>	<i>C. foliosum</i>	<i>C. bonus-henricus</i>
<i>S. tetrandra</i> 'Allele C'	5.379***				
<i>M. nuttalliana</i>	8.534***	8.705***			
<i>C. foliosum</i>	8.957***	9.085***	2.529**		
<i>C. bonus-henricus</i>	8.516***	8.869***	4.077***	4.981***	
<i>C. album</i>	11.022***	11.671***	9.840***	10.431***	10.083***

**Table 24. Matrix for *UFO* test of purifying selection.**

Test statistic (dS - dN) is shown for each species pair. Significance is shown at the 5% level (\*), at the 1% level (\*\*), and the 0.1% level (\*\*\*).



	<i>S. oleracea</i> cv <i>America</i>	<i>M. nuttalliana</i>	<i>C. foliosum</i>	<i>C. bonus-henricus</i> Allele 1	<i>C. bonus-henricus</i> Allele 2
<i>M. nuttalliana</i>	0.97 (0.32518)				
<i>C. foliosum</i>	5.43 (0.01974)*	3.00 (0.8326)			
<i>C. bonus-henricus</i> A1	2.17 (0.14077)	0.08 (0.78151)	2.69 (0.10105)		
<i>C. bonus-henricus</i> A2	3.83 (0.05020)	1.75 (0.18588)	0.04 (0.84597)	1.11 (0.29206)	

**Table 25. Tajima's test statistics for *AGAMOUS*.**

Test statistics from the *AG* relative rates tests (Tajima's) for each selected species pair using *C. album* as the outgroup, with the P-values given in parentheses. A P-value less than 0.05 is often used to reject the null hypothesis of equal rates between lineages & is indicated by an asterisk (\*).

	<i>S. oleracea cv America</i>	<i>M. nuttalliana</i>	<i>C. foliosum</i>	<i>C. bonus-henricus</i> Allele 1	<i>C. bonus-henricus</i> Allele 2
<i>M. nuttalliana</i>	-4.670***				
<i>C. foliosum</i>	-4.146***	-1.764			
<i>C. bonus-henricus</i> A1	-4.060***	-2.168*	-1.222		
<i>C. bonus-henricus</i> A2	-3.975***	-2.765**	-2.056*	-1.934	
<i>C. album</i>	-6.006***	-6.337***	-5.806***	-6.079***	-5.824***

**Table 26. Matrix for AG test of neutral selection.**

Test statistic (dN – dS) is shown for each species pair. Significance is shown at the 5% level (\*), at the 1% level (\*\*), and the 0.1% level (\*\*\*).

	<i>S. oleracea</i> <i>cv America</i>	<i>M. nuttalliana</i>	<i>C. foliosum</i>	<i>C. bonus-</i> <i>henricus</i> Allele 1	<i>C. bonus-</i> <i>henricus</i> Allele 2
<i>M. nuttalliana</i>	4.670***				
<i>C. foliosum</i>	4.146***	---			
<i>C. bonus-</i> <i>henricus</i> A1	4.060***	2.168*	---		
<i>C. bonus-</i> <i>henricus</i> A2	3.975***	2.765**	2.056*	---	
<i>C. album</i>	6.006***	6.337***	5.806***	6.079***	5.824***

**Table 27. Matrix for AG test of purifying selection.**

Test statistic (dS - dN) is shown for each species pair. Significance is shown at the 5% level (\*), at the 1% level (\*\*), and the 0.1% level (\*\*\*). Dashes (---) indicate a species pair for which the null hypothesis of strict neutrality between sequences could not be rejected.

	<i>S. oleracea</i> cv <i>America</i>	<i>M. nuttalliana</i>	<i>C. foliosum</i>
<i>M. nuttalliana</i>	2.82 (0.09322)		
<i>C. foliosum</i>	5.44 (0.01966)*	0.88 (0.34897)	
<i>C. bonus-henricus</i>	11.79 (0.00059)*	9.32 (0.00226)*	6.34 (0.01181)*

**Table 28. Tajima's test statistics for AP3.**

Test statistics from the AP3 relative rates tests (Tajima's) for each selected species pair using *C. album* 'Allele 1' as the outgroup, with the P-values given in parentheses. A P-value less than 0.05 is often used to reject the null hypothesis of equal rates between lineages & is indicated by an asterisk (\*).

	<i>S. oleracea</i> cv <i>America</i>	<i>M. nuttalliana</i>	<i>C. foliosum</i>	<i>C. bonus-henricus</i>
<i>M. nuttalliana</i>	-5.067***			
<i>C. foliosum</i>	-5.075***	0.425		
<i>C. bonus-henricus</i>	-5.222***	0.634	-0.856	
<i>C. album</i> ‘Allele 1’	-5.097***	-0.533	-0.739	-1.512

**Table 29. Matrix for AP3 test of neutral selection.**

Test statistic (dN – dS) is shown for each species pair. Significance is shown at the 5% level (\*), at the 1% level (\*\*), and the 0.1% level (\*\*\*).

	<i>S. oleracea</i> cv <i>America</i>
<i>M. nuttalliana</i>	5.067***
<i>C. foliosum</i>	5.075***
<i>C. bonus-henricus</i>	5.222***
<i>C. album</i> 'Allele 1'	5.097***

**Table 30. Matrix for AP3 test of purifying selection.**

Test statistic (dS - dN) is shown for each species pair. Significance is shown at the 5% level (\*), at the 1% level (\*\*), and the 0.1% level (\*\*\*). Not shown are those species pairs for which the null hypothesis of strict neutrality between sequences could not be rejected.

	<i>S. oleracea</i> cv <i>America</i>	<i>M. nuttalliana</i>	<i>C. foliosum</i>	<i>C. bonus-henricus</i>
<i>M. nuttalliana</i>	3.41 (0.06471)			
<i>C. foliosum</i>	5.70 (0.01692)*	0.83 (0.36213)		
<i>C. bonus-henricus</i>	5.95 (0.01468)*	0.13 (0.72367)	0.03 (0.85968)	

**Table 31. Tajima's test statistics for *PI*.**

Test statistics from the *PI* relative rates tests (Tajima's) for each selected species pair using *C. album* as the outgroup, with the P-values given in parentheses. A P-value less than 0.05 is often used to reject the null hypothesis of equal rates between lineages & is indicated by an asterisk (\*).

	<i>S. oleracea cv America</i>	<i>M. nuttalliana</i>	<i>C. foliosum</i>	<i>C. bonus-henricus</i>
<i>M. nuttalliana</i>	-4.121 (0.000)***			
<i>C. foliosum</i>	-4.136 (0.000)***	-0.235 (0.815)		
<i>C. bonus-henricus</i>	-3.459 (0.001)***	-0.692 (0.490)	-0.918 (0.361)	
<i>C. album</i>	-4.111 (0.000)***	-3.745 (0.000)***	-4.030 (0.000)***	-3.326 (0.001)***

**Table 32. Matrix for *PI* test of neutral selection.**

Test statistic (dN – dS) is shown for each species pair. Significance is shown at the 5% level (\*), at the 1% level (\*\*), and the 0.1% level (\*\*\*).



	<i>S. oleracea</i> cv <i>America</i>	<i>M. nuttalliana</i>	<i>C. foliosum</i>	<i>C. bonus-</i> <i>henricus</i>
<i>M. nuttalliana</i>	4.121 (0.000)***			
<i>C. foliosum</i>	4.136 (0.000)***	---		
<i>C. bonus-</i> <i>henricus</i>	3.459 (0.000)***	---	---	
<i>C. album</i>	4.111 (0.000)***	3.745 (0.000)***	4.030 (0.000)***	3.326 (0.001)***

**Table 33. Matrix for *PI* test of purifying selection.**

Test statistic (dS - dN) is shown for each species pair. Significance is shown at the 5% level (\*), at the 1% level (\*\*), and the 0.1% level (\*\*\*). Dashes (---) indicate a species pair for which the null hypothesis of strict neutrality between sequences could not be rejected.

	<i>S. oleracea</i> cv <i>America</i>	<i>M. nuttalliana</i>	<i>C. foliosum</i>	<i>C. bonus-henricus</i>
<i>M. nuttalliana</i>	25.45 (0.00000)***			
<i>C. foliosum</i>	39.54 (0.00000)***	7.35 (0.00670)**		
<i>C. bonus-henricus</i>	53.91 (0.00000)***	14.07 (0.00018)***	1.99 (0.15853)	

**Table 34. Tajima’s test statistics for the ‘super-gene’ alignment, including *rbcL*.**

Rates testing of the entire concatenated sequence alignment consisting of *rbcL*, *rbcS*, *matK*, *UFO*, *AG*, *AP3*, and *PI* genes from Figure 36, taking into account all codon-positions as well as non-coding regions. Test statistics from the relative rates tests (Tajima’s) are shown for each selected species pair using *C. album* as the outgroup, with the P-values given in parentheses. A P-value less than 0.05 is often used to reject the null hypothesis of equal rates between lineages & is indicated by an asterisk (\*).

	<i>S. oleracea</i> cv <i>America</i>	<i>M. nuttalliana</i>	<i>C. foliosum</i>	<i>C. bonus-henricus</i>
<i>M. nuttalliana</i>	22.80 (0.00000)***			
<i>C. foliosum</i>	31.92 (0.00000)***	4.05 (0.04417)*		
<i>C. bonus-henricus</i>	45.36 (0.00000)***	10.28 (0.00134)**	2.18 (0.13941)	

**Table 35. Tajima’s test statistics for the ‘super-gene’ alignment, not including *rbcL*.**

Rates testing of the concatenated sequence alignment consisting of *rbcS*, *matK*, *UFO*, *AG*, *AP3*, and *PI* genes, taking into account all codon-positions as well as non-coding regions. Test statistics from the relative rates tests (Tajima’s) are shown for each selected species pair using *C. album* as the outgroup, with the P-values given in parentheses. A P-value less than 0.05 is often used to reject the null hypothesis of equal rates between lineages & is indicated by an asterisk (\*).

	<i>S. oleracea cv America</i>	<i>M. nuttalliana</i>	<i>C. foliosum</i>	<i>C. bonus-henricus</i>
<i>M. nuttalliana</i>	3.06 (0.08018)			
<i>C. foliosum</i>	11.05 (0.00089)***	4.79 (0.02578)*		
<i>C. bonus-henricus</i>	11.05 (0.00089)***	3.21 (0.07337)	0.03 (0.86671)	

**Table 36. Tajima’s test statistics for the ‘super-gene’ alignment that included *rbcS*, *matK*, and *AG*.**

Rates testing of the concatenated sequence alignment consisting of genes not thought to be related to dioecy (*rbcS*, *matK*, and *AG*), taking into account all codon-positions as well as non-coding regions. The *rbcL* gene was not included. Test statistics from the relative rates tests (Tajima’s) are shown for each selected species pair using *C. album* as the outgroup, with the P-values given in parentheses. A P-value less than 0.05 is often used to reject the null hypothesis of equal rates between lineages & is indicated by an asterisk (\*).

