Evolutionary Relationships and Salt Tolerance within *Medicago*

by

Andrew Hopkins

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved April 2023 by the
Graduate Supervisory Committee:

Martin Wojciechowski, Chair
Yujin Park
Kelly Steele

ARIZONA STATE UNIVERSITY

May 2023

ABSTRACT

Genome-wide, single nucleotide polymorphisms (SNPs) and germination data were analyzed to better understand species delimitation and salt-tolerance within the legume genus *Medicago*. Molecular phylogenies revealed that the widely-used, genomic model line R108 and two deeply divergent accessions of *Medicago truncatula* are in fact more closely related to *Medicago littoralis* than to other accessions representing *Medicago truncatula*. This result was supported by germination data wherein the two accessions representing deeply divergent *Medicago truncatula* demonstrated salt-tolerance that was more similar to *Medicago littoralis* than to other accessions of *Medicago truncatula*. Molecular phylogenies revealed that two additional accessions representing deeply divergent *Medicago truncatula* appear to be more closely related to *Medicago italica* than to other accessions representing *Medicago truncatula*. The results of the present study elucidate complex evolutionary relationships and contribute to the present understanding of existing salt-tolerance within *Medicago*.

DEDICATION

This project would not have been possible without the support of my incredible mentors, family, and friends. Dr. Kelly Steele fostered my interest in research and academics from the first day I walked into her classroom, opening doors to me that I never thought possible. Dr. Steele's passion for seeking truth and sharing knowledge with others will always be an inspiration. Dr. Yujin Park has a talent for recognizing potential in her students, often before they recognize it in themselves. I am constantly striving to bring the same creativity and insight that Dr. Park brings to the laboratory and classroom every day. Dr. Martin Wojciechowski has shown me true generosity with his time, talents, and experience. My life outside of graduate school has been tumultuous at times, however Dr. W's steadfast nature and good humor have always been a great comfort. I cannot finish this dedication without thanking my mother, Caroline Hopkins, for the late nights and early morning we spent building the foundation of my education and character; and my father, David Hopkins, for his patience and commitment to whatever goals I set out to achieve. The guidance of my parents has been, and always will be, paramount to whatever successes I may find. I am so grateful to my dear partner Marissa for her unending support as I absorbed myself in this academic pursuit. I am excited to meet whatever life offers next by your side. Finally, enough cannot be said for the sense of love and community offered by my beloved brothers, Tyler and Evan, and my cherished friends, Christopher and Mike.

Thank you all.

ACKNOWLEDGMENTS

Thank you to: Jean-Marie Prosperi (INRA-Montpellier) , Laurent Gentzbittel (École Nationale Supérieure Agronomique de Toulouse), Nevin Young (University of Minnesota), Peter Tiffin (University of Minnesota), Connor Cameron (National Center for Genome Resources), and Andrew Farmer (National Center for Genome Resources) for their tireless efforts in maintaining the Medicago HapMap germplasm and supplying the *Medicago* research community with valuable genomic resources.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

ANALYZING COMPLEX RELATIONSHIPS WITHIN *MEDICAGO* SUBSECT. *PACHYSPIREAE* USING NUCLEAR SNP DATA

Introduction

Whole Genome Sequencing of Medicago truncatula *and Relatives*

The genus *Medicago* L. has been subject to extensive genomic study. Interest in factors underlying root nodulation, forage suitability, and abiotic stress, coupled with advancements in sequencing technology have resulted in several chromosome-level assemblies and hundreds of sequenced accessions (e.g. Young et al., 2011; Stanton-Geddes et al., 2013; Yoder et al., 2013; Moll et al., 2017; Zhou et al., 2017; Pecrix et al., 2018; Chen et al., 2020). *Medicago truncatula* Gaertn. garnered special interest as a model organism due its small, diploid genome (approximately 500 Mbp), self-compatibility, and short generation time (summarized by Cook, 1999). Young et al. (2011) published the first draft assembly of the *M. truncatula* genome. The project utilized bacterial artificial chromosomes (BAC clones) from the *M. truncatula* line A17 (Young et al., 2011). This clone-by-clone sequencing strategy struggled to assemble highly repetitive regions and would be improved as sequencing technologies progressed (Young et al., 2011). An updated version of the *M. truncatula* genome was published by Tang et al. (2014; Mt4.0), where next generation sequencing (NGS), combined with RNA-seq data, uncovered 50,894 genes. Pecrix et al. (2018) used PacBio long-read sequencing and optical mapping to publish the most current version of the genome at the time of writing (Mt5.0). This version resolved highly repetitive regions and produced the assembly in 64 contigs (Pecrix et al., 2018).

Interest in the *Medicago truncatula* genome is not limited to line A17. Hoffman et al. (1997) described the increased nitrogen fixing ability of *M. truncatula* ecotype 108-1 and subsequently developed line R108-1, a derivative with increased transformation efficiency. The suitability of line R108-1 for transformation using the tobacco retrotransposon *Tnt1* was a boon to the functional genomic study of *Medicago* (Cerbah et al., 1999, Tadege, Ratet, & Mysore, 2005; Tadege et al., 2008). Moll et al. (2017) published the first whole genome assembly of line R108 utilizing PacBio long-read sequencing and Dovetail optical mapping. The R108 assembly was subsequently improved using Hi-C optical mapping (Kaur et al., 2021; Li et al., 2022). This updated R108 assembly was used to re-map *Tnt1* insertions that were previously mapped to line A17 (Kaur et al., 2021) while Zhou et al. (2017) published a draft pangenome of *M. truncatula* based on 15 *de novo* assemblies – including both A17 and R108.

To date, three additional species of *Medicago* have been assembled at the chromosome-level. Sequencing the *Medicago sativa* L. genome was initially frustrated by autotetraploidy and self-incompatibility. However, PacBio long-read technology and Hi-C optical mapping enabled an allele-aware, chromosome-level assembly by Chen et al. (2020). The *M. sativa* assembly was followed by chromosome-level assemblies of both *Medicago polymorpha* L. and *Medicago ruthenica* L. (Cui et al., 2021; Wang et al., 2021).

Beyond the nuclear genome, researchers have successfully assembled whole-chloroplast genomes for a range of species within *Medicago*. Gurdon and Maliga (2014) sequenced and assembled 10 plastomes representing different lines of *M. truncatula*, including Jemalong (A17) and R108. Gurdon and Maliga (2014) identified a 45kb plastid in R108

relative to A17. The origin and implications of this inversion have been explored via whole plastid sequencing and assembly (Jiao et al., 2022; Choi et al., 2022). Jiao et al. (2022) sequenced and assembled plastid genomes representing 18 species across *Medicago* and reported the inversion in several accessions representing *Medicago littoralis*. Choi et al. (2022) assembled 54 plastid genomes from publicly available sequence data and reported the plastid inversion in a clade formed by *M. littoralis* and several accessions representing *M. truncatula*.

Insights into Evolutionary Relationships within Medicago

Prior to the availability of whole-genome sequence data, a series of molecular phylogenies were inferred using individual molecular markers. Downie et al. (1998) sequenced the nuclear ribosomal DNA internal transcribed spacer (nrDNA ITS) of 68 accessions, representing 65 species of *Medicago*. The aligned sequence data was analyzed by maximum parsimony and maximum likelihood and the resulting gene trees were used to trace the loss of an intron in the chloroplast gene: *rpo*C1 (Downie et al., 1998). Bena (2001) sequenced nrDNA ITS along with the ribosomal external transcribed spacer (ETS) across 53 *Medicago* species. Maureira-Butler et al. (2008) sequenced nuclear genes CNGC5 and *β-cop* for 60 accessions, representing 56 species of *Medicago*. Interestingly, different nuclear genes produced incongruent phylogenetic signals, attributed to incomplete lineage sorting and hybridization (Maureira-Butler et al., 2008). Steele et al. (2010) sequenced over 70 species of *Medicago* for the nuclear gene gibberellin 3-β-hydroxylase (*GA3ox1*) and the chloroplast *trnK intron/matK gene* region. Based on phylogenetic analyses of these genes, *Medicago truncatula*, *Medicago littoralis*, and *Medicago italica* formed a strongly supported subclade within the

Medicago subsect. *Pachyspireae* clade (Steele et al., 2010). However, conflicting topologies and varying levels of resolution for this group were produced by analyses of *GA3ox1*, ITS, CNGC5, and *trnK/matK* sequence data (Steele et al., 2010).

Publication of the *M. truncatula* genome hastened phylogenetic study of *M. truncatula* by enabling map-based assemblies for an array of species and individuals across the genus. Branca et al. (2011) sequenced 26 accessions of *M. truncatula* to an average depth of ~15x. When aligned to the A17 genome approximately 3 million single nucleotide polymorphisms (SNPs) were identified and used to calculate nucleotide diversity for the species (Branca et al., 2011). Stanton-Geddes et al. (2013) sequenced 288 accessions from the INRA core collection of *M. truncatula*, to an average depth of ~8x. The reads were aligned to the A17 genome and 5,000 random SNPs were used to produce a neighbor-joining tree. Yoder et al. (2013) added to the breadth of species studied by sequencing 29 taxa, representing 24 generally recognized species of *Medicago*, to an average depth of ~10.2x. When aligned to the A17 genome, more than 87,000 SNPs were identified and subsequently used in parsimony analysis (Yoder et al., 2013). A recurring group of deeply divergent *M. truncatula* accessions is identified in each of the phylogenetic studies reviewed above (Branca et al., 2011; Stanton-Geddes et al., 2013; Yoder et al., 2013).

The studies described above utilize plastid structure and nuclear SNP data to indicate the misidentification of line R108 and other deeply divergent accessions as *M. truncatula*. Jiao et al. (2022) and Choi et al. (2022) report that line R108 demonstrates plastid structure that is more similar to *M. littoralis* than *M. truncatula*. Furthermore, analysis of nuclear SNP data indicates that deeply divergent accessions of *M. truncatula* are more

closely related to *M. littoralis* and *M. italica* (Branca et al., 2011; Stanton-Geddes et al., 2013; Yoder et al., 2013). The goal of the present study is to examine these misidentified accessions using an updated, comprehensive nuclear SNP dataset. I hypothesize that R108, and other deeply divergent accessions of *M. truncatula*, are in fact more closely related to *Medicago littoralis* and *Medicago italica*, than to other accessions of *M. truncatula*.

