2022

Disentangling the Complex Evolutionary History of Invasive Russian Thistle: Phylogenetic Relationships, the Role of Hybridization, and Public Engagement

Deva Raj Khanal
South Dakota State University, deva.khanal@jacks.sdstate.edu

DISENTANGLING THE COMPLEX EVOLUTIONARY HISTORY OF INVASIVE RUSSIAN THISTLE: PHYLOGENETIC RELATIONSHIPS, THE ROLE OF HYBRIDIZATION, AND PUBLIC ENGAGEMENT

BY

DEVA RAJ KHANAL

A thesis submitted in partial fulfillment of the requirements for the

Master of Science

Major in Biological Sciences

Specialization in Natural Resource Management

South Dakota State University

2022

THESIS ACCEPTANCE PAGE

deva raj khanal

This thesis is approved as a creditable and independent investigation by a candidate for the master's degree and is acceptable for meeting the thesis requirements for this degree. Acceptance of this does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Maribeth Latvis
Advisor Date

Michele R. Dudash
Department Head Date

Nicole Lounsbery, PhD
Director, Graduate School Date

I would like to dedicate this thesis to my parents, Shiva Raj Khanal and Bishnu Maya Khanal. This work would not have been possible without your love, blessings, and encouragement.

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my advisor, Dr. Maribeth Latvis, for this great opportunity and supervision during my MS journey. Thank you for being a wonderful advisor who supported, guided, and encouraged me in my every step. I also extend my sincere gratitude towards my committee members: Dr. David Tank, Dr. Krista Ehlert, and graduate faculty representative Dr. Aritra Banerjee for serving on my committee and for their valuable suggestions and support.

I would also like to sincerely thank Drs. Jose Gonzalez and Yeyan Qiu from the South Dakota State University Genome Sequencing Facility for their guidance and training, and Dr. Anton Semenchenko, Rachael Auch and Luke Gassman from Information Technology at SDState for their tremendous help regarding High Performance Computing troubleshooting and optimization. I thank the broad Consortium for Plant Invasion Genomics (CPING) network for their expertise, support, and guidance throughout this project.

I am grateful to former and current members from the Latvis Lab (Sarah Herzog, Braedon Honetschlager, Kaitlin Schieuer, Eric Puetz, and Sierra Sattler), and the entire Department of Natural Resource Management for their friendly assistance and input during my time at SDState. I would also like to thank the herbarium institutions (BRIT, CDA, DAV, HUH, NYBG), the Roalson Lab at Washington State University, and the volunteer participants in the CPING *Salsola* Community Science Program for providing sampling resources for this research.

I would like to acknowledge, with gratitude, the funding agencies that made this project possible. I recognize CPING (NSF OIA 1920858) and USDA (USDA NIFA

1016764) for the majority of financial support, the American Society of Plant Taxonomists (ASPT) for providing a 2021 Graduate Student Research Grant 2021 and ASPT BIPOC Excellency in Systematic Botany Award, and the Department of Natural Resource Management for providing me with the 2022 Outstanding Biological Science (M.S.) Graduate Student Award.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

DISENTANGLING THE COMPLEX EVOLUTIONARY HISTORY OF INVASIVE RUSSIAN THISTLE: PHYLOGENETIC RELATIONSHIPS, THE ROLE OF HYBRIDIZATION, AND PUBLIC ENGAGEMENT

DEVA RAJ KHANAL

2022

Invasive species are forces of environmental, economic, and social change with detrimental impacts on ecosystems, commerce, agriculture, and human health. However, we still have a poor understanding of the processes that could be driving invasive potential. Hybridization may contribute to the establishment and spread of invasive lineages, as it may enhance genetic variation and facilitate rapid adaptation. Focusing on Russian thistle (*Salsola tragus s.l.*, family Amaranthaceae) and closely related species, my thesis aims to clarify phylogenetic relationships and investigate how hybridization has shaped the invasion dynamics of these lineages. Known as one of the fastest plant invasions in the United States, these species are of management concern, yet taxonomic confusion impedes communication among stakeholders. My thesis aims to: 1) describe the history of *Salsola* invasion into the United States, provide a glimpse into the taxonomic confusion of these lineages, describe management strategies, and touch on how genetic information may clarify our understanding of their biology, and 2) build a phylogenetic framework for tribe Salsoleae and genus *Salsola s.l.* to understand the role of hybridization in plant invasions. I conduct a phylogenomic study using targeted sequencing of the Angiosperms353 probe set with the goal of resolving relationships at several taxonomic scales. Taken together,

these chapters broadly communicate aspects of invasion biology to diverse audiences and will result in an improved understanding of this challenging group of invasive plants.

INTRODUCTION

Invasive species are forces of environmental, economic, and social change with detrimental impacts on ecosystems, commerce, agriculture, and human health. Of the 100 worst invasive species identified by the International Union for Conservation of Nature, 34 are noxious weeds, with the U.S. spending $36 billion to prevent and mitigate the spread of invasive plants. In spite of this investment, we still have a poor understanding of the processes that could be driving invasive potential, and there is a lack of genetic data for most invasive plants.

In 2021, the Consortium for Plant INvasion Genomics (CPING) held a "Botany of Invasions" symposium at the annual Botany conference (19 July 2021, virtual meeting) with the goal of identifying research and outreach synergies that would move the field forward. These include:

1) Linking studies that span different temporal scales from evolutionary/phylogenetic questions to contemporary population genetics.
2) Combining "Big Data" in its various forms, including genomics, museum collections, spatial data, etc.
3) Connecting experimental studies in model systems to non-model invasions.
4) Clarify terminology among researchers and stakeholders to improve communication—for example, what is meant by "weed" or "exotic"?
5) The need to include land managers, stakeholders, members of the public, and facilitate interdisciplinary collaborations among ecologists, agronomists, geneticists, physiologists, etc.

Focusing on invasive Russian thistle (*Salsola s.l.*, family Amaranthaceae) and closely related species, my thesis aims to address several of these points. The genus *Salsola s.l.* contains several notorious invasive species in North America, including Prickly Russian thistle (the *Salsola tragus* species complex), which has been considered as one of the fastest plant invasions in the United States. Following introduction into North America, invasive species of *Salsola* are suspected to hybridize readily with each other, complicating taxonomy and subsequent communication about invasive biology and management. It has also been suggested that hybridization between invading lineages may facilitate their establishment and spread.

In Chapter 1, published as two Extension papers via South Dakota State University, I aim to describe the history of *Salsola* invasion into the United States, provide a glimpse into the taxonomic confusion of these lineages, describe management strategies, and touch on how genetic information may clarify our understanding of their biology. Written for a broad audience, the goal of this contribution was intended to engage stakeholders regarding invasive species biology, with a focus on management. In Chapter 2, entitled "BUILDING A PHYLOGENETIC FRAMEWORK FOR THE TRIBE SALSOLEAE AND GENUS *SALSOLA s.l.*, TO UNDERSTAND THE ROLE OF HYBRIDIZATION IN PLANT INVASIONS", I conduct a phylogenomic study using targeted sequencing of the Angiosperms353 probe set with the goal of resolving relationships at several taxonomic scales: broad relationships across tribe Salsoleae, within *Salsola s.l.*, and between closely related invasive lineages in *Salsola* section Kali with a particular focus on the *Salsola tragus* species complex. I then use this phylogenetic framework to look for patterns of colonization and hybridization, along with

visualization of individual gene trees, and assessment of locus heterozygosity and allele divergence. Taken together, these chapters broadly communicate aspects of invasion biology to diverse audiences and will result in an improved understanding of this challenging group of invasive plants.

CHAPTER 1: PRICKLY RUSSIAN THISTLE - INVASION HISTORY, BIOLOGY, IDENTIFICATION, AND MANAGEMENT STRATEGIES

1.1. Overview

The goal of this chapter is to provide an introduction of the notorious invasive plant, Prickly Russian thistle. It provides an overview of the history and spread of this invasive species into North America, ecological and economic harm, taxonomy, identification, and management considerations.

1.2. Impact of Invasive Plants

Invasive plants are those that are non-native to the ecosystem and cause economic and environmental harm. They threaten biodiversity and often have traits that allow them to outcompete our native species, such as large seed producing capacity, fast growth, and allelopathic effects (Bais et al., 2003). It is estimated that invasive species cost the United States ~$120 billion annually through decreased agricultural yield, mitigation efforts, and property damage (Pimentel et al., 2005). They may also restructure ecosystems through soil disturbances and increasing the regularity of fires (Fusco et al., 2019). In some cases, these plants are intentionally introduced for forage, biocontrols, or horticultural/agricultural uses, while others are accidentally transported. With increasing global transportation of goods and services, the problem of invasive species has been growing, with one third of them first appearing between 1970 and 2014 (Seebens et al., 2017).

1.3. Prickly Russian thistle (*Salsola tragus*)

1.3.1. An Incredibly Fast Plant Invasion

The native range of *Salsola tragus*, or prickly Russian thistle, is in the arid steppes of eastern Russian and Ukraine through China, extending southward through northern Africa and the Middle East. The earliest documentation of *S. tragus* North America is from 1877 in Bon Homme county, South Dakota, with a suspected introduction with contaminated flax seeds from Russia brought by settlers to the area (Shinn, 1895). Its initial spread throughout the Great Plains states was catastrophic, fueling abandonment of infested land and anti-immigrant sentiment within the region (Young, 1988). It also directly informed public policy by prompting state laws regulating imported seeds and suggestions of statewide fencing to contain the spread (Young, 1988). The transcontinental railroad system greatly facilitated the spread of *S. tragus* beyond the Great Plains, and within twenty years of the initial introduction, *S. tragus* was widespread in California. Its expansion has been noted as one of the fastest plant invasions on record (Rilke, 1999). As *S. tragus* did not readily colonize untilled prairie, early agricultural reports made an association between infestation and environmental degradation due to poor farming/grazing practices. Molecular studies of plants from California show similarities with *S. tragus* from the Ukraine (Gaskin et al., 2006; Ryan et al., 2007), but since its initial appearance in South Dakota, it is likely that additional introductions from the native range have also occurred (Ayres et al., 2009). Although *S. tragus* is most well-known from arid areas in western North America, it has been reported in all states with the exception of Alaska and Florida (Bernau and Eldredge, 2018).

Figure 1. Prickly Russian thistle. Photo credit: Maribeth Latvis

1.3.2. Identification

Salsola tragus, or prickly Russian thistle, (Figure 1), is an herbaceous annual species that is one of the earliest emerging weeds in the Spring. The species is now recognized as part of a broader definition of amaranth family, Amaranthaceae, where it is related to other

rangeland invasives such as *Halogeton* (i.e., saltlover) and *Kochia*. It is capable of quickly establishing an extensive root system that may extend 2 m deep and 5 m laterally before shoot expansion into the characteristic rounded growth form of the mature plants (Holm et al., 1977). The rounded growth form comes from multiple, erect, branching stems that bow outward. Stems are often striped with a red or pink color. *S. tragus* uses the C_4 photosynthetic pathway, has a high water-use efficiency, and is tolerant of arid, saline habitats, which helps it outcompete native species (Fowler et al., 1992). When the plants are young, leaves are linear, fleshy, 2-5 cm in length, and with a soft point at the apex. Leaves lose their succulence later in the season and become shorter (to 6 mm) and harder, with a broad base (1-2 mm) and a sharp tip. The overall spinyness of the mature plants is a characteristic of *S. tragus*, but this feature is not as pronounced in young plants, making them more challenging to identify. Reproductive features of the plant (i.e., flowers, fruits) are very helpful in distinguishing between closely related species and genera. Flowers develop in the axils of the leaves, lack petals, with five sepals that form papery wings (4-8 mm across) appearing white, pale green, to pink. After senescence, the entire plant breaks above the ground to form a tumbleweed, which may contain thousands of seeds and are capable of traveling hundreds of miles (Shinn, 1895). This architecture is similar to that of kochia (*Kochia scoparia*) (Figure 2), but kochia tends to have leaves that are lighter green in color and lacking in spines.

Figure 2. Kochia. Photo credit: Maribeth Latvis

S. tragus is considered one of the worst agricultural weeds in North America, costing the United States $50 million annually. This economic harm is the result of lowered crop yield, higher production costs, injury to livestock, structural damage, environmental degradation (e.g., through water depletion) and related mitigation efforts. Dead plants have been noted to cause road accidents, and they are extremely flammable, which exacerbates economic losses due to the promotion of wildfire, especially in the western

US (Bernau and Eldredge, 2018). *S. tragus* is also a vector of the curly top virus that affects several crops, including beets, tomatoes, beans, and squash.

1.3.3. What's in a (Scientific) Name? *Kali* vs. *Salsola* and the *Salsola tragus* "Species Complex"

Plants known as "prickly Russian thistle" have also been called Russian cactus, burning bush, windwitch, tumbleweed, and common saltwort, and their taxonomic classification has been the subject of much rearrangement and confusion. For example, the species is recognized in some databases as *Salsola tragus* (e.g. iNaturalist, USDA Plants, Integrated Taxonomic Information System) and *Kali tragus* in others (e.g. Global Biodiversity Information Facility, Wikipedia) based on an updated (yet controversial) classification schemes (Akhani et al., 2007; Akhani et al., 2014).

The Flora of North America currently recognizes 6 species of *Salsola* within the United States, Canada, and Mexico (*S. collina, S. kali, S. paulsenii, S. soda, S. tragus,* and *S. vermiculata*) (FNA, 2020). At the species level, pricky Russian thistle (*Salsola tragus*) is highly polymorphic, with some forms appearing to be geographically distinct, while others appear to be morphological variants with no taxonomic significance (Mosyakin, 2004). A survey of 600 Canadian herbarium records revealed differences in plant size and growth form from plants collected from different environments (Beckie and Francis, 2009). In 1996, one study (Mosyakin, 1996) represented *S. tragus* as a broader name for all the existing tumbleweeds in North America, lumping together *S. tragus*, *S. australis, S. iberica*, *S. kali*, *S. pestifera*, and *S. runthenica* as synonyms of each other.

In addition to morphological variation within *S. tragus*, hybridization between species appears to be common (Beatley, 1972; Hrusa and Gaskin, 2008; Welles and Ellstrand,

2020), which makes identification challenging and further complicates taxonomic classification. Additionally, differing chromosome numbers (i.e., 2n= 18, 36, and 54) between populations of *S. tragus* suggests that whole genome doubling, or polyploidy, may play a role in the biology of these plants. This process alters the genetic makeup of plants through hiding recessive mutations and may affect physiology and morphology within as little as one generation. This may allow polyploids to colonize new ranges and increase their invasive potential (Ellstrand and Schierenbeck, 2000; te Beest et al., 2012). Polyploid species also tend to have higher survival rates and fitness in early stages of adaptation and this might play a crucial role by restoring sexual reproduction after hybridization (te Beest et al., 2012).

1.4. A Closely Related Species in South Dakota

Slender Russian thistle (*Salsola collina*; also spineless Russian thistle, slender saltwort)

Slender Russian thistle (Figure 3) tends to be slender and soft when the plant is younger and is either spineless or weakly spined compared to prickly Russian thistle. The two species more closely resemble each other when they are younger, before spiny leaves develop on prickly Russian thistle. When the plants are reproductive, slender Russian thistle will lack the membranous sepal wings that are characteristic of prickly Russian thistle. The two species both produce tumbleweeds following senescence. Slender Russian thistle is native to arid regions of Central Asia, easternmost Europe and southern Siberia. It was first reported from Dakota County, Minnesota in 1937 (Moore, 1938). Currently, it is mostly distributed across central North America and is not as widespread as prickly Russian thistle (Figure 4).

Figure 3. Slender Russian thistle (*Salsola collina*) Photo credit: bugwood.org

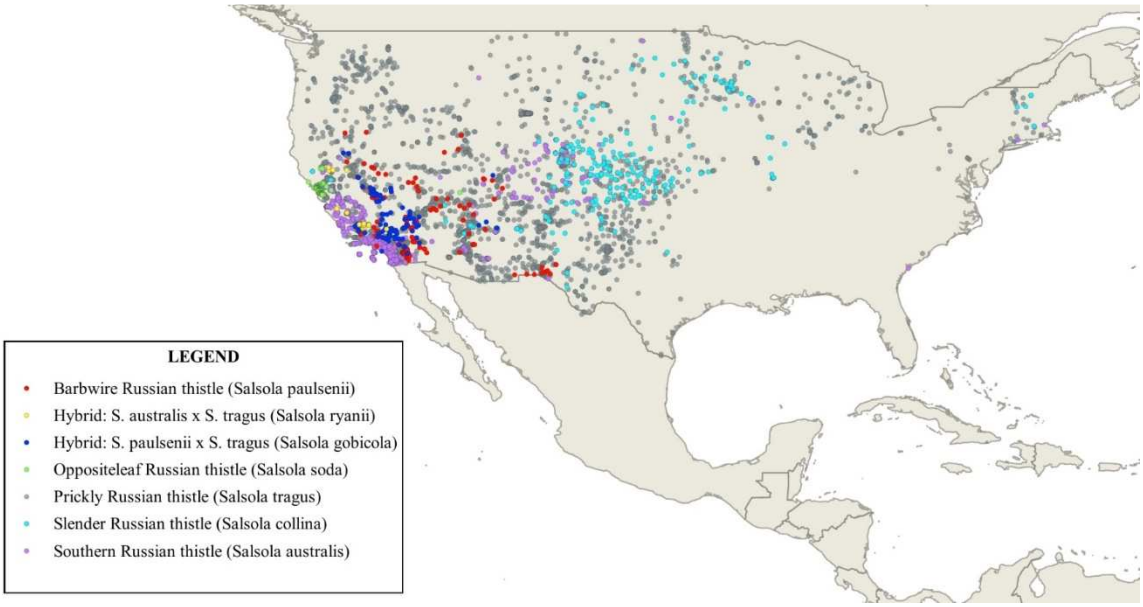

Figure 4. Occurrence map showing different species of Salsola in the U.S. from 1870 to 2021 (Credit Deva Raj Khanal). Data points were taken from Global Biodiversity Information Facility (GBIF.org). Numerous species and hybrids overlap in Southern California, which obscures some of the data points.

1.5. Other Described Species of *Salsola* in the US

Barbwire Russian thistle (*Salsola paulsenii*)

Barbwire Russian thistle was first collected in 1913 near Barstow, California; however, that collection had gone unrecognized until it was first reported in 1968 (Munz, 1968; Mosyakin, 1996). This species is commonly found in arid and semi-arid ecosystems in Texas, New Mexico, Colorado, Arizona, Utah, Nevada, California, and Oregon. Native to central Asia to southwest Russia (Rilke, 1999), it has been recently documented as an introduced species in Ukraine in 2017 (Mosyakin, 2017). This species readily hybridizes with *S. tragus*, forming the species *S. gobicola* (Rilke, 1999), which is recognized by the Jepson Manual of the Vascular Plants of California.

Hybrid: *S. australis* x *S. tragus* (*Salsola ryanii*)

S. ryanii is a recently formed polyploid hybrid between *S. tragus* and *S. australis* and has a chromosome number of 2n= 54. This new species is thought to have formed recurrently over the last 25-100 years (Ayres et al., 2009). Polyploidy produces plants of a larger size, which produce an even larger seed set, sparking concerns that *S. ryanii* may become the next "super weed" (Welles and Ellstrand, 2020). Since its formation, *S. ryanii* has rapidly expanded its distribution in California.

Hybrid: *S. paulsenii* x *S. tragus* (*Salsola gobicola*)

S. gobicola was reported to be a hybrid species formed between *S. tragus* and *S. paulsenii* (Rilke, 1999), which have fruits with wings at some or all lower nodes like *S. australis* but differing in other morphological characteristics (Hrusa and Gaskin, 2008). A native

species to Eurasia, *S. gobicola* is mostly found in sandy and disturbed places of the California Floristic Province, the Great Basin Province, and the Desert Province (Hrusa, 2012).

Oppositeleaf Russian thistle (*Salsola soda*)

S. soda is a native species to Eurasia and North Africa and was first reported in Alameda County, California in 1971 (Thomas and JH, 1975). With the existence in several locations of central California, it was expected to scatter in the coastal saline habitats, but its distribution is mostly concentrated in central California.