	<i>S. oleracea cv America</i>	<i>M. nuttalliana</i>	<i>C. foliosum</i>	<i>C. bonus-henricus</i>
<i>M. nuttalliana</i>	20.17 (0.00001)***			
<i>C. foliosum</i>	21.40 (0.00000)***	0.54 (0.46099)		
<i>C. bonus-henricus</i>	34.33 (0.00000)***	7.10 (0.00771)**	3.79 (0.05148)	

**Table 37. Tajima’s test statistics for the ‘super-gene’ alignment which included *UFO*, *AP3*, and *PI*.**

Rates testing of the concatenated sequence alignment consisting of genes that are thought to be related to dioecy (*UFO*, *AP3*, and *PI*), taking into account all codon-positions as well as non-coding regions. Test statistics from the relative rates tests (Tajima’s) are shown for each selected species pair using *C. album* as the outgroup, with the P-values given in parentheses. A P-value less than 0.05 is often used to reject the null hypothesis of equal rates between lineages & is indicated by an asterisk (\*).

## DISCUSSION

Mutations leading to significant changes in the number and rate of development of reproductive organs have the immediate potential to effect individual fecundity and thus fitness. Additionally, mutations leading either to dichogamy (*C. bonus-henricus*), gynomonoeicy (*C. foliosum* and *M. nuttalliana*), and dioecy (*Spinacia*) change outcrossing rates; hence genotype distribution is changed as is recombination within populations. Lastly, recent domestication such as has occurred in *Spinacia* causes population bottlenecks that result in drift, linkage disequilibria, and hitchhiking effects under selection. Thus, the patterns, modes and rates of evolution of the genes in this group are expected to vary among these species. The species within the *Spinacia* and *Blitum* groups are distinctive in demonstrating extensive evolution in floral morphology and reproductive strategy, and for having experienced relatively recent domestication. Our results describe a complex pattern of substantial sequence deviations driven primarily by rapid insertion/deletion events in both exons and introns. Though there is evidence of increased rates of nucleotide substitutions particularly in *Spinacia*, there is no evidence of positive selection in coding regions. It appears, however, that significant changes in amino acid sequences in the floral developmental genes *AP3*, *PI*, and *UFO* are driven by patterns of indels that are not taken into account by tests based solely on ratios of non-synonymous/synonymous substitution.

### Phylogenetic Relationships

Molecular phylogenetic studies with dense taxonomic sampling have consistently proposed that the broad sense genus *Chenopodium* is highly paraphyletic [11, 13-15]. Kadereit et al. [14] sampled nine species of *Chenopodium*, plus *Monolepis nuttalliana* and *Spinacia oleracea* in a large phylogenetic study using *rbcL* in the Chenopodiaceae and the Amaranthaceae. The genus

*Chenopodium* was polyphyletic and may be divided into three clades (Chenopodieae I, Chenopodieae II, and Chenopodieae III) that may represent three tribes. Together with Atripliceae, these form a monophyletic Chenopodioideae subfamily. Chenopodieae II included *C. bonus-henricus* as a basal taxon to *Scleroblitum atriplicinum*-*C. foliosum* and *Spinacia oleracea*-*Monolepis nuttalliana* nested clades. In an expanded study utilizing sequences from *rbcL*, the *ATPB-rbcL* spacer, and the nuclear *ITS* region, Kadereit et al. [15] presented further support for the Chenopodieae II subfamily using Bayesian approaches. The *rbcL* and *ATPB-rbcL* trees are incongruent within this subfamily. As reported from the *rbcL* tree in the previous study, *C. bonus-henricus* is basal to the clade and *Spinacia* and *Monolepis* are sister taxa. In contrast, *Spinacia* is basal in the *ATPB-rbcL* spacer tree with *Monolepis* and *C. foliosum* as sister taxa to *C. bonus-henricus*. They note that the clade including *Spinacia* (Irano-Turkestan region), *Scleroblitum* (Australia), *Monolepis* (Siberia/North America), and *Chenopodium* subgenus *Blitum* forms a geographically and chromosomally diverse group, and may require more extensive taxonomic sampling [15]. More recently, Fuentes-Bazan et al. [11] used the plastid *TRNL-F* and nuclear *ITS* regions in Bayesian and Maximum parsimony analyses. They confirm the paraphyletic structure of the genus *Chenopodium*, and retrieve a well-supported clade (Clade 3) comparable to Chenopodieae II [14, 15]. Within the clade, the topologies are congruent with the non-*rbcL* based trees, with *Spinacia* being the basal lineage in the *TRNL-F* /*ITS* study versus *C. bonus-henricus* as basal in the *rbcL* study. Fuentes-Bazan et al. suggest recognizing this clade at the tribal level with the name *Spinacieae*, a designation that they attribute first to Moquin-Tandon [68]. In an additional study, Fuentes-Bazan et al. [12] suggest using *Blitum* as the genus name to the clade of *C. bonus-henricus*, *C. capitatum*, *C. hastatum*, *M. nuttalliana*, *S. atriplicinum* and *C. virgatum* (syn. *C. foliosum*). They further propose that the tribal level

designation that would include *Spinacia* and *Blitum* be designated *Anserineae* Dumort, a name having precedence to *Spinacidae* Moquin.

With the exception of *rbcL*, phylogenetic hypotheses generated from all genes in this study consistently support *Spinacia* as basal to a *Chenopodium* subspecies *Blitum* group. There is no consistent differentiation between nuclear-encoded genes and chloroplast-encoded genes. *rbcS*, which was selected due to the functional interaction of its encoded protein with that of *rbcL*, supported the basal *Spinacia* position within the clade. Similarly, analyses of the floral developmental genes *AP3*, *PI*, *AG*, and *UFO* strongly support a basal position of *Spinacia*. When the concatenated data set is analyzed, the (*Spinacia* (*C. bonus-henricus* (*C. foliosum*, *Monolepis nuttalliana*))) hierarchy is supported with 100% support.

Indels were not used for phylogenetic inferences in this study but can be highly informative when found in coding as well as noncoding regions. *Spinacia*-specific deletions in *rbcS*, simple repeat sequences in the hypervariable region of *UFO*, and deletions in *AG* and *AP3* consistently support a distinction of this genus from the *Blitum* group. Shared indels in *rbcS*, *UFO*, *AG*, and *AP3* support the sister taxa relationship of *C. foliosum* and *M. nuttalliana* within a defined *Blitum* clade. Thus, even with the low taxon density, the combined data indicate a consistent pattern.

In contrast to the apparent consistency with the *Anserineae* subfamily, no clear pattern of phylogenetic relationship can be determined within the genus *Spinacia*. Variation among accessions is limited. In the case of *UFO* where we increased the sample size to study the polymorphisms within species, and *AG* where genes were sequenced from multiple individuals, *S. oleracea*, *S. turkestanica*, and *S. tetrandra* were not identifiable as monophyletic species. Among the *AG* sequences there is one autapomorphic SNP found in two *S. oleracea* individuals sampled. However, there is one shared SNP that is found in one individual of *S. tetrandra* and



one individual of *S. turkestanica*. This observation is consistent with gene flow between the two taxa. One *S. tetrandra* UFO allele found in a female was distinct in that it had only two hexanucleotide repeats, 34 SNPs, six of which are non-synonymous substitutions, and one nine nucleotide insertion in the F box region in comparison with other alleles found commonly in all three *Spinacia* species. As will be discussed later, the degree of sequence differentiation might indicate at least a partial isolation from the other species or retention of an ancient allele. More intensive sampling of wild populations rather than reliance on collected accessions would be necessary to support this hypothesis.

This lack of genetic distinction among *Spinacia oleracea*, *Spinacia turkestanica*, and *Spinacia tetrandra* is actually consistent with the somewhat muddled taxonomic designations. According to Sneep [69], Linnaeus identified *S. oleracea* as having five stamens, identical to the common Chenopod stamen number. To contrast this, *S. tetrandra* was named based on having four stamens. This is, of course, the actual number of stamens in *S. oleracea*. *S. turkestanica* was differentiated from *S. tetrandra* based on petiole length, a trait that is selected in cultivated spinach, and therefore, a trait that has genetic variation within populations. Additional species identifications in *S. glabra* Miller, *S. spinosa* Moench, and *S. inermis* Moench were based on prickly vs. smooth pericarp morphology. Linnaeus recognized this pericarp morphology as a variable trait of *S. oleracea* and it was later determined to be controlled by a single dominant allele [70]. Additionally, no wild populations of *S. oleracea* are recognized today [71] which brings into question whether cultivated *S. oleracea* originated from an independent species. The lack of molecular differentiation among the three recognized *Spinacia* species is remarkable and may be grounds for reviewing the species designation for each form.

### **Floral Morphology and Reproductive Strategy**

Floral morphological variation and hence reproductive strategies are extensive among the Anserineae. Within our study, *Chenopodium bonus-henricus* is the only species that retains perfect flowers with five sepal/five stamen composition common within the Chenopodiaceae. This species, however, has developed a strong protogynous growth pattern wherein stamens are covered by sepals during a time when the pistil and its two stigmas are exposed. This pattern of development is expected to increase outcrossing rates. The other two *Blitum* group species in this study, *Monolepis nuttalliana* and *Chenopodium foliosum*, have reduced organ numbers and are gynomonocious. *Monolepis nuttalliana* has a single sepal and a single opposite stamen in perfect flowers with no perianth or stamen in pistillate flowers. *Chenopodium foliosum* has three sepals in both its perfect and pistillate flowers. In its perfect flowers, *Chenopodium foliosum* has a single stamen, similar to *Monolepis nuttalliana*. This would suggest that stamen number reduction is synapomorphic and is independent of sepal number. Also, *Chenopodium foliosum* appears to be protogynous in its perfect flowers with its two stigmas protruding above the sepals while the single stamen is still enclosed by the sepals, similar to the developmental progression seen in *Chenopodium bonus-henricus*. Our taxonomic sampling is too limited to do more than speculate that dichogamy, and thus increased outcrossing, precedes organ loss and gynomonocoe in this group. This is, however, an intriguing hypothesis that could be challenged with an expanded sample.