Materials and Methods

Sampling

The present study samples 24 accessions, with one accession representing line A17 (HM101) and two accessions representing line R108 (HM029 and HM340) (Appendix A). Note that HM029 is identified as *Medicago tricycla* in the Medicago Analysis Portal (Appendix A). Seventeen accessions are identified as *M. truncatula* in Stanton-Geddes (2013), four of which are categorized as highly divergent (HM274, HM250, HM255, and HM258) (Appendix A). HM250 and HM274 are identified as *Medicago murex* in the Medicago Analysis Portal and are used as an outgroup for subsequent analyses (Appendix A). Yoder et al. (2013) discusses two highly divergent accessions which are likely misidentified as *M. truncatula* (HM017 and HM022) both of which are included here (Appendix A). One accession is identified as *Medicago italica* (HM324) and another accession is identified as *M. littoralis* (HM030) in the Medicago Analysis Portal (Appendix A). HM018 was excluded from analyses, as sequence data likely represents a mixture of 2 different accessions (Nevin Young, University of Minnesota, personal

communication) and is no longer included in available germplasm by the Medicago Analysis Portal.

Dataset Construction

Variant call data was downloaded from the Medicago Analysis Portal (https://medicago.legumeinfo.org). The Epstein-Burghardt variant call dataset (unpublished) aligns available HapMap sequence data to Mt5.0, the most current version of the *M. truncatula* genome available. The *Mt5_qual30_primitives.2021-09-29.bcf* dataset was used as a starting point for dataset filtering. Filtering with bcftools was performed in the following order (Li et al., 2011). First, invariant sites were removed by filtering for a minimum quality score of 30 (*bcftools view --include "QUAL>30"*). Genotypes without depth statistics were set to missing (*bcftools +setGT -- -t q -n . -i FMT/DP=="."*). Genotypes with coverage less than 5x, or greater than 200x, were set to missing (*bcftools +setGT -- -t q -n . -i FMT/DP>5*; *bcftools +setGT -- -t q -n . -i FMT/DP>200*). Genotypes with a phred-scaled quality score less than 30 (i.e. accuracy < 99.9%) were also set to missing (*bcftools +setGT -- -t q -n . -i FMT/GQ<30*). Subsequently, any site missing data for more than 3 samples was removed (*bcftools view -e COUNT(FORMAT/GT="mis")>3*). Indels were removed, limiting downstream analyses to SNPs (*bcftools view -v snps*). From the remaining SNPs, multivariant sites and heterozygous genotypes were removed, limiting downstream analysis to biallelic, homozygous variants (*bcftools view -m2 -M2* and *bcftools view -g ^het*).

The filtered SNP dataset was pruned for linkage disequilibrium using PLINK (Purcell et al., 2007). A window size of 50 bp and a step size of 5 bp were used as sliding window

parameters. The inflation factor was conservatively set to 7 (*plink --indep 50 5 7 --allow-extra-chr*). Finally, the SNP dataset was limited to nuclear sites using bcftools (*bcftools view -r medtr.gnm5.A17.MtrunA17Chr1, medtr.gnm5.A17.MtrunA17Chr2, medtr.gnm5.A17.MtrunA17Chr3, medtr.gnm5.A17.MtrunA17Chr4 ,medtr.gnm5.A17.MtrunA17Chr5, medtr.gnm5.A17.MtrunA17Chr6, medtr.gnm5.A17.MtrunA17Chr7, medtr.gnm5.A17.MtrunA17Chr8*).

Phylogenetic Analyses

Maximum parsimony analysis of the concatenated nuclear SNP dataset was performed using PAUP* version 4.0a (Swofford, 2002). HM250 and HM274, *M. murex*, were set as outgroup taxa. A heuristic search for the most parsimonious trees was executed with no topological constraints. To estimate branch support, 100 non-parametric bootstrap replicates were executed with no topological constraints and Maxtrees set to 1,000. The same parameters were applied for maximum parsimony analysis of each chromosome individually. The resulting Newick tree files were visualized with ggtree (Yu et al., 2017).

Maximum likelihood analysis of the concatenated SNP dataset was performed using RAxML version 8.2.12 (Stamatakis, 2014). HM250 and HM274, *M. murex*, were set as outgroup taxa. A rapid bootstrap analysis using 50 replicates (determined by autoMRE) was followed by a maximum likelihood search. A GTR substitution matrix with Lewis correction for ascertainment bias was used, as the SNP dataset does not contain invariant sites. The same parameters were applied to maximum likelihood analysis of each

chromosome individually. The resulting Newick tree files were visualized using ggtree (Yu et al., 2017).

Detecting translocation between chromosomes 4 and 8

Pecrix et al. (2018) localized breakpoints of a translocation in line A17 to positions 46,925,611 of Chromosome 4 and 37,031,558 of Chromosome 8. Sequence from genes on either side of each breakpoint were queried against the Medicago_truncatula_Genomic_Sequence_Collection using BLAST (https://medicago.legumeinfo.org/). The search returned scaffold-level positioning of the genes in R108_HiC, HM056, HM058, and HM125, which were tested for synteny across the breakpoints (Table 2).

Results

After filtering and pruning for linkage disequilibrium, the concatenated nuclear SNP dataset comprised 63,773 SNPs with a mean coverage of ~50 reads per site (Table 1).

Parsimony Analysis with PAUP*

PAUP* removed 28,124 uninformative characters, leaving 35,649 informative characters. A heuristic search of the concatenated SNP dataset tried 4,311 rearrangements and retained one most parsimonious tree. An 80% majority-rule consensus tree of the bootstrap replicates is presented in Figure 1a. *Medicago littoralis* (HM030) and *M. italica* (HM324) form a strongly supported clade (100% BS, bootstrap) sister to *M. truncatula* (HM034, HM010, HM005 HM004, HM056, HM060, HM058, HM185, HM050, HM095, HM125, HM129, HM023, and HM101). *Medicago littoralis* (HM030) forms a strongly supported clade (100% BS) with two accessions representing line R108 (HM340 and

HM029) and two accessions representing deeply divergent *M. truncatula* (HM017 and HM022). *Medicago italica* (HM324) forms a strongly supported clade (100% BS) with two accessions representing deeply divergent *M. truncatula* (HM255 and HM258). Within the *M. truncatula* clade, line A17 (HM101) forms a subclade (95% BS) with three other accessions representing *M. truncatula* (HM056, HM058, and HM125). Five accessions (HM185, HM095, HM060, HM034, and HM129) form a second subclade within *M. truncatula* (100% BS).

The 80% majority rule consensus trees produced by separate analysis of each chromosome are presented in Figure 1b. The clade formed by *M. littoralis* (HM030) and *M. italica* (HM324), sister to *M. truncatula* (HM034, HM010, HM005 HM004, HM056, HM060, HM058, HM185, HM050, HM095, HM125, HM129, HM023, and HM101), is resolved by separate analysis of SNP data from all eight chromosomes, as is the clade formed by *M. littoralis* (HM030), R108 (HM029 and HM340), HM017, and HM022 and the clade formed by *M. italica* (HM324), HM255, and HM258. The subclade formed by A17 (HM101), HM056, HM058, and HM125 is resolved by separate analysis of chromosomes two, three, four, five, and seven. The subclade formed by HM185, HM095, HM060, HM034, and HM129 is not resolved by any separate analysis of any individual chromosome.

Maximum Likelihood Analysis with RAxML

The best-scoring maximum likelihood tree (Final ML Optimization Likelihood: 521902.9), with estimated branch support from 50 bootstrap replicates, is presented in Figure 2a. The results produced by maximum likelihood analysis were consistent with the

results produced by maximum parsimony. *Medicago littoralis* (HM030) and *M. italica* (HM324) form a strongly supported clade (100% BS) sister to *M. truncatula*. *Medicago littoralis* (HM030) forms a strongly supported clade (100% BS) with line R108 (HM029 and HM340), HM022, and HM017. *Medicago italica* (HM324) forms a strongly supported clade (100% BS) with HM255 and HM258. Line A17 (HM101) forms a strongly supported subclade (100% BS) with HM056, HM058, and HM125, with slightly higher support than the results produced by maximum parsimony analysis. HM185, HM095, HM060, HM034, and HM129 form a strongly supported subclade (100% BS).

Best-scoring maximum likelihood trees produced by separate analysis of each chromosome are presented in Figure 2b. The clade formed by *M. littoralis* (HM030) and *M. italica* (HM324) sister to *M. truncatula* is resolved by separate analysis of all eight chromosomes as is the clade formed by *M. littoralis* (HM030), R108 (HM029 and HM340), HM022, and HM017 and the clade formed by *M. italica* (HM324), HM255, and HM258. The subclade formed by A17 (HM101), HM056, HM058, and HM125 is resolved by separate analysis chromosomes two, three, four, five, and seven, similar to the results obtained with the maximum parsimony analyses. The subclade formed by HM185, HM095, HM060, HM034, and HM129 is not resolved by separate analysis of any individual chromosomes.

Translocation between Chromosomes 4 and 8

Synteny across the translocation breakpoints are presented for A17, R108, HM056, HM058, and HM125 in Table 2. As expected, genes are syntenic across the same breakpoint for A17, while genes are syntenic across opposite breakpoints for R108_HiC,

indicating that it does not share the translocation with A17. Similarly, genes were syntenic across opposite breakpoints for HM056, HM058, and HM125.

Discussion

Clade formed by M. littoralis *and R108*

Line R108 is typically described as a variety, ecotype, or subspecies of *Medicago truncatula* (Yoder et al., 2013; Gurdon & Maliga, 2014; Wang et al., 2014; Li et al., 2022). Common designations for R108 are "*M. tricycla*" or "*M. truncatula* var *tricycla*". However a well-known authority on *Medicago*, E. Small (2011), treats *M. tricycla* DC. as a heterotypic synonym ("based on a type … that is different from the accepted scientific name" p. 704) for *M. italica*, *M. tricycla* Senn. nom. nud. as a heterotypic synonym for *M. littoralis*, and *M. truncatula* Gaertn. var. *tricycla* as a heterotypic synonym for *M. truncatula*.

In the present study, line R108 (HM029 and HM340) appears to be more closely related to *Medicago littoralis* (HM030) than to other accessions of *M. truncatula* (Figure 1; Figure 2). This result is consistent with previous morphological, cytological, and molecular comparisons between R108 and *M. truncatula*. Hoffman et al. (1997) documented the morphological and physiological differences between R108 and Jemalong (A17), including shorter seed-to-seed germination times, leaf shape, leaf pigmentation, seed pod morphology, and seed shape. The increased transformation efficiency of R108 compared to A17 is widely recognized in the functional study of the *Medicago* genome (Tadege, Ratet, & Mysore, 2005).