Southern Russian thistle (*Salsola australis*)

S. australis, native to Australia and South Africa, have been found throughout the San Joaquin valley in disturbed places and railway tracks but was previously reported as *S. kali* L. subsp. *austroafricana* and *S. kali* subsp. *pontica* (Pall) (Hrusa and Gaskin, 2008). While this species is recognized by the Jepson Manual of the Vascular Plants of California, it is considered a synonym of *S. tragus* in Flora of North America. However, *S. australis* is considered a diploid (2n= 18), while the widespread *S. tragus* is considered to be a tetraploid (2n= 36) (Ayres et al., 2009).

1.6. Challenges and Consequences

There are several challenges associated with prickly Russian thistle. First, it uses the C_4 photosynthetic pathway, has a high water-use efficiency, and is tolerant of arid, saline habitats, which helps it outcompete native species (Fowler et al., 1992). It is considered one of the worst agricultural weeds in North America, costing the United States $50

million annually (Young, 2006). The economic harm is the result of lowered crop yield, higher production costs, injury to livestock, structural damage, environmental degradation (e.g., through water depletion) and related mitigation efforts. Dead plants have been noted to cause road accidents, and they are extremely flammable, exacerbating economic losses due to the promotion of wildfire, especially in the western U.S. (Bernau and Eldredge, 2018). Prickly Russian thistle is also a vector of the curly top virus that affects several crops, including beets, tomatoes, beans, and squash.

1.7. Management

Biological

Even though biological control agents have been established, there are not any significant control approaches for Russian thistle. There are biocontrol agents currently under development: an Eriophyid mite and two fungal pathogens (Bernau and Eldredge, 2018).

Mechanical

For small infestations, hand pulling is found to be effective, but mowing is not recommended as it might disseminate the seed. Although tillage helps in controlling larger plants and seedlings, it also increases disturbance, which may promote prickly Russian thistle's germination and growth (DiTomaso et al., 2013).

Prescribed Fire

As prickly Russian thistle is extremely flammable, prescribed fire is not considered an appropriate method. This might increase the risk of wildfire, especially considering the mobility of the tumbleweeds.

Grazing

Prior to reproductive maturity, prickly Russian thistle is considered a source of forage with ample nutrition. Seed production could be altered if grazing occurs earlier than the flowering stage. However, excessive consumption of Russian thistle might negatively affect animals due to the presence of oxalates, as has been noted in sheep (Boerboom, 1993).

Chemical

Preemergence herbicide treatments are effective in late winter to early spring. However, when the plants are hard and spiny, postemergence applications may not be effective in controlling prickly Russian thistle (DiTomaso et al., 2013). The chances of Russian thistle becoming established, and invading is higher if a non-selective herbicide negatively impacts non-target species. There should be strong competition from other vegetation in area to counter the recolonization of *Salsola* species.

Populations of *Salsola* have acquired resistance to herbicides in as little as a few generations. Thus, excessive use or overuse of a single herbicide over multiple years might contribute to the development of resistant populations. Previously, 2,4-D and glyphosate were found to be very effective for the control of Russian thistle (Young et al., 2008) but repeated application resulted in the development of resistant plants. Glyphosate-resistant Russian thistle were first reported in Choteau County, Montana (Kumar et al., 2017; Heap, 2021). Similarly, resistance has been noted or suspected in the following chemical herbicides: sulfonylurea (DiTomaso et al., 2013), sulfonylurea and

imidazolinone (Morrisson and Devine, 1994), and triazines (DiTomaso et al., 2013). Because the genus *Salsola* has complex taxonomy, with several species that are difficult to differentiate, questions arise if the species reported to be resistant to certain herbicides are indeed prickly Russian thistle (*Salsola tragus*) or another closely related species.

A combination of approaches — e.g. strategic tillage methods, rotating field and site-specific herbicide applications, weed sanitation, pre- and post-herbicide applications, competition from desired species — may be effective for controlling and managing herbicide-resistant invasive plants like Russian thistle (Beckie and Harker, 2017). Promoting perennial plants with mycorrhizal associations in areas invested with Russian thistle can eliminate roots of young thistle seedlings through fungi (Barroso et al., 2019).

1.8. Conclusions

Overall, the information provided on the classification of *Salsola* species sheds light on the importance of identification, and often, genetic testing, to help distinguish between species. It also reveals that scientists are always learning and classifying species correctly helps guide research. There are several related species to *Salsola tragus* (pricky Russian thistle) that are found in South Dakota and throughout the U.S. Understanding nuances — morphology, genetics, distribution — about each related species aids in its identification, enhances knowledge of invasion, and informs management strategies.

CHAPTER 2: BUILDING A PHYLOGENETIC FRAMEWORK FOR THE TRIBE SALSOLEAE AND GENUS *SALSOLA s.l.*, TO UNDERSTAND THE ROLE OF HYBRIDIZATION IN PLANT INVASIONS

2.1. Abstract

Premise: Hybridization and polyploidy have been suggested to facilitate successful plant invasions through range expansion, increased vigor, and enhanced genetic variation, allowing invading lineages to gain a foothold in novel environments. Phylogenetic studies of invasive species and their relatives provide a needed framework to not only clarify taxonomic relationships, but to identify putative hybridization and introduction events. The genus *Salsola s.l.* contains one of the most notorious invasive species complexes in North America — prickly Russian thistle (*S. tragus s.l.*) — in which hybridization, introgression, and shifts in ploidy have been noted. However, there is a need for a much-improved phylogenetic understanding of the genus *Salsola s.l.*, including multiple accessions of invasive species, to understand how pervasive these processes are and their consequences for the process of invasion. *Salsola s.l* is taxonomically complicated genus in the tribe Salsoleae (family Amaranthaceae). While many relationships in the tribe are strongly supported in previous studies, there has been nomenclatural instability between *Salsola s.l.* or *Kali sensu* Akhani and other segregate genera in the tribe, and limited sampling of invasive species in North America. This study employs the Angiosperms353 probe set and phylogenomic analyses to 1) clarify relationships at several taxonomic scales and 2) explore evidence of hybridization among invasive species in North America.

Methods: We selected 96 accessions within the tribe Salsoleae for sequencing for target capture enrichment using the Angiosperms353 probe set. We included broad sampling across the tribe (15 representative taxa, and more dense sampling for *Salsola s.l.* (57 accessions, including several individuals for the *Salsola tragus s.l.* species complex from the native and invaded range, and population-level sampling for three populations in North America (24 accessions; eight individuals per population). HybPiper was used to map reads to target genes and produce 'supercontigs' that included targeted exons with variable flanking introns, followed by HybPhaser to reduce missing data, identify and remove paralogs, and to assess levels of locus heterozygosity and allele divergence to aid in hybrid identification. RAxML was used to estimate individual gene trees and for a concatenated gene analysis in a Maximum Likelihood framework. We also used multispecies coalescent approaches SVDQuartets and ASTRAL-Pro to infer species trees. Following manual inspection of discordance in each gene tree, we found a clear pattern among suspected hybrids *S. ryanii* and *S. gobicola.* 68% and 57% of the gene trees designating closely associated with parental species of *S. ryanii* and *S. gobicola* respectively, and *S. tragus s.l.* species complex showing discordance in their patterns. Results: All phylogenetic analyses based on the Angiosperms353 probe set (RAxML, SVDQuartets, and ASTRAL-Pro) show improved resolution (BS > 90% for RAxML and SVDQuartets, and LPP > 0.90 for ASTRAL-Pro) in the tribe Salsoleae and genus *Salsola s.l.* compared to previous studies. Within tribe Salsoleae, our results strongly support a monophyletic *Salsola* section Kali, which contains the invasive Russian thistle species in North America, within a polyphyletic *Salsola s.l*, however, some of the lineages represents poor scores in internal clade support for closely related species. Phylogenetic

relationships are generally congruent among the RAxML concatenated analyses and species tree approaches (SVDQuartets and ASTRAL-Pro). Among the invasive lineages in North America, including *S. tragus s.l.* and closely related species, we detect high levels of gene tree discordance and high levels of locus heterozygosity and allele divergence, particularly for *S. ryanii* and *S. gobicola*, two taxa previously suggested to be of hybrid origin. Accessions of *S. tragus s.l.* from North America (invaded range) do not form a clade within Eurasian accessions (native range) and instead come out in several places with Eurasian accessions with moderate support, suggesting several introductions of *S. tragus s.l.* into North America.

Conclusions: Our results demonstrate the utility of the Angiosperms353 probe set for resolving phylogenetic relationships at different taxonomic scales, including deeper relationships within tribe Salsoleae and also among closely related lineages of the *Salsola tragus* species complex. At the tribe level, our findings lend support for nomenclatural changes regarding *Salsola s.l.* Relationships between *Salsola tragus s.l.* from both the native and invaded range suggest several introductions into North America from Eurasia and that colonization has been an ongoing process rather than a one-time event. Multiple introductions of *Salsola* section Kali were likely followed by multiple hybridization events with closely related species, as evidenced by gene tree discordance and high levels of locus heterozygosity and allele divergence among North American lineages.

Keywords: invasive species, hybridization, phylogenomics, Angiosperms353, *Salsola*, Salsoleae, Amaranthaceae, Russian thistle

2.2. Introduction

Hybridization may serve as an evolutionary stimulus, leading to the formation of new species, the collapse of reproductive isolation between lineages, and allowing for new trait combinations (Anderson and Stebbins Jr, 1954; Arnold, 1997; Rieseberg et al., 2003; Soltis and Soltis, 2009; Eberlein et al., 2019). Hybridization — either between closely related species or other invading congeners from different parts of the native range (Vallejo-Marín et al., 2021) — may contribute to the establishment and spread of invasive lineages, with enhanced genetic variation facilitating adaptation to novel environments (Ellstrand and Schierenbeck, 2000; Rieseberg et al., 2007; Schierenbeck and Ellstrand, 2009). In addition to rapid adaptation, this process may also provide a mechanism for overcoming Allee effects, or reduced fitness at low population densities following invasion (Mesgaran et al., 2016). Within a few generations, hybrids have demonstrated increased fecundity (Hovick and Whitney, 2014), and robust alleles introduced from the hybridization process can contribute to range expansion (Aïnouche et al., 2009; Hovick and Whitney, 2014), suggesting natural selection can play a vital role in shaping hybrid performance and invasiveness over time. Invasive species cost the United States US$ 26 billion annually through ecosystem degradation, reduced agricultural yields, and management efforts (Crystal-Ornelas et al., 2021), and hybridization may directly influence mitigation efforts, such as the evolution of herbicide resistance (Williams et al., 2014).

Robust phylogenetic hypotheses of invasive species and their relatives provide an important framework to understand the number of colonization events (Soltis and Soltis, 2003; Hopley et al., 2021) and traits that might enhance invasion potential (Wood et al.,

2009). Additionally, a phylogenetic framework is an important step in understanding gene flow between closely related species, which may enhance invasion success or form new invasive plant species (Ellstrand and Schierenbeck, 2000; Welles and Ellstrand, 2016). Inclusion of hybrid taxa in phylogenetic analyses may result in conflicting signals, or discordance, between datasets (Funk, 1985). When considering genetic datasets, hybridization may manifest as conflicting relationships between individual gene trees, with the hybrid taxon grouping with one parental lineage in some of the datasets, and grouping with the other parental lineage in other datasets, depending on the genomic contribution of the parental species (Maddison, 1997; Crowl et al., 2017). Observations of the conflicting behavior of hybrids in individual gene trees helps to reveal their placement and putative parental lineages (Holder et al., 2001; Crowl et al., 2017). With this in mind, phylogenetic network analysis may account for discordant signals resulting from hybridization and estimates reticulation events under a coalescent model (e.g. (Crowl et al., 2017; Morales□Briones et al., 2018)). For an example, a recent phylogenomic study focusing on the evolution of photosynthetic pathways found considerable gene tree incongruence and reticulations in the tribe Salsoleae, providing evidence of a hybrid origin of *Salsola divaricata agg*., which utilizes the C_2 photosynthetic pathway, from C_3 and C_4 parental lineages (Tefarikis et al., 2021).

Salsola L. *sensu lato* (family Amaranthaceae, tribe: Salsoleae), is a cosmopolitan genus that includes 130-150 species with a center of diversity in central Asia (Mosyakin, 1996). *Salsola s.l.* encompasses both annual and perennial species and includes examples of C_3, C_4 and intermediate (C_2) photosynthetic pathways. These species can tolerate water, heat and salt stresses (Toderich et al., 2012), and are mostly found in arid and

semi-arid regions (Beckie and Francis, 2009). *Salsola s.l.* also includes a notorious invasive species complex, *Salsola tragus* L. (prickly Russian thistle or tumbleweed), which has been noted to be the fastest plant invasion on record in the history of the United States following introduction via contaminated flax seeds into South Dakota in the 1870s (Rilke, 1999). Introgression and hybridization appear to be common between species invasive species of *Salsola s. l.*, and different ploidal levels (Hrusa and Gaskin, 2008), and rapidly spreading recent hybrid species have been observed in California (Welles and Ellstrand, 2020). For example, *S. gobicola* is a recognized putative hybrid species formed from *S. tragus* and *S. paulsenii* (Rilke, 1999), and *S. ryanii* (Hrusa and Gaskin, 2008) is a recently described allohexaploid species formed from recurrent hybridization between *S. tragus s.s.*, a tetraploid, and *S. australis*, a diploid. Within a decade of its formation, *S. ryanii* seems to be spreading rapidly compared to its progenitors, and it can form much larger tumbleweeds with a greater seed set (Welles and Ellstrand, 2020).

More broadly, *Salsola s.l.* is part of the large tribe Salsoleae, which consists of one third of the genera of Amaranthaceae *s.l.* (including Chenopodiaceae (Hernández-Ledesma et al., 2015; Angiosperm Phylogeny Group, 2016), and is mostly distributed in Central Asia and the Middle East with radiations into the Mediterranean, Africa, Australia, and with some introductions into the Americas (Akhani et al., 2007). Within Salsoleae, several genera have been the subject of nomenclatural and taxonomic revisions, some of which have been controversial. A treatment from Akhani et al. (2007) based on nuclear ITS and chloroplast *psbB-psbH* regions found *Salsola s.l.* to be polyphyletic and reduced *Salsola s.s.* to 25-27 species while transferring others to

separate genera *Caroxylon*, *Turania*, *Xylosalsola*, and *Kaviria* (Akhani et al., 2007; Akhani et al., 2014). Wen et al. (2010) echoed the finding of a polyphyletic *Salsola s.l.* in their investigation of Salsoleae based on three gene regions (nuclear ITS and chloroplast *psb*B–*psb*H and *rbcL)* and expanded sampling of Asian *Salsola s.l.* Based on these findings, several invasive lineages in North America were transferred from *Salsola* to *Kali* (e.g. *K. ryanii*, *K. australe*, *K. paulsenii*, and *K. tragus* (Akhani et al., 2007; Akhani et al., 2014). However, Mosyakin et al. (2014) and Mosyakin et al. (2017) make arguments for the conservation of *Salsola s.l.*, based on *Salsola* with *S. kali* as the conserved type that best agrees with formal typification criteria (Rec. 9A) (Mosyakin et al., 2014). The Nomenclatural Committee for Vascular Plants narrowly supported conserving *Salsola s.l.* with *S. kali* as the conserved type (Wilson, 2017), and we maintain this classification scheme throughout the text. The resulting taxonomic instability has created much confusion among applied conservation scientists and land managers, as some biodiversity databases and references (e.g., Canadensys, GBIF, Wikipedia) use *Kali* for several invasive North American species, while others (e.g., USDA PLANTS, iNaturalist) use *Salsola s.l.*

Species concepts for several invasive lineages in North America have also been the subject of much disagreement and confusion. The Flora of North America recognizes six different species of *Salsola s.l.*: *S. collina, S. kali* (subsp. Salsola and pontica), *S. paulsenii*, *S. soda*, *S. tragus* and *S. vermiculata* (Mosyakin, 1996). Under this treatment, *S. tragus* L. is a broader species concept for all the existing tumbleweeds in the North America. However, a broad *S. tragus* includes several ploidal levels and potentially cryptic species. The widespread form of *S. tragus* is presumed to be a tetraploid (2n=36)

(Ryan and Ayres, 2000). The discovery of a more restricted diploid form (2n=18) provided evidence of a cryptic species (Ryan and Ayres, 2000), which was later supported by molecular studies (Gaskin et al., 2006). The diploid form of *S. tragus* L. was recognized as *S. australis* due to similarities with this species thought to be native in Australia (Hrusa and Gaskin, 2008). Thus, some references (Jepson eFlora, 2022) recognize additional species of *Salsola s.l.* (e.g., *S. australis*, *S. gobicola*, *S. ryanii*) compared to the Flora of North America.

For the past two decades, much of our understanding of phylogenetic relationships within the tribe Salsoleae*,* genus *Salsola s.l.*, and species complexes within *Salsola s.l.* were based on smaller molecular datasets generated using Sanger sequencing approaches, limited taxonomic sampling, and not taking different gene histories (i.e. due to hybridization) into account (Pyankov et al., 2001; Akhani et al., 2007; Wen et al., 2010), however two recent studies with larger genomic datasets have provided an updated phylogenic framework for the family Amaranthaceae and tribe Salsoleae, and revealing patterns of hybridization with implications for the development of the C_2 photosynthetic pathway (Morales-Briones et al., 2021; Tefarikis et al., 2021). Poor bootstrap support along the backbone in Sanger sequencing-based studies might be due to limited phylogenetic signal with smaller datasets, data processing (e.g., accidental inclusion of paralogous gene copies), or lack of hybrid assessment (Morales-Briones et al., 2021; Nauheimer et al., 2021). While recent phylogenomic studies from Morales-Briones et al. (2021) and Tefarikis et al. (2021) greatly clarified many relationships in Salsoleae and emphasized the role of hybridization in the evolution of this clade, there was little sampling of *Salsola* section Kali (or *Kali sensu* Akhani et al., 2007) and other

segregate genera of *Salsola s.l.*, which limits inference into relationships among the highly invasive species in North America. Given the enormous economic toll the *Salsola tragus s.l.* species complex in North America, it would be beneficial to include these lineages in a robust phylogenomic framework to investigate the role of hybridization in invasion success and greatly clarify nomenclature.

For inferring relationships at multiple taxonomic scales target enrichment has proven to be a cost-effective and versatile high-throughput sequencing (HTS) approach for both fresh and herbarium specimens (McKain et al., 2018). The angiosperm-wide Angiosperms353 probe kit (Johnson et al., 2019) has emerged as a workhorse for phylogenetic studies in flowering plants (Baker et al., 2021), capable of retrieving gene targets from degraded herbarium specimens (Brewer et al., 2019) and with the potential to recover highly supported relationships between closely related species (Larridon et al., 2020; Murphy et al., 2020; Larridon et al., 2021; Starr et al., 2021). Non-coding sequences that flank the targeted exons (i.e. within the "splash zone") are particularly useful for resolving phylogenetic relationships at lower taxonomic levels (Weitemier et al., 2014). Additionally, the targeted regions from the Angiosperms353 probe set may produce enough SNP data to infer within-species relationships and estimate population genetic parameters (Beck et al., 2021; Slimp et al., 2021).

While the Angiosperms353 probe kit was developed to target single copy, orthologous genes, paralogous gene copies may exist and can confound phylogenetic inference if accidentally included in downstream analyses. However, they may also provide a valuable source of information if they are identified and if duplicated copies are analyzed separately (Frost and Lagomarsino, 2021). HybPiper (Johnson et al., 2016), a

bioinformatics pipeline commonly used to map reads from Angiosperms353 to their gene targets, has a tool to detect paralogs, but it may miss them and accidentally combine them into chimeric contigs (Kates et al., 2018; Nauheimer et al., 2021). This is because HybPiper employs a *de novo* assembly method that makes it difficult to combine heterozygous sites accurately. The issue of unknown paralogs may be ubiquitous among all Hyb-Seq datasets, necessitating the need for approaches to identify them (Zhou et al. 2022). Therefore, for a more rigorous approach to paralog detection, we use HybPhaser, a newly developed bioinformatics pipeline that removes sequences over a certain heterozygosity threshold (Nauheimer et al., 2021). We also use HybPhaser's estimation of locus heterozygosity (LH) and allele divergence (AD) to aid in identifying putative hybrid taxa in our datasets.