The *Spinacia* species display the most extreme variation in floral development and reproductive strategy being predominantly dioecious. As with *Chenopodium foliosum* and *Monolepis nuttalliana*, there is a reduction in sepal number to two in pistillate flowers and four in staminate flowers. Developmental studies [60] indicate that the sepals develop in pairs of two, the pistillate flowers having a single pair that expand around the circumference of the floral

meristem, whereas the males develop two pairs of sepals sequentially. The staminate flowers have four stamens that similarly arise in sequential pairs opposite the sepals. The carpel (fourth whorl organs) do not initiate at all in these flowers. The pistillate flowers lack the complete third whorl (stamens) and develop a single gynoecium, similar to the other species in this study, except having four or five stigma instead of two. Monoecious individuals do occasionally appear in populations and have been used in cultivar breeding [69, 72, 73]. Perfect flowers as well as homeotically-transformed flowers can be generated by affecting B class floral organ identity gene expression [46]. The genetic studies of monoecy in spinach [73-75] suggest that dioecy may have evolved from monoecy [51, 76].

Kadereit et al. [15] argues that the structures that enclose the gynoecium in *Spinacia* are not sepals as previously noted [60, 77, 78]. Based on the tissue structure, they argue that these organs are bracts and hence the female flowers lack a perianth, similar to the naked gynoecium in the *Monolepis nuttalliana* pistillate flowers. As previously mentioned, Sather et al. [60] identified a sequential development of pairs of perianth organs (sepals) in staminate flowers. If the first pair of organ primordia is homologous to the primordia pair that develops in pistillate flowers, then perhaps the second pair alone in staminate flowers is sepals whereas the shared organs would be bracts. However, given the development of the two pairs of stamen opposite all four perianth organs in the staminate flowers, it appears more likely that all four perianth organs are sepals. Based on this, the two organs surrounding the gynoecium in pistillate organs must also be considered sepals.

### **Rates of Sequence Evolution**

The patterns of nucleotide substitutions present a consistent phylogenetic hierarchy, with the exception of the *rbcL* data, with *Spinacia* species being a sister taxon to the *Blitum* species, and,

within the *Blitum* group, *C. bonus-henricus* sister to *C. foliosum* and *Monolepis nuttalliana*. The individual tests of rates of nucleotide substitution do not, however, give as consistent a pattern. Rates of nucleotide substitution tended to be faster in *Spinacia* species in comparison to *Blitum* species in *rbcL*, *UFO*, *AP3*, and *PI*. *rbcL* is an apparent outlier as indicated by the gene trees. In the *rbcL*-based tree, both *S. oleracea* and *C. album* group with *M. nuttalliana* and *C. foliosum*, respectively, totally incongruent with trees generated in all other data sets. This gene phylogeny does not match other chloroplast gene phylogenies in this and other studies [11, 12, 15] and probably is the result of convergent evolution driven by increased substitution rates and purifying selection. In contrast, *AP3*, *PI*, and *UFO* all support the consensus phylogeny. They have also all been either directly or indirectly linked to dioecious development in spinach. *PI* and *AP3* are normally required for stamen development, but, in spinach, have been shown to be involved in the suppression of fourth whorl organs [46, 47]. *UFO* is required for normal floral development [29] and has been demonstrated to interact with the floral identity transcription factor *LFY* in *Arabidopsis* [79]. Furthermore, *UFO* is an F box protein that forms a part in SCF complexes [34, 35, 79]. It may be the functional link that ties the GA sex determination effect [80, 81] to the regulation of B class floral organ gene expression in spinach [46]. In contrast to these three genes, the fourth gene in this study that is involved in floral development, *AG*, shows no strong signal for increased rates of nucleotide substitution in *Spinacia* compared to the *Blitum* species. This is noteworthy given that *AG* does not appear to have a role in the development of dioecy in spinach [46, 60]. When the *UFO*, *AP3*, and *PI* sequences are concatenated and introns are excluded, rates of nucleotide substitution does not vary significantly within the *Blitum* group, whereas *Spinacia* evolves significantly faster than the *Blitum* group species. If *matK*, *rbcS*, and *AG* are concatenated and non-coding regions are excluded, *Spinacia* does evolve marginally

faster than *C. foliosum* and *C. bonus-henricus*, but not in comparison with *M. nuttalliana*. *C. bonus-henricus* and *C. foliosum* also vary marginally in their rates. The cumulative pattern that is emerging from these analyses suggests increased rates of sequence evolution occurring in the *Spinacia* lineage, which could be potentially correlated with the evolution of dioecy as a reproductive strategy.

Tajima's relative-rates test does not measure rates of insertion/deletions (indels) in modeling sequence evolution. These types of mutational events prove to be common and extensive among many of the genes studied here, both in introns and exons. The two chloroplast-encoded genes *rbcL* and *matK* have no introns and show very little sequence length variation with the exception of a single in-frame 18 bp insertion in *C. album*. The nuclear-encoded chloroplast protein gene *rbcS* does have introns and does show indel variation within and among species. *Monolepis nuttalliana* is polymorphic for a 17 bp insertion in intron 1. There is some variation among the *Blitum* species within the introns; however *Spinacia* species show the greatest differentiation with a reduction in intron 1 from 541 bp in *C. album* to 150 bp in spinach.

The remaining three nuclear encoded genes with introns in this study, *AG*, *AP3*, and *PI*, show even more extensive variations in intron size within *Spinacia* in comparison with other species. For example, the *C. album* *AG* intron 5 is estimated to be 1526 nucleotides long. In comparison, the same intron in *S. oleracea* is 228 nucleotides long. As noted in the results, often islands of relatively high sequence similarity can be identified even when sequences are very different in size, and these islands tend to have sequences that are identical to transcription factor binding motifs. In the case of the *AG* intron 5, the reduced *Spinacia* results in a loss of one conserved CCAAT motif relative to other species, which could potentially affect transcription rates or patterns. Intron lengths, and, of course, sequences in *Spinacia* *AP3* and *PI* also vary

greatly from the *C. album* and *Blitum* group sequences. *Spinacia AP3* Intron 2 and 6 are 2522 and 943 nucleotides long, respectively, while the same introns are 1201 nucleotides and 277 in *C. bonus-henricus*, 1552 and 276 in *C. foliosum*, and 1645 and 277 in *M. nuttalliana*. Similarly, introns 2, 5, and 6 in *PI* have size ranges within the *Blitum* group of 763 to 779, 346 to 378, and 659 to 666 nucleotides, respectively. The same introns in *Spinacia sp.* are 1504, 912, and 253 nucleotides, respectively. Thus, the extent of intron sequence acquisition and loss in *Spinacia* greatly exceeds the scale seen among the three *Blitum* group species.

*UFO* is also a good comparison to the chloroplast genes as there are no introns in the sequence. Despite this, there is substantial sequence variation due to insertions in the hypervariable linker region between the F-box domains and the Kelch propeller domain. In comparison with the *C. album* sequences, the *Blitum* group species are 9 or 12 nucleotides shorter, thus preserving the reading frame. The deviations are not clearly due to obvious insertions or deletions. The *C. album* sequence is enriched with repeated asparagine codons (AAY). The *Blitum* group species have a restricted number of these repeats and have a conserved HLN<sub>X</sub>GE sequence. In *C. bonus-henricus*, the last four codons of this motif are AATGATGGAGAA. In the shortest *UFO* allele of *S. tetrandra*, this region is GGAGATGGAGAA. By a single transversion substitution of T to A, the sequence becomes two identical hexanucleotide repeats GGAGAAGGAGAA that can lead to the predominant hexanucleotide repeat alleles found among the *Spinacia* species and cultivars. The persistence of the indel polymorphism among the wild and cultivated accessions of spinach supports either rapid, persistent regeneration of variable repeat number alleles, or preservation of the polymorphism through balancing selection. The “ancestral” *S. tetrandra UFO* allele also differs from the other *Spinacia UFO* alleles at 34 substitutions (6 of which are non-synonymous) and by

three additional tri-nucleotide repeats in a repeated region found in all *Spinacia* alleles. This deviation, plus the location of this allele in the *UFO* gene tree, would suggest that this *S. tetrandra* allele is ancient, and support the balancing selection hypothesis. If this is true, then the length variation in the linker region between the F-box and Kelch domains must affect the functionality of the translated proteins. In the case of *UFO*, it is tantalizing to speculate that such variation could result in differential sensitivity to GA concentrations, as has been hypothesized as a mechanism to develop dioecy in *Spinacia* [46, 76].

The two B class floral identity genes also demonstrate the role of indels in coding sequences in this group. Both *PI* and *AP3* have been shown to act as masculinizing factors in *Spinacia* and it was hypothesized that deletions of conserved motifs in the carboxyl coding region could be the mutational events that are linked to masculinization [46, 47]. The phylogenetic analysis indicates that the deletion in the *PI* gene is a shared character within the *Anserineae*, and therefore predates the evolution of dioecy in *Spinacia*. In contrast, the two deletion events in the *Spinacia AP3* gene that result in the loss of the euAP3 and *PI* motifs are unique to *Spinacia*. All of the *Blitum* species sampled share the canonical and ancestral euAP3 and *PI* motif sequences that are also found in *C. album*.

### **Domestication of Spinach, Genetic Variability, and Species Differentiation**

Domestication and development of cultivated varieties of crop plants is expected to reduce genetic diversity within populations and can lead to significant differentiation between cultivars and wild progenitors due to founder effects. Furthermore, strong selection programs for Agriculturally important characteristics can add to population differentiation in selected and closely linked genes. Spinach is a relatively recent cultivated crop with records of domestication of *S. oleracea* only going back to the sixth century when it was introduced into China (Laufer

1919[82, 83]). It is reasonable to assume that it must have been cultivated or foraged to some extent in its native Persia prior to its Agricultural spread. Records and archeological remains of identifiable cultivated varieties in Europe go back to the 12<sup>th</sup> or 13<sup>th</sup> century [84], and by the 16<sup>th</sup> through 18<sup>th</sup> century there are three or four cultivated varieties in the literature, two of which Epinard de Hollande (round pericarp) and Epinard d'Angleterre (prickly pericarp) are dominant [69].

Sneep notes that cultivars developed since the early 19<sup>th</sup> century can be segregated into categories based on three predominant selected traits: prickly-seeded (prickly-pericarp) versus smooth-seeded, smooth-leaved versus savoy-leaved, and late bolting versus early bolting. Secondary traits including petiole length, leaf width, leaf color, and disease resistance are also targets for selection. With the possible exception of *UFO* which may be involved in the GA response pathway during flowering and hence could potentially be involved in time of bolting, there is no obvious link between the genes investigated in this study and traits involved in modern spinach cultivar development. Consistent with this expectation, we did not identify strong genetic differentiation among the cultivated and wild varieties of spinach in this study, the single exception being a *UFO* allele identified in *S. tetrandra* that was discussed earlier. Thus, there is no indication that domestication has left a distinguishing signature in our study and therefore differences between spinach species and *Blitum* group species must reflect the evolutionary history before domestication.