The first assembly of the R108 genome revealed that 5.7% of R108 sequence could not be identified in the A17 genome, indicating novel R108 sequence and R108-specific genes that were either deleted from A17 or derived in R108 (Moll et al., 2017). Gurdon and Maliga (2014) identified a 45kbp plastid inversion in R108 relative to Jemalong (A17). Jiao et al. (2022) identified the same 45kbp plastid inversion in two accessions of *M. littoralis*. Choi et al. (2022) present a whole-plastid phylogeny in which R108 forms a clade with *M. littoralis* (HM030). The accessions with sufficient sequence data in the *M. littoralis* clade presented by Choi et al. (2022) demonstrate the same 45kbp plastid inversion described by Gurdon and Maliga (2014). Li et al. (2022) describe structural variants with potential phenotypic effects between R108 and A17.

In addition to R108, two accessions representing deeply diverged *Medicago truncatula* (HM017 and HM022) appear to be more closely related to *M. littoralis* than to other accessions of *M. truncatula* (Figures 1, 2). This result is consistent with previous molecular phylogenies of *Medicago*. Branca et al. (2011) removed HM017 and HM022 from their analysis based on a plastid phylogeny that indicated HM017 and HM022 "are more closely related to *M. tricycla* (HM029) and *M. littoralis* (HM030)" than to *M. truncatula* (p. E869).

In a parsimony analysis presented by Yoder et al. (2013), HM017 and HM022 were placed in a clade with *M. littoralis* (HM030), indicating that "their identification as *M. t. truncatula* is incorrect" (p. 434).

Clade formed by **M. italica**

In the present study, two accessions representing deeply divergent *M*. *truncatula* (HM255 and HM258) appear to be more closely related to *M*. *italica* (HM324) than other accessions of *M*. *truncatula* (Figures 1, 2). This result is consistent with previous molecular phylogenies of *Medicago*. Stanton-Geddes et al. (2013) removed 18 accessions (including HM255 and HM258) from an association study of the *M*. *truncatula* genome because the accessions were highly divergent from other *M*. *truncatula*. In a *trnK intron/matK gene* tree presented by Choi et al. (2022), HM255 and HM258 were placed in a clade with *M*. *italica*.

M. truncatula *subclades*

Phylogenetic analyses of the concatenated SNP dataset revealed that HM058, HM125, and HM056 are more closely related to A17 (HM101) than to other accessions of *M*. *truncatula*. Interestingly, a defining feature of this clade appears to be an original collection location in Spain (Ronfort et al. 2006) (Appendix A) (Figure 3). This result is consistent with population structure analysis performed by Ronfort el al. (2006), which assigned Jemalong (A17) to the Spanish-Moroccan group. However, accessions HM058, HM125, and HM056 do not appear to share the A17 translocation between chromosomes 4 and 8 (Table 2). This result is consistent with previous studies that indicate A17 is the only accession of *M*. *truncatula* which possesses this translocation (Cerbah et al., 1999; Kamphuis et al., 2007; Young et al., 2011; Tang et al., 2014; Pecrix et al., 2018; Moll et al., 2017; Kaur et al., 2021; Li et al., 2022).

The translocation was first suggested by Cerbah et al. (1999) based on fluorescent *in-situ* hybridization (FISH) analysis of 5S rDNA. The translocation was confirmed by Kamphuis et al. (2007) based on pollen viability and linkage analysis. Hybrids involving A17 and other lines of *M. truncatula* showed less than 50% pollen viability, as opposed to hybrids excluding A17, which showed nearly 100% pollen viability; suggesting that the translocation is private to A17 (Kamphuis et al., 2007). Genetic mapping of hybrids involving A17 produced seven linkage groups, as opposed to the expected eight (Kamphuis et al., 2007). Markers from groups 4 and 8 formed a single linkage group in these hybrids (Kamphuis et al., 2007). The A17 translocation was further supported by whole-genome assemblies of A17, R108, *M. sativa*, and *M. ruthenica* (Young et al., 2011; Tang et al., 2014; Pecrix et al., 2018; Moll et al., 2017; Kaur et al., 2021; Li et al., 2022; Shen et al., 2020; Wang et al., 2021).

Based on phylogenetic analysis of the concatenated SNP dataset, a second subclade was formed within *M. truncatula* by five accessions collected in France (HM185, HM095, HM060, HM034, and HM129) (Ronfort et al., 2006) (Appendix A; Figure 3).

Conclusion

Accurate species identification is important for the design and interpretation of comparative and functional genomics. The goal of the present study was to analyze complex evolutionary relationships within *Medicago* subsect. *Pachyspireae*, focusing on accessions of *M. truncatula* with uncertain species identification. Phylogenetic analyses of nuclear SNP data support the hypothesis that such accessions are more closely related to *M. littoralis* and *M. italica*, than to other accessions of *M. truncatula*. Notably, line

R108 forms a strongly supported clade with *M. littoralis*, calling into question its classification as a subspecies, variety, or ecotype of *M. truncatula*. This conclusion is consistent with previous studies of whole-plastome sequence data and structure (Gurdon and Maliga, 2014; Jiao et al., 2022; Choi et al., 2022).

Within the clade formed by *M. truncatula*, two subclades emerge with a geographic basis. Line A17 formed a subclade with three accessions originally collected in Spain. This result is consistent with microsatellite data and suggests a geographic origin of the Jemalong cultivar (Ronfort et al., 2006). Interestingly, A17 is the only member of this clade that demonstrates the characteristic translocation between chromosomes 4 and 8. This result supports the hypothesis that the translocation is private to A17 (Kamphuis et al., 2007).

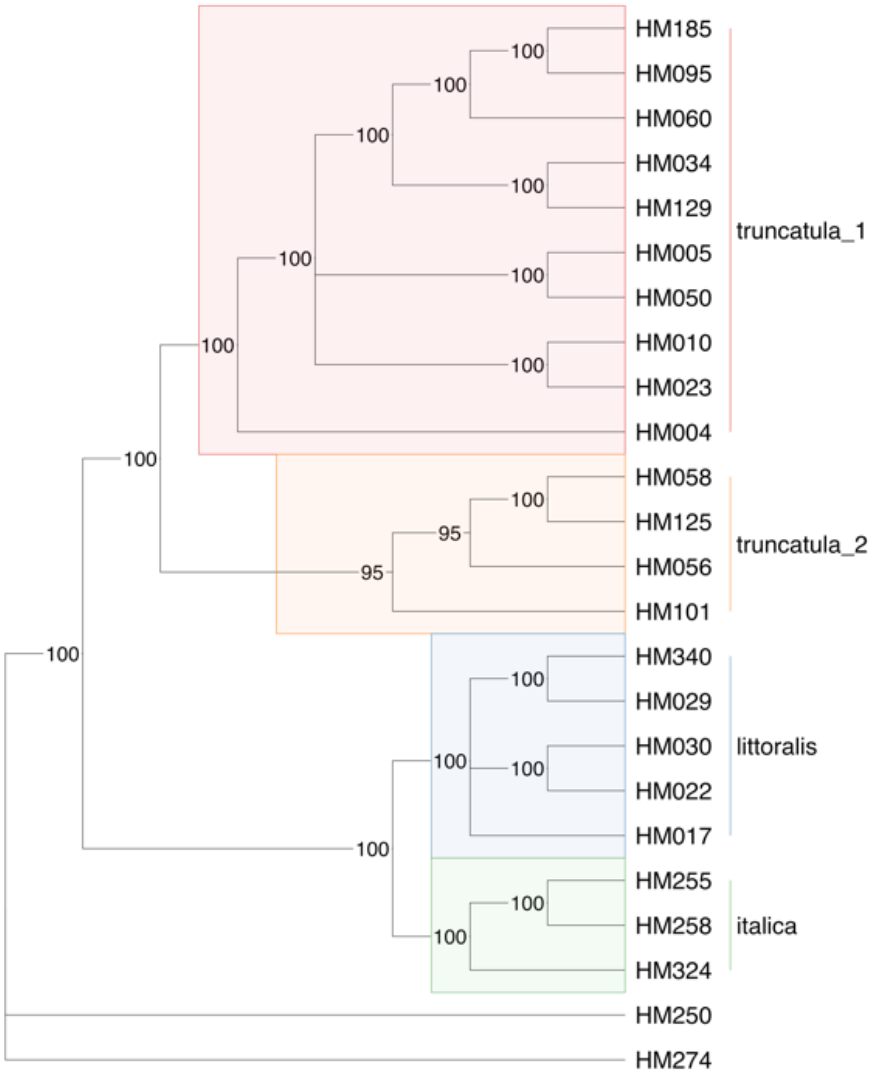
Table 1. Number of polymorphic sites in the nuclear SNP dataset by chromosome, proportion of sites missing data, mean coverage (number of reads per site), standard deviation of coverage, minimum coverage (depth), and maximum coverage (depth), by chromosome.