Our study aims to generate an updated phylogeny of tribe Salsoleae and genus *Salosla s.l.* using the Angiosperms353 probe set, with a particular focus on representing the invasive lineages within North America. Although we use HybPiper for initial data processing, we employ HybPhaser to remove putative paralogs and detect hybrids to generate an improved phylogenetic framework. In addition to inspection of individual gene trees, we use estimates of LH and AD to corroborate identification of hybrid taxa in our datasets. Our sampling strategy also spans several taxonomic levels (i.e., tribe, genus, closely related species, within population), allowing us to test the utility of the universal Angiosperms353 probe set for phylogenetic inferences at differing levels of divergence.

2.3. Materials and Methods

2.3.1. Sampling

To investigate phylogenetic relationships and investigate the utility of the Angiosperms353 probe set at multiple taxonomic scales, we used a tiered sampling strategy for tribe Salsoleae and shallower relationships within *Salsola s.l.*, particularly focusing on *Salsola* section Kali, which includes many of the notorious invasive tumbleweeds in North America (Table 1). Our sampling represents taxonomic breadth of tribe Salsoleae (15 species; 13 individuals sampled per species and 2 multiple individuals) and genus *Salsola s.l.* (57 accessions; multiple individuals sampled per species), with a particular focus on sampling members of the *Salsola tragus* species complex. Multiple accessions were included of *Salsola tragus s.l.* from across its geographic breadth, including 4 samples from its native range in Eurasia. *Caroxylon sensu* Akhani (2007) served as our outgroup. We also included population-level sampling for 2 populations of *S. tragus* and 1 population of *S. aff. australis sensu* (Hrusa and Gaskin, 2008) (24 accessions; 8 individuals per population). We retain the name *Salsola* for *Salsola* section Kali following Wilson (2017) but adopt updated taxonomy from Akhani et al. (2007) for some segregate genera (Table 1). The herbarium specimens were contributed by CDFA, DAV, NYBG, HUH, KANU, BRIT, and leaf material came from collaborators and field collections (Appendix S1 for a complete list of accessions in the supplemental data).

Table 1. 96 samples representing herbarium collections and fresh tissues of tribe Salsoleae, *Salsola s.l.*, population samples and outgroup species

Tribe samples (15)

Species	Numbers	Proposed taxonomy (Akhani et al., 2007)
Anabasis jaxartica	1	
Anabasis rausskeehtii	1	
Cornulaca aucheri	1	
Cornulaca monacantha	1	
Girgensohnia minima	1	
Girgensohnia oppositifolia	1	
Halogeton glomeratus	1	
Haloxylon articulatum	1	
Haloxylon persicum	1	
Horaninowia platyptera	1	
Horaninowia pungens	1	
Raphidophyton regelii	1	
Sympegma regelii	2	
Turania miloatee	1	

Salsola s.l. samples (48)

Species	Numbers	Proposed taxonomy (Akhani et al., 2007)
Salsola arbuscula	2	*Xylosalsola arbuscula*
Salsola arbusculiformis	1	
Salsola australis	3	*Kali australis*
Salsola chivensis	1	*Xylosalsola chiwensis*
Salsola collina	3	*Kali collina*
Salsola damascena	1	*Kali damascena*
Salsola florida	2	
Salsola foliosa	1	
Salsola gobicola	1	*Kali gobicola*
Salsola jacquemontii	1	*Kali jacquemontii*
Salsola laricifolia	1	
Salsola longifolia	2	
Salsola monoptera	2	*Kali monopterum*
Salsola montana	2	
Salsola oppositifolia	1	
Salsola oreophila	1	
Salsola paulsenii Litv.	2	*Kali paulsenii*
Salsola richterii	2	*Xylosalsola richteri*
Salsola rosmarinus	2	
Salsola ryanii	2	*Kali ryanii*
Salsola setifera	1	
Salsola soda	2	

Salsola tianschanica	1	
Salsola webii	1	
Salsola zygophylla	2	
Outgroup samples (9)		
Salsola albisepala	1	*Caroxylon albisepalum*
Salsola aphylla	2	*Caroxylon aphyllum*
Salsola ericoides	1	*Caroxylon ericoides*
Salsola gemmascens	2	*Caroxylon gemmascens*
Salsola imbricata	1	*Caroxylon imbricatum*
Salsola vermiculata	2	*Caroxylon vermiculatum*
Population-level samples (24)		
Salsola tragus	8	*Kali tragus*
Salsola tragus	8	*Kali tragus*
Salsola c.f. tragus	8	*Kali. aff. australis*

2.3.2. DNA Extraction, Library Preparation, Target Enrichment and Sequencing

We extracted genomic DNA from 72 herbarium specimens and 24 field-collected, silica-dried samples using a modified CTAB protocol (Doyle and Doyle, 1987) optimized for herbarium material. The DNA extractions were visualized for quality (e.g., degree of fragmentation and concentration) and concentration was assessed with Qubit 3 and Qubit 4 fluorometer using broad range reagents. Sufficient concentrations of DNA (more than 4.0 ng/μL) were obtained to proceed with genomic library construction for most of the material, including several old herbarium samples. 7 samples (including 1 fresh tissue) yielded low concentration (> 4.0 ng/μL), however these samples were included in our sampling as some of them were important for broad, representative taxonomic sampling. All samples were visualized on the Agilent Tapestation platform at 1 ng/μL using a High Sensitivity D1000 screen tape and were successful for the amplification of the targeted loci, including 72 samples that were from relatively degraded herbarium specimens.

Following quantification of DNA extractions, library preparation and sequencing using targeted regions were performed at Daicel Arbor Biosciences, Ann Arbor, Michigan, USA. Libraries preparations were carried out using an Illumina TruSeq-stile library preparation method optimized for targeted capture. Unique dual-index combinations were added to each sample via 6-10 cycles of PCR amplification. The indexed libraries were quantified with both a spectrofluorimetric assay and a quantitative PCR assay. To prepare for target capture, libraries were pooled in equimolar ratios in 12-plex captures (fresh samples, 100 ng for each sample) or 6-plex captures (degraded herbarium samples, 200 ng for each sample) and each capture pool was dried down to 7 μL by vacuum centrifugation. Samples were grouped by taxonomy, where possible,

keeping samples with similar yields in the same reaction. Captures were performed using the myBaits Expert Angiosperms 353 v1 kit following the myBaits v5 protocol with an overnight hybridization and washes at 62C. Post-capture, the reactions were amplified for 10 cycles and were quantified again with both a spectrofluorimetric assay and a quantitative PCR assay, both as above. The captures were pooled in approximately equilmolar ratios based on the number of libraries in each capture. Targets were paired-end sequenced on the Illumina NovaSeq 6000 platform on a partial S4 PE150 lane to approximately 1Gbp per library. Demultiplexed FASTQ data were generated from Illumina sequencing was uploaded to BaseSpace.

2.3.3. Sequence Data Processing

HybPiper target mapping and contig assembly – Following sequencing, our targeted regions were assembled using HybPiper 1.3.1 (Johnson et al., 2016), a flexible bioinformatics pipeline that can be adjusted for broad or shallow-scale phylogenetic applications and has been used for processing Angiosperms353 data (Johnson et al., 2016; Johnson et al., 2019; Larridon et al., 2020). Reads were mapped to their target genes and assembled into contigs using the the "read_first.py" script using the expanded 'mega353' target file (https://github.com/chrisjackson-pellicle/NewTargets). Scripts "get_seq_lengths.py" and "gene_recovery_heatmap.R" were used to summarize and visualize sequencing success and gene recovery. Similarly, "hybpiper_stats.py" was used to summarize the target enrichment and gene recovery efficiency for a set of samples. Introns and intergenic sequences flanking target exons were extracted using "intronerate.py". Exons and supercontigs were retrieved using "retrieve_sequence.py". Supercontigs include targeted exons in addition to non-coding sequences adjacent to the

exons in a region known as the “splash zone”, which are useful for inferring phylogenetic relationships among closely related species (Weitemier et al., 2014; Frost and Lagomarsino, 2021).

HybPhaser data cleaning – We used HybPhaser 2.0 (Nauheimer et al., 2021), a recently developed bioinformatics pipeline, which helps in the removal of paralogs genes and helps in detection of putative hybrid taxa using the HybPiper assemblies following three major steps; estimating heterozygous sites for putative hybrid detection, clade association and phasing of read files according to the clade association. To identify SNPs, sequencing reads were remapped to the assembled contigs from HybPiper using the bash script “1_generate_consensus_sequence.sh”. This step generated consensus sequences that contain IUPAC ambiguity codes for heterozygous sites. To minimize missing data and improve the quality of the datasets, we removed all loci that had sequence recovery for < 20% of samples and < 50% of target sequence length recovered, and we removed all samples that had < 20% of loci and < 45% of the target sequence length recovered. These thresholds were adjusted in the configuration script “config.txt” and RScript “1b_assess_dataset.R” was executed to implement the thresholds.

Removal of putative paralogous genes – Paralogous loci are predicted to have a higher proportion of SNPs when compared to other loci, so we assessed the SNP distribution for each locus and generated boxplots and graphs across multiple samples, to visualize the output from the HybPhaser consensus step. For the detection of putative paralogs for all samples, we used a statistical outlier method to flag a locus with a frequency of SNPs

greater than 1.5x the interquartile range as paralogs (Nauheimer et al., 2021). We then manually adjusted the configuration script "config.txt" to remove samples and loci that were statistical outliers.

Assessment of locus heterozygosity and allele divergence – As the product of crossing between two different species, hybrid samples inherit divergent alleles from both parents, which leads to a higher proportion of loci consisting of SNPs (locus heterozygosity or LH) and a higher proportion of SNPs across all loci (allele divergence or AD) (Nauheimer et al., 2021). The RScript "1c_generate_sequence_lists.R" was executed to generate the summary of locus heterozygosity and allele divergence of each sample to recognize the putative hybrid samples and to generate summary lists for the HybPhaser output (i.e. loci for each sample and samples for each locus).

2.3.4. Phylogenetic Analyses

Individual Gene Tree and Concatenated Inference – Following the removal of missing data and putative paralogs, and using the consensus sequences from the HybPhaser output (with ambiguity code for heterozygous sites), we inferred the evolutionary history of 281 individual loci in a Maximum Likelihood framework using RAxML 8.2.12 (Stamatakis, 2014). Prior to analysis, sequences for each locus were aligned with MAFFT 7.475 (Katoh and Standley, 2013), and sites containing 50% gaps missing data were removed using Phyutility 2.7.1 (Smith and Dunn, 2008). Each alignment was visualized in Geneious Prime 2022.1.1 (Kearse et al., 2012) and manually adjusted. Each locus was analyzed under the GTR model with optimization of substitution rates and sites specific evolutionary rates (option -f a -m GTRCAT) and 200 rapid bootstrap replicates to

estimate clade support. In addition to the individual gene trees, we also used RAxML 8.2.12 to conduct maximum likelihood analysis on the 281 concatenated loci. Concatenation of trimmed alignment files was performed using Phyutility 2.7.1 (Smith and Dunn, 2008). We then performed the RAxML analysis similarly using the GTR model with optimization of substitution rates and sites specific evolutionary rates (option -f a -m GTRCAT) and 200 bootstrap replicates to estimate clade support. *Caroxylon vermiculatum* served as the outgroup taxon.

Gene tree binning – We manually examined the topologies of 279 RAxML individual gene trees for known hybrids and suspected hybrid species. The known hybrid species include *S. ryanii* and *S. gobicola*, which have been suggested based upon previous taxonomic and/or molecular studies. Each gene tree was examined for placement of these taxa with either of their suspected parental species (*S. tragus* or *S. australis* in the case of *S. ryanii*, and *S. tragus* and *S. paulsenii* in the case of *S. gobicola*) or some other relationship. These relationships were scored and quantified as a percentage of the total gene tree topologies. Similarly, closely related species in *Salsola* species complex and other species recorded higher number of locus heterozygosity (LH) and allele divergence (AD) were also manually investigated.

Species tree inference – The process of incomplete lineage sorting may result in gene trees that differ from the true pattern of inheritance among species (i.e. the species tree), and the gene concatenation method (described above) may lead to the inference of incorrect phylogenies with high statistical support. Coalescent-based species tree approaches accommodate differences in the evolutionary histories of individual genes

and are more robust to the presence of incomplete lineage sorting. For these reasons, we used two coalescent-based approaches in addition to our concatenated RAxML analysis: SVDQuartets (Chifman and Kubatko, 2014) and ASTRAL-Pro (Zhang et al., 2020). As SVDQuartets treats each SNP as an independent sample with a coalescent history within the species tree, we used the same concatenated multi-locus dataset as our RAxML concatenated analysis. Within the input nexus file, we defined 78 species blocks, with each one including multiple accessions per species whenever possible. 281 genes were included for a total length of 205,134 bp, with each individual gene defined by gene partitions blocks within the input nexus file. The tree was reconstructed using exhaustively evaluating quartets (evalquartets=all) and support for relationships estimated with 100 bootstrap replicates. *Caroxylon* served as the outgroup. SVDQuartets was executed in command line as implemented in the software PAUP* 4.0a (Wilgenbusch and Swofford, 2003).

While SVDQuartets infers species-tree from SNPs within the sequence data and does not estimate branch lengths within the phylogeny, ASTRAL-Pro (Zhang et al., 2020) relies on the topologies of individual gene trees as input and can also estimate the branch lengths in coalescent units, yielding an estimation of uncertainty in local posterior probability (LPP). ASTRAL-Pro was run with 279 individual gene trees estimated from RAxML analyses for the species tree reconstruction. LPP for quartet support, as well as normalized quartet score for the species tree were calculated.

All the trees (RAxML individual gene trees, RAxML concatenated gene tree, SVDQuartets and ASTRAL-Pro) were visualized and exported using FigTree (Rambaut,

2020). Our taxonomic comparisons, results and discussion were based on both gene trees and species trees.

2.4. Results

Sampling and Target Sequence Capture Success – Out of a total of 96 genomic DNAs sent for sequencing, 72 were from herbarium specimens and were degraded, and 24 were from recent silica-dried field collections. When visualized on the Agilent Tapestation platform at 1 ng/uL using a High Sensitivity D1000 screen tape, they were successful for the amplification of the targeted loci. From the 72 herbarium specimens sequenced, between 30,410 and 22,397,799 (mean: 9,117,947) reads were produced per sample, and between 8,022 and 15,033,359 (mean: 4,232,811) were successfully mapped to their target. 351 genes (out of 353 targeted) were successfully retrieved from our herbarium samples. Similarly, from the 24 silica-dried specimens sequenced, in between 261,186 and 14,429,359 (mean: 8,046,421) reads were produced, with in between 108,225 and 5,920,335 (mean: 3,435,694) mapping to their target. We found that on average, 352 genes were successfully retrieved from silica-dried specimens.

HybPiper sequence assembly – Overall, HybPiper sequence was successfully performed across 353 loci within the 96 samples (Appendix S2). Paralog warnings were issued for 74 samples (average 10.83 paralogous genes per sample). Some loci were flagged repeatedly across all samples, where highest number of paralogs written was found for locus 6128.

Reducing proportions of missing data – Two samples were below the threshold of < 20% of loci recovered and 18 samples were < 45% of the target sequence length recovered. Similarly, one locus had sequence recovered for < 20% and 58 loci were < 50% of the target sequence length recovered. In total, 18 samples and 59 loci were excluded based on all criteria for missing data. Thus, the cleaned datasets were composed of 78 samples (out of 96) and 294 loci (out of 353), following the removal of missing data (Figure 5 and Appendix S3).

Identification and removal of paralogous genes with HybPhaser – Based on the statistical outlier analysis completed on the HybPhaser consensus sequences, 13 loci out of 294 loci were flagged as putative paralogs with a mean proportion of SNPs of 0.02859 (Appendix S3). On average, 12.5 loci per sample were flagged as putative paralogs, slightly higher than detected using HybPiper. This approach differs from HybPiper because it uses the proportion of heterozygous sites to detect paralogs. Compared to orthologous genes, the paralogous genes are expected to have greater divergence resulting in significantly higher rates of heterozygous sites.

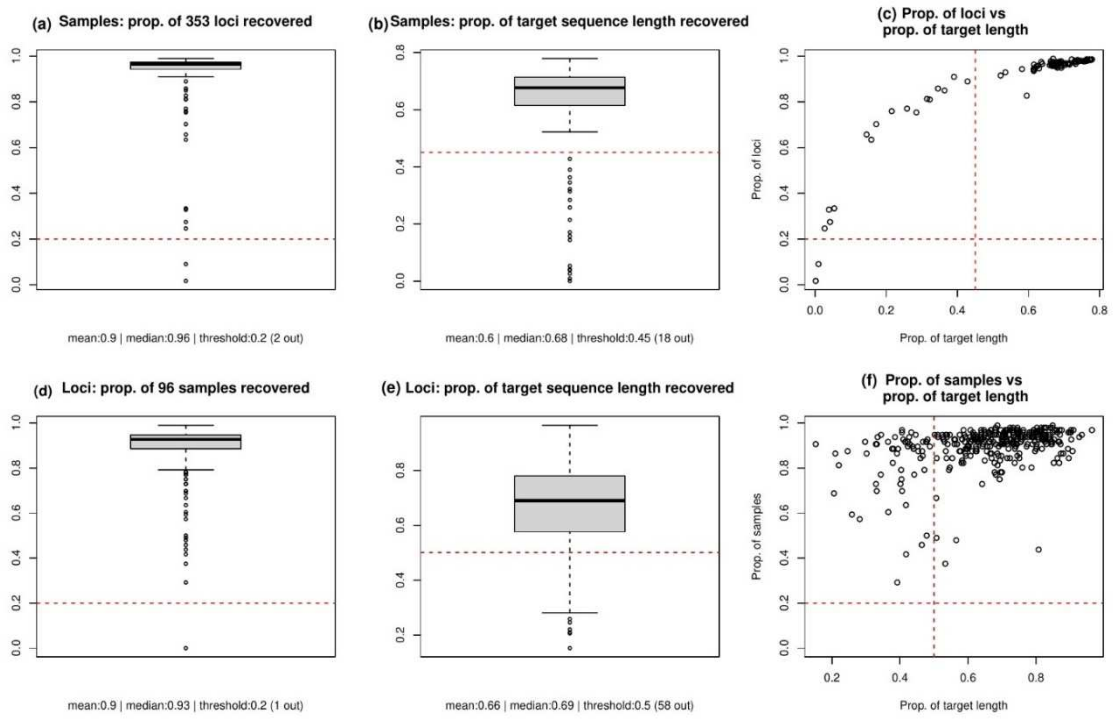

Figure 5. Graph showing threshold in red line and recovered data for removing samples and loci to reduce the missing data. (a) samples below the threshold of 20% of loci recovered (b) samples below the threshold of 45% of the target sequence length recovered (c) proportions of loci vs. proportion of target length (d) loci below the threshold of 20% of the samples recovered (e) loci below the threshold of 50% of the target sequence length recovered, and (f) proportion of samples vs. proportion of target length

Assessment of locus heterozygosity and allele divergence – Among the 78 analyzed samples, AD varied between 0.29% and 4.10%, and LH varied in between 47.67% and 99.64%. Higher locus heterozygosity was generally associated with higher allele divergence percentage (Figure 6). Some of the samples visualized to be higher values of LH and AD (Appendix S4). The known North American hybrid species *Salsola gobicola* and *S. ryanii* have higher percentages of AD and LH (Figure 6). *Salsola_soda_L20048* had low values of AD (0.44%) and LH (47.67%) compared to other samples. Among the multiple individuals of *S. tragus s.l.,* AD varied in between 2.14% and 0.38% and LH varied in between 98.91% and 54.58%. *Salsola_jacquemontii_L20034*, *Girgensohnia_oppositifolia_L20421*, *Caroxylon_vermiculatum_L20314*, *Salsola_tragus_L21432* and *Salsola_ryanii_L20047* are additional accessions found with higher LH and AD, similar to the levels found in *Salsola gobicola* and *S. ryanii*.