In the cases where we sampled different spinach cultivars, we did not detect great differentiation that could reflect founder effects during cultivar development or the effects of direct phenotypic selection. However, based on the genealogy of the cultivars used in our study [69], these cultivars may be closely related and would not reflect earlier independent gene pools.



Our most common source is the cultivar America, which is a round-pericarp, savoy-leaved variety introduced in 1945 and developed from a cross of Bloomsdale Longstanding and Viking. Viking is a cultivar developed from a cross of Monstrueux de Viroflay (Viroflay) and Koning Denemarken (King of Denmark) both of which are round-pericarp, smooth-leaved varieties. Géant d'Hiver may be a selection from Viroflay, whereas Nobel is likely derived from Gaudry, which, like Viroflay, is derived from Epinard de Flanders in the mid-19<sup>th</sup> century. Thus, with the exception of the Bloomsdale Longstanding, which is itself derived from a round-pericarp, savoy-leaved variety Bloomsdale, all of these varieties are potentially derived from a single gene pool selected in the 19<sup>th</sup> century, and perhaps we should not anticipate great genetic variation among them. It is worth noting, however, that the persistent polymorphisms that were found in the spinach *UFO* gene in domesticated cultivars and wild accessions is the exception, and therefore may reflect some form of balancing selection.

## CONCLUSIONS

We have shown that the *rbcL* gene is inconsistent in resolving the organismal phylogeny of the species in *Anserineae*. Our analyses of the *rbcS*, *matK*, *UFO*, *AG*, *AP3*, and *PI* genes reveal a phylogeny that is consistent with those generated from plastid *TRNL-F* and nuclear *ITS* data [11] and one which most likely represents the true organismal phylogeny for these species (Figure 39). This gives us a better insight into the evolution of dioecy in *Spinacia*. Our phylogenetic analysis indicates that the deletion in the *PI* gene is a shared character within the *Anserineae*, and therefore predates the evolution of dioecy in *Spinacia*. In contrast, the two deletion events in the *Spinacia AP3* gene that result in the loss of the euAP3 and *PI* motifs are unique to *Spinacia*. All of the *Blitum* species sampled share the canonical and ancestral euAP3 and *PI* motif sequences in the *AP3* gene similar to that found in *C. album*. We also see a pattern of polymorphism unique to *Spinacia* in the *UFO* gene, which regulates the B-class genes *AP3* and *PI*.

It is likely that the combination of insertion/deletion events in key floral developmental genes described here plays an integral role in the appearance of unisexual flowers in *Spinacia*. Both products of the B-class genes *AP3* and *PI* are thought to interact with each other and with other proteins involved in development of floral organs [85]. We know that the *UFO* gene regulates the products of the *AP3* and *PI* genes, and that perfect flowers as well as homeotically-transformed flowers can be generated in *Spinacia* by affecting B-class gene expression [46]. Both *PI* and *AP3* have been shown to act as masculinizing factors in *Spinacia* and it was hypothesized that deletions of conserved motifs at the carboxyl-end could be the mutational events that are linked to masculinization, via suppression of gynoecium development [46, 47]. That the *AP3* mutation is not shared in the *Blitum* group supports this idea, as the *Blitum* species

have pistillate and hermaphroditic flowers, but not staminate flowers. It is possible that the *PI* mutation shared among members of *Anserineae* represents pre-adaptation [9], allowing for a rapid transition of characters in *Spinacia* following the mutational event in the *AP3* gene.

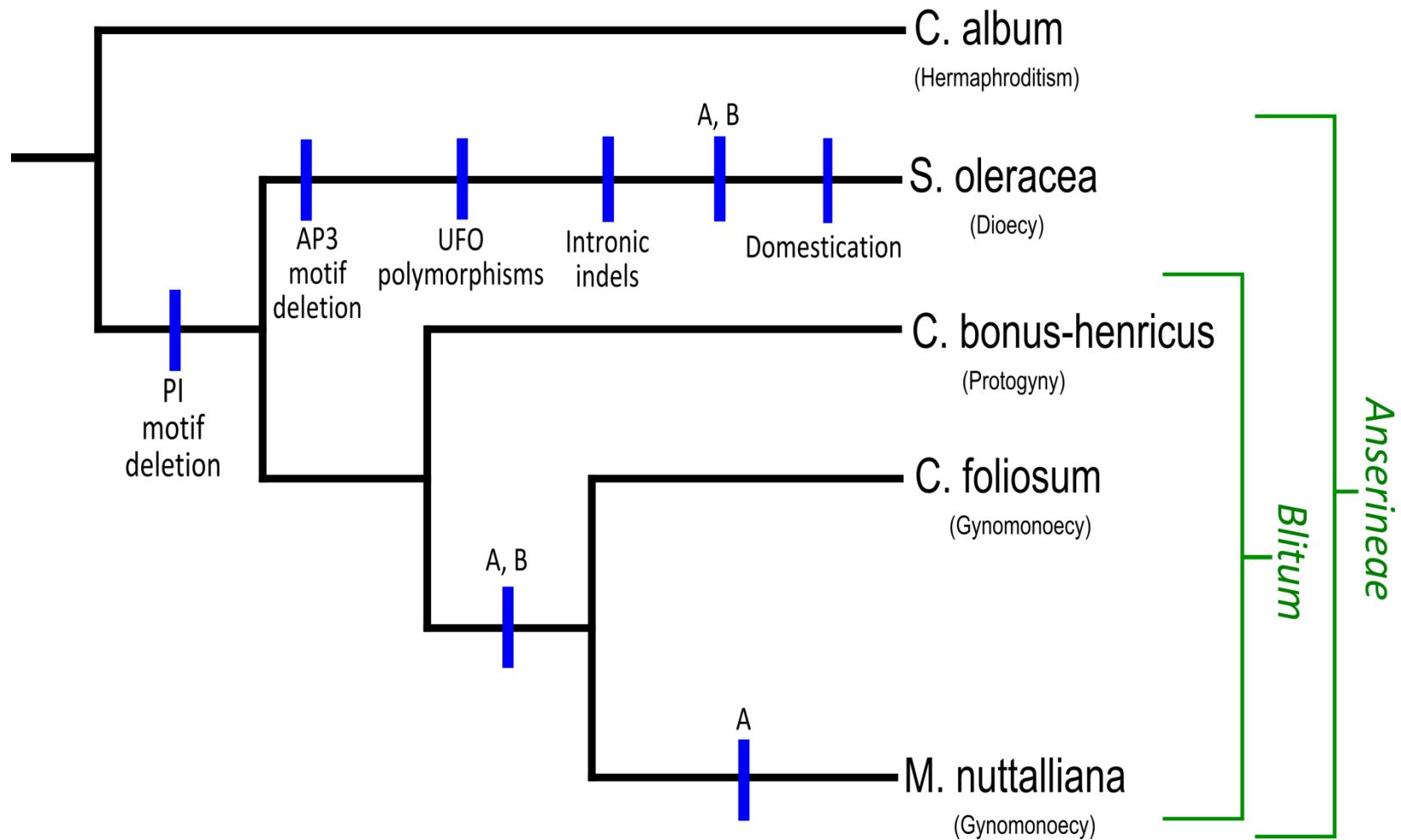
The *UFO* gene is necessary for floral organ and floral meristem identity and is a key regulator of the *AP3* and *PI* genes [29, 31, 37, 86]. Length variation seen in the linker region between the F-box and Kelch domains of the *UFO* sequences in *Spinacia* may affect the functionality of the translated proteins. It is tantalizing to speculate that such variation could result in differential sensitivity to GA concentrations, as has been hypothesized as a mechanism to develop dioecy in *Spinacia* [46, 76].

We performed tests of selection and did not find any evidence of positive selection acting on any of the sequences in this study. However, we did find evidence of purifying selection occurring in the *Spinacia* lineage and on many of the genes in the *Blitum* clade, indicating that amino acid sequence evolution is not being driven by nucleotide-substitution but by indel mutations.

Rates analysis indicates a general pattern of increased rates in the *Spinacia* lineage, although these tests do not account for non-nucleotide substitution variation, which can be seen extensively in the *Spinacia* sequences in the form of insertion/deletion events. The effect that these mutations have on overall rates of sequence change is unclear, although their effects on the expressed amino acid sequences are evident. Most of the variation between gene orthologs in this study is seen as insertion/deletion events in non-coding as well as coding regions. Alignments of these sequences also reveal “islands” of conservation occurring in introns that have high sequence similarity to regulatory motifs. These hint at regulatory similarities shared among species but also to the possibility that some are differentially regulated, possibly leading

to morphological effects. In contrast, the lack of molecular differentiation among the three recognized *Spinacia* species is remarkable considering their morphological dissimilarity. Our analysis indicates that the domestication of *Spinacia* has had no effect on the rates or modes of sequence evolution in the genes sampled.

We do not see any molecular evidence of domestication in any of the *Spinacia* species sampled, but we do see evidence for increased rates of sequence change in *Spinacia* compared to its closest relatives. Amino acid sequence evolution among the genes studied is being driven by insertion/deletion events and not by nucleotide substitutions, suggesting that new models of sequence evolution be used for *Spinacia*. We note extensive changes in floral morphology and reproductive strategy in this group, and we find that sequence evolution in the genes sampled is driven primarily by major insertion/deletion changes that are unique to *Anserineae* and which are linked to the development of dioecy in *Spinacia*. The *rbcL* gene is evolving inconsistently with other gene sequences presented here and with other novel sequence data recently used for phylogenetic analysis of the Chenopodiaceae. Trees generated from the *rbcS*, *matK*, *UFO*, *AG*, *AP3*, and *PI* genes most likely represent the actual evolutionary history of the *Anserineae*, while at the same time providing key insights into the evolution of reproductive strategy in this tribe.



**Figure 39. Horizontal cladogram of characters along evolutionary branches in *Anserineae*.**

This tree follows the consensus phylogeny of our concatenated dataset.. Reduction in sepal number (A) and anther number (B) is seen multiple times within *Anserineae*.