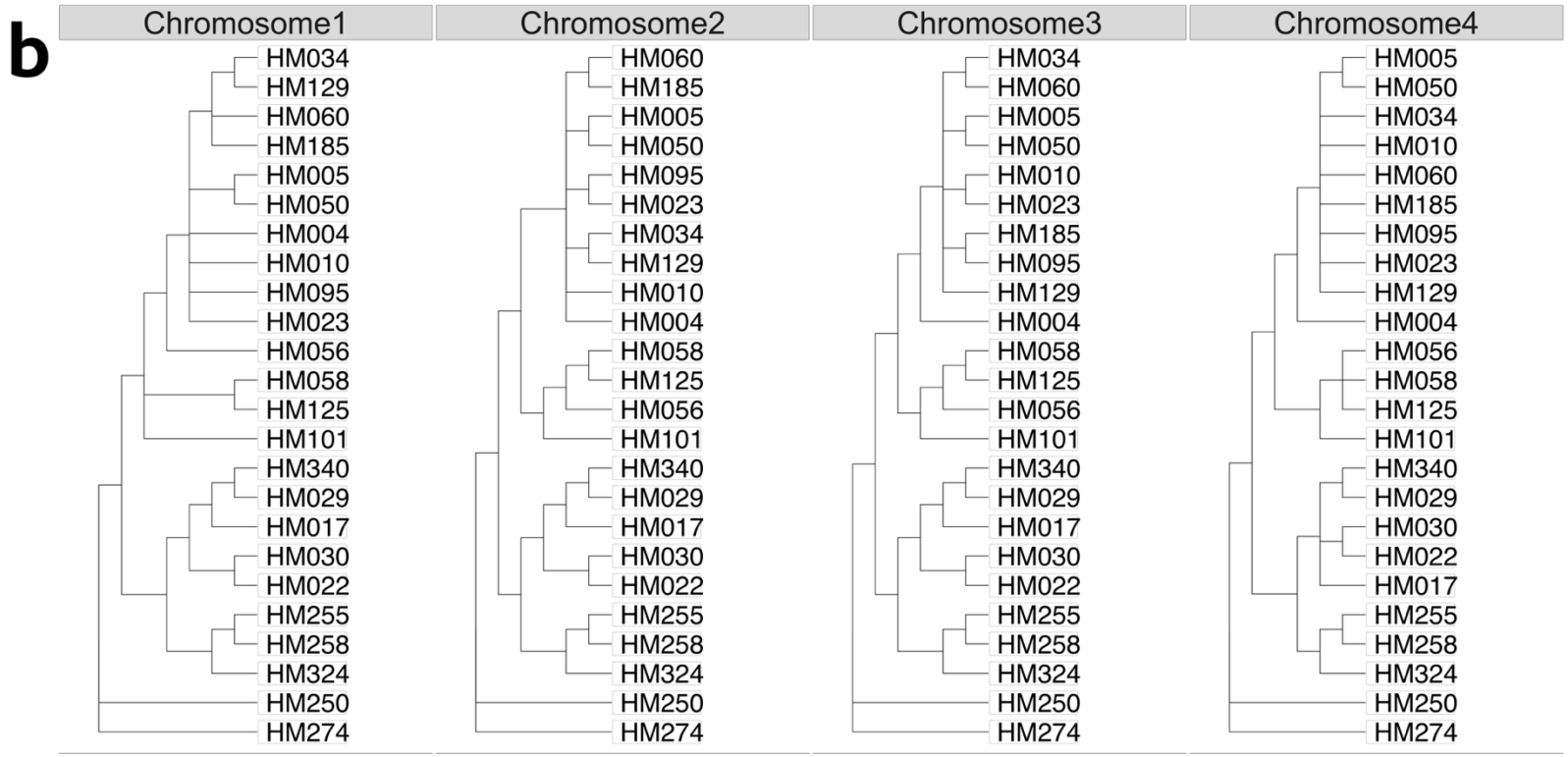
Chromosome	Number of Sites	Proportion of Sites Missing Data	Coverage Mean (number of reads per site)	Coverage Standard Deviation	Minimum Depth	Maximum Depth
1	9596	0.12	49.78	40.39	5	200
2	7444	0.12	49.76	40.51	5	200
3	8765	0.12	50.06	40.64	5	200
4	10207	0.12	49.99	40.56	5	200
5	7734	0.12	50.09	40.69	5	200
6	4220	0.12	50.05	40.85	5	200
7	7913	0.12	50.04	40.60	5	200
8	7894	0.12	49.86	40.52	5	200
Total	**63773**	**0.12**	**49.95**	**40.57**	**5**	**200**

Table 2. Synteny of genes across the breakpoints of the A17 translocation between chromosomes 4 and 8 in *Medicago truncatula*.. Syntenic pairs are highlighted in yellow. Genes are syntenic across the same breakpoint in A17. Genes are syntenic across opposite breakpoints for R108_HiC_1, HM056, HM058, and HM125.

	Gene Upstream	Gene Downstream
breakpoint 1	**MtrunA17_Chr8g0375101**	**MtrunA17_Chr8g0375151**
A17	Chromosome 8	Chromosome 8
HM056	scaffold_267	scaffold_139
HM058	scaffold_030	scaffold_1
HM125	scaffold_133	scaffold_102
R108_HiC_1	scaffold_8	scaffold_4
breakpoint 2	**MtrunA17_Chr4g0052941**	**MtrunA17_Chr4g0053011**
A17	Chromosome 4	Chromosome 4
HM056	scaffold_139	scaffold_267
HM058	scaffold_1	scaffold_30
HM125	scaffold_102	scaffold_133
R108_HiC_1	scaffold_4	scaffold_8

a


HM185
HM095
HM060
HM034
HM129
HM005
HM050
HM010
HM023
HM004
truncatula_1
HM058
HM125
HM056
HM101
truncatula_2
HM340
HM029
HM030
HM022
HM017
littoralis
HM255
HM258
HM324
italica
HM250
HM274
100
95


b
Chromosome1
HM034
HM129
HM060
HM185
HM005
HM050
HM004
HM010
HM095
HM023
HM056
HM058
HM125
HM101
HM340
HM029
HM017
HM030
HM022
HM255
HM258
HM324
HM250
HM274
Chromosome2
HM060
HM185
HM005
HM050
HM095
HM023
HM034
HM129
HM010
HM004
HM058
HM125
HM056
HM101
HM340
HM029
HM017
HM030
HM022
HM255
HM258
HM324
HM250
HM274
Chromosome3
HM034
HM060
HM005
HM050
HM010
HM023
HM185
HM095
HM129
HM004
HM058
HM125
HM056
HM101
HM340
HM029
HM017
HM030
HM022
HM255
HM258
HM324
HM250
HM274
Chromosome4
HM005
HM050
HM034
HM010
HM060
HM185
HM095
HM023
HM129
HM004
HM056
HM058
HM125
HM101
HM340
HM029
HM030
HM022
HM017
HM255
HM258
HM324
HM250
HM274