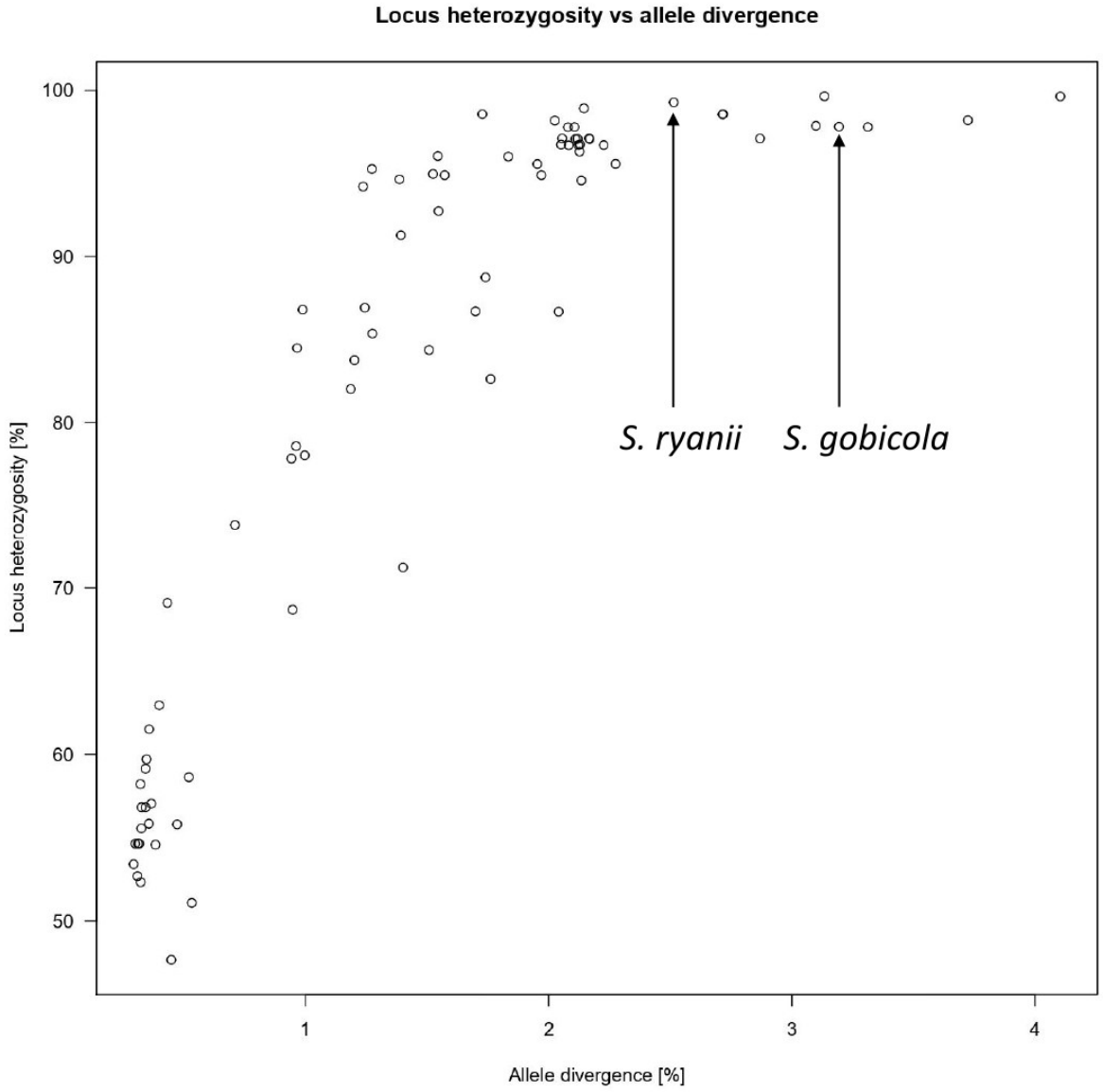

Figure 6. Scatter plot displaying the overview of locus heterozygosity (LH) on the y-axis and allele diversity (AD) on the x-axis of all the samples. High percentages of LH and AD indicates the suspected hybrid samples including known hybrid species S. gobicola and S. ryanii. Other species also represents in higher AD (>2.5%) and LH (>95%).

Phylogenetic analyses

Concatenated RAxML topology –The total length of the concatenated alignment was 205,134 characters. Support for the monophyly of tribe Salsoleae and backbone relationships within the tribe have 100% bootstrap support, a marked improvement from previous studies, however some of the internal nodes among closely related species of *Salsola* section Kali have lower bootstrap values in their branches (Figure 7). *Salsola s.l.* is found to be polyphyletic with strong support. *Salsola* section Kali, including the invasive tumbleweeds in North America, is sister to a clade comprised of *Xylosalsola, Turania,* and *S. arbusculiformis* (BS = 100%), while other species are more closely related to various other genera. For example, *S. longifolia, S. zygopylla, S. rosmarinus, S. soda* and *S. florida* (comprising *Salsola s.s. sensu* Akhani et al. 2007) formed a clade with *Cornulaca, Girgensohnia, Halogeton, Haloxylon* and *Horaninowia* (BS = 100%), and *Anabasis* was resolved as the sister lineage to this clade. *Salsola arbusculiformis*, which was recognized as *Collinosalsola* in Akhani et al. 2007, is nested within the *Xylosalsola* clade (BS = 100%). *Salsola monoptera, Salsola damascena*, and *Salsola webii* fall out in strongly supported relationships with other genera distant from *Salsola s.l.* (BS = 100%). Monotypic *Sympegma* is inferred as the earliest diverging lineage of Salsoleae *s.s.* with strong support (BS = 100%).

Within the *Salsola* section Kali, *S. australis*, a diploid taxon that had been lumped into *S. tragus* by FNA, was found to be highly distinctive and more closely associated to *S. monoptera* than *S. tragus* (BS = 100%). The clade formed by *S. australis* with *S. monoptera* is weakly associated with a highly supported clade formed by *S. gobicola*, *S. paulsenii* and *S. jacquemontii.* The placement of *S. jacquemontii* is of interest, because it

had one of the highest AD (3.13%) and LH (99.64%) of our samples, and it is found in Asia but not North America, yet is very closely associated with *S. paulsenii*, one of the putative parental species of *S. gobicola*. Several species for which multiple accessions were analyzed (e.g., *S. tragus, S. collina* and *S. paulsenii*) are polyphyletic with moderately high support (BS > 80%). The known hybrid species from North America, *S. gobicola* and *S. ryanii* have poor bootstrap support (BS = 35% and BS = 28% respectively) in their sub-clade node.

When examining the population samples with multiple individuals, eight *Salsola c.f. australis* collected from the field in California population cluster very closely with herbarium accessions of *Salsola australis*. However, for *Salsola tragus*, the population samples and herbarium accessions are not as tightly clustered, suggesting greater levels of divergence. The individuals from an Oregon population of *S. tragus* did not form a clade with each other, but accessions of *S. tragus* from a South Dakota population clustered together. Our sampling also included four herbarium accessions of *S. tragus* from Eurasia (native range). The North American accessions of *S. tragus* did not form a clade within the Eurasian samples, but instead fell in different places, with the Eurasian accessions more closely related to different groups of *S. tragus s. l.*

Gene tree inference and binning – Individual gene tree topologies will be deposited in Dryad (accession number xx). From manual observations of each gene tree, the placement of known hybrid species *S. ryanii* and *S. gobicola* were observed to be strongly discordant across datasets in a non-random fashion. Of the topologies with sufficient resolution to be scored, *S. ryanii* grouped with *S. tragus* 55.55% of the time

with *S. australis* 12.54% of the time, and with accessions split between either *S. tragus* and *S. australis* 16.48% of the time. Similarly, *S. gobicola* grouped with *S. tragus* 25.08% of the time, with *S. paulsenii* 32.25% of the time, and with accessions split between either *S. tragus* or *S. paulsenii* 3.58% of the time. Within the *S. tragus* species complex, other accessions were observed to vary in their locations between the gene trees, but it was difficult to discern a pattern amongst the different relationships. Similarly, when examining other species with high percentages of AD and LH, such as *Salsola_jacquemontii_L20034, Girgenshohnia_oppositifolia_L20421, Caroxylon_vermiculatum_L20314, Xylosalsola_richteri_L20311,* and *Salsola_arbusculiformis_L20514*, it was difficult to identify a pattern upon which to base gene binning.

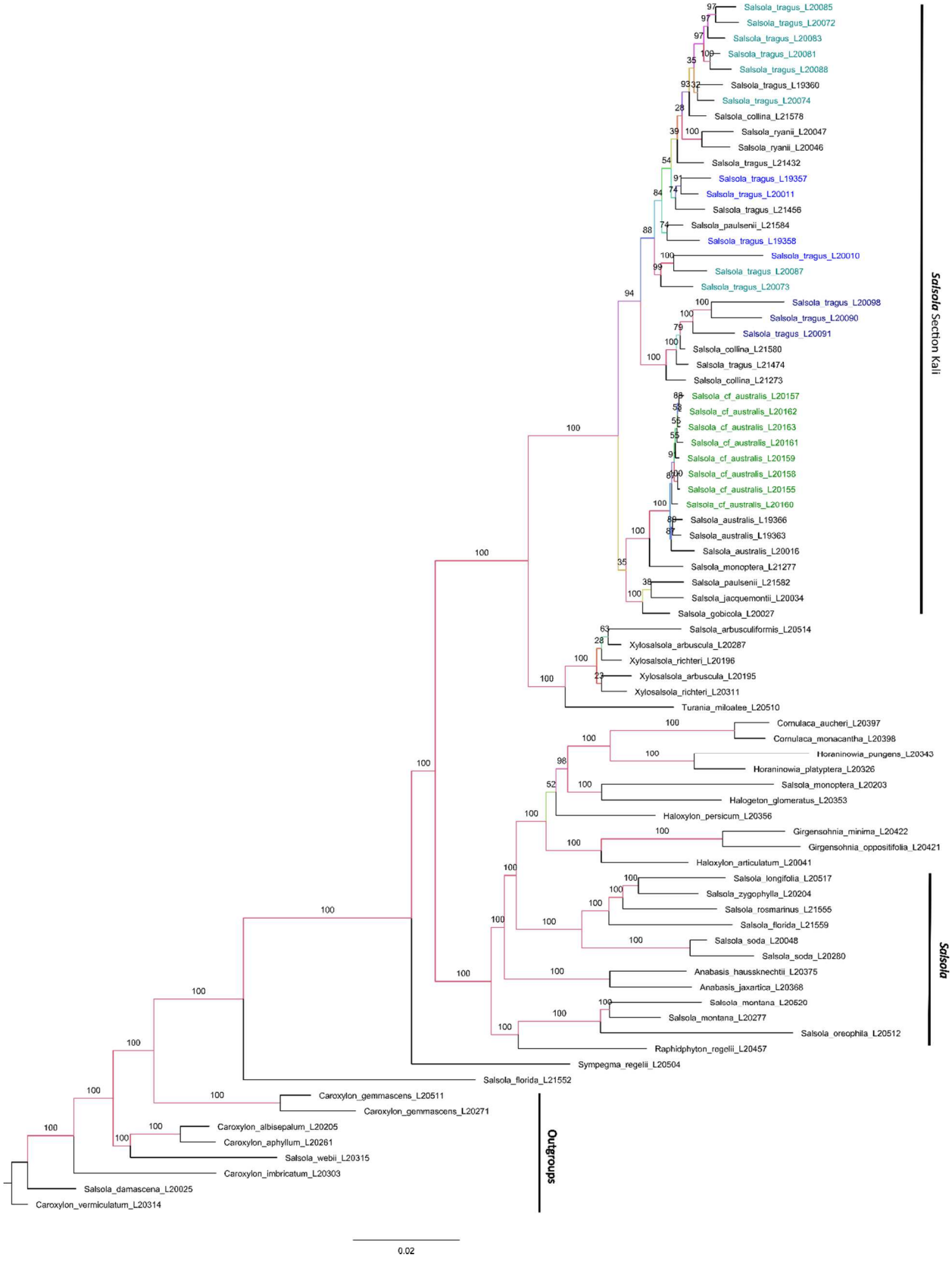
Salsola_tragus_L20085
Salsola_tragus_L20072
Salsola_tragus_L20083
Salsola_tragus_L20081
Salsola_tragus_L20088
Salsola_tragus_L19360
Salsola_tragus_L20074
Salsola_collina_L21578
Salsola_ryanii_L20047
Salsola_ryanii_L20046
Salsola_tragus_L21432
Salsola_tragus_L19357
Salsola_tragus_L20011
Salsola_tragus_L21456
Salsola_paulsenii_L21584
Salsola_tragus_L19358
Salsola_tragus_L20010
Salsola_tragus_L20087
Salsola_tragus_L20073
Salsola_tragus_L20098
Salsola_tragus_L20090
Salsola_tragus_L20091
Salsola_collina_L21580
Salsola_tragus_L21474
Salsola_collina_L21273
Salsola_cf_australis_L20157
Salsola_cf_australis_L20162
Salsola_cf_australis_L20163
Salsola_cf_australis_L20161
Salsola_cf_australis_L20159
Salsola_cf_australis_L20158
Salsola_cf_australis_L20155
Salsola_cf_australis_L20160
Salsola_australis_L19366
Salsola_australis_L19363
Salsola_australis_L20016
Salsola_monoptera_L21277
Salsola_paulsenii_L21582
Salsola_jacquemontii_L20034
Salsola_gobicola_L20027
Salsola_arbusculiformis_L20514
Xylosalsola_arbuscula_L20287
Xylosalsola_richteri_L20196
Xylosalsola_arbuscula_L20195
Xylosalsola_richteri_L20311
Turania_miloatee_L20510
Cornulaca_aucheri_L20397
Cornulaca_monacantha_L20398
Horaninowia_pungens_L20343
Horaninowia_platyptera_L20326
Salsola_monoptera_L20203
Halogeton_glomeratus_L20353
Haloxylon_persicum_L20356
Girgensohnia_minima_L20422
Girgensohnia_oppositifolia_L20421
Haloxylon_articulatum_L20041
Salsola_longifolia_L20517
Salsola_zygophylla_L20204
Salsola_rosmarinus_L21555
Salsola_florida_L21559
Salsola_soda_L20048
Salsola_soda_L20280
Anabasis_haussknechtii_L20375
Anabasis_jaxartica_L20368
Salsola_montana_L20520
Salsola_montana_L20277
Salsola_oreophila_L20512
Raphidphyton_regelii_L20457
Sympegma_regelii_L20504
Salsola_florida_L21552
Caroxylon_gemmascens_L20511
Caroxylon_gemmascens_L20271
Caroxylon_albisepalum_L20205
Caroxylon_aphyllum_L20261
Salsola_webii_L20315
Caroxylon_imbricatum_L20303
Salsola_damascena_L20025
Caroxylon_vermiculatum_L20314
Salsola Section Kali
Salsola
Outgroups
0.02

Figure 7. Concatenated RAxML gene tree of tribe Salsoleae and genus *Salsola s.l.* Branch color corresponds to bootstrap support values with warmer colors conveying higher support and cooler colors representing lower bootstrap support. Teal, navy, and green colored taxa represents population specimens from Oregon, South Dakota, and California, respectively. Blue colored taxa represent herbarium accessions from Eurasia. *Caroxylon vermiculatum* is an outgroup species. Bootstrap support values are reflected above the branch of the tree.

Species tree inference – The results from the analyses using SVDQuartets and ASTRAL-Pro were largely congruent between the two datasets (Figure 8 & 9). The topology inferred from the coalescent-based species tree analyses using SVDQuartets analyses in PAUP* 4.0a with 100 bootstrap replicates found broader congruence of tribal and genus-level relationships. The clade support has been found consistent as compared to the RAxML concatenated analysis with most BS = 100% (Figure 7). Results from SVDQuartets (Figure 8) shows *Salsola s. l*. as a polyphyletic group. *Salsola s. s* (*sensu* Akhani et al. 2007) forms a clade, albeit with moderate support (BS = 72%) with several other segregate genera in the tribe such as *Cornulaca, Horaninowia, Halogeton, Haloxylon*, *Girgensohnia* and *Anabasis*. *Salsola* section Kali is recognized as a separate clade of *Salsola s. s.* (*sensu* Akhani et al. 2007) including *Salsola arbusculiformis* with BS= 100%. *Xylosalsola* and *Turania*, two genera form a clade together that is highly supported (BS = 98%) as a sister lineage to *Salsola* section Kali. Remaining genera *Sympega* and *Caroxylon* and species have also strong support found in the nodes and their backbones.

Similarly, our result inferred from the coalescent-based species tree topology using ASTRAL-Pro (Figure 9) shows local posterior probability (LPP) higher than 0.90 in majority of the branches. Inferred relationships are similar with those found in the SVDQuartets analyses (Figure 10). Among all the clades, 72.37% have LPP = 1, 22.93%

have LPP 0.70 - 0.9 and 4.7% have LPP < 0.70. *Salsola s.s.* formed a group with strong support of internal branches (LPP = 1) but very low score with genus *Anabasi*s (LPP = 0.44) in the same clade. However, the related genera and species have strong support (LPP = 0.99), and strong support in their internal relationships i.e., *Horaninowia, Cornulaca, Halogeton, Haloxylon* and *Girgensohnia*. *Salsola monopterum, Salsola damascena*, and *Salsola webii* were inferred in different group than their genera, as similar as in the RAxML concatenated gene tree. *Salsola s.s.* (LPP = 0. 99) and *Salsola* section Kali (LPP = 0.98) are inferred in two different clades, supporting *Salsola s.l.* as clearly a polyphyletic group. *S. australis* is clearly distinguished as distinctive lineage with high support (LPP = 0.98).

Among the known two North American hybrid species, *S. gobicola* and *S. ryanii, S. gobicola* have moderately higher support score (BS = 35%, BS = 71%, LPP = 0.98 from RAxML concatenated, SVDQuartets and ASTRAL-Pro phylogenies respectively) however, *S. ryanii* have been found with low support score (BS = 28%, BS = 60%, LPP = 0.62 from RAxML concatenated, SVDQuartets and ASTRAL-Pro phylogenies respectively) in comparison. This might reflect the discordance in the gene trees. Similarly, the Angiosperms353 probe set confidently resolved the backbone of the tribe Salsoleae, and partially resolves the relationship between the closely related species of *Salsola* section Kali.