## FUTURE WORK

Species within the Anserineae display a wide range of floral morphological variation. Floral organ development genes in classes A, B, C, and E, as well as their regulators such as *LFY* and *UFO*, also exhibit a wide range of sequence variation within this group. Investigating the difference in function and interactivity between the products of the *UFO*, *AP3*, and *PI* orthologs sequenced in this study is the logical next step after this work. Whether or not the unique combination of sequences found only in *Spinacia* is causally linked to dioecy is the next main question to be answered by functional testing. Understanding the link between dioecy, sexual lability, and hormone stress is also necessary to uncover the basic mechanisms behind unisexual flower development in *Spinacia* and other species. Co-immunoprecipitation (Co-IP), chromatin immunoprecipitation (ChIP), quantitative PCR (qPCR), and transgenic protocols taking advantage of CRISPR-dependent methods could help to casually link the development of unisexual flowers with certain gene mutations highlighted in this work. Additional sampling of species and populations *in natura* could also illuminate other novel mechanisms by which unisexual flowers develop from initiation. We hypothesize that several key mutations described in this work play a key role in the development of unisexual flowers in *Spinacia*, and that a mutation shared by members of Anserineae is an example of mutational preadaptation. Our aims are: 1.) to determine the developmental effects of observed mutations in *Spinacia*, 2.) to determine if observed mutations in key floral organ development genes are linked to developmental changes in floral organs in Anserineae, and 3.) to further characterize the interactivity of these genes and their products, and their roles in floral organ development in Chenopodiaceae.

In order to determine the physiological and developmental effects the mutations described in this work have on floral organs in *Spinacia*, we must first understand the functional difference (if any) between the protein products of the B-class gene sequences observed in *Spinacia* and those observed in purely hermaphroditic species, e.g. *Chenopodium album*. It has been hypothesized that the products of floral organ identity genes in the A, B, C, and E-classes interact to form quartets or complexes of four proteins [87]. It is the composition of these quartets that are thought to affect the downstream activation of genes related to the development of floral organs. The carboxyl-ends of these gene products have been shown to be an essential determinant in how these products interact to form these complexes [85 Floral Quartets]. We hypothesize that variations observed in the carboxyl-end of the *APETALA3* and *PISTILLATA* genes within the Anserineae could affect formation of these complexes. Co-immunoprecipitation techniques can be used to determine the functional differences among the protein products of the various gene orthologs revealed in this work, as they relate to protein-protein interaction and quartet assembly. Whether or not the combination of key motif deletions unique to *Spinacia* is directly related to the development of unisexual flowers remains to be casually determined, and experiments utilizing Co-IP are a logical next step.

We hypothesize that mutation in key floral organ development genes observed to be shared within the Anserineae act as mutational preadaptation to certain unique mutations observed solely in *Spinacia*. In order to characterize the effects these shared mutations have on floral organ development, it would be useful to establish some transgenic protocols for the Chenopodiaceae. In the past, development of transgenic protocols for *Spinacia* has proven difficult. *Agrobacterium* transformation and callus regeneration protocols similar to those used in *Arabidopsis* [88, 89] need to be advanced to a state where they can consistently produce

results in *Spinacia*. With these protocols in place, CRISPR (clustered regularly interspaced short palindromic repeats) could also provide a novel method to replace genes by a true “cut-and-paste” technique [90]. If adapted to the Chenopodiaceae, this method could be used to determine the effect each unique sequence change has on floral organ morphology in *Spinacia*, by knocking-in floral organ development sequences ancestral to the onset of dioecy in this group.

In this work, we also discovered an astonishing amount of variation between orthologs in the form of insertion/deletion events, not only in non-coding regions but also in coding regions. The interplay between genes involved with floral organ development and their regulators has been described [76] and the intron sequence differences described in this work highlight the importance of characterizing these relationships. Chromatin immunoprecipitation could be used to determine whether putative regulatory motifs in non-coding regions play a role in the differential regulation of these genes. In-depth analysis of the differential expression of genes following hormone treatments, particularly gibberellic acid, could lead to a better understanding of how the floral organ development genes have evolved and are regulated in this group of plants.

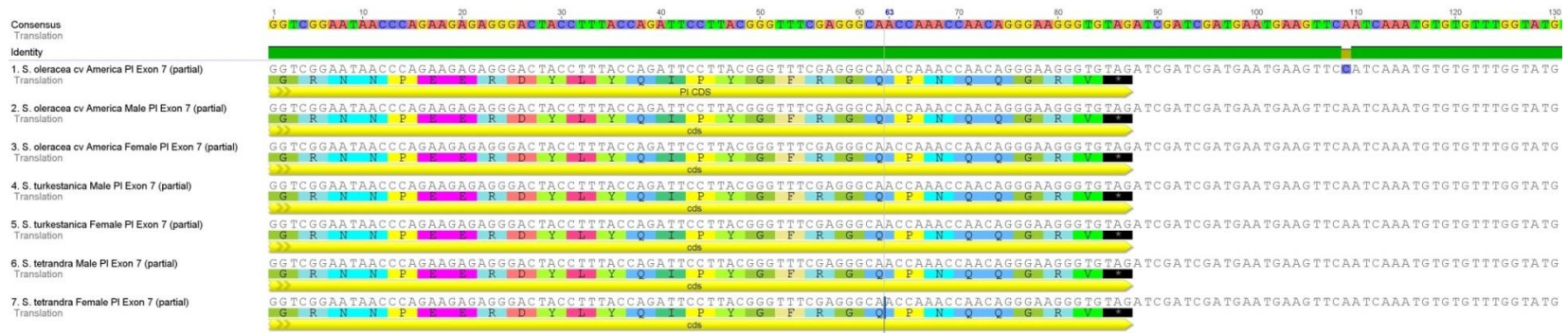
What is the nature of the common ancestor to the members of *Anserineae*? Is this ancestor hermaphroditic? Is this ancestor monoecious or dioecious? Or some other intermediate state? Greater taxonomic sampling and morphological analysis could help to determine how dioecy evolved as a sexual strategy in *Spinacia*, and determine if dioecy evolved by means of similar mechanisms in other close relatives like *Atriplex* and *Grayia*. The rapidly falling cost of sequencing [91] makes it very reasonable to propose that the entire genomes of various key Amaranths be sequenced and analyzed. This method brings with it unique computational challenges, but the prospect of quickly obtaining all the sequences for the floral organ development genes across many closely-related genera is an exciting possibility. New models of



evolution that take into account variation in the form of insertion/deletion events could also shed some light on the evolution of mating systems in this group of plants [92]. There are monoecious cultivars of *Spinacia* which should have their genomes sequenced and analyzed alongside that of dioecious spinach. Additionally, populations of “wild” *Spinacia tetrandra* and *Spinacia turkestanica* should be sampled more extensively as they occur *in natura*. This would help to determine if *Spinacia* is monospecific and also would further clarify the role that domestication has played in this plant’s evolutionary history, if any.

The task of characterizing the functional mechanisms behind the development of unisexual flowers in *Spinacia* is part of a vexatious puzzle that has not yet been fully solved. Future work in this area should focus on key differences between dioecious *Spinacia* and its closest non-dioecious relatives. The sequence differences described in this paper provide a launching point for these functional studies, but it is likely that a confluence of gene sequence mutations is involved. Further sequencing within *Spinacia* and its closest relatives might provide other targets for functional testing.

**Supplemental Figure 1.** Sequences from the *PI* C-terminal domain all the species of *Spinacia* (male and female).



**Supplemental Table 1.** Primers used in this study.

Name	Sequence	Gene
JRBCS.1.F	ATGGCTTCCTCMRTSMTCTCYTC	<i>rbcS</i>
JRBCS.580.Rev	TTAGWAGCCWGSRGGCTTGTAGGC	<i>rbcS</i>
AP Short	GGCCACGCGTCGACTACTTTTTTTTTTTTTTT	n/a
AuAP Short	GGCCACGCGTCGACTACT	n/a
AuAP Short (Nest)	CCACGCGTCGACTACTTT	n/a
SpUFO.58F	ACATTCAGCCCTCCTTTTCAGC	<i>UFO</i>
SpUFO.1279Rev	ACGGCATCCAAATCCACCTTTT	<i>UFO</i>
ChUFO.328F	CGKRCACGKTSYGTCTGCAAGAGR	<i>UFO</i>
ChUFO.aa237F	CCWMGDCTYTTYCCTTCVRTWGG	<i>UFO</i>
ChUFO.aa333R	CTBARRAAYCTYCTCATVGGWGC	<i>UFO</i>
ChUFO.aa356R	ARCTTRCTYTTYTCRACAGCKGC	<i>UFO</i>
CbhUFO.117F	ACCCAACACGAGCATCGTTTACCG	<i>UFO</i>
CbhUFO.287F	CCGCTACCTCCTCGGGTGGG	<i>UFO</i>
CaUFO.61F	CTCTACCTTCAAGTCTCCCCTCGCC	<i>UFO</i>
CaUFO.268F	CCCCTAGGGTTTTTCGCCCCG	<i>UFO</i>
CaUFO.313F	TGGGTGTCCGATGAACCCGGC	<i>UFO</i>
MnutUFO.267F	TTCACCCGCTACCTCCTCGGG	<i>UFO</i>
MnutUFO.475F	GCAGTCAAAAACCTCACCGCCG	<i>UFO</i>
ChUFO.aa651R	SNCWRAAAGDGAARARSTAGC	<i>UFO</i>
ChUFO.1325R	CCMTARATTMTGAWARTTAACAC	<i>UFO</i>
JChAG.275.F	ATTGASAGGTAYAAGAARGC	<i>AG</i>
JChAG.557.Rev	TTYTGCATRWASTCRATYTCRGC	<i>AG</i>
ChAG.E2.degen.F	TGCTCTGATCAGACYGGTKCTGG	<i>AG</i>
ChAG.234.E3.F	CTATCAGCAAGAGGCTGCCAAGC	<i>AG</i>
ChAG.395.E4.F	TGATGGGYGAAGGGCTGAGC	<i>AG</i>
ChAG.1351.E7.degen.Rev	AGMGSRGTYTGGTCTTGCGG	<i>AG</i>
ChAG.1362.E7degen.Rev	ACACTAACTGGAGMGSRGTYTGG	<i>AG</i>
AG_F_seq1	GCCTTATCCATCTAAGTTAGTACTCCG	<i>AG</i>
AG_F_seq2	TCTAGTCTTCCCCTTTCATTGGG	<i>AG</i>
CBH AG_seq 1	AATCAAGCTTATGTCCATTTGATCCG	<i>AG</i>
CBH AG_seq 2	TGTTTATTGGAAGTACTACAAAGAGCC	<i>AG</i>
CBH AG_seq 3	ACTATTCAATATTGCTCATAACATACC	<i>AG</i>
CBH AG_A2_seq 1	GGGATGGAGGAAAGCATAACC	<i>AG</i>
CBH AG_A2_seq 2	TGGAGTACTTATTGGAAATAAAAGGGC	<i>AG</i>
JChAG.584.F	GATCGAGTTCATGCAGAAAAGGG	<i>AG</i>
JChAG.1395.Rev	CTAGTGGCTTAAACATTGCAAAGC	<i>AG</i>
JChAP3.aa1.E1.F	GGSRAGAGGMAARWTMSARATWAAG	<i>AP3</i>
SpAP3.226R	CAWAKATCAACCCCYAWAGC	<i>AP3</i>
ChAP3.493.E2.Rev	TCRTAAAYCTSCTTCGTCCTAYACAC	<i>AP3</i>