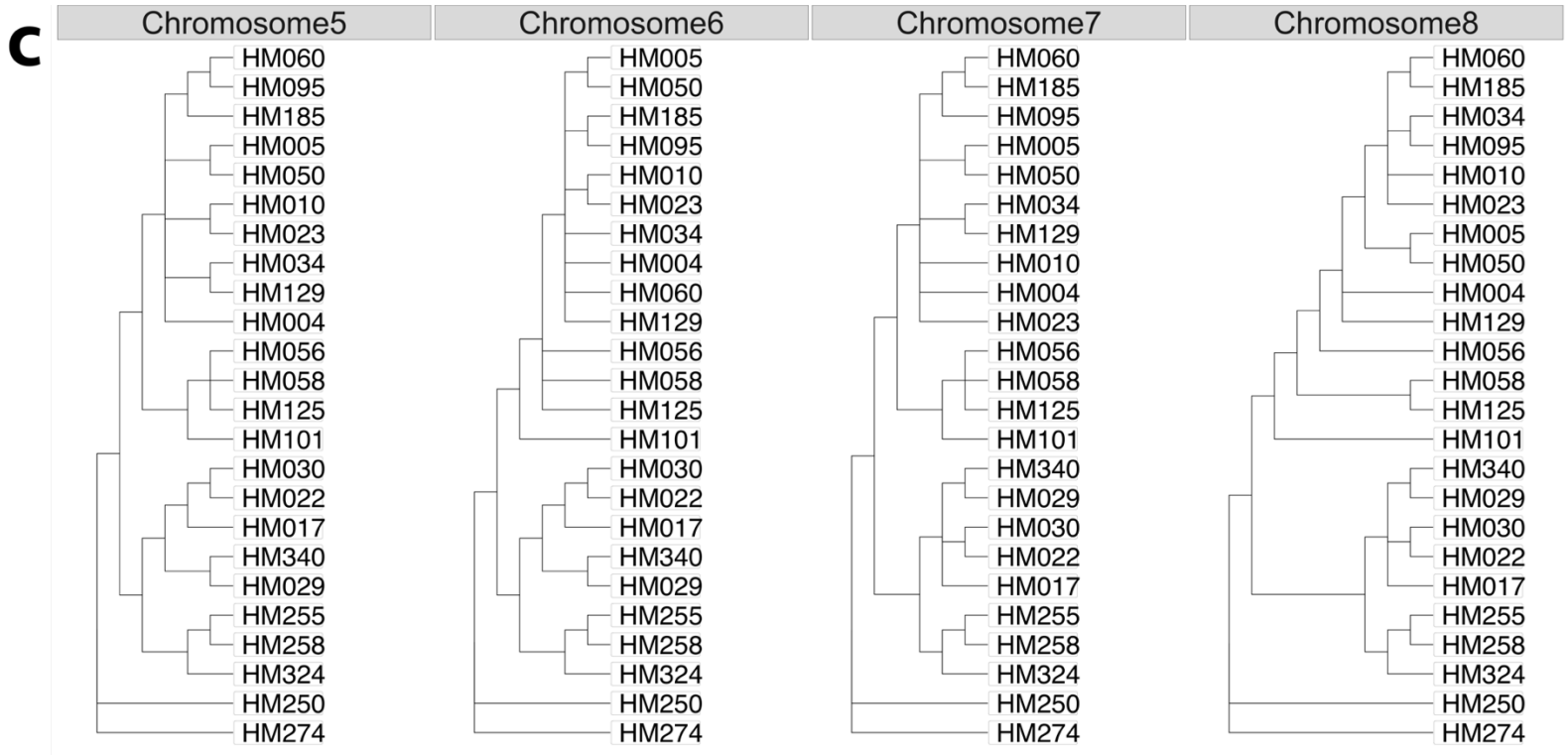
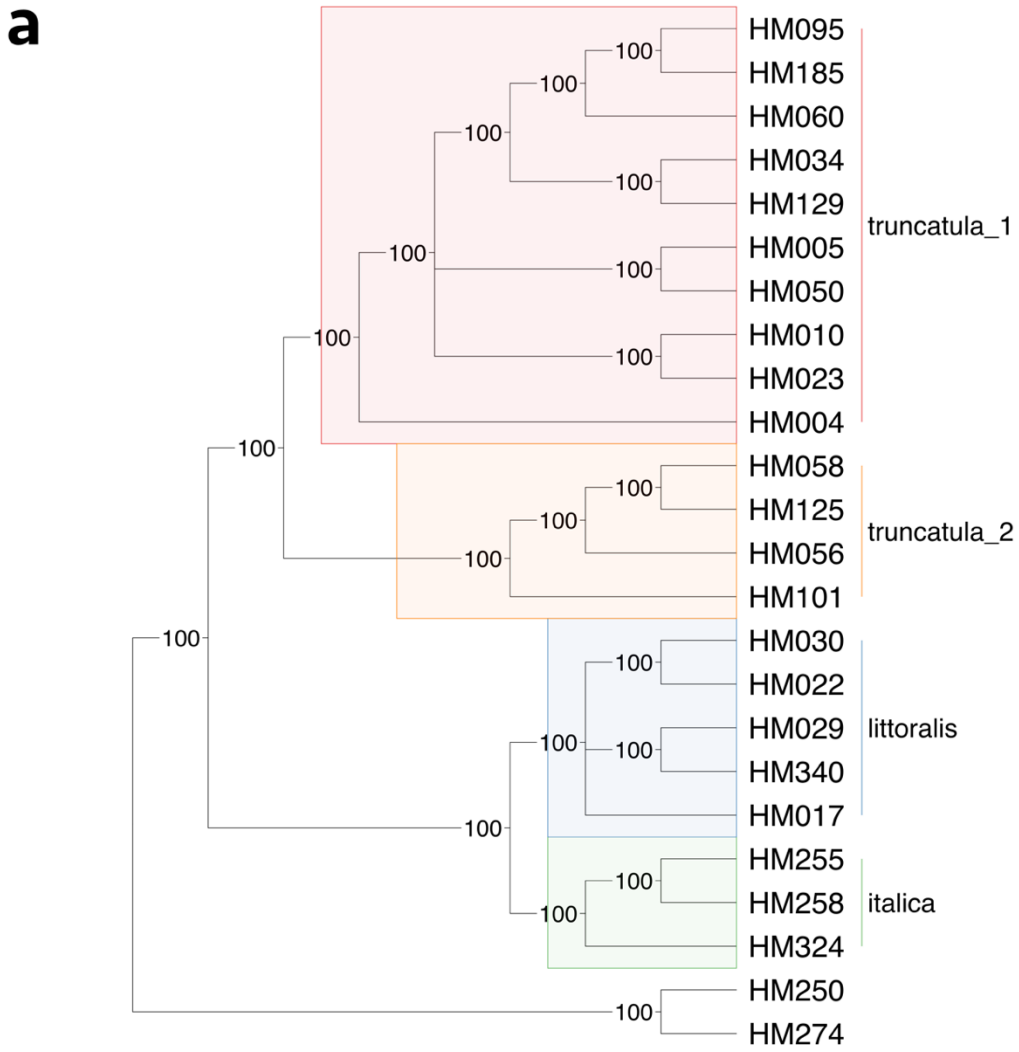
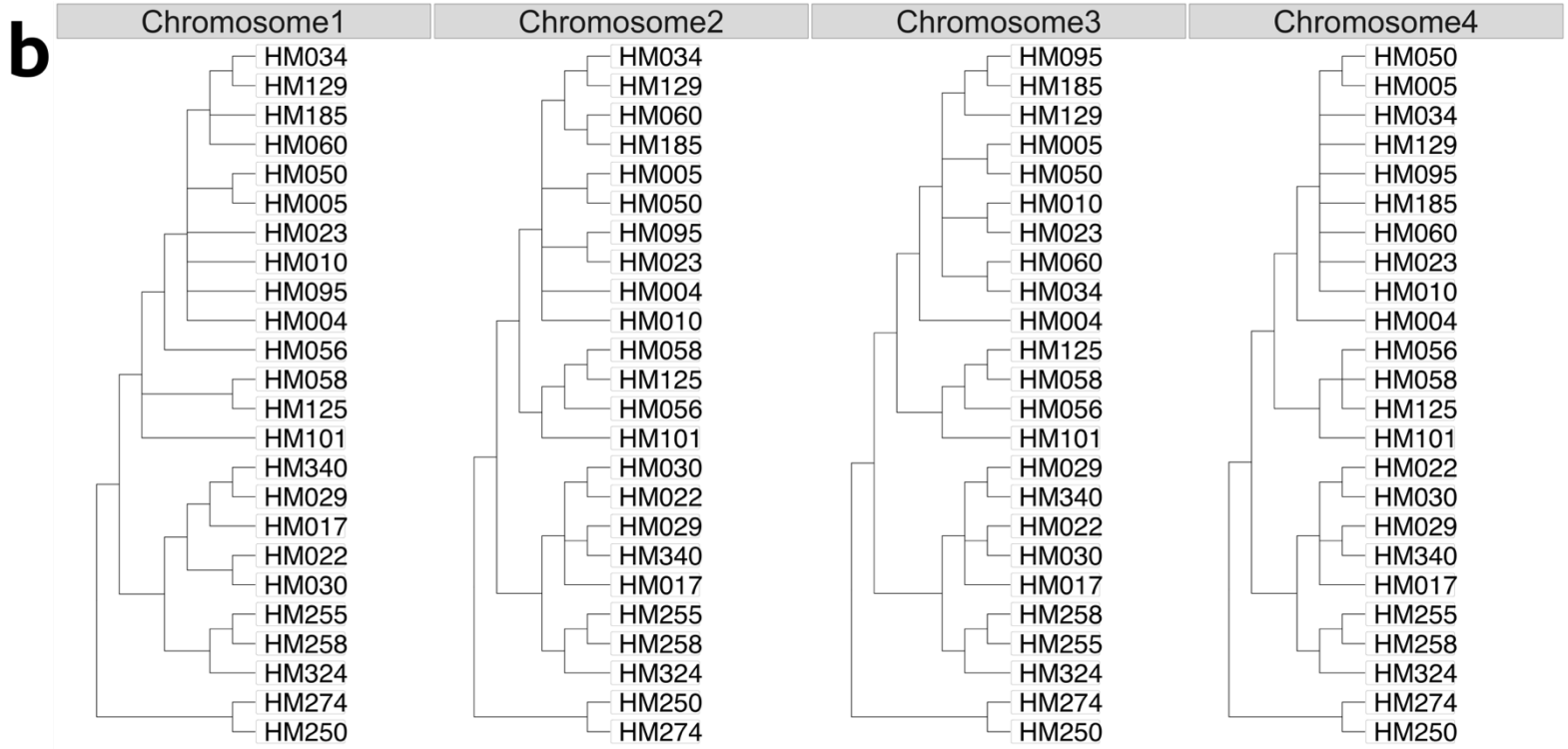


Figure 1. (**a**) Phylogenetic relationships within *Medicago* subsect. *Pachyspireae* based on maximum parsimony analysis of the concatenated, nuclear SNP dataset. Tree shown is an 80% majority-rule consensus of 100 bootstrap replicates. Numbers along branches indicate bootstrap support. The *M. truncatula* subclade formed by A17 (HM101) and three additional accessions of *M. truncatula* is highlighted as "truncatula_2". The clade formed by *M. littoralis* (HM030), R108 (HM029 and HM340), and two highly divergent *M. truncatula* (HM022 and HM017) is highlighted as "littoralis". The clade formed by *M. italica* and two highly divergent *M. truncatula* (HM255 and HM258) is highlighted as "italica". (**b-c**) Maximum parsimony analyses of nuclear SNP data by chromosome. Trees shown are 80% majority-rule consensus trees of bootstrap replicates. Nodes with less than 80% bootstrap support are collapsed.


a
HM095
HM185
HM060
HM034
HM129
HM005
HM050
HM010
HM023
HM004
truncatula_1
HM058
HM125
HM056
HM101
truncatula_2
HM030
HM022
HM029
HM340
HM017
littoralis
HM255
HM258
HM324
italica
HM250
HM274
100


b
Chromosome1
HM034
HM129
HM185
HM060
HM050
HM005
HM023
HM010
HM095
HM004
HM056
HM058
HM125
HM101
HM340
HM029
HM017
HM022
HM030
HM255
HM258
HM324
HM274
HM250
Chromosome2
HM034
HM129
HM060
HM185
HM005
HM050
HM095
HM023
HM004
HM010
HM058
HM125
HM056
HM101
HM030
HM022
HM029
HM340
HM017
HM255
HM258
HM324
HM250
HM274
Chromosome3
HM095
HM185
HM129
HM005
HM050
HM010
HM023
HM060
HM034
HM004
HM125
HM058
HM056
HM101
HM029
HM340
HM022
HM030
HM017
HM258
HM255
HM324
HM274
HM250
Chromosome4
HM050
HM005
HM034
HM129
HM095
HM185
HM060
HM023
HM010
HM004
HM056
HM058
HM125
HM101
HM022
HM030
HM029
HM340
HM017
HM255
HM258
HM324
HM274
HM250