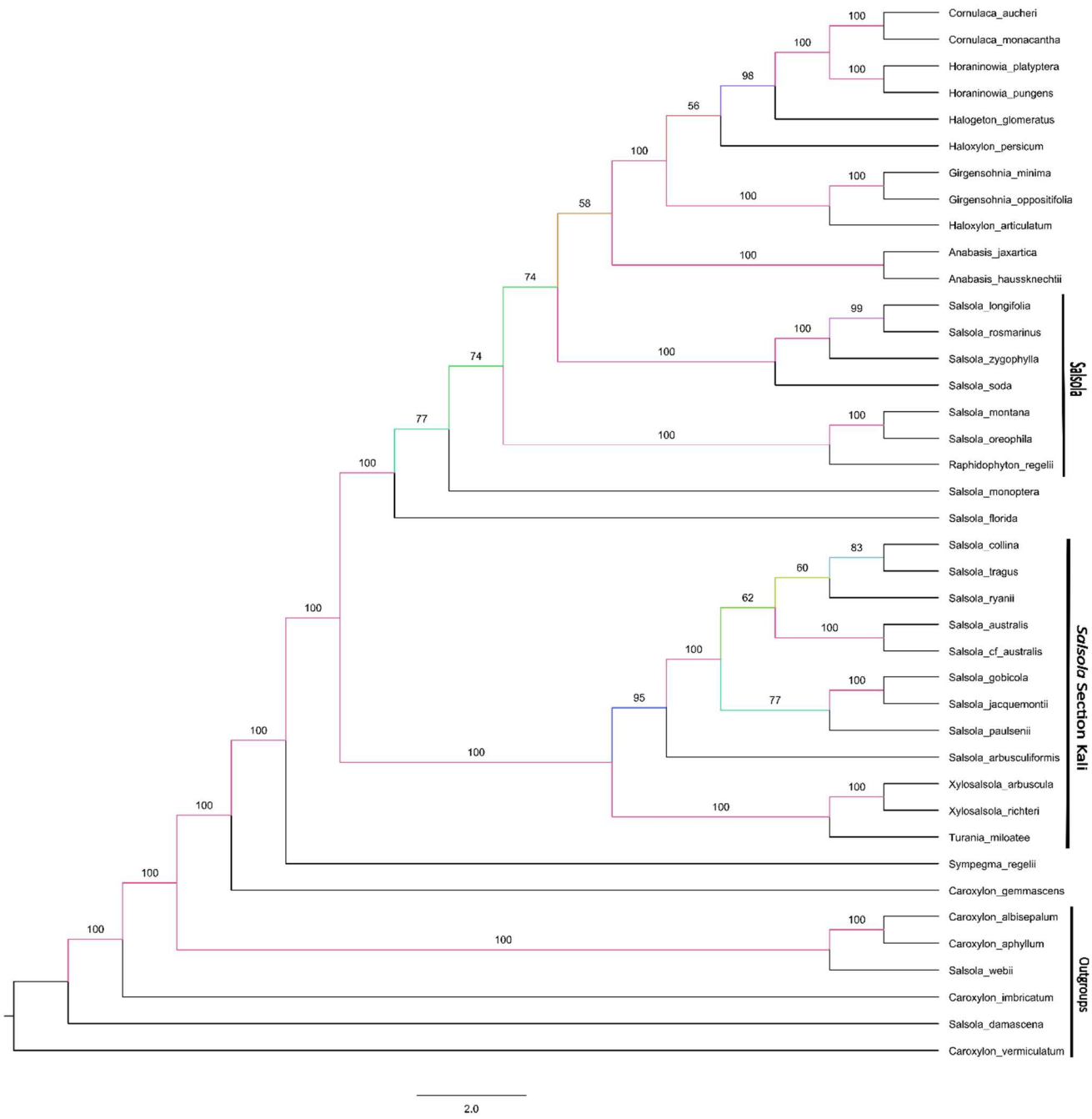

Figure 8. SVDQuartets species tree illustrating relationships in tribe Salsoleae and genus *Salsola s.l.* Maximum likelihood bootstrap support values are shown above the branches. Branch color corresponds to bootstrap support values with warmer colors conveying higher support and cooler colors representing lower bootstrap support. Bootstrap Support (BS) are overlaid over the branches. *Caroxylon vermiculatum* is an outgroup species.

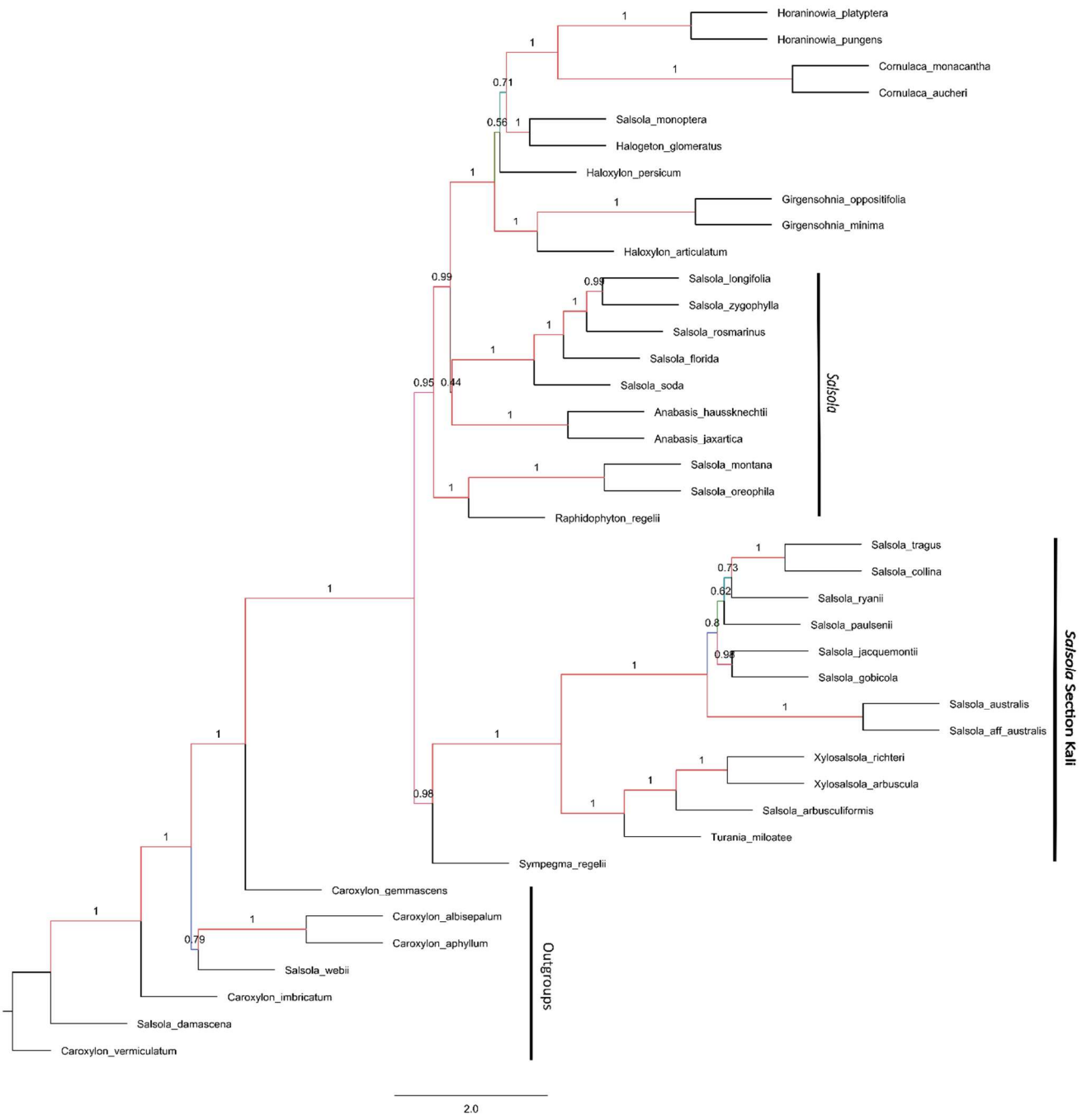

Figure 9. ASTRAL-Pro species tree illustrating relationships in tribe Salsoleae and genus *Salsola s.l.* using 279 trimmed nuclear loci from the Angiosperms353 probe set. Local posterior probability (LPP) is overlaid over the branches. Branch lengths are in coalescent units. Branch color corresponds to LPP values with warmer colors conveying higher LPP and cooler colors representing lower LPP. *Caroxylon vermiculatum* is an outgroup species.

2.5. Discussion

The process of hybridization has been linked to successful invasions, and rigorous phylogenetic studies represent an important tool for identifying these patterns and characterizing lineages involved in gene flow. Careful analyses of gene tree discordance across datasets is an important first step in investigating hybridization patterns (Crowl et al., 2017; Morales□Briones et al., 2018). Our study system, the taxonomically confusing genus *Salsola s.l.*, has a history of hybridization in Eurasia, its native range. and in North America (Beatley, 1972; Rilke, 1999). Here, we use phylogenomic data and improved sampling across invasive lineages in North America to clarify evolutionary relationships, identify colonization events, identify signatures of hybridization, and provide an improved framework for broader relationships in the tribe Salsoleae.

The tribe Salsoleae s.s. is diverse, both physiologically (C_3, C_4 and intermediate C_2) and morphologically (Akhani et al., 2007; Morales-Briones et al., 2021). Previous taxonomic work revealed a polyphyletic *Salsola*, which resulted in it getting split into several segregate genera (Akhani et al., 2007; Akhani et al., 2014), followed by multiple debates over the conservation of the *Salsola* over *Kali* (Mosyakin et al., 2014; Mosyakin et al., 2017; Wilson, 2017). Our results, based on phylogenomic analyses of the Angiosperms353 probe set strongly echo the findings of a polyphyletic *Salsola s.l.* and highly supported *Salsola* section Kali, which contains the invasive tumbleweeds in North America, and reiterates that nomenclatural should be made to reflect our understanding of evolutionary relationships. Our results also greatly clarify previously unresolved relationships in Akhani et al. (2007). For example, we found *Salsola arbusculiformis*,

recognized by Akhani et al. (2007) as *Collinosalsola*, was nested in a clade with *Xylosalsola* in RAxML concatenated tree (BS = 100%) and ASTRAL-Pro tree (LPP = 1), and as a sister lineage of *Xylosalsola* in SVDQuartets tree with strong support (BS = 100%). Our results also show a resolved placement of *Anabasis* relative to other genera in Salsoleae as a sister lineage to a clade comprised of *Cornulaca, Girgensohnia, Halogeton, Haloxylon* and *Horaninowia*. Numerous other internal relationships have been clarified compared to Akhani et al. 2007 and corroborate findings from other recent high-throughput studies across the family Amaranthaceae (Morales-Briones et al., 2021).

In our inclusion of numerous accessions representing invasive lineages from *Salsola* section Kali, we have also begun to disentangle the complicated evolutionary history of closely related species with a likely history of hybridization and polyploidy following colonization of North America. These results may also clarify species concepts among lineages that have been variously lumped into or split from *Salsola tragus*. For example, *Salsola australis* is a morphologically distinct and diploid species (2n = 18) identified as a cryptic lineage distinct from *S. tragus* (Hrusa and Gaskin, 2008). Our results show that all accessions of *S. australis* form a distinct and highly supported clade that is separate from *Salsola tragus* and warrants widespread recognition as *S. australis*. When examining the relationships among multiple accessions of *S. tragus*, including from the native and invaded range, we find that the North American accessions (invaded range) fall into a few moderately supported clades with Eurasian accessions (native range), suggesting multiple independent colonization events. While the initial landfall of *S. tragus* into North America occurred in South Dakota via contaminated seeds, multiple introductions have been suspected, however, the exact numbers of introduction events are unknown (Young,

1988). With emerging phylogenetic and population genetics studies, multiple introductions (vs. one-time events) are thought to be the rule rather than the exception (Ellstrand and Schierenbeck, 2000), and admixture between colonizers from different regions of the native range can dramatically increase genetic variation (Vallejo-Marín et al., 2021). In our study system, multiple introductions of *S. tragus s.l.* were reported in North America (Young, 1988). Additionally, species with multiple accessions did not generally cluster together into supported clades, but occurred in different places within *Salsola* section Kali, with moderate to high support. These various placements of these accessions could be the result of several ongoing processes, including hybridization, introgression, polyploidy, and incomplete lineage sorting, and underscore the unclear species boundaries among these invasive plants.

Inspection of individual gene trees, together with assessment of LH and AD, paints a picture of hybridization that occurred between different lineages of *S. tragus* and other closely related species occurring after their introduction into North America. Patterns of gene tree discordance between parental species were particularly clear for *S. gobicola* and *S. ryanii*, two putative hybrids from previous taxonomic literature and with high levels of LH and AD. *S. ryanii*, is an allohexaploid hybrid species formed from the recurrent combination of *S. tragus* and *S. australis* (Hrusa and Gaskin, 2008), and the placement of *S. ryanii* across individual gene trees is closely aligned with the parental species, seemingly with a greater contribution from *S. tragus* in the two accessions we included in our study. In the RAxML concatenated tree, *S. ryanii* is only closely related to *S. tragus* with low support (BS = 28%), masking some of the underlying history of hybridization with *S. australis*. Support for this relationship in the SVDQuartets tree and ASTRAL-Pro tree is

higher (BS =62% and LPP = 1) where *S. tragus* is within the same group and *S. australis* is placed as a sister lineage (Figure 8 & 9). *S. gobicola* was suggested to be a hybrid species of *S. tragus* and *S. paulsenii* (Rilke, 1999). Similarly, the overall placement of *S. gobicola* in the individual gene trees is closely aligned with the suspected parental species, with low support score in the clade backbone of the RAxML concatenated gene tree (BS = 35%), well supported score in the clade backbone of SVDQuartets tree and strong support score in the backbone of ASTRAL-Pro tree (LPP = 0.8) (Figure 7).

Although we generated much improved resolution of phylogeny within the tribe Salsoleae and focusing the complex genus *Salsola s.l.*, while attempting to accommodate other sources of discordance in our datasets stemming from paralogy and incomplete lineage sorting, we did not estimate polyploidy for our samples. Polyploid along with hybridization are two important factors for the evolutionary understanding of invasiveness (Abbott, 1992; Ellstrand and Schierenbeck, 2000; te Beest et al., 2012). *Salsola* section Kali consists of at least three known ploidal levels- *S. tragus* L. (tetraploid), *S. australis* (diploid) and recently formed hybrid species *S. ryanii* (allohexaploid) (Hrusa and Gaskin, 2008). Thus, polyploidy is a crucial factor to consider that could develop a different signature in our phylogenetic analyses. Beyond relying on laborious chromosome counts or flow cytometry, which are impractical for herbarium specimens, a promising future direction might be to use a modeling approach that estimates copy number from target capture sequence data directly (Viruel et al., 2019). Additionally, we only relied on assessment LH-AD to provide clues regarding known and putative hybrid species, however for the complete detection of hybrid events between the species, using the full functionality of the HybPhaser pipeline by mapping the reference sequences and phasing them would be

beneficial (Nauheimer et al., 2021). Similarly, to recover the phylogenetic signal with gene flow, Bayesian Concordance Analysis (or BUCKy) could be helpful for our analyses (Larget et al., 2010), in addition to phylogenetic network analyses (PhyloNetworks) (Solís-Lemus et al., 2017) to investigate the possibility of hybridization events for the high LH-AD samples. These additional explorations of our datasets are an avenue for future research.

In spite of underlying complexity in our data due to hybridization and/or polyploidy, we generally found Angiosperms353 datasets are capable of resolving phylogenetic relationships at multiple taxonomic scales, including between closely related species (Baker et al., 2021), although clear resolution was not obtained in our analyses for population-level accessions. In our study, by including herbarium specimens and fresh tissues, we were able to recover the large datasets from all the 96 samples. This demonstrates the successful generation of gene sequence data from both herbarium specimens and silica-dried samples alike (Appendix S5, heatmap). We also found relationships were largely concordant across the concatenated and coalescent-based topologies. Target sequence capture is a promising method for obtaining low-cost genomic information and represents a large step forward towards understanding relationships within *Salsola* section Kali.

Our study reveals that hybridization stemming from multiple colonization events likely shaped the evolution and adaptive capacity of invasive *Salsola s.l.* in North America. Hybridization may allow plants to colonize to a new habitat or more extreme habitats, and grow larger and produce more seed, (e.g. *Salsola ryanii*, (Welles and Ellstrand, 2020)). Understanding patterns of hybridization also has applied management implications, as

newly formed hybrid species could be resistant to herbicide applications (DiTomaso et al., 2013; Heap, 2021).

2.6. Conclusions

The phylogenetic analyses in our individual gene trees, concatenated gene tree and two different species trees provides a strong support for the previous and ongoing hybridization events in *Salsola* L. Previously suggested hybrid species *Salsola gobicola* and *Salsola ryanii* in show clear patterns of hybrid origin when surveying gene tree topologies and have a higher percentage of allele divergence (AD) and locus heterozygosity (LH). In addition, our sampling of *Salsola tragus s.l.* from Eurasia and North America highly support a scenario of multiple introductions of this notorious invasive species complex into North America. However, a further study is needed for the investigation of ploidy and phylogenetic network analyses for the hybrid events, and understanding the related hybrid species within the group, which could be a great effort for the effective management of *Salsola tragus* L. (Russian thistle) in a natural habitat and agricultural land.

LITERATURE CITED

Abbott, R. J. 1992. Plant invasions, interspecific hybridization and the evolution of new plant taxa. *Trends in ecology & evolution* 7: 401-405.

Aïnouche, M. L., P. Fortune, A. Salmon, C. Parisod, M.-A. Grandbastien, K. Fukunaga, M. Ricou, and M.-T. Misset. 2009. Hybridization, polyploidy and invasion: lessons from Spartina (Poaceae). *Biological Invasions* 11: 1159-1173.

Akhani, H., G. Edwards, and E. H. Roalson. 2007. Diversification of the old world Salsoleae sl (Chenopodiaceae): molecular phylogenetic analysis of nuclear and chloroplast data sets and a revised classification. *International Journal of Plant Sciences* 168: 931-956.

Akhani, H., W. Greuter, and E. H. Roalson. 2014. Notes on the typification and nomenclature of Salsola and Kali (Chenopodiaceae). *Taxon* 63: 647-650.

Anderson, E., and G. L. Stebbins Jr. 1954. Hybridization as an evolutionary stimulus. *Evolution*: 378-388.

Angiosperm Phylogeny Group. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical journal of the Linnean Society* 181: 1-20.

Arnold, M. L. 1997. Natural hybridization and evolution. Oxford University Press on Demand.

Ayres, D., F. J. Ryan, E. Grotkopp, J. Bailey, and J. Gaskin. 2009. Tumbleweed (Salsola, section Kali) species and speciation in California. *Biological Invasions* 11: 1175-1187.

Bais, H. P., R. Vepachedu, S. Gilroy, R. M. Callaway, and J. M. Vivanco. 2003. Allelopathy and exotic plant invasion: from molecules and genes to species interactions. *Science* 301: 1377-1380.

Baker, W. J., S. Dodsworth, F. Forest, S. W. Graham, M. G. Johnson, A. J. McDonnell, L. Pokorny, et al. 2021. Exploring Angiosperms353: An open, community toolkit for collaborative phylogenomic research on flowering plants. *American Journal of Botany* 108(7): 1059–1065.

Barroso, J., D. J. Lyon, and T. S. Prather. 2019. Russian Thistle: Management in a Wheat-fallow Crop Rotation. Oregon State University Extension Service.

Beatley, J. 1972. RUSSIAN-THISTLE (SALSOLA) SPECIES IN WESTERN UNITED STATES. California Univ., Los Angeles. Lab. of Nuclear Medicine and Radiation Biology.

Beck, J. B., M. L. Markley, M. G. Zielke, J. R. Thomas, H. J. Hale, L. D. Williams, and M. G. Johnson. 2021. Are Palmer's Elm-Leaf Goldenrod and the Smooth Elm-Leaf Goldenrod Real? The Angiosperms353 Kit Provides Within-Species Signal in Solidago ulmifolia sl. *Systematic botany* 46: 1107-1113.

Beckie, H. J., and A. Francis. 2009. The biology of Canadian weeds. 65. Salsola tragus L.(Updated). *Canadian journal of plant science* 89: 775-789.

Beckie, H. J., and K. N. Harker. 2017. Our top 10 herbicide□resistant weed management practices. *Pest management science* 73: 1045-1052.

Bernau, C. R., and E. P. Eldredge. 2018. Plant Guide for Prickly Russian Thistle (Salsola tragus L.). USDA-Natural Resource

Conservation Service, Great Basin Plant Materials Center, Fallon, Nevada 89406.

Boerboom, C. 1993. Russian thistle (Salsola iberica Sennen & Pau).

Brewer, G. E., J. J. Clarkson, O. Maurin, A. R. Zuntini, V. Barber, S. Bellot, N. Biggs, et al. 2019. Factors affecting targeted sequencing of 353 nuclear genes from herbarium specimens spanning the diversity of angiosperms. *Frontiers in plant science* 10: 1102.

Chifman, J., and L. Kubatko. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30: 3317-3324.

Crowl, A. A., C. Myers, and N. Cellinese. 2017. Embracing discordance: Phylogenomic analyses provide evidence for allopolyploidy leading to cryptic diversity in a Mediterranean Campanula (Campanulaceae) clade. *Evolution* 71: 913-922.

Crystal-Ornelas, R., E. J. Hudgins, R. N. Cuthbert, P. J. Haubrock, J. Fantle-Lepczyk, E. Angulo, A. M. Kramer, et al. 2021. Economic costs of biological invasions within North America. *NeoBiota* 67: 485.

DiTomaso, J., G. Kyser, S. Oneto, R. Wilson, S. Orloff, L. Anderson, S. Wright, et al. 2013. Weed control in natural areas in the western United States. *Weed Research and Information Center, University of California* 544.

Doyle, J. J., and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue.