Name	Sequence	Gene
ChAP3.463.E2.F	KKRYTGTGTRTAGGACGAAGSAG	AP3
ChAP3.480.E2.F	AAGSAGRRTTAYGAYAARTAYCAGGC	AP3
JChAP3.492.E2.F	AGGMCSAAGSAGRRTTAYG	AP3
CaAP3.311.E2.F	TTTTTCTTTGATGTTTGCAGGCC	AP3
CaAP3.319.E2.F	TGATGTTTGCAGGCCCAAGCAGG	AP3
CaAP3.2998.E4.Rev	AGCCAAATCCTCCAAACAATCTCCC	AP3
CaAP3.3027.E4.Rev	ATTCACCATCTCATCTTCAAGACAGCG	AP3
CaAP3.3053.E4.Rev	TTGCGTTCACGAATGACCTTCACAGC	AP3
JChAP3.4861.E4.Rev	GYCARRTCCTCCAAACAATCTCCC	AP3
JChAP3.4921.E4.Rev	YCACGAATGACSTTCACAGAGTYCACC	AP3
SpAP3.311	ATGGGAGATTGTTTGGAGGA	AP3
SpAP3.588.R	GCAAAGCAAGTATGCGAGAG	AP3
ChAP3.aa211R	SAKGGYKGCANCKCAAAGC	AP3
ChPI.27.E1.UD.F	GAGGATHGARAACDCMAVHAAYAGRS	PI
ChPI.154.E1.UD.F	GGNAARATGMMYGMKTAYWRYWGYC	PI
ChPI.992.E3.UD.Rev	GCAWRYYRTCRTTYTCYTTYTTGATYC	PI
ChPI.1223.E4.UD.Rev	WYBDDCRAKVCCAWKHTCAAKRGC	PI
ChPI.759.E2.edge.F	TGGGATGCTAAGCATGAGGTG	PI
ChPI.1402.I2.F	CATWTGAACAGATCTATAAATRGG	PI
CbhPI.2948.E4.Rev	GCCTGCCTGTCATTTACGCC	PI
ChPI.3928.E6.Rev	AMTYGAGCTGGTTWYTTTCCTC	PI
CaPI.3primeUTR.outer.Rev	GCTAGCTAGGATCTTGTTCCGG	PI
CaPI.3primeUTR.inner.Rev	CGGGTATATATGGTATGGCCGG	PI
ChPI.E4.edge.intron7bridge.F	TAGGCACTTGAATGGAGAGG	PI
ChPI.intron7bridge.F	CTTGWAAWTKRWCAGACAWCGATGC	PI
ChPI.exon7inner.bridge.Rev	GGAATCTGGTAWAGGTAGTCC	PI
ChPI.exon7outer.bridge.Rev	CTCGRAAVCCGTAAGGAATCTGG	PI
CaPI.E4.196.F	CAATACCCAGATTTGATTCGCC	PI
CaPI.E4.219.F	CGAAGAAGCTCTGGATATTGGC	PI
SpPI.E7.4234.s.degen.F	AATGAAATGATGGACKGGCCG	PI
SpPI.3UTR.4420.Rev	TGAACAAGTTGATAAACAAGCC	PI
CbhPI.3137.E5.Rev	TACGCGTACATTCTTTTCGTGC	PI
Cbh.Eu-PI.I5.F	CACAACGTATTATTGCTTTGGCG	PI
Cbh.Eu-PI.I6.Rev	AACATTCTACTGTGAGATTTAGTGCG	PI
Cbh.PI-like-2.E3.F	GCAATGAGATAGAAAGGATCAAGAAGG	PI-like
Cbh.PI-like-2.E3.ds.F	TCAAGAAGGAGAATGACAACATGC	PI-like

**Supplemental Table 2.** List of sequences generated by the authors with tissue sample and GenBank numbers.

Scientific Name	Sample	Accession	<i>rbcS</i>	AG	AP3	PI	UFO
<i>Spinacia oleracea</i> cv <i>America</i>	gDNA #026, 9.27.11 (Male)	Twilley Seed Co	KJ920220	KM052154	GQ120478	GQ120477	KM076618
<i>Spinacia oleracea</i> cv <i>America</i>	gDNA #027, 9.27.11 (Male)	Twilley Seed Co	----	----	----	----	KM076619
<i>Spinacia oleracea</i> cv <i>America</i>	gDNA #032, 9.30.11 (Female)	Twilley Seed Co	----	KM052153	----	----	----
<i>Spinacia oleracea</i> cv <i>America</i>	SpUFOcomplete (EMG, unpublished)	Twilley Seed Co	----	----	----	----	KM076628
<i>Spinacia oleracea</i> cv <i>Giant Noble</i>	gDNA #024, 8.31.11 (Male)	Trade Winds Fruit	KM030280	----	----	----	----
<i>Spinacia oleracea</i> cv <i>Viroflay</i>	gDNA #043, 12.28.11 (Female)	Trade Winds Fruit	KM030281	----	----	----	----
<i>Spinacia tetrandra</i>	gDNA #039, 12.21.11 (Male)	Ames 23664	KM030282	KM052155	----	----	KM076620, KM076621
<i>Spinacia tetrandra</i>	gDNA #037, 12.21.11 (Female)	Ames 23664	----	KM052156	----	----	----
<i>Spinacia tetrandra</i>	gDNA # 017, 8/18/11 (Female)	Ames 23664	----	----	----	----	KM076622
<i>Spinacia tetrandra</i>	gDNA # 046, 12.28.11 (Female)	PI 608712	----	----	----	----	KM076623
<i>Spinacia turkestanica</i>	gDNA #009, 8.25.11 (Male)	PI 647862	KM030283	KM052157	----	----	KM076624
<i>Spinacia turkestanica</i>	gDNA #016, 8.18.11 (Male)	Ames 23666	----	----	----	----	KM076625
<i>Spinacia turkestanica</i>	gDNA #006, 7.25.11 (Female)	Ames 23666	----	----	----	----	KM076626
<i>Spinacia turkestanica</i>	gDNA #010, 8.18.11 (Female)	Ames 23666	----	KM052158	----	----	KM076627
<i>Chenopodium foliosum</i>	gDNA #Cf003 7.17.12	Ames 26923	KM030287	KM052159	KM076605	KM076606	----
<i>Chenopodium foliosum</i>	gDNA #Cf001 10.3.11, cDNA #Cf002 4.4.12	Ames 26923	----	----	----	----	KM076614
<i>Chenopodium album</i>	gDNA #Ca001 9.5.12, cDNA #Ca001 2.22.12	Ames 28345, wild collected	KM030286	KM052160	KM076602, KM076603	KM076607	KM076615
<i>Chenopodium bonus-henricus</i>	gDNA #Cbh002 7.17.12	PI 662294	KM030288	KM052162, KM052163	KM076601	KM076609	----
<i>Chenopodium bonus-henricus</i>	gDNA #Cbh041 (12.21.11), cDNA #Cbh001 2.22.12	PI 662294	----	----	----	KM076612, KM076613	KM076617
<i>Monolepis nuttalliana</i>	gDNA #Mnut002 (12.10.12), gDNA #047 (3.27.12)	PI 658757	KM030284, KM030285	KM052161	KM076604	KM076608	KM076616

**Supplemental Table 3.** List of sequences analyzed in this study that were obtained from the nucleotide database at the NCBI.

Scientific Name	Common Name	<i>rbcL</i>	<i>matK</i>
<i>Spinacia oleracea</i> cv <i>Geant d'hiver</i> & cv <i>Monatol</i>	Cultivated spinach	AJ400848.1	NC_002202.1
<i>Spinacia turkestanica</i>	Wild spinach	----	HE855620.1
<i>Spinacia tetrandra</i>	Wild spinach	----	HE855619.1
<i>Chenopodium foliosum</i>	Strawberry spinach	AY270081.1	HE855617.1
<i>Chenopodium album</i>	Lamb's Quarters	HM849888.1	HE855644.1
<i>Chenopodium bonus-henricus</i>	Lincolnshire spinach	AY270079.1	HE855613.1
<i>Monolepis nuttalliana</i>	Nuttall's povertyweed	AY270108.1	HE855621.1
<i>Mesembryanthemum crystallinum</i>	Common ice plant	HM850175.1	HM850877.1
<i>Beta vulgaris</i>	Beet	AY270065.1	AY514832.1

**Supplemental Table 4.** List of sequences included in the course of this study, but which were considered homologs.

Scientific Name	Common Name	Sample	Accession	GenBank #	Name
<i>Chenopodium album</i>	Lamb's Quarters	gDNA #Ca001 9.5.12, cDNA #Ca001 2.22.12	Ames 28345, & wild collected	KM076610	<i>PI</i> -like 1
<i>Chenopodium bonus-henricus</i>	Lincolnshire spinach	cDNA #Cbh001 (2.22.12)	<i>PI</i> 662294	KM076611	<i>PI</i> -like 1

## REFERENCES

1. Avise, J.C., W.S. Nelson, and H. Sugita, *A Speciation History of "Living Fossils": Molecular Evolutionary Patterns in Horseshoe Crabs*. *Evolution*, 1994. **48**(6): p. 1986-2001.
2. Sturmbauer, C. and A. Meyer, *Genetic divergence, speciation and morphological stasis in a lineage of African cichlid fishes*. *Nature (London)*, 1992. **358**(6387): p. 578-581.
3. Elmer, K.R., et al., *Local variation and parallel evolution: morphological and genetic diversity across a species complex of neotropical crater lake cichlid fishes*. *Philosophical transactions. Biological sciences*, 2010. **365**(1547): p. 1763-1782.
4. Cubas, P., C. Vincent, and E. Coen, *An epigenetic mutation responsible for natural variation in floral symmetry*. *Nature (London)*, 1999. **401**(6749): p. 157-161.
5. Golenberg, E.M., et al., *Evolution of a noncoding region of the chloroplast genome*. *Molecular Phylogenetics and Evolution*, 1993. **2**(1): p. 52-64.
6. Borsch, T. and D. Quandt, *Mutational dynamics and phylogenetic utility of noncoding chloroplast DNA*. *Plant systematics and evolution*, 2009. **282**(3-4): p. 169-199.
7. Easteal, S. and C. Collet, *Consistent variation in amino-acid substitution rate, despite uniformity of mutation rate: protein evolution in mammals is not neutral*. *Molecular Biology and Evolution*, 1994. **11**(4): p. 643-647.
8. Vawter, L. and W.M. Brown, *Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in the molecular clock*. *Science (New York, N.Y.)*, 1986. **234**(4773): p. 194-196.
9. Wright, S., *Tempo and Mode in Evolution: A Critical Review*. *Ecology*, 1945. **26**(4): p. 415-419.
10. Goldschmidt, R., *The Material Basis of Evolution*. 1940: Yale University Press.