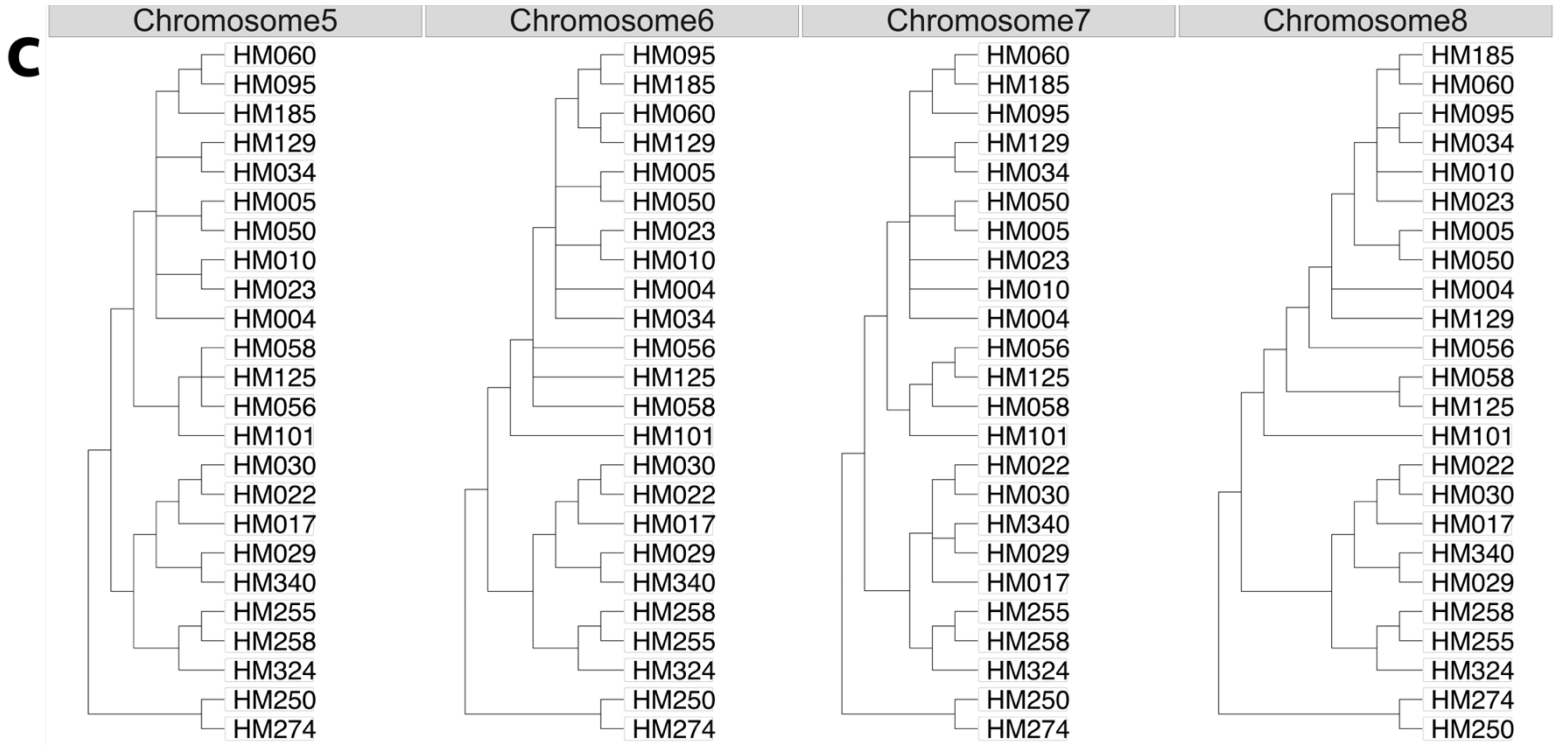


Figure 2. (**a**) Phylogenetic relationships within *Medicago* subsect. *Pachyspireae* based on maximum likelihood analysis of the concatenated, nuclear SNP dataset. Tree shown is the best-scoring maximum likelihood tree (Final ML Optimization Likelihood: -521902.9) with estimated branch support from 50 bootstrap replicates. The *M. truncatula* subclade formed by A17 (HM101) and three additional accessions of *M. truncatula* is highlighted as "truncatula_2". The clade formed by *M. littoralis* (HM030), R108 (HM029 and HM340), and two highly divergent *M. truncatula* (HM022 and HM017) is highlighted as "littoralis". The clade formed by *M. italica* and two highly divergent *M. truncatula* (HM255 and HM258) is highlighted as "italica". (**b-c**) Maximum likelihood analysis of nuclear SNP data by chromosome. Trees shown are best-scoring maximum likelihood trees. Nodes with less than 80% bootstrap support are collapsed.

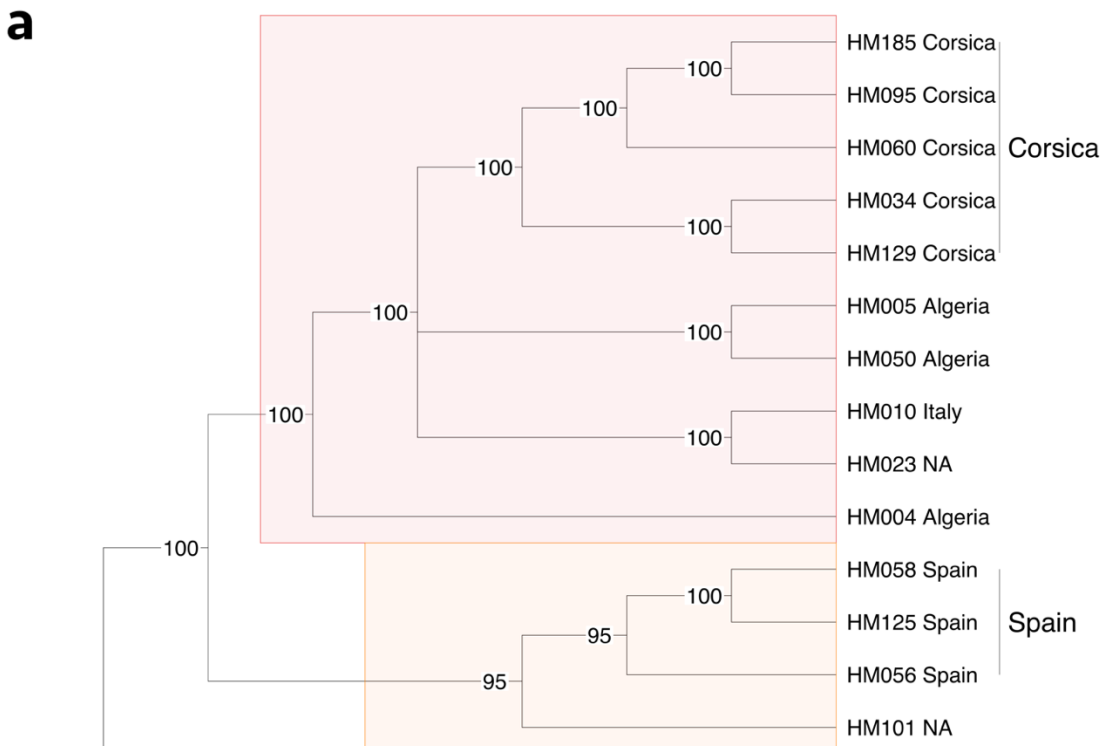

a
HM185 Corsica
HM095 Corsica
HM060 Corsica
HM034 Corsica
HM129 Corsica
Corsica
HM005 Algeria
HM050 Algeria
HM010 Italy
HM023 NA
HM004 Algeria
HM058 Spain
HM125 Spain
HM056 Spain
HM101 NA
Spain
100
100
100
100
100
100
100
100
100
100
95
95

Figure 3. (**a**) Phylogenetic relationships within the *Medicago truncatula* clade based on maximum parsimony analysis (see Figure 1). (**b**)Tip labels are mapped based on coordinate data published by Ronfort et al. (2006). A subclade formed by accessions collected in Corsica is highlighted and colored red. A subclade formed by accessions collected in Spain and A17 (HM101) is highlighted and colored in orange.

CHAPTER 2

EVALUATING GERMINATION SALT-TOLERANCE WITHIN AND BETWEEN *MEDICAGO TRUNCATULA* AND CLOSE RELATIVES

Introduction

Saline and salt-affected soils are found in over 100 countries, affecting an estimated one billion hectares (ha) worldwide (FAO-ITS-GSP, 2015). High levels of salinity lead to osmotic pressure, ion toxicity, and oxidative stress; affecting plants at all stages of development (Rengasamy, 2010; Liang et al., 2018). Mechanisms of soil salinization are categorized as primary (natural causes) or secondary (anthropogenic causes) (Hassani et al., 2021). Primary mechanisms of salinization include natural rock weathering and oceanic salt deposition through rainfall and wind (FAO-ITPS-GSP, 2015; Hassani et al., 2021). Secondary mechanisms of salinization include: excessive fertilizer application, irrigation with saline water, seawater intrusion, and overexploitation of underground freshwater (Rengasamy, 2010; FAO-ITPS-GSP, 2015; Eswar et al., 2021; Hassani et al., 2021). Secondary soil salinization currently affects 76 million ha worldwide, however human-induced salinization is expected to increase with climate change and rising surface temperatures (FAO-ITS-GSP, 2015; Hassani et al., 2021).

Climate models predict that the land area experiencing extreme-to-exceptional drought will more than double, from 3% to 7%, over the next century (Pokhrel et al., 2021). Increased evapotranspiration and reduced rainfall lead to an accumulation of salts in upper soil layers (FAO-ITPS-GSP, 2015; Eswar et al., 2021; Hassani et al., 2021). Drought conditions increase reliance on brackish and waste-water for irrigation, further

increasing soil salinization (Eswar et al., 2021; Hassani et al., 2021). Salinization by seawater intrusion is expected to increase as an estimated 70% of coastlines experience sea-level rise over the next century (Eswar et al., 2021).

Approximately 1.5 million ha of farmland are taken out of production each year due to problems arising from salinity, with an additional 20-45 million ha losing production potential (FAO-ITPS-GSP, 2015). As of 2013, losses in crop yield due to salinity was estimated to be \$441 ha^{-1}, totaling \$27 billion lost in production annually (Qadir et al., 2014).

Established Salt Tolerance in **Medicago truncatula** *and relatives*

One strategy to address soil salinization is to identify mechanisms of salt-tolerance and to develop more salt tolerant crops. An array of salt-tolerant species, populations, and cultivars of the legume genus *Medicago* L. ("alfalfa" and relatives) have been identified through laboratory, greenhouse, and field experiments. For example, Monirifar & Barghi (2009) tested five alfalfa (*M. sativa*) cultivars,identified two salt-tolerant cultivars 'Malekan' 'Ahar-Hourand', that showeded minimal reductions in forage yield and chlorophyll content under salt stress. Scasta et al. (2012) tested 12 alfalfa cultivars, identified one salt-tolerant cultivar (*M. sativa* 'Salado'), and demonstrateded a higher mean germination percentage compared to all other alfalfa cultivars under high salt-stress. Sarri et al. (2021) report that *M. arborea* is more tolerant to 100 mM NaCl application than *M. sativa*, showing a 10% reduction in growth size compared to 20% in alfalfa. Guan et al. (2009) demonstrated the salt-tolerance of wild *M. ruthenica* seed,

showing consistent germination percentages up to 100 mM treatments with NaCl, and roughly 50% reduction in germination percentage under 200 mM treatment with NaCl.

Extensive research has been conducted on *Medicago* annuals growing in Tunisia, where approximately 9% of the total surface and 25% of the cultivated area is affected by salinity (Aloui et al., 2022). The natural salinity gradient provides an excellent system for researchers to search for salt-tolerant populations (Lazrek et al., 2009; Arraouadi et al., 2011; Castro et al., 2013; Cordeiro et al., 2014; Friesen et al., 2014; Mbarki et al., 2020; Aloui et al., 2022). Cordeiro et al. (2014) demonstrate that Tunisian accessions of *Medicago truncatula*, originating from saline environments, tolerate saline treatments during germination better than accessions originating from non-saline environments, consistent with results presented by Friesen et al. (2014). Mbarki et al. (2020) evaluated nine wild populations of *Medicago*, representing three species (*M. ciliaris*, *M. scutellata*, and *M. intertexta*). Three populations of *M. ciliaris* and one population of *M. intertexta* were identified as salt-tolerant according to germination percentages under 200mM treatment with NaCl (Mbarki et al., 2020). Khalil et al. (2011) report higher salt tolerance of *M. ciliaris* compared to *M. polymorpha* in germination under saline conditions.