Eberlein, C., M. Hénault, A. Fijarczyk, G. Charron, M. Bouvier, L. M. Kohn, J. B. Anderson, and C. R. Landry. 2019. Hybridization is a recurrent evolutionary stimulus in wild yeast speciation. *Nature communications* 10: 1-14.

Ellstrand, N. C., and K. A. Schierenbeck. 2000. Hybridization as a stimulus for the evolution of invasiveness in plants? *Proceedings of the National Academy of Sciences* 97: 7043-7050.

FNA. 2020. FNA | Family List | FNA Vol. 4 | Chenopodiaceae 4: 260,261,340,351.

Fowler, J. L., J. H. Hageman, K. J. Moore, M. Suzukida, H. Assadian, and M. Valenzuela. 1992. Salinity effects on forage quality of Russian thistle. *Rangeland Ecology & Management/Journal of Range Management Archives* 45: 559-563.

Frost, L., and L. Lagomarsino. 2021. More-curated data outperforms more data: Treatment of cryptic and known paralogs improves phylogenomic analysis and resolves a northern Andean origin of Freziera (Pentaphylacaceae). *bioRxiv*.

Fusco, E. J., J. T. Finn, J. K. Balch, R. C. Nagy, and B. A. Bradley. 2019. Invasive grasses increase fire occurrence and frequency across US ecoregions. *Proceedings of the National Academy of Sciences* 116: 23594-23599.

Gaskin, J. F., F. J. Ryan, G. F. Hrusa, and J. P. Londo. 2006. Genotype diversity of Salsola tragus and potential origins of a previously unidentified invasive Salsola from California and Arizona. *Madroño* 53: 244-251.

Heap, I. 2021. The International Survey of Herbicide Resistant Weeds. www.weedscience.org. [accessed October 27 2021].

Hernández-Ledesma, P., W. G. Berendsohn, T. Borsch, S. Von Mering, H. Akhani, S. Arias, I. Castañeda-Noa, et al. 2015. A taxonomic backbone for the global synthesis of species diversity in the angiosperm order Caryophyllales. *Willdenowia* 45: 281-383.

Holder, M. T., J. A. Anderson, and A. K. Holloway. 2001. Difficulties in detecting hybridization. *Systematic Biology* 50: 978-982.

Holm, L. G., D. L. Plucknett, J. V. Pancho, and J. P. Herberger. 1977. The world's worst weeds. Distribution and biology. University press of Hawaii.

Hopley, T., B. L. Webber, S. Raghu, L. Morin, and M. Byrne. 2021. Revealing the Introduction History and Phylogenetic Relationships of Passiflora foetida sensu lato in Australia. *Frontiers in plant science*: 1453.

Hovick, S. M., and K. D. Whitney. 2014. Hybridisation is associated with increased fecundity and size in invasive taxa: meta□analytic support for the hybridisation□ invasion hypothesis. *Ecology letters* 17: 1464-1477.

Hrusa, G., and J. Gaskin. 2008. The Salsola tragus complex in California (Chenopodiaceae): characterization and status of Salsola australis and the autochthonous allopolyploid Salsola ryanii sp. nov. *Madroño* 55: 113-131.

Hrusa, G. F. 2012. Salsola gobicola, in Jepson Flora Project (eds.). Jepson eFlora [accessed March 29, 2022 https://ucjeps.berkeley.edu/eflora/eflora_display.php?tid=82452 DOI.

Jepson eFlora. 2022. Jepson Flora Project (eds.) 2022. [accessed 1 June 2022 https://ucjeps.berkeley.edu/eflora/ DOI.

Johnson, M. G., E. M. Gardner, Y. Liu, R. Medina, B. Goffinet, A. J. Shaw, N. J. Zerega, and N. J. Wickett. 2016. HybPiper: Extracting coding sequence and introns for phylogenetics from high□throughput sequencing reads using target enrichment. *Applications in plant sciences* 4: 1600016.

Johnson, M. G., L. Pokorny, S. Dodsworth, L. R. Botigue, R. S. Cowan, A. Devault, W. L. Eiserhardt, et al. 2019. A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. *Systematic Biology* 68: 594-606.

Kates, H. R., M. G. Johnson, E. M. Gardner, N. J. Zerega, and N. J. Wickett. 2018. Allele phasing has minimal impact on phylogenetic reconstruction from targeted nuclear gene sequences in a case study of Artocarpus. *American Journal of Botany* 105: 404-416.

Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* 30: 772-780.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28: 1647-1649.

Kumar, V., J. F. Spring, P. Jha, D. J. Lyon, and I. C. Burke. 2017. Glyphosate-resistant Russian-thistle (Salsola tragus) identified in Montana and Washington. *Weed Technology* 31: 238-251.

Larget, B. R., S. K. Kotha, C. N. Dewey, and C. Ané. 2010. BUCKy: gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics* 26: 2910-2911.

Larridon, I., T. Villaverde, A. R. Zuntini, L. Pokorny, G. E. Brewer, N. Epitawalage, I. Fairlie, et al. 2020. Tackling rapid radiations with targeted sequencing. *Frontiers in plant science*: 1655.

Larridon, I., A. R. Zuntini, É. Léveillé□Bourret, R. L. Barrett, J. R. Starr, A. M. Muasya, T. Villaverde, et al. 2021. A new classification of Cyperaceae (Poales) supported by phylogenomic data. *Journal of Systematics and Evolution* 59: 852-895.

Maddison, W. P. 1997. Gene trees in species trees. *Systematic Biology* 46: 523-536.

McKain, M. R., M. G. Johnson, S. Uribe□Convers, D. Eaton, and Y. Yang. 2018. Practical considerations for plant phylogenomics. *Applications in plant sciences* 6: e1038.

Mesgaran, M. B., M. A. Lewis, P. K. Ades, K. Donohue, S. Ohadi, C. Li, and R. D. Cousens. 2016. Hybridization can facilitate species invasions, even without enhancing local adaptation. *Proceedings of the National Academy of Sciences* 113: 10210-10214.

Moore, J. W. 1938. Euphorbia dentata and Salsola collina in Minnesota. *Rhodora* 40: 135-136.

Morales-Briones, D. F., G. Kadereit, D. T. Tefarikis, M. J. Moore, S. A. Smith, S. F. Brockington, A. Timoneda, et al. 2021. Disentangling sources of gene tree discordance in phylogenomic data sets: testing ancient hybridizations in Amaranthaceae sl. *Systematic Biology* 70: 219-235.

Morales□Briones, D. F., A. Liston, and D. C. Tank. 2018. Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus Lachemilla (Rosaceae). *New Phytologist* 218: 1668-1684.

Morrisson, I., and M. Devine. 1994. Herbicide resistance in the Canadian prairie provinces: five years after the fact. *Phytoprotection* 75: 5-16.

Mosyakin, S. L. 1996. A taxonomic synopsis of the genus Salsola (Chenopodiaceae) in North America. *Annals of the Missouri Botanical Garden*: 387-395.

Mosyakin, S. L. 2004. SL. *FNA* 4: 260,261,340,351.

Mosyakin, S. L. 2017. The first record of Salsola paulsenii (Chenopodiaceae) in Ukraine, with taxonomic and nomenclatural comments on related taxa. *Український ботанічний журнал*.

Mosyakin, S. L., S. Rilke, and H. Freitag. 2014. (2323) Proposal to conserve the name Salsola (Chenopodiaceae s. str.; Amaranthaceae sensu APG) with a conserved type. *Taxon* 63: 1134-1135.

Mosyakin, S. L., H. Freitag, and S. Rilke. 2017. Kali versus Salsola: the instructive story of a questionable nomenclatural resurrection. *Israel Journal of Plant Sciences* 64: 18-30.

Munz, P. A. 1968. Supplement to a California Flora. University of California Press.

Murphy, B., F. Forest, T. Barraclough, J. Rosindell, S. Bellot, R. Cowan, M. Golos, et al. 2020. A phylogenomic analysis of Nepenthes (Nepenthaceae). *Molecular Phylogenetics and Evolution* 144: 106668.

Nauheimer, L., N. Weigner, E. Joyce, D. Crayn, C. Clarke, and K. Nargar. 2021. HybPhaser: A workflow for the detection and phasing of hybrids in target capture data sets. *Applications in plant sciences* 9.

Pimentel, D., R. Zuniga, and D. Morrison. 2005. Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological economics* 52: 273-288.

Pyankov, V. I., E. G. Artyusheva, G. E. Edwards, C. C. Black Jr, and P. S. Soltis. 2001. Phylogenetic analysis of tribe Salsoleae (Chenopodiaceae) based on ribosomal ITS

sequences: implications for the evolution of photosynthesis types. *American Journal of Botany* 88: 1189-1198.

Rambaut, A. 2020. FigTree v. 1.4. 4. 2018.

Rieseberg, L. H., S.-C. Kim, R. A. Randell, K. D. Whitney, B. L. Gross, C. Lexer, and K. Clay. 2007. Hybridization and the colonization of novel habitats by annual sunflowers. *Genetica* 129: 149-165.

Rieseberg, L. H., O. Raymond, D. M. Rosenthal, Z. Lai, K. Livingstone, T. Nakazato, J. L. Durphy, et al. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301: 1211-1216.

Rilke, S. 1999. Revision der Sektion Salsola SL der Gattung-Salsola (Chenopodiaceae). *Bibliotheca Botanica* 189.

Ryan, F. J., and D. R. Ayres. 2000. Molecular markers indicate two cryptic, genetically divergent populations of Russian thistle (Salsola tragus) in California. *Canadian Journal of Botany* 78: 59-67.

Ryan, F. J., S. L. Mosyakin, and M. J. Pitcairn. 2007. Molecular comparisons of Salsola tragus from California and Ukraine. *Botany* 85: 224-229.

Schierenbeck, K. A., and N. C. Ellstrand. 2009. Hybridization and the evolution of invasiveness in plants and other organisms. *Biological Invasions* 11: 1093-1105.

Seebens, H., T. M. Blackburn, E. E. Dyer, P. Genovesi, P. E. Hulme, J. M. Jeschke, S. Pagad, et al. 2017. No saturation in the accumulation of alien species worldwide. *Nature communications* 8: 1-9.

Shinn, C. H. 1895. The Russian thistle in California. Agricultural Experiment Station.

Slimp, M., L. D. Williams, H. Hale, and M. G. Johnson. 2021. On the potential of Angiosperms353 for population genomic studies. *Applications in plant sciences* 9.

Smith, S. A., and C. W. Dunn. 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. *Bioinformatics* 24: 715-716.

Solís-Lemus, C., P. Bastide, and C. Ané. 2017. PhyloNetworks: a package for phylogenetic networks. *Molecular biology and evolution* 34: 3292-3298.

Soltis, D. E., and P. S. Soltis. 2003. The role of phylogenetics in comparative genetics. *Plant physiology* 132: 1790-1800.

Soltis, P. S., and D. E. Soltis. 2009. The role of hybridization in plant speciation. *Annual review of plant biology* 60: 561-588.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312-1313.

Starr, J. R., P. Jiménez□Mejías, A. R. Zuntini, É. Léveillé□Bourret, I. Semmouri, M. Muasya, W. J. Baker, et al. 2021. Targeted sequencing supports morphology and embryo features in resolving the classification of Cyperaceae tribe Fuireneae sl. *Journal of Systematics and Evolution* 59: 809-832.

te Beest, M., J. J. Le Roux, D. M. Richardson, A. K. Brysting, J. Suda, M. Kubešová, and P. Pyšek. 2012. The more the better? The role of polyploidy in facilitating plant invasions. *Annals of botany* 109: 19-45.

Tefarikis, D. T., D. F. Morales-Briones, Y. Yang, G. Edwards, and G. Kadereit. 2021. On the hybrid origin of the C2 Salsola divaricata agg.(Amaranthaceae) from C3 and C4 parental lineages. *bioRxiv*.

Thomas, J. H., and T. JH. 1975. Salsola soda L.(Chenopodiaceae) in central California.

Toderich, K., E. Shuyskaya, F. Taha, S. Ismail, L. Gismatullina, and E. Li. 2012. Adaptive fruit structural mechanisms of Asiatic Salsola species and its germplasm conservation and utilization. *Journal of. Arid Land Studies* 22: 73-76.

Vallejo-Marín, M., J. Friedman, A. D. Twyford, O. Lepais, S. M. Ickert-Bond, M. A. Streisfeld, L. Yant, et al. 2021. Population genomic and historical analysis suggests a global invasion by bridgehead processes in Mimulus guttatus. *Communications biology* 4: 1-12.

Viruel, J., M. Conejero, O. Hidalgo, L. Pokorny, R. F. Powell, F. Forest, M. B. Kantar, et al. 2019. A target capture-based method to estimate ploidy from herbarium specimens. *Frontiers in plant science* 10: 937.

Weitemier, K., S. C. Straub, R. C. Cronn, M. Fishbein, R. Schmickl, A. McDonnell, and A. Liston. 2014. Hyb□Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Applications in plant sciences* 2: 1400042.

Welles, S. R., and N. C. Ellstrand. 2016. Genetic structure reveals a history of multiple independent origins followed by admixture in the allopolyploid weed Salsola ryanii. *Evolutionary applications* 9: 871-878.

Welles, S. R., and N. C. Ellstrand. 2020. Evolution of increased vigour associated with allopolyploidization in the newly formed invasive species Salsola ryanii. *AoB PLANTS* 12: plz039.

Wen, Z.-B., M.-L. Zhang, G.-L. Zhu, and S. C. Sanderson. 2010. Phylogeny of Salsoleae sl (Chenopodiaceae) based on DNA sequence data from ITS, psbB–psbH, and rbcL, with emphasis on taxa of northwestern China. *Plant Systematics and Evolution* 288: 25-42.

Wilgenbusch, J. C., and D. Swofford. 2003. Inferring evolutionary trees with PAUP. *Current protocols in bioinformatics*: 6.4. 1-6.4. 28.

Williams, W. I., J. M. Friedman, J. F. Gaskin, and A. P. Norton. 2014. Hybridization of an invasive shrub affects tolerance and resistance to defoliation by a biological control agent. *Evolutionary applications* 7: 381-393.

Wilson, S., Karen L. 2017. Report of the General Committee: 20. *Taxon* 66: 981-981.

Wood, T. E., N. Takebayashi, M. S. Barker, I. Mayrose, P. B. Greenspoon, and L. H. Rieseberg. 2009. The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences* 106: 13875-13879.

Young, F. L. 2006. Russian thistle (Salsola spp.) biology and management. 15th Australian Weeds Conference, Papers and Proceedings, Adelaide, South Australia: 24-28.

Young, F. L., J. P. Yenish, G. K. Launchbaugh, L. L. McGrew, and J. R. Alldredge. 2008. Postharvest control of Russian thistle (Salsola tragus) with a reduced herbicide applicator in the Pacific Northwest. *Weed Technology* 22: 156-159.

Young, J. A. 1988. The public response to the catastrophic spread of Russian thistle (1880) and halogeton (1945). *Agricultural history* 62: 122-130.

Zhang, C., C. Scornavacca, E. K. Molloy, and S. Mirarab. 2020. ASTRAL-Pro: quartet-based species-tree inference despite paralogy. *Molecular biology and evolution* 37: 3292-3307.

APPENDIX

Appendix S1: Supplementary table for a complete list of accessions

Species	DNA no.	*Updated taxonomy*	Voucher details	Collection date	Herbarium	Notes
Anabasis jaxastica	L20368		C25	N/A		N/A
Anabasis rausskeehtii	L20375		C66	N/A		N/A
Cornulaca aucheri	L20397		C41	N/A		N/A
Cornulaca monacantha	L20398		C82	N/A		N/A
Girgensohnia minima	L20422		C38	N/A		N/A
Girgensohnia oppositifolia	L20421		C2	N/A		N/A
Halogeton glomeratus	L20353		C172	N/A		N/A
Haloxylon articulatum	L20041		M. Nazir Sankary s.n.	15-Jul-71	DAV	Yolo Co., CA
Haloxylon persicum	L20356		C281	N/A		N/A
Horaninowia platyptera	L20326		C283	N/A		N/A
Horaninowia pungens	L20343		C368	N/A		N/A
Raphidophyton regelii	L20457		C50	N/A		N/A
Sympegma regelii	L20503		C54	N/A		N/A
Sympegma regelii	L20504		C293	N/A		N/A
Turania miloatee	L20510		C90	N/A		N/A
Salsola albisepala	L20205	*Caroxylon albisepalum*	NA	N/A		N/A
Salsola aphylla	L20261	*Caroxylon aphyllum*	C370	N/A		N/A
Salsola aphylla	L20264	*Caroxylon aphyllum*	C170	N/A		N/A
Salsola arbuscula	L20195	*Xylosalsola arbuscula*	NA	N/A		N/A
Salsola arbuscula	L20287	*Xylosalsola arbuscula*	C288	N/A		N/A
Salsola arbusculiformis	L20514		N/A	13-Nov-96		Uzbekistan

Salsola australis	L19363	*Kali australis*	D. Fritz s.n.	13-Oct-08	CDA	SanDiego Co., CA
Salsola australis	L19366	*Salsola australis*	M. O'Brien s.n.	3-Nov-04	CDA	Los Angeles Co., CA
Salsola australis	L20016	*Kali australis*	G.F. Hrusa 16196	14-Oct-03	CDA	San Luis Obispoco., CA
Salsola chivensis	L20210	*Xylosalsola chiwensis*	NA	N/A		NA
Salsola collina	L21273	*Kali collina*	D.E. Bouflord et al., 26975	4-Jul-95	HUH	Qinghai, China
Salsola collina	L21580	*Kali collina*	G. Rink, 6273	2-Aug-07	NyBG	Cibola County, NM
Salsola collina	L21578	*Kali collina*	D.C. Thornburg, 1345	13-Jun-14	NYBG	Yavapai County, AZ
Salsola damascena	L20025	*Kali damascena*	T.C. Fuller 20249	19-Jul-78	DAV	San Luis Obispo Co., CA
Salsola ericoides	L21576	*Caroxylon ericoides*	M. Khutsishvili, 11	21-Sep-03	NYBG	Rustavi, Georgia
Salsola florida	L21552		Ibadullayeva et al., 4	13-Jun-04	NYBG	Nakhchivan AR District, Azerbaijan
Salsola florida	L21559		Behboudi et al, sn	16-Oct-49	NYBG	Prov. Azerbeidjan, Iran
Salsola foliosa	L20265		C222	N/A		N/A
Salsola gemmascens	L20271	*Caroxylon gemmascens*	C15	N/A		N/A
Salsola gemmascens	L20511	*Caroxylon gemmascens*	N/A	N/A		Uzbekistan
Salsola gobicola	L20027	*Kali gobicola*	Hrusa et. al., 16707	27-Sep-05	CDA	Inyo Co., CA
Salsola imbricata	L20303	*Caroxylon imbricatum*	C378	N/A		N/A
Salsola jacquemontii	L20034	*Kali jacquemontii*	M. Cristoforo KZ-04-11-3	29-Jun-04	CDA	Baqanas, Kazakhstan
Salsola laricifolia	L20518		N/A			N/A
Salsola longifolia	L20307		C381	N/A		N/A
Salsola longifolia	L20517		N/A	N/A		N/A