11. Fuentes-Bazan, S., G. Mansion, and T. Borsch, *Towards a species level tree of the globally diverse genus *Chenopodium* (Chenopodiaceae)*. *Molecular Phylogenetics and Evolution*, 2012. **62**(1): p. 359-374.
12. Fuentes-Bazan, S., P. Uotila, and T. Borsch, *A novel phylogeny-based generic classification for *Chenopodium* sensu lato, and a tribal rearrangement of *Chenopodioideae* (Chenopodiaceae)*. *Willdenowia*, 2012. **42**(1): p. 5-24.
13. Muller, K. and T. Borsch, *Phylogenetics of amaranthaceae based on matK/trnK sequence data - Evidence from Parsimony, likelihood, and Bayesian analyses*. *Annals of the Missouri Botanical Garden*, 2005. **92**(1): p. 66-102.
14. Kadereit, G., et al., *Phylogeny of Amaranthaceae and Chenopodiaceae and the evolution of C-4 photosynthesis*. *International Journal of Plant Sciences*, 2003. **164**(6): p. 959-986.
15. Kadereit, G., et al., *MOLECULAR PHYLOGENY OF ATRIPLICEAE (CHENOPODIOIDEAE, CHENOPODIACEAE): IMPLICATIONS FOR SYSTEMATICS, BIOGEOGRAPHY, FLOWER AND FRUIT EVOLUTION, AND THE ORIGIN OF C-4 PHOTOSYNTHESIS*. *American Journal of Botany*, 2010. **97**(10): p. 1664-1687.
16. Steele, K.P. and R. Vilgalys, *Phylogenetic analyses of Polemoniaceae using nucleotide-sequences of the plastid gene matk*. *Systematic Botany*, 1994. **19**(1): p. 126-142.
17. Johnson, L.A. and D.E. Soltis, *matK DNA Sequences and Phylogenetic Reconstruction in Saxifragaceae s. str.* *Systematic Botany*, 1994. **19**(1): p. 143-156.
18. Hilu, K.W. and H.P. Liang, *The matK gene: Sequence variation and application in plant systematics*. *American Journal of Botany*, 1997. **84**(6): p. 830-839.
19. Chase, M.W., et al., *PHYLOGENETICS OF SEED PLANTS - AN ANALYSIS OF NUCLEOTIDE-SEQUENCES FROM THE PLASTID GENE RBCL*. *Annals of the Missouri Botanical Garden*, 1993. **80**(3): p. 528-580.



20. Wilson, M.A., B. Gaut, and M.T. Clegg, *CHLOROPLAST DNA EVOLVES SLOWLY IN THE PALM FAMILY (ARECACEAE)*. *Molecular Biology and Evolution*, 1990. **7**(4): p. 303-314.
21. Rettig, J.H., H.D. Wilson, and J.R. Manhart, *PHYLOGENY OF THE CARYOPHYLLALES - GENE SEQUENCE DATA*. *Taxon*, 1992. **41**(2): p. 201-209.
22. Martin, W., et al., *MOLECULAR PHYLOGENIES IN ANGIOSPERM EVOLUTION*. *Molecular Biology and Evolution*, 1993. **10**(1): p. 140-162.
23. Doebley, J., et al., *EVOLUTIONARY ANALYSIS OF THE LARGE SUBUNIT OF CARBOXYLASE (RBCL) NUCLEOTIDE-SEQUENCE AMONG THE GRASSES (GRAMINEAE)*. *Evolution*, 1990. **44**(4): p. 1097-1108.
24. Albert, V.A., S.E. Williams, and M.W. Chase, *CARNIVOROUS PLANTS - PHYLOGENY AND STRUCTURAL EVOLUTION*. *Science*, 1992. **257**(5076): p. 1491-1495.
25. Hilu, K.W., et al., *Angiosperm phylogeny based on matK sequence information*. *American Journal of Botany*, 2003. **90**(12): p. 1758-1776.
26. Gaut, B.S., et al., *RELATIVE RATES OF NUCLEOTIDE SUBSTITUTION AT THE RBCL LOCUS OF MONOCOTYLEDONOUS PLANTS*. *Journal of Molecular Evolution*, 1992. **35**(4): p. 292-303.
27. Kellogg, E.A. and N.D. Juliano, *The structure and function of RuBisCO and their implications for systematic studies*. *American Journal of Botany*, 1997. **84**(3): p. 413-428.
28. Hollingsworth, P.M., et al., *A DNA barcode for land plants*. *Proceedings of the National Academy of Sciences - PNAS*, 2009. **106**(31): p. 12794-12797.
29. Wilkinson, M.D. and G.W. Haughn, *UNUSUAL FLORAL ORGANS CONTROLS MERISTEM IDENTITY AND ORGAN PRIMORDIA FATE IN ARABIDOPSIS*. *Plant Cell*, 1995. **7**(9): p. 1485-1499.
30. Pouteau, S., et al., *Transcription pattern of a FIM homologue in Impatiens during floral development and reversion*. *Plant Journal*, 1998. **14**(2): p. 235-246.

31. Levin, J.Z. and E.M. Meyerowitz, *UFO: an Arabidopsis gene involved in both floral meristem and floral organ development*. The Plant cell, 1995. **7**(5): p. 529-548.
32. Souer, E., et al., *Patterning of inflorescences and flowers by the F-box protein DOUBLE TOP and the LEAFY homolog ABERRANT LEAF AND FLOWER of petunia*. Plant Cell, 2008. **20**(8): p. 2033-2048.
33. Zhao, D.Z., et al., *The ASK1 gene regulates B function gene expression in cooperation with UFO and LEAFY in Arabidopsis*. Development, 2001. **128**(14): p. 2735-2746.
34. Wang, X.P., et al., *The COP9 signalosome interacts with SCFUFO and participates in Arabidopsis flower development*. Plant Cell, 2003. **15**(5): p. 1071-1082.
35. Ni, W.M., et al., *Regulation of flower development in Arabidopsis by SCF complexes*. Plant Physiology, 2004. **134**(4): p. 1574-1585.
36. Laufs, P., et al., *Separable roles of UFO during floral development revealed by conditional restoration of gene function*. Development, 2003. **130**(4): p. 785-796.
37. Hepworth, S.R., J.E. Klenz, and G.W. Haughn, *UFO in the Arabidopsis inflorescence apex is required for floral-meristem identity and bract suppression*. Planta, 2006. **223**(4): p. 769-778.
38. Mandel, M.A., et al., *MANIPULATION OF FLOWER STRUCTURE IN TRANSGENIC TOBACCO*. Cell, 1992. **71**(1): p. 133-143.
39. Kempin, S.A., M.A. Mandel, and M.F. Yanofsky, *CONVERSION OF PERIANTH INTO REPRODUCTIVE-ORGANS BY ECTOPIC EXPRESSION OF THE TOBACCO FLORAL HOMEOTIC GENE NAG1*. Plant Physiology, 1993. **103**(4): p. 1041-1046.
40. Bowman, J.L., D.R. Smyth, and E.M. Meyerowitz, *GENES DIRECTING FLOWER DEVELOPMENT IN ARABIDOPSIS*. Plant Cell, 1989. **1**(1): p. 37-52.
41. Bowman, J.L., G.N. Drews, and E.M. Meyerowitz, *EXPRESSION OF THE ARABIDOPSIS FLORAL HOMEOTIC GENE AGAMOUS IS RESTRICTED TO*

- SPECIFIC CELL-TYPES LATE IN FLOWER DEVELOPMENT*. Plant Cell, 1991. **3**(8): p. 749-758.
42. Saedler, H. and P. Huijser, *MOLECULAR-BIOLOGY OF FLOWER DEVELOPMENT IN ANTIRRHINUM-MAJUS (SNAPDRAGON)*. Gene, 1993. **135**(1-2): p. 239-243.
  43. Riechmann, J.L. and E.M. Meyerowitz, *Determination of floral organ identity by Arabidopsis MADS domain homeotic proteins AP1, AP3, PI, and AG is independent of their DNA-binding specificity*. Molecular Biology of the Cell, 1997. **8**(7): p. 1243-1259.
  44. Kramer, E.M. and V.F. Irish, *Evolution of the petal and stamen developmental programs: Evidence from comparative studies of the lower eudicots and basal angiosperms*. International Journal of Plant Sciences, 2000. **161**(6): p. S29-S40.
  45. Kramer, E.M., R.L. Dorit, and V.F. Irish, *Molecular evolution of genes controlling petal and stamen development: Duplication and divergence within the APETALA3 and PISTILLATA MADS-box gene lineages*. Genetics, 1998. **149**(2): p. 765-783.
  46. Sather, D.N., M. Jovanovic, and E.M. Golenberg, *Functional analysis of B and C class floral organ genes in spinach demonstrates their role in sexual dimorphism*. BMC Plant Biology, 2010. **10**.
  47. Pfent, C., et al., *Characterization of SpAPETALA3 and SpPISTILLATA, B class floral identity genes in Spinacia oleracea, and their relationship to sexual dimorphism*. Development Genes and Evolution, 2005. **215**(3): p. 132-142.
  48. Yang, Y.Z. and T. Jack, *Defining subdomains of the K domain important for protein-protein interactions of plant MADS proteins*. Plant Molecular Biology, 2004. **55**(1): p. 45-59.
  49. Yang, Y.Z., L. Fanning, and T. Jack, *The K domain mediates heterodimerization of the Arabidopsis floral organ identity proteins, APETALA3 and PISTILLATA*. Plant Journal, 2003. **33**(1): p. 47-59.