Differences in salt-tolerance have also been reported for two important reference lines of *Medicago truncatula*: A17 and R108. Wang et al. (2014) conclude that R108 is more susceptible to mineral toxicity and deficiency than A17. Specifically, R108 experienced a greater reduction in K^+ concentration relative to A17 (Wang et al., 2014). Increasing the ratio of potassium to sodium ions in plant tissues to resist osmotic pressure is an important adaptive response to elevated (Wang et al., 2014).

Analysis of nuclear SNP data presented in Chapter 1 supports the hypothesis that line R108 and other deeply divergent accessions of *Medicago truncatula* are in fact more closely related to the sister species *Medicago littoralis* and *Medicago italica*. *Medicago littoralis* ("shore medic") and *M. italica* are found in moderate, coastal climates on well drained, sandy soils (Small, 2011). Moderate salt-tolerance has been reported for *M. littoralis* (Small, 2011). *M. truncatula* grows inland under a range of climatic conditions. Limited salt-tolerance has been reported for populations of *M. truncatula* growing in Tunisia. The goal of the present study is to collect phenotypic data (i.e. germination salt-tolerance) to further study the complicated identification of line R108 and other deeply divergent accessions of *M. truncatula*. I predict that the misidentified accessions will exhibit salt-tolerance more similar to *M. littoralis* than to other accessions of *M. truncatula.*

Original collection locations for many of the accessions sampled here are well-documented (Ronfort et al., 2006). It has been reported that populations of *M. truncatula* collected from environments with high soil salinity exhibit greater salt-tolerance than populations collected from environments with lower salinity (Cordeiro et al., 2014; Friesen et al., 2014). Using predicted soil salinity data published by Ivushin et al. (2019), I will also test the hypothesis that accessions collected from locations with high predicted salinity exhibit greater germination salt-tolerance than accessions collected from locations with lower predicted salinity.

Materials and Methods

Sampling

The present study includes 13 accessions from the Medicago HapMap germplasm. One accession represents *M. littoralis* (HM030), one accession represents *M. italica* (HM324), and the remaining accessions are typically treated as *M. truncatula* (HM017, HM022, HM029, HM023, HM050, HM095, HM129, HM185, HM058, HM101, and HM125) (https://medicago.legumeinfo.org). Phylogenetic analysis suggests that three of the accessions (HM017, HM022, and HM029) are more closely related to *M. littoralis* than they are to other accessions identified as *M. truncatula* (Branca et al., 2011; Yoder et al., 2012; Choi et al., 2022; Appendix B; Figure 4a).

Seed preparation

Seeds were manually removed from pods with forceps to avoid damage (Garcia et al., 2006). Obviously discolored and malformed seeds were removed from analysis at this stage (Supplementary Figure 1). After removal from pods, seeds were acid scarified following a modified procedure from Garcia et al. (2006). Seeds were submerged in concentrated, anhydrous sulfuric acid with intermittent agitation. Progress was monitored by observing for small black dots on the tegument surface, with seeds removed from the sulfuric acid after 19 minutes. After decanting the sulfuric acid, seeds were rinsed 3 times using chilled, distilled water.

Following scarification, seeds were sterilized using a 1% ZeroTol 2.0 solution (27.1% hydrogen peroxide and 2.0% peroxyacetic acid). Per documentation by the Office of

Chemical Safety and Pollution Prevention (2018), seeds were soaked in the ZeroTol 2.0 solution for two minutes and were not rinsed following sterilization.

Germination experiment

Scarified and sterilized seeds were immediately transferred to 100 mm petri dishes. Thirty seeds from each accession were divided evenly between a control group and two treatment groups. Seeds were placed between two pieces of 100 mm, autoclaved Whatman filter paper with 4.5 mL of the appropriate treatment solution. Autoclaved water was applied to seeds in the control group, 25 mM NaCl solution was applied to seeds in the first treatment group, and 50 mM NaCl solution was applied to seeds in the second treatment group. Petri dishes were sealed with Parafilm to maintain humidity. The sealed dishes were placed in a dark growth chamber and maintained at 25.0 ºC and monitored daily for germination. In a pilot experiment, it was noted that the filter paper went completely dry after five days. In all subsequent germination experiments, half of the original volume (2.25 mL) of the appropriate treatment solution was applied to each dish. After 10 days, ungerminated seeds were transferred to a recovery dish, where 4.5 mL of autoclaved water was applied. Germination in the recovery dishes was monitored daily over four days. Following the recovery period, germinated seeds were transferred to soil and grown to a small size (10-15 cm). The seedlings were pressed to produce voucher specimens from which DNA could be extracted at a future time if desired. The experiment was replicated twice, providing two observations for most accessions.

Determining experimental and environmental soil salinity

A common measure of soil salinity is electrical conductivity (EC). EC measures the ability of soil water to carry an electrical current, and is influenced by the concentration of dissolved ions such as sodium and chloride among others. A pH and electrical conductivity meter (HI9814; Hanna Instruments, Woonsocket, RI, USA) was used to measure the EC for the control and each treatment solutions.

Original collection coordinates for each accession, published by Ronfort et al. (2006), were used to create a point layer in QGIS (QGIS.org, 2023; Figure 4a). The original collection point layer was joined with published shapefiles containing soil salinity predictions for the years 2000, 2002, 2005, and 2009 (Ivushkin et al., 2019; Figure 4b). Soil salinity is categorized and scored by Ivushkin et al. (2019) as follows: non-saline (EC <2 $ms \cdot cm^{-1}$, score=0), slightly saline (EC 2-4 $ms \cdot cm^{-1}$, score=1), moderately saline (EC 4-8 $ms \cdot cm^{-1}$, score=2), highly saline (EC 8-16 $ms \cdot cm^{-1}$, score=3), and extremely saline (EC >16 $ms \cdot cm^{-1}$, score=4). The average salinity score was recorded for each of the accessions (Appendix B). Accessions with an average score greater than zero were placed into a category of high salinity, while accessions with an average score of zero were placed into a category of low salinity (Appendix B). In other words, accessions that scored as saline in at least one year were grouped together in the high salinity category.

Measures of germination success

Seeds were counted as germinated when radicle length met or exceeded 3 mm. Seeds that did not germinate after four days of recovery were treated as inviable. The number of inviable seeds for each replicate was subtracted from 10 to calculate a corrected total

germination number (Rumbaugh, 1991). Final germination proportion was calculated by dividing the number of germinated seeds by the corrected total (Rumbaugh, 1991). Germination rate was calculated following a modified formula from Esechie (1994) and Khalil et al. (2011). The index was calculated as follows:

$$\textit{Germination Rate Index} = (G1 / 1) + (G2 / 2) + \ldots + (Gx / x)$$

where *Gx* is the proportion of seeds that germinated *x* days after sowing.

Finally, germination inhibition was calculated by dividing the proportion of seeds that germinated under a given treatment by the proportion of seeds that germinated under control conditions of the same replication (Rumbaugh, 1991). Germination proportions, rates, and inhibition reported in Appendix B represent the mean between two replicates.

Results

The measured EC of the 0 mM NaCl (control), 25 mM NaCl, and 50 mM NaCl solutions were 0.03 $ms \cdot cm^{-1}$, 2.74 $ms \cdot cm^{-1}$, and 5.42 $ms \cdot cm^{-1}$, respectively.

Differences by treatment

The mean germination percentage under control conditions was 94% (Table 3a). The mean germination percentage of seeds treated with 25 mM and 50 mM NaCl was 91% and 75%, respectively (Table 3a). The difference in germination percentage between the control and 25 mM NaCl treatment was not significant (Table 3b). Mean germination percentage was significantly lower in the 50 mM NaCl treatment groups compared to both the control and 25 mM treatment groups ($p<0.05$) (Table 3b; Figure 5a).

The mean germination rate was 1.44 under control conditions, 0.77 in the 25 mM NaCl treatment group, and 0.53 in the 50 mM NaCl treatment group (Table 3a). Differences in germination rates were significant between all treatment groups ($p<0.01$) (Table 3b; Figure 5b). Mean germination inhibition was 0.98 in the 25 mM NaCl treatment group and 0.80 in the 50 mM NaCl treatment group (Table 3a). Inhibition was significantly lower in the 25 mM treatment group compared to the 50 mM treatment group ($p<0.05$) (Table 3b; Figure 5c).

The mean percentage of seeds that germinated in recovery was 11.1% for the control group, 32.5% for the 25 mM NaCl treatment groups, and 38.7% for the 50m M NaCl treatment groups (Table 3a). The difference in germination proportion under recovery conditions was not significant between treatment groups (Table 3b).

Differences by environmental salinity

Accessions collected from environments with low predicted soil salinity had a mean germination proportion of 0.90 under control conditions, 0.91 under treatment with 25 mM NaCl, and 0.70 under treatment with 50 mM NaCl (Table 4a). Accessions collected from environments with high predicted soil salinity had mean germination proportions of 0.95 under control conditions, 0.89 under treatment with 25 mM NaCl, and 0.76 under treatment with 50 mM NaCl (Table 4a). Differences in mean germination proportion were not significant between accessions collected from environments with low versus high predicted soil salinity (Table 4b; Figure 6a).

The mean germination rates of accessions collected from environments with low predicted soil salinity were 1.42, 1.00, and 0.59 in the control, 25 mM NaCl, and 50 mM

NaCl treatment groups, respectively (Table 4a). The mean germination rates of accessions collected from environments with high predicted soil salinity were 1.36, 0.59, and 0.48 in the control, 25 mM NaCl, and 50 mM NaCl treatment groups, respectively (Table 4a). Differences in germination rate between accessions collected from low versus high predicted soil salinity were not significant in the control or the 50 mM NaCl treatment groups (Table 4b). However, germination rates in the 25 mM NaCl treatment group were significantly higher for accessions collected from regions with low predicted soil salinity compared to accessions collected from regions with high predicted salinity (Table 4b; Figure 6b).

Mean germination inhibition for accessions collected from locations with low predicted soil salinity was 1.01 and 0.79 in the 25 mM and 50 mM NaCl treatment groups, respectively (Table 4a). Mean germination inhibition for accessions collected from high predicted soil salinity was 0.96 and 0.78 in the 25 mM and 50 mM NaCl treatment groups, respectively (Table 4a). Differences in germination inhibition between accessions collected from locations with low versus high predicted soil salinity were not significant in either treatment group (Table 4b, Figure 6c).