Salsola monoptera	L20203	*Kali monopterum*	NA	N/A		N/A
Salsola monoptera	L21277	*Kali monopterum*	T.N. Ho et al., 1533	21-Aug-93	HUH	Qinghai, China
Salsola montana	L20277		C1	N/A		N/A
Salsola montana	L20520		N/A	N/A		N/A
Salsola oppositifolia	L20044		Kelch et. al., 96.096	5-May-96	CDA	N of Almeria, Spain
Salsola oreophila	L20512		N/A	N/A		N/A
Salsola paulsenii Litv.	L21582	*Kali paulsenii*	A. Tiehm, 15089	6-Sep-05	NYBG	Nye County, NV
Salsola paulsenii Litv.	L21584	*Kali paulsenii*	N.H. Holmgren, 16757	8-Oct-17	NYBG	Box Elder County, UT
Salsola richterii	L20196	*Xylosalsola richteri*	NA	N/A		N/A
Salsola richterii	L20311	*Xylosalsola richteri*	C290	N/A		N/A
Salsola rosmarinus	L21555		Hikmat Abbas, 81	16-Nov-61	NYBG	Bahral Milh, Iraq
Salsola rosmarinus	L21556		Barkley et al., 37	26-Oct-62	NYBG	Abu Ghraib, Iraq
Salsola ryanii	L20046	*Kali ryanii*	Hrusa et. al., 16788	29-Sep-05	CDA	Kern Co., CA
Salsola ryanii	L20047	*Kali ryanii*	Hrusa et. al., 16808	29-Sep-05	CDA	Kern Co., CA
Salsola setifera	L21557		NA	1928	NYBG	USSR
Salsola soda	L20048		DG Kelch 09.452	24-Jun-09	CDA	Solano Co., CA
Salsola soda	L20280		C46	N/A		N/A
Salsola tianschanica	L20211		NA	N/A		N/A
Salsola vermiculata	L20263	*Caroxylon vermiculatum*	C58	N/A		N/A
Salsola vermiculata	L20314	*Caroxylon vermiculatum*	C257	N/A		N/A
Salsola webii	L20315		C384	N/A		N/A
Salsola zygophylla	L20204		NA	N/A		N/A
Salsola zygophylla	L20324		C92	N/A		N/A
Salsola tragus	L19357	*Kali tragus*	S. Mosayakin s.n.	14-Sep-00	CDA	Kiev bank, Ukraine
Salsola tragus	L19358	*Kali tragus*	S.L. Mosayakin s.n.	23-Sep-00	CDA	Kyiv, Ukraine
Salsola tragus	L19360	*Kali tragus*	GF Hrusa et al., 16841	14-Nov	CDA	San Joaquin Co., CA

Salsola tragus	L20010	*Kali tragus*	A. Paolini TU-04-11-02	22-Nov-04	CDA	Sfax,Tunisia
Salsola tragus	L20011	*Kali tragus*	P. Toth 3121	Sep-09	CDA	Hungarian border, Slovakia
Salsola tragus	L21432	*Kali tragus*	Quayle, 844	17-Aug-04	BRIT	Roberts County, TX
Salsola tragus	L21456	*Kali tragus*	Stuckey, 3052	19-Jul-66	BRIT	Ontario, Canada
Salsola tragus	L21474	*Kali tragus*	Winter, 1692	25-Jun-06	BRIT	Woodward County, OK
Salsola tragus	L20072	*Kali tragus*	Nic Kooyers pop 2-4	18-Aug-20		Eugene, OR
Salsola tragus	L20073	*Kali tragus*	Nic Kooyers pop 2-5	18-Aug-20		Eugene, OR
Salsola tragus	L20074	*Kali tragus*	Nic Kooyers pop 2-6	18-Aug-20		Eugene, OR
Salsola tragus	L20081	*Kali tragus*	Nic Kooyers pop 2-13	18-Aug-20		Eugene, OR
Salsola tragus	L20083	*Kali tragus*	Nic Kooyers pop 2-15	18-Aug-20		Eugene, OR
Salsola tragus	L20085	*Kali tragus*	Nic Kooyers pop 2-17	18-Aug-20		Eugene, OR
Salsola tragus	L20087	*Kali tragus*	Nic Kooyers pop 2-19	18-Aug-20		Eugene, OR
Salsola tragus	L20088	*Kali tragus*	Nic Kooyers pop 2-20	18-Aug-20		Eugene, OR
Salsola tragus	L20089	*Kali tragus*	Patricia Johnson 1-1	10/10/2020		Rapid City, SD
Salsola tragus	L20090	*Kali tragus*	Patricia Johnson 1-2	10/10/2020		Rapid City, SD
Salsola tragus	L20091	*Kali tragus*	Patricia Johnson 1-3	10/10/2020		Rapid City, SD
Salsola tragus	L20094	*Kali tragus*	Patricia Johnson 1-6	10/10/2020		Rapid City, SD
Salsola tragus	L20095	*Kali tragus*	Patricia Johnson 1-7	10/10/2020		Rapid City, SD
Salsola tragus	L20096	*Kali tragus*	Patricia Johnson 1-8	10/10/2020		Rapid City, SD
Salsola tragus	L20098	*Kali tragus*	Patricia Johnson 1-10	10/10/2020		Rapid City, SD
Salsola tragus	L20099	*Kali tragus*	Patricia Johnson 1-11	10/10/2020		Rapid City, SD
Salsola c.f. australis	L20155		Richard Rachman and Gabriel Valbuena 2-1	N/A		Ventura Co., CA
Salsola c.f. australis	L20157		Richard Rachman and Gabriel Valbuena 2-3	N/A		Ventura Co., CA
Salsola c.f. australis	L20158		Richard Rachman and Gabriel Valbuena 2-4	N/A		Ventura Co., CA

Salsola c.f. australis	L20159	Richard Rachman and Gabriel Valbuena 2-5	N/A	Ventura Co., CA
Salsola c.f. australis	L20160	Richard Rachman and Gabriel Valbuena 2-6	N/A	Ventura Co., CA
Salsola c.f. australis	L20161	Richard Rachman and Gabriel Valbuena 2-7	N/A	Ventura Co., CA
Salsola c.f. australis	L20162	Richard Rachman and Gabriel Valbuena 2-8	N/A	Ventura Co., CA
Salsola c.f. australis	L20163	Richard Rachman and Gabriel Valbuena 2-9	N/A	Ventura Co., CA

APPENDIX S2. Summary table with statistics retrieved for the normal contigs (not those including intron regions)

Name	NumReads	ReadsMapped	PctOnTarget	GenesMapped	GenesWithContigs	GenesWithSeqs	ParalogWarnings
Anabasis_jaxartica_L20368	7813502	2959582	0.379	353	353	345	7
Anabasis_haussknechtii_L20375	9041909	3842588	0.425	353	353	341	16
Cornulaca_aucheri_L20397	11253845	5167617	0.459	353	353	344	5
Cornulaca_monacantha_L20398	10312811	4203244	0.408	353	353	341	4
Girgensohnia_minima_L20422	6706223	2963686	0.442	353	351	340	12
Girgensohnia_oppositifolia_L20421	7927776	3570106	0.45	353	353	337	23
Halogeton_glomeratus_L20353	8724131	3965243	0.455	353	353	342	2
Haloxylon_articulatum_L20041	7269421	3235519	0.445	353	353	334	12
Haloxylon_persicum_L20356	10499116	4950118	0.471	353	353	347	8
Horaninowia_platyptera_L20326	9770213	4181555	0.428	353	352	340	9
Horaninowia_pungens_L20343	7109451	3990654	0.561	353	353	341	3
Raphidphyton_regelii_L20457	9369879	4874910	0.52	353	353	341	4
Sympegma_regelii_L20503	10386362	6116054	0.589	353	346	266	0
Sympegma_regelii_L20504	7607446	4360189	0.573	353	353	344	4
Turania_miloatee_L20510	9604124	4665921	0.486	352	352	343	14
Caroxylon_albisepalum_L20205	8900652	3896782	0.438	353	352	342	4
Caroxylon_aphyllum_L20261	9408226	4534803	0.482	353	352	342	2
Caroxylon_aphyllum_L20264	7035299	4336920	0.616	353	348	287	0
Xylosalsola_arbuscula_L20195	7188161	3424795	0.476	352	352	348	1
Xylosalsola_arbuscula_L20287	9057133	3835127	0.423	352	351	347	2

Salsola_arbusculiformis_L20514	10906538	3915455	0.359	353	353	345	35
Salsola_australis_L19363	9657234	4647362	0.481	353	353	345	2
Salsola_australis_L19366	7996488	3904954	0.488	353	353	346	2
Salsola_australis_L20016	7189054	3985655	0.554	353	353	344	1
Xylosalsola_chiwensis_L20210	30410	8022	0.264	222	89	7	0
Salsola_collina_L21273	5801042	2781219	0.479	353	353	344	1
Salsola_collina_L21580	8784745	4121813	0.469	353	353	345	2
Salsola_collina_L21578	7982914	3489011	0.437	353	353	339	26
Salsola_damascena_L20025	7747807	2861079	0.369	353	353	332	5
Caroxylon_ericoides_L21576	12824926	7864588	0.613	352	335	97	0
Salsola_florida_L21552	10086149	4953978	0.491	353	353	339	9
Salsola_florida_L21559	15334578	8779303	0.573	353	353	337	3
Salsola_foliosa_L20265	18996007	8630794	0.454	353	343	224	0
Caroxylon_gemmascens_L20271	8789415	3793048	0.432	352	352	335	6
Caroxylon_gemmascens_L20511	9783607	2847468	0.291	353	352	340	2
Salsola_gobicola_L20027	8937976	3592441	0.402	353	352	341	18
Caroxylon_imbricatum_L20303	10557239	5230132	0.495	353	352	339	2
Salsola_jacquemontii_L20034	8359421	3914617	0.468	353	351	344	30
Salsola_laricifolia_L20518	11858338	6812409	0.574	352	348	314	0
Salsola_longifolia_L20307	5966630	2049221	0.343	353	346	300	0
Salsola_longifolia_L20517	9923160	4090345	0.412	353	352	330	20
Salsola_monopterum_L20203	9487387	4122770	0.435	353	353	339	1
Salsola_monopterum_L21277	12322296	5831789	0.473	353	353	345	1
Salsola_montana_L20277	9477759	5119048	0.54	353	353	338	4
Salsola_montana_L20520	9563998	5016305	0.524	353	352	340	5
Salsola_oppositifolia_L20044	3500074	2035149	0.581	353	337	286	4
Salsola_oreophila_L20512	11790131	6608079	0.56	353	352	342	8
Salsola_paulsenii_L21582	7452619	3053824	0.41	353	353	349	22

Salsola_paulsenii_L21584	7927045	3136434	0.396	353	353	339	27
Xylosalsola_richteri_L20196	9347015	4524334	0.484	352	352	346	2
Xylosalsola_richteri_L20311	9311339	2933726	0.315	352	351	345	6
Salsola_rosmarinus_L21555	6946424	3220777	0.464	352	351	338	4
Salsola_rosmarinus_L21556	7702729	4116202	0.534	353	339	118	0
Salsola_ryanii_L20046	8557122	3368887	0.394	353	353	344	23
Salsola_ryanii_L20047	12246389	3824501	0.312	353	353	343	19
Salsola_setifera_L21557	22397799	15033359	0.671	352	307	32	0
Salsola_soda_L20048	7792140	3028948	0.389	353	350	341	5
Salsola_soda_L20280	6107110	2878177	0.471	353	351	338	0
Salsola_tianschanica_L20211	4254272	2287760	0.538	352	343	272	0
Caroxylon_vermiculatum_L20263	12141059	5820048	0.479	353	351	303	0
Caroxylon_vermiculatum_L20314	9108581	3089980	0.339	353	352	339	10
Salsola_webii_L20315	8837676	4129746	0.467	352	351	323	0
Salsola_zygophylla_L20204	11333924	4326803	0.382	353	352	335	4
Salsola_zygophylla_L20324	1723744	545860	0.317	353	334	87	0
Salsola_tragus_L19357	9276062	2718224	0.293	353	353	342	25
Salsola_tragus_L19358	13259150	4665616	0.352	353	353	343	20
Salsola_tragus_L19360	10512877	4584071	0.436	353	353	342	25
Salsola_tragus_L20010	5202390	2085074	0.401	353	349	328	4
Salsola_tragus_L20011	7265966	3141061	0.432	353	353	337	22
Salsola_tragus_L21432	9160272	4517357	0.493	353	352	338	22
Salsola_tragus_L21456	8810795	4443600	0.504	353	353	344	18
Salsola_tragus_L21474	11176751	5206565	0.466	353	353	347	2
Salsola_tragus_L20072	10458909	4467120	0.427	353	353	341	27
Salsola_tragus_L20073	6287490	2817467	0.448	353	352	342	23
Salsola_tragus_L20074	12170133	4897655	0.402	353	353	342	30
Salsola_tragus_L20081	13858576	5342008	0.385	353	353	335	30

Salsola_tragus_L20083	8488911	3402269	0.401	353	353	339	27
Salsola_tragus_L20085	11637596	5288786	0.454	353	353	344	24
Salsola_tragus_L20087	4353896	2096467	0.482	353	352	342	19
Salsola_tragus_L20088	14429359	5920335	0.41	353	353	339	32
Salsola_tragus_L20089	651263	281868	0.433	352	273	116	0
Salsola_tragus_L20090	3799714	2126328	0.56	353	352	333	0
Salsola_tragus_L20091	4464888	2397561	0.537	353	352	338	0
Salsola_tragus_L20094	2694995	1439403	0.534	353	350	321	0
Salsola_tragus_L20095	994749	470495	0.473	352	334	268	0
Salsola_tragus_L20096	1351368	681951	0.505	353	332	248	0
Salsola_tragus_L20098	3535915	1843924	0.521	353	352	333	0
Salsola_tragus_L20099	261186	108225	0.414	352	305	232	0
Salsola_c.f._australis_L20155	11130928	4552652	0.409	353	353	348	1
Salsola_c.f._australis_L20157	13416005	5442485	0.406	353	353	348	2
Salsola_c.f._australis_L20158	13226260	5757699	0.435	353	353	346	2
Salsola_c.f._australis_L20159	12576563	5349032	0.425	353	353	346	1
Salsola_c.f._australis_L20160	11212774	4656429	0.415	353	353	348	2
Salsola_c.f._australis_L20161	8582903	3565554	0.415	353	353	348	1
Salsola_c.f._australis_L20162	11093168	4523573	0.408	353	352	347	1
Salsola_c.f._australis_L20163	12436559	5027390	0.404	353	353	347	1

APPENDIX S3: Data set optimization

1. Samples and loci removed to reduce missing data

Samples

Two samples are below the threshold (0.2) for proportion of recovered loci:

Sample	Proportion of loci recovered
Xylosalsola_chiwensis_L20210	0.017
Salsola_setifera_L21557	0.091

Eighteen samples are below the threshold (0.45) for recovered target sequence length:

Sample	Recovered length as proportion of target sequence length
Sympegma_regelii_L20503	0.285
Caroxylon_aphyllum_L20264	0.314
Xylosalsola_chiwensis_L20210	0.002
Caroxylon_ericoides_L21576	0.042
Salsola_foliosa_L20265	0.158
Salsola_laricifolia_L20518	0.428
Salsola_longifolia_L20307	0.364
Salsola_oppositifolia_L20044	0.322
Salsola_rosmarinus_L21556	0.054
Salsola_setifera_L21557	0.009
Salsola_tianschanica_L20211	0.258
Caroxylon_vermiculatum_L20263	0.345
Salsola_zygophylla_L20324	0.027
Salsola_tragus_L20089	0.039

Salsola_tragus_L20094	0.39
Salsola_tragus_L20095	0.215
Salsola_tragus_L20096	0.172
Salsola_tragus_L20099	0.145

In total 18 samples were removed:

Xylosalsola_chiwensis_L20210	Salsola_setifera_L21557
Sympegma_regelii_L20503	Caroxylon_aphyllum_L20264
Caroxylon_ericoides_L21576	Salsola_foliosa_L20265
Salsola_laricifolia_L20518	Salsola_longifolia_L20307
Salsola_oppositifolia_L20044	Salsola_rosmarinus_L21556
Salsola_tianschanica_L20211	Caroxylon_vermiculatum_L20263
Salsola_zygophylla_L20324	Salsola_tragus_L20089
Salsola_tragus_L20094	Salsola_tragus_L20095
Salsola_tragus_L20096	Salsola_tragus_L20099

Loci

One locus is below the threshold (0.2) for the proportion of recovered samples:

Locus	Proportion of samples recovered
6514	0

Fifty-Eight loci are below the threshold (0.5) for proportion of recovered target sequence length:

Locus	Proportion of target sequence length
6995	0.404
6933	0.467
5489	0.415
4889	0.478

5670	0.423
5177	0.392
5348	0.406
6389	0.282
6406	0.206
6450	0.446
5104	0.365
4744	0.44
6034	0.406
6864	0.152
6114	0.385
6366	0.246
6270	0.401
6893	0.331
6056	0.297
5842	0.409
7331	0.377
7583	0.379
6128	0.22
5703	0.394
5980	0.46
7024	0.442
5299	0.427
6565	0.464
4890	0.428
6713	0.334
6559	0.417
7021	0.381
4989	0.406
5032	0.391
5981	0.471
5968	0.453
5123	0.423

6563	0.433
5596	0.452
7013	0.343
6064	0.467
6685	0.478
6430	0.259
6544	0.344
5354	0.329
6733	0.488
4793	0.329
6746	0.21
6448	0.437
5944	0.481
6457	0.418
6792	0.366
5866	0.332
7361	0.403
6507	0.415
6175	0.476
6398	0.465
6540	0.302

In total 59 loci were removed:

6514, 6995, 6933, 5489, 4889, 5670, 5177, 5348, 6389, 6406, 6450, 5104, 4744, 6034, 6864,
6114, 6366, 6270, 6893, 6056, 5842, 7331, 7583, 6128, 5703, 5980, 7024, 5299, 6565, 4890, 6713, 6559, 7021, 4989, 5032, 5981, 5968,
5123, 6563, 5596, 7013, 6064, 6685, 6430, 6544, 5354, 6733, 4793, 6746, 6448, 5944, 6457, 6792, 5866, 7361, 6507, 6175, 6398, 6540

Data set optimization: Putative paralogous genes

Putative paralogs removed for all samples
13 loci were removed:

Locus	Mean proportion SNPs
4724	0.0832
6526	0.0354
6639	0.0325
5347	0.0355
6303	0.0532
5660	0.0418
5950	0.0361
6051	0.048
6570	0.03
4806	0.0462
5853	0.0458
5469	0.0351
5551	0.0365

Sample	Threshold	No. removed	Name of removed loci
Anabasis_jaxartica_L20368	0.0603	12	5404, 6946, 6376, 5770, 7136, 6494, 6705, 6527, 5398, 6227, 5940, 5406
Anabasis_haussknechtii_L20375	0.06942	10	6376, 5220, 4802, 6705, 6004, 5328, 6527, 6791, 6227, 5406
Cornulaca_aucheri_L20397	0.0298	13	4757, 6620, 6946, 6532, 5318, 5434, 5639, 6909, 6494, 6883, 6068, 6954, 6782
Cornulaca_monacantha_L20398	0.02842	20	6532, 5318, 7067, 7324, 6652, 5434, 6026, 5639, 6000, 6494, 5941, 6004, 6320, 6557, 6527, 5398, 5893, 6068, 6782, 6631
Girgensohnia_minima_L20422	0.06551	11	5404, 6072, 6979, 5357, 6439, 6119, 6875, 6538, 5428, 6407, 6227
Girgensohnia_oppositifolia_L20421	0.09869	5	4954, 6004, 6875, 5271, 5355
Halogeton_glomeratus_L20353	0.02834	19	6412, 6459, 5702, 5913, 5163, 6652, 5802, 5945, 5894, 6496, 6825, 6909, 7135, 6533, 4796, 5843, 5528, 5406, 6631
Haloxylon_articulatum_L20041	0.07827	11	7141, 6459, 5326, 6705, 6550, 6961, 5343, 5398, 6401, 5271, 5257
Haloxylon_persicum_L20356	0.03984	20	6620, 5578, 6738, 6946, 5168, 6459, 4527, 5639, 6164, 6705, 7325, 6533, 6462, 4796, 6527, 6886, 5940, 6717, 6068, 5406
Horaninowia_platyptera_L20326	0.05889	14	5404, 6532, 6393, 5318, 7136, 6459, 5960, 6238, 5744, 6282, 6854, 5919, 5355, 5406
Horaninowia_pungens_L20343	0.0546	19	5404, 6532, 6393, 7141, 6459, 5702, 6048, 5945, 5644, 6494, 7577, 6460, 6282, 6527, 6854, 5919, 5355, 6797, 5406
Raphidphyton_regelii_L20457	0.05533	17	4757, 6738, 5168, 5770, 5702, 4527, 6865, 6955, 6494, 6705, 6601, 6462, 6557, 5188, 5398, 5531, 6947
Sympegma_regelii_L20504	0.04319	14	6072, 6459, 5913, 6003, 5513, 5430, 5859, 5958, 6961, 5188, 5893, 6401, 6284, 6968