50. Krizek, B.A. and E.M. Meyerowitz, *The Arabidopsis homeotic genes APETALA3 and PISTILLATA are sufficient to provide the B class organ identity function*. *Development*, 1996. **122**(1): p. 11-22.
51. Diggle, P.K., et al., *Multiple developmental processes underlie sex differentiation in angiosperms*. *Trends in Genetics*, 2011. **27**(9): p. 368-376.
52. Katoh, K. and H. Toh, *Recent developments in the MAFFT multiple sequence alignment program*. *Briefings in Bioinformatics*, 2008. **9**(4): p. 286-298.
53. Higo, K., et al., *Plant cis-acting regulatory DNA elements (PLACE) database: 1999*. *Nucleic Acids Research*, 1999. **27**(1): p. 297-300.
54. Kelley, L.A. and M.J.E. Sternberg, *Protein structure prediction on the Web: a case study using the Phyre server*. *Nature Protocols*, 2009. **4**(3): p. 363-371.
55. Gadagkar, S.R., M.S. Rosenberg, and S. Kumar, *Inferring species phylogenies from multiple genes: Concatenated sequence tree versus consensus gene tree*. *Journal of Experimental Zoology Part B-Molecular and Developmental Evolution*, 2005. **304B**(1): p. 64-74.
56. Guindon, S., et al., *New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0*. *Systematic biology*, 2010. **59**(3): p. 307-321.
57. Tamura, K., et al., *MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods*. *Molecular Biology and Evolution*, 2011. **28**(10): p. 2731-2739.
58. Subrahmanyam, N.S., *Modern Plant Taxonomy, 1E*. 2009: Vikas Publishing House Pvt Limited.
59. Kumar, A.B.D.A., *A Text Book Of Practical Botany 2*. 2009: Rastogi Publications.

60. Sather, D.N., et al., *Sequence evolution and sex-specific expression patterns of the C class floral identity gene, SpAGAMOUS, in dioecious Spinacia oleracea L.* *Planta*, 2005. **222**(2): p. 284-292.
61. Gilmartin, P.M., et al., *Molecular light switches for plant genes.* *The Plant Cell Online*, 1990. **2**(5): p. 369-78.
62. Terzaghi, W.B. and A.R. Cashmore, *Light-Regulated Transcription.* *Annual Review of Plant Physiology and Plant Molecular Biology*, 1995. **46**(1): p. 445-474.
63. Craig, K.L. and M. Tyers, *The F-box: a new motif for ubiquitin dependent proteolysis in cell cycle regulation and signal transduction.* *Progress in Biophysics & Molecular Biology*, 1999. **72**(3): p. 299-328.
64. Kaufmann, K., R. Melzer, and G. Theissen, *MIKC-type MADS-domain proteins: structural modularity, protein interactions and network evolution in land plants.* *Gene*, 2005. **347**(2): p. 183-198.
65. Wenkel, S., et al., *CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of Arabidopsis.* *Plant Cell*, 2006. **18**(11): p. 2971-2984.
66. Borsch, T., et al., *Noncoding plastid trnT-trnF sequences reveal a well resolved phylogeny of basal angiosperms.* *Journal of evolutionary biology*, 2003. **16**(4): p. 558-576.
67. Xiang, Q.Y., et al., *Timing the eastern Asian-eastern North American floristic disjunction: molecular clock corroborates paleontological estimates.* *Molecular Phylogenetics and Evolution*, 2000. **15**(3): p. 462-472.
68. Moquin-Tandon, A., *Chenopodearum monographica enumeratio; auctore A. Moquin-Tandon.* 1840, Parisiis: apud P.-J. Loss.
69. Sneep, J., *The Domestication of Spinach and the Breeding History of Its Varieties.* 1982: Euphytica.

70. Nohara, S., *Genetic Studies on Spinacia*. 1923.
71. van der Vossen, H.A.M., *Spinacia oleracea L.*, in *Prota 2: Vegetables/Légumes*, G.J.H. Grubben and O.A. Denton, Editors. 2004, PROTA: Wageningen, Netherlands.
72. Prohens, J., F. Nuez, and M.J. Carena, *Handbook of plant breeding*. 2008: Springer.
73. Janick, J. and E.C. Stevenson, *Genetics of the Monoecious Character in Spinach*. *Genetics (Austin)*, 1955. **40**(4): p. 429-437.
74. Onodera, Y., et al., *Mapping of the genes for dioecism and monoecism in Spinacia oleracea L.: evidence that both genes are closely linked*. *Plant Cell Reports*, 2011. **30**(6): p. 965-971.
75. Mahoney, D.L., J. Janick, and E.C. Stevenson, *Sex Determination in Diploid-Triploid Crosses of Spinacia oleracea*. *American Journal of Botany*, 1959. **46**(5): p. 372-375.
76. Golenberg, E.M. and N.W. West, *Hormonal interactions and gene regulation can link monoecy and environmental plasticity to the evolution of dioecy in plants*. *American Journal of Botany*, 2013. **100**(6): p. 1022-1037.
77. Eichler, A.W., *Blüthendiagramme /construirt und erläutert von A.W. Eichler*. Vol. v. 2. 1878, Leipzig :: W. Engelmann.
78. Sherry, R.A., K.J. Eckard, and E.M. Lord, *Flower Development in Dioecious Spinacia oleracea (Chenopodiaceae)*. *American Journal of Botany*, 1993. **80**(3): p. 283-291.
79. Chae, E., et al., *An Arabidopsis F-box protein acts as a transcriptional co-factor to regulate floral development*. *Development (Cambridge)*, 2008. **135**(7): p. 1235-1245.
80. Heslop-Harrison, J., *THE EXPERIMENTAL MODIFICATION OF SEX EXPRESSION IN FLOWERING PLANTS*. *Biological reviews of the Cambridge Philosophical Society*, 1957. **32**(1): p. 38-90.
81. Chailakhyan, M.K. and V.N. Khryanin, *Effect of growth regulators and role of roots in sex expression in spinach*. *Planta*, 1978. **142**(2): p. 207-210.

82. Laufer, B., *Sino-Iranica: Chinese contributions to the history of civilization in ancient Iran, with special reference to the history of cultivated plants and products*. 1919: Field Museum of Natural History.
83. Simoons, F.J., *Food in China: A Cultural and Historical Inquiry*. 1991: CRC Press.
84. Hallavant, C. and M.-P. Ruas, *The first archaeobotanical evidence of *Spinacia oleracea* L. (spinach) in late 12th–mid 13th century a.d. France*. *Vegetation History and Archaeobotany*, 2014. **23**(2): p. 153-165.
85. Theissen, G., *Development of floral organ identity: stories from the MADS house*. *Current opinion in plant biology*, 2001. **4**(1): p. 75-85.
86. Durfee, T., et al., *The F-box-containing protein UFO and AGAMOUS participate in antagonistic pathways governing early petal development in Arabidopsis*. *Proceedings of the National Academy of Sciences - PNAS*, 2003. **100**(14): p. 8571-8576.
87. Honma, T. and K. Goto, *Complexes of MADS-box proteins are sufficient to convert leaves into floral organs*. *Nature*, 2001. **409**(6819): p. 525-529.
88. Tinland, B., et al., *The Agrobacterium tumefaciens virulence D2 protein is responsible for precise integration of T-DNA into the plant genome*. *The EMBO journal*, 1995. **14**(14): p. 3585-3595.
89. Puchta, H., B. Dujon, and B. Hohn, *Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination*. *Proceedings of the National Academy of Sciences - PNAS*, 1996. **93**(10): p. 5055-5060.
90. Auer, T.O., et al., *Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair*. *Genome research*, 2014. **24**(1): p. 142-153.
91. Wetterstrand, K.A. *DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)*. 6/9/2014]; Available from: [www.genome.gov/sequencingcosts](http://www.genome.gov/sequencingcosts).

92. Hollister, J.D., J. Ross-Ibarra, and B.S. Gaut, *Indel-Associated Mutation Rate Varies with Mating System in Flowering Plants*. *Molecular Biology and Evolution*, 2010. **27**(2): p. 409-416.



**ABSTRACT****RATES AND MODES OF SEQUENCE EVOLUTION IN VARIOUS LINEAGES  
WITHIN CHENOPODIACEAE**

by

**JAMES A. NAEGER****August 2014****Advisor:** Dr. Edward Golenberg**Major:** Biological Sciences**Degree:** Master of Science

Sexual dimorphism in domesticated spinach, *Spinacia oleracea*, is thought to be determined by differential expression of floral organ identity genes. We examined the floral organ morphologies of seven species in the Chenopodiaceae, including two wild species of spinach, in order to gain insight into the evolution of floral reproductive strategy in this clade. The species within the Anserineae demonstrate extensive evolution in floral morphology and reproductive strategy, and *Spinacia* is unique for having been domesticated rather recently and for being dioecious. We found *C. album* to be hermaphroditic, *C. foliosum* and *M. nuttalliana* to be gynomonocious, while *C. bonus-henricus* exhibits protogyny. For each species and cultivar examined, we isolated sequences for the genes *UNUSUAL FLORAL ORGANS (UFO)*, *AGAMOUS (AG)*, *APETALA3 (AP3)*, *PISTILLATA (PI)* and for the gene coding for the small subunit of the RuBisCO complex (*rbcS*), which is involved in carbon fixation during photosynthesis. We also analyzed published sequences for the chloroplast genes *rbcL* and *matK*. The *rbcS*, *matK*, *UFO*, *AG*, *AP3*, and *PI* genes reveal a consistent phylogeny that most likely represents the true organismal phylogeny for these species, providing key insights into the

evolution of reproductive strategy in this tribe. Phylogenetic analysis of the *rbcL* gene shows *Spinacia* as sister to *M. nuttalliana*, and placing *C. album* within a cluster with *C. bonus-henricus* and *C. foliosum*. This tree topology deviates from those generated from *TRNL-F* /*ITS* data as well as all others generated in this study. All of these other gene analyses consistently give *Spinacia* as sister to a clade that includes *C. bonus-henricus*, *C. foliosum*, and *M. nuttalliana*, recently recognized as the genus *Blitum*. We demonstrate evidence for increased rates of nucleotide substitutions in *Spinacia*, though there is no molecular evidence for domestication or positive selection. Our findings indicate that sequence evolution in this tribe is being driven mostly by insertion/deletion mutations for the genes sampled, and show that *Spinacia* has a unique combination of mutations in the coding regions of the floral identity genes *AP3* and *PI*. We also describe unique polymorphisms in a hypervariable region of the *UFO* gene that are unique to *Spinacia*.

**AUTOBIOGRAPHICAL STATEMENT**

Benjamin Franklin is quoted as saying, “If you would not be forgotten as soon as you are dead, either write something worth reading or do something worth writing.” I hope that with this thesis I will have finally done something that falls into the first category, although I hope to have many more somethings in both.