Differences by clade

The germination proportions for the accession representing *Medicago italica* (HM324) was: 0.95 in the control group, 0.70 in the 25 mM NaCl treatment group, and 0.90 in the 50 mM treatment group (Table 5a). The mean germination proportion for the accessions assigned to the 'littoralis' clade (HM030, HM029, HM017, and HM022) was: 0.91 in the control group, 0.89 in the 25 mM NaCl treatment group, and 0.49 in the 50 mM NaCl

treatment group (Table 5a). The mean germination proportion for accessions assigned to the 'truncatula_1' clade (HM023, HM050, HM095, HM129, and HM185) was: 0.94 in the control group, 0.93 in the 25 mM NaCl treatment group, and 0.81 in the 50m M treatment group (Table 5a). The mean germination proportion for assigned to the 'truncatula_2' clade (HM058, HM101, and HM125) was: 0.97 in the control group, 1.00 in the 25 mM NaCl treatment group, and 0.95 in the 50 mM NaCl treatment group (Table 5a).

The mean germination rates for the accession representing *M. italica* were 1.88, 0.83, and 0.58 in the control, 25 mM, and 50 mM NaCl treatment groups, respectively (Table 5a). The mean germination rates for accessions assigned to the 'littoralis' clade were 1.21, 0.58, and 0.30 in the control, 25 mM, and 50 mM NaCl treatment groups, respectively (Table 5a). The mean germination rates for accessions assigned to the 'truncatula_1' clade were 1.68, 0.90, and 0.56 in the control, 25 mM, and 50 mM NaCl treatment groups, respectively (Table 5a).

Mean germination inhibition for the accession representing *M. italica* was 0.73 in the 25 mM NaCl treatment group and 1.00 in the 50 mM NaCl treatment group (Table 5a). Mean germination inhibition for accessions assigned to the 'littoralis' clade was 0.98 in the 25 mM NaCl treatment group and 0.50 in the 50 mM treatment group (Table 5a). Mean germination inhibition for accessions assigned to the 'truncatula_1' clade was 1.00 in the 25 mM NaCl treatment group and 0.88 in the 50 mM NaCl treatment group (Table 5a). Mean germination inhibition for accessions assigned to the 'truncatula_2' clade was

1.02 in the 25 mM NaCl treatment group and 0.98 in the 50 mM treatment group (Table 5a).

Insufficient data were collected to test for significance between 'italica' and other clades. There was no significant difference in germination proportion, germination rate, or germination inhibition between the 'truncatula_1' and 'truncatula_2' clades (Table 5b; Figure 7). The difference in germination proportion between 'littoralis' and 'truncatula_1' in the 25 mM NaCl treatment group was weakly significant ($p<0.1$) (Table 5b; Figure 7). Accessions assigned to the 'littoralis' clade had significantly lower germination proportions and mean germination rates in the 50 mM NaCl treatment group than accessions assigned to the 'truncatula_2' clade ($p<0.05$) (Table 5b; Figure 7). Accessions assigned to the 'littoralis' clade demonstrated a significantly higher level of germination inhibition compared to the 'truncatula_2' clade in the 50 mM NaCl treatment group ($p<0.1$) (Table 5b; Figure 7).

Discussion

Proposed mechanism of inhibition

The percentage of seeds that germinated in recovery was not significantly different between the control and treatment groups. This suggests that treatment with 25mM and 50mM NaCl solution did not change seed viability. Successful germination in recovery suggests that inhibition was due to osmotic pressure (reversible) rather than ion toxicity (irreversible) (Khalil et al., 2011). It should be noted that the number of observations in recovery were limited, as not every sample had ungerminated seeds after 10 days.

Effect of environmental soil salinity on salt-tolerance

Final germination proportions and germination inhibition were seemingly unaffected by predicted environmental salinity (Table 3; Figure 6). However, accessions collected from regions with low predicted salinity had significantly faster germination rates than accessions collected from regions with high predicted salinity (Table 3; Figure 6). This difference was only observed in the 25 mM NaCl treatment group. This result may be explained by the delayed germination of salt-tolerant plants to avoid seedling mortality when salinity is high (Khalil et al., 2011; Cordeiro et al., 2014). However, it is unclear then why this trend is lost in the 50 mM NaCl treatment group. Alternatively, increased germination percentages and rates in *Medicago* have been demonstrated by Mbarki et al. (2020) using 25 mM NaCl treatment. However, within accessions collected from environments with low predicted salinity, germination rates were higher under control conditions than in the 25 mM NaCl treatment group. A third explanation for this result may be limited resolution of predicted soil salinity. Ivushkin et al. (2019) predicted global soil salinity based on thermal infrared imagery combined with auxillary soil data (pH, organic content, texture, and water retention). The authors acknowledge that the model has difficulty estimating degrees of salinity, though the model is useful in distinguishing between salt-affected and non-affected lands (Ivushkin et al., 2019). However, the model is limited by a 250 m resolution, meaning that the precise salinity at each collection location may be lost (Ivushkin et al., 2019).

Difference in germination salt-tolerance between clades

Mean germination proportions, rates, and inhibition were significantly different between accessions assigned to the 'littoralis' clade and accessions assigned to the 'truncatula_2' clade (Table 4; Figure 7, Figure S2). This result is consistent with the hypothesis that HM022 and HM017 are more closely related to *M. littoralis* than to other accessions of *M. truncatula*. *Medicago littoralis* and *Medicago italica* are expected to demonstrate greater salt-tolerance than *Medicago truncatula*. The significantly decreased germination proportion and rate, combined with increased inhibition may be explained by an adaptive response to delay germination under saline conditions to prevent seedling mortality (Khalil et al., 2011; Cordeiro et al., 2014). It should be noted that the germination proportion, rate, and inhibition for line R108 (HM029) are outliers within the 'littoralis' clade. This may be explained by presence of heavy mold in the 50 mM treatment of HM029, which prevented proper replication.

Limitations

The *in vitro* germination experiments presented here do not capture the complex soil dynamics found in the field. Scasta et al. (2012) report that field results do not necessarily correlate with laboratory and greenhouse experiments using alfalfa. Ion composition is one explanation offered for incongruent laboratory and field results. While sodium and chloride are predominantly found in saline soils, the presence and effect of magnesium, calcium, sulfate, and carbonate cannot be ignored (Rengasamy, 2010). The presence of gypsum, for example, significantly decreases the solubility of NaCl (Rengasamy, 2010).

Transgenerational effects of parental environment are also unaccounted for in the present study. There is evidence that parental exposure to salinity induces changes in the response of offspring to salinity (Castro et al., 2013; Cordeiro et al., 2014). In the present study, the seeds evaluated for salt tolerant germination came directly from the HapMap collection and parental exposure to salinity, or other abiotic stresses, are unknown.

Conclusion

Climate change is expected to increase both the natural and cultivated land area affected by soil salinity, threatening ecosystem stability and food security (FAO-ITPS-GSP, 2015; Hassani et al., 2021; Pokhrel et al., 2021). Along with more sustainable agricultural practices, it is becoming increasingly important to identify salt-tolerant crops and crop wild relatives. Germination is one of the first traits exposed to natural selection and is a valuable phenotype in screening for salt-tolerance (Cordeiro et al., 2014). In the present study, accessions collected from regions with high predicted soil salinity demonstrated a lower germination rate than accessions collected from regions with low predicted soil salinity. This result may be explained by an adaptive response to delay germination under saline conditions to avoid seedling mortality. This result provides tentative support for the hypothesis that environmental soil salinity influences germination rates under saline conditions for accessions found in the *Medicago* HapMap germplasm. The present study provides phenotypic evidence to support the hypothesis that HM017 and HM022 are likely misidentified as *M. truncatula*, as their response to germination under saline conditions is more similar to *M. littoralis* than to other accessions of *M. truncatula.* This conclusion is consistent with the phylogenetic analyses presented in Chapter 1, which

indicate that line R108 and other deeply divergent accessions of *M*. *truncatula* are more closely related to *M*. *littoralis* and *M*. *italica*.

Table 3a. Mean germination proportion, mean germination rate, mean germination inhibition, mean germination recovery, and the number of observations for each treatment group (0 mM, 25 mM, and 50 mM NaCl).

Treatment (mM NaCl)	Mean Germination Proportion	Mean Germination Rate	Mean Germination Inhibition	Mean Germination Proportion in Recovery	Number of Observations
0	0.94	1.44	1.00	0.11	26
25	0.91	0.77	0.98	0.33	26
50	0.75	0.53	0.80	0.39	23

Table 3b. Significance testing between treatment groups for mean germination proportion and mean germination rate (**p<0.05 and *** p <0.01).

Measure	p-value	test
0 mM to 25 mM		
Germination Proportion	0.867	Wilcox
Germination Rate	4.02E-06***	t-test
Germination Recovery	0.267	Wilcox
0 mM to 50 mM		
Germination Proportion	0.012**	Wilcox
Germination Rate	1.15E-08***	t-test
Germination Recovery	0.157	Wilcox
25 mM to 50 mM		
Germination Proportion	0.019**	Wilcox
Germination Rate	0.00793***	t-test
Germination Recovery	0.696	Wilcox

Table 4a. Mean germination proportion, mean germination rate, mean germination inhibition, and the number of observations for accessions collected from environments with low vs high predicted soil salinity and corresponding treatment (0 mM, 25 mM, and 50 mM NaCl). Germination inhibition is not shown for the control group (0 mM NaCl), as the control group serves as the denominator in this calculation (not applicable, NA).

Environmental Soil salinity	Treatment (mM NaCl)	Mean Germination Proportion	Mean Germination Rate	Mean Germination Inhibition	Number of Observations
high	0	0.95	1.36	NA	11
high	25	0.89	0.59	0.96	11
high	50	0.76	0.48	0.78	10
low	0	0.90	1.42	NA	10
low	25	0.91	1.00	1.01	10
low	50	0.70	0.59	0.79	10

Table 4b. Significance testing for differences between mean germination proportion, rate, and inhibition for accessions collected from environments with low versus high predicted soil salinity and corresponding treatment group (***p<0.01).

Treatment (mM NaCl)	Measure	p-value	test
0	Germination Proportion	0.3115	Wilcox
25	Germination Proportion	0.8805	Wilcox
50	Germination Proportion	0.1918	Wilcox
0	Germination Rate	0.5730	Wilcox
25	Germination Rate	0.0030***	Tukey
50	Germination Rate	0.4490	Tukey
0	Germination Inhibition	NA	NA
25	Germination Inhibition	0.6109	Wilcox
50	Germination Inhibition	0.4819	Wilcox