Turania_miloatee_L20510	0.07542	13	6110, 6393, 6459, 6979, 6492, 6882, 5454, 6550, 5859, 6875, 5343, 5919, 5926
Caroxylon_albisepalum_L20205	0.03335	14	7174, 6492, 7313, 4527, 6955, 6439, 6531, 6462, 6282, 6538, 6527, 5772, 5531, 4951
Caroxylon_aphyllum_L20261	0.03185	17	6393, 6498, 6048, 6882, 5326, 6150, 4527, 6660, 5821, 6955, 5918, 6462, 6527, 5772, 5531, 5656, 6572
Xylosalsola_arbuscula_L20195	0.04502	8	6946, 5168, 6404, 5733, 6860, 5949, 5857, 5406
Xylosalsola_arbuscula_L20287	0.04421	6	5460, 5859, 5620, 5949, 6274, 5406
Salsola_arbusculiformis_L20514	0.09094	2	6955, 6785
Salsola_australis_L19363	0.01407	17	6738, 6978, 5664, 7174, 6048, 5733, 6865, 5945, 5840, 6439, 6139, 5865, 6538, 6447, 7628, 6038, 6572
Salsola_australis_L19366	0.01034	22	5404, 6738, 6978, 5664, 5260, 7174, 6048, 5463, 5460, 6865, 5945, 5840, 6439, 5644, 6139, 5865, 5936, 6483, 6538, 7628, 6038, 6572
Salsola_australis_L20016	0.0123	20	5404, 6738, 6978, 5664, 6048, 5463, 5733, 6865, 5945, 5840, 6439, 5644, 6139, 5865, 5936, 6483, 6538, 7628, 6038, 6572
Salsola_collina_L21273	0.01514	13	6110, 6992, 6506, 4527, 5945, 5644, 6226, 6732, 5531, 5940, 6038, 7333, 6572
Salsola_collina_L21580	0.01524	13	6110, 5990, 5460, 6226, 5936, 6732, 5531, 5940, 6038, 7333, 6068, 5464, 6572
Salsola_collina_L21578	0.05289	8	5220, 6528, 7273, 6401, 6284, 6038, 6449, 6962
Salsola_damascena_L20025	0.0524	13	5933, 5910, 4527, 6003, 6955, 6494, 6550, 6679, 5958, 7577, 7367, 6954, 5406
Salsola_florida_L21552	0.04786	12	5664, 6412, 6506, 5733, 5894, 7325, 5772, 5977, 4951, 6954, 5406, 5257
Salsola_florida_L21559	0.05089	10	5318, 4471, 7067, 6825, 6557, 5188, 6527, 5843, 6797, 5406
Caroxylon_gemmascens_L20271	0.03842	27	5404, 5578, 5664, 6176, 4932, 5702, 5434, 5716, 5821, 6955, 6494, 5018, 7028, 6732, 7628, 5398, 5772, 6284, 5919, 6068, 5857, 5304, 6962, 5116, 5406, 7296, 6968

Caroxylon_gemmascens_L20511	0.04522	19	6393, 6432, 5434, 6955, 5018, 7028, 6705, 7135, 6679, 7577, 7367, 5343, 5398, 5772, 5531, 4951, 6068, 5406, 5257
Salsola_gobicola_L20027	0.07277	4	6738, 6528, 6401, 5721
Caroxylon_imbricatum_L20303	0.0216	18	5366, 6376, 7324, 5933, 5463, 4527, 6026, 6955, 7577, 6875, 6483, 6282, 6969, 6447, 7628, 5531, 5406, 5257
Salsola_jacquemontii_L20034	0.07206	3	6299, 6528, 5721
Salsola_longifolia_L20517	0.06765	15	6620, 6780, 5404, 6459, 6914, 7067, 5463, 6496, 6439, 6705, 5744, 6527, 5940, 5919, 5304
Salsola_monopterum_L20203	0.01691	18	5404, 5168, 4932, 6412, 5960, 5913, 7111, 5460, 6496, 6909, 6958, 7273, 6460, 6961, 6527, 6552, 6284, 5464
Salsola_monopterum_L21277	0.01209	18	5404, 6978, 6992, 5913, 6979, 6048, 5463, 5460, 5945, 5264, 6439, 5018, 5841, 6732, 6538, 5398, 6488, 6038
Salsola_montana_L20277	0.02935	24	6738, 6978, 5318, 6148, 6432, 6459, 5913, 5477, 6404, 4527, 5264, 6955, 5644, 6494, 5958, 5865, 5974, 5843, 5271, 6227, 5822, 6947, 6068, 7296
Salsola_montana_L20520	0.03637	21	6738, 5168, 5318, 6848, 5913, 4527, 6265, 6003, 6955, 6494, 5958, 5974, 5271, 5822, 6947, 5919, 6068, 6454, 5464, 4848, 7296
Salsola_oreophila_L20512	0.04	22	4757, 6532, 5318, 6848, 6404, 4527, 5335, 6955, 6494, 5958, 5974, 6462, 6483, 6557, 6552, 7029, 6883, 5271, 6227, 6947, 6068, 6968
Salsola_paulsenii_L21582	0.07217	3	6299, 6528, 5721
Salsola_paulsenii_L21584	0.05846	8	6738, 7241, 6528, 5644, 7273, 6679, 6401, 6962
Xylosalsola_richteri_L20196	0.04255	14	6738, 6946, 6376, 5639, 6000, 6660, 7135, 7194, 6791, 6636, 5949, 5977, 5926, 5406
Xylosalsola_richteri_L20311	0.04483	8	6946, 5168, 5733, 5639, 5859, 6883, 5949, 5406
Salsola_rosmarinus_L21555	0.03301	20	5404, 6384, 6532, 5318, 6412, 7067, 7324, 6000, 6825, 5821, 5849, 6139, 6705, 6961, 5188, 6538, 5398, 6068, 5304, 5257
Salsola_ryanii_L20046	0.06541	5	6299, 7241, 6528, 7273, 6962
Salsola_ryanii_L20047	0.06747	5	6299, 6528, 5733, 7273, 6401

Salsola_soda_L20048	0.01985	16	5038, 5702, 6487, 7067, 6026, 5335, 5422, 4942, 5859, 5188, 6636, 5949, 5271, 6954, 5406, 6318
Salsola_soda_L20280	0.02045	16	5404, 6738, 6992, 6487, 6299, 7067, 6048, 6026, 6958, 5859, 5188, 6636, 5949, 5271, 6954, 5406
Caroxylon_vermiculatum_L20314	0.05825	11	6738, 5463, 4527, 6003, 6955, 7367, 6527, 6886, 6447, 6954, 5406
Salsola_webii_L20315	0.0447	9	6738, 6384, 4527, 5894, 6164, 6955, 6649, 6462, 5343
Salsola_zygophylla_L20204	0.03861	12	6780, 6459, 5913, 7067, 5463, 6496, 6909, 6320, 6527, 6227, 5304, 5257
Salsola_tragus_L19357	0.05617	6	6528, 6506, 7273, 6401, 6284, 6962
Salsola_tragus_L19358	0.05291	7	6738, 5220, 6528, 7273, 5328, 6636, 6962
Salsola_tragus_L19360	0.05558	5	6738, 7241, 6528, 6506, 6962
Salsola_tragus_L20010	0.05204	8	6110, 6780, 5220, 5644, 6679, 6636, 6401, 6038
Salsola_tragus_L20011	0.05402	6	6110, 7241, 6528, 5843, 6449, 6962
Salsola_tragus_L21432	0.05603	9	7241, 5220, 6528, 5536, 7273, 6636, 6401, 6449, 6962
Salsola_tragus_L21456	0.05627	9	6110, 6848, 6528, 6506, 5536, 7273, 6226, 6401, 6038
Salsola_tragus_L21474	0.01761	14	6110, 5366, 6412, 5460, 5945, 6660, 6226, 6732, 5531, 5656, 5940, 6038, 5464, 6631
Salsola_tragus_L20072	0.05723	5	7241, 6528, 7273, 6449, 6962
Salsola_tragus_L20073	0.0595	4	6528, 5536, 7273, 6962
Salsola_tragus_L20074	0.05599	5	4757, 6299, 7241, 6528, 7273
Salsola_tragus_L20081	0.05577	7	7241, 6528, 5945, 6636, 6038, 6449, 6962
Salsola_tragus_L20083	0.05568	6	4757, 7241, 6528, 7273, 6401, 6962
Salsola_tragus_L20085	0.05362	5	5260, 6528, 7273, 6636, 6962
Salsola_tragus_L20087	0.05294	4	6528, 7273, 6449, 6962
Salsola_tragus_L20088	0.05354	8	6738, 7241, 5220, 6528, 5791, 6636, 6449, 6962
Salsola_tragus_L20090	0.01666	12	6110, 6992, 6882, 6216, 5460, 4527, 5945, 6732, 6282, 5531, 5656, 5464
Salsola_tragus_L20091	0.01679	17	6110, 6992, 6412, 5477, 6216, 5460, 5945, 6226, 5936, 6282, 6557, 5531, 5656, 6854, 5940, 6038, 6068
Salsola_tragus_L20098	0.01933	8	6110, 6992, 5460, 6732, 6913, 6038, 6068, 5464
Salsola_c.f._australis_L20155	0.01314	17	6738, 6978, 5664, 7174, 6048, 6865, 5945, 5840, 6439, 5644, 6139, 5865, 5936, 6483, 6538, 6038, 6572

Salsola_c.f._australis_L20157	0.01164	17	5404, 6738, 6978, 5664, 5260, 6048, 5463, 6865, 5840, 6439, 5644, 6139, 5865, 6483, 6538, 6038, 6572
Salsola_c.f._australis_L20158	0.01246	15	6738, 6978, 5664, 6048, 5463, 6865, 5840, 6439, 6139, 5865, 6483, 6538, 7628, 6038, 6572
Salsola_c.f._australis_L20159	0.01202	20	6738, 6978, 5664, 5260, 7174, 6048, 5463, 5460, 6865, 5945, 5840, 6439, 6139, 5865, 6483, 6538, 7628, 5843, 6038, 6572
Salsola_c.f._australis_L20160	0.01181	20	5404, 6738, 6978, 5664, 7174, 6048, 5463, 5460, 6865, 5945, 5840, 6439, 5644, 6139, 5865, 6483, 6538, 7628, 6038, 6572
Salsola_c.f._australis_L20161	0.01149	21	5404, 6738, 6978, 5664, 5260, 6992, 6048, 5463, 5460, 6865, 5945, 5840, 6439, 5644, 6139, 5865, 6483, 6913, 7628, 6038, 6572
Salsola_c.f._australis_L20162	0.01478	14	6738, 6978, 5664, 5260, 6048, 6865, 5945, 5840, 6439, 5644, 6139, 5865, 7628, 6038
Salsola_c.f._australis_L20163	0.01431	16	6738, 6978, 5664, 6048, 5463, 6865, 5945, 5840, 6439, 5644, 6139, 5865, 6538, 7628, 6038, 6572

APPENDIX S4: Supplementary of LH and AD summary

sample	bp	bpoftarget	paralogs_each	nloci	allele_divergence	locus_heterozygosity	loci with >0.5% SNPs	loci with >1% SNPs	loci with >2% SNPs
Salsola_jacquemontii_L20034	184818	74.4	3	276	3.134	99.64	98.19	94.2	77.17
Girgensohnia_oppositifolia_L20421	170118	68.5	5	270	4.104	99.63	99.26	97.04	83.33
Caroxylon_vermiculatum_L20314	184005	74.1	11	274	2.514	99.27	97.08	86.86	57.66
Salsola_tragus_L21432	187587	75.5	9	275	2.145	98.91	94.55	78.55	50.18
Salsola_ryanii_L20047	189090	76.1	5	277	2.717	98.56	96.75	90.97	62.82
Xylosalsola_richteri_L20311	201408	81.1	8	278	1.727	98.56	92.81	73.74	32.37
Salsola_ryanii_L20046	186159	74.9	5	275	2.715	98.55	96.73	93.09	62.55
Salsola_arbusculiformis_L20514	179040	72.1	2	278	3.724	98.2	96.4	93.17	77.7
Salsola_tragus_L19357	183816	74	6	274	2.025	98.18	92.34	77.01	44.89
Salsola_paulsenii_L21582	184278	74.2	3	280	3.099	97.86	97.14	93.93	74.29
Salsola_gobicola_L20027	183699	74	4	274	3.194	97.81	97.08	95.26	76.28
Haloxylon_articulatum_L20041	173118	69.7	11	272	3.312	97.79	95.96	90.81	68.75
Salsola_tragus_L20088	185832	74.8	8	272	2.106	97.79	93.38	78.68	46.32
Salsola_tragus_L20011	183570	73.9	6	270	2.079	97.78	93.33	77.78	47.78
Salsola_tragus_L21456	184758	74.4	9	277	2.166	97.11	93.86	79.42	50.18
Turania_miloatee_L20510	180801	72.8	13	276	2.869	97.1	93.12	84.42	57.97
Salsola_tragus_L20087	182514	73.5	4	276	2.055	97.1	92.39	76.45	48.19
Salsola_tragus_L20072	187344	75.4	5	274	2.119	97.08	91.97	79.2	47.81
Salsola_tragus_L20073	186657	75.1	4	273	2.168	97.07	93.04	80.95	51.65
Salsola_tragus_L20074	185256	74.6	5	273	2.11	97.07	91.58	76.56	47.62
Salsola_tragus_L19358	186684	75.2	7	276	2.129	96.74	90.22	80.43	47.83

Salsola_tragus_L20085	184221	74.2	5	275	2.051	96.73	92.36	76.73	49.45
Salsola_paulsenii_L21584	185502	74.7	8	273	2.226	96.7	92.67	79.12	50.18
Salsola_tragus_L20083	189513	76.3	6	273	2.122	96.7	91.58	77.66	46.15
Salsola_collina_L21578	183759	74	8	272	2.082	96.69	92.28	79.04	48.16
Salsola_tragus_L20081	186996	75.3	7	271	2.127	96.31	90.41	77.12	45.02
Xylosalsola_arbuscula_L20287	201054	80.9	6	278	1.544	96.04	86.33	62.23	29.14
Salsola_florida_L21552	182622	73.5	12	276	1.834	96.01	86.96	71.01	33.7
Salsola_damascena_L20025	168411	67.8	13	271	2.275	95.57	92.62	83.03	47.97
Salsola_tragus_L20010	148929	60	8	270	1.953	95.56	89.26	73.33	42.96
Caroxylon_aphyllum_L20261	182634	73.5	17	274	1.273	95.26	82.12	47.08	17.88
Xylosalsola_richteri_L20196	198648	80	14	278	1.524	94.96	82.01	60.07	28.78
Horaninowia_platyptera_L20326	193884	78.1	14	274	1.97	94.89	80.29	64.96	39.05
Salsola_oreophila_L20512	192006	77.3	22	274	1.572	94.89	75.18	51.82	22.26
Xylosalsola_arbuscula_L20195	197679	79.6	8	280	1.386	94.64	75.71	54.29	27.5
Salsola_tragus_L19360	186090	74.9	5	276	2.134	94.57	90.22	78.26	47.1
Caroxylon_albisepalum_L20205	192462	77.5	14	276	1.237	94.2	77.17	48.19	17.39
Sympegma_regelii_L20504	173418	69.8	14	275	1.547	92.73	78.55	58.55	25.45
Salsola_florida_L21559	187188	75.4	10	275	1.392	91.27	66.18	45.82	27.64
Girgensohnia_minima_L20422	186990	75.3	11	275	1.74	88.73	74.18	57.45	35.27
Caroxylon_gemmascens_L20511	185418	74.6	19	275	1.244	86.91	59.27	37.45	21.82
Cornulaca_aucheri_L20397	200694	80.8	13	280	0.988	86.79	62.86	36.07	14.29
Horaninowia_pungens_L20343	175383	70.6	19	278	1.699	86.69	65.83	46.76	28.42
Salsola_longifolia_L20517	168912	68	15	270	2.041	86.67	71.85	59.26	36.3
Caroxylon_gemmascens_L20271	169608	68.3	27	273	1.275	85.35	61.17	38.1	19.41
Cornulaca_monacantha_L20398	193956	78.1	20	277	0.965	84.48	57.4	33.94	11.55
Raphidphyton_regelii_L20457	184728	74.4	17	275	1.508	84.36	59.27	41.82	24.73
Salsola_montana_L20520	167757	67.5	21	277	1.201	83.75	59.57	33.94	17.69

Anabasis_haussknechtii_L20375	186531	75.1	10	276	1.761	82.61	62.32	44.2	30.07
Haloxylon_persicum_L20356	198741	80	20	278	1.186	82.01	56.47	34.17	19.42
Salsola_rosmarinus_L21555	162045	65.2	20	238	0.961	78.57	48.32	26.89	16.39
Salsola_zygophylla_L20204	188349	75.8	12	273	0.997	78.02	48.72	28.57	16.48
Salsola_montana_L20277	183012	73.7	24	275	0.942	77.82	44.36	27.64	12.36
Caroxylon_imbricatum_L20303	180057	72.5	18	275	0.71	73.82	42.55	20	7.27
Anabasis_jaxartica_L20368	189612	76.3	12	278	1.401	71.22	46.4	33.45	22.66
Salsola_tragus_L20091	185145	74.5	17	275	0.432	69.09	30.91	12	4.36
Salsola_webii_L20315	143871	57.9	9	265	0.947	68.68	40.38	24.91	16.98
Salsola_tragus_L21474	206208	83	14	278	0.398	62.95	25.54	11.15	3.6
Salsola_monopterum_L21277	205215	82.6	18	278	0.357	61.51	19.42	7.91	2.88
Salsola_c.f._australis_L20159	208521	83.9	20	278	0.346	59.71	18.35	7.55	2.52
Salsola_c.f._australis_L20158	207540	83.6	15	279	0.342	59.14	18.64	7.53	2.51
Salsola_monopterum_L20203	186432	75.1	18	278	0.52	58.63	24.1	12.59	6.12
Salsola_c.f._australis_L20157	210993	84.9	17	280	0.321	58.21	17.5	6.43	1.79
Salsola_collina_L21273	199023	80.1	13	277	0.366	57.04	20.22	11.55	4.33
Salsola_australis_L20016	201543	81.1	20	278	0.342	56.83	18.35	7.91	2.88
Salsola_collina_L21580	206412	83.1	13	278	0.326	56.83	19.78	8.63	3.6
Salsola_tragus_L20090	168987	68	12	274	0.356	55.84	22.63	9.12	3.28
Salsola_soda_L20280	171246	68.9	16	276	0.472	55.8	22.1	12.68	6.52
Salsola_c.f._australis_L20163	209406	84.3	16	279	0.325	55.56	18.28	8.6	2.51
Salsola_c.f._australis_L20155	209163	84.2	17	280	0.316	54.64	18.21	7.86	2.14
Salsola_c.f._australis_L20160	207423	83.5	20	280	0.311	54.64	16.79	7.14	2.86
Salsola_c.f._australis_L20161	208194	83.8	21	280	0.3	54.64	15.36	7.5	1.79
Salsola_tragus_L20098	161220	64.9	8	273	0.383	54.58	26.01	11.72	2.56
Salsola_australis_L19366	205488	82.7	22	279	0.293	53.41	14.34	8.24	3.23
Salsola_c.f._australis_L20162	209397	84.3	14	279	0.309	52.69	19	7.17	2.87

Salsola_australis_L19363	207600	83.6	17	279	0.322	52.33	16.49	7.89	2.87
Halogeton_glomeratus_L20353	195318	78.6	19	276	0.532	51.09	20.29	14.49	9.78
Salsola_soda_L20048	199716	80.4	16	279	0.448	47.67	16.85	11.11	5.73

APPENDIX S5: Supplementary image demonstrates the successful generation of gene sequence data from both herbarium specimens and silica-dried samples. The x-axis represents individual genes sequenced from the Angiosperms353 probe set and the y-axis represents the 96 samples submitted for sequencing. Shade amount in each box represents gene recovered for that samples, relative to the length of the target reference. Samples with low, moderate, and highly recovered genes are represented in white, light grey and black color respectively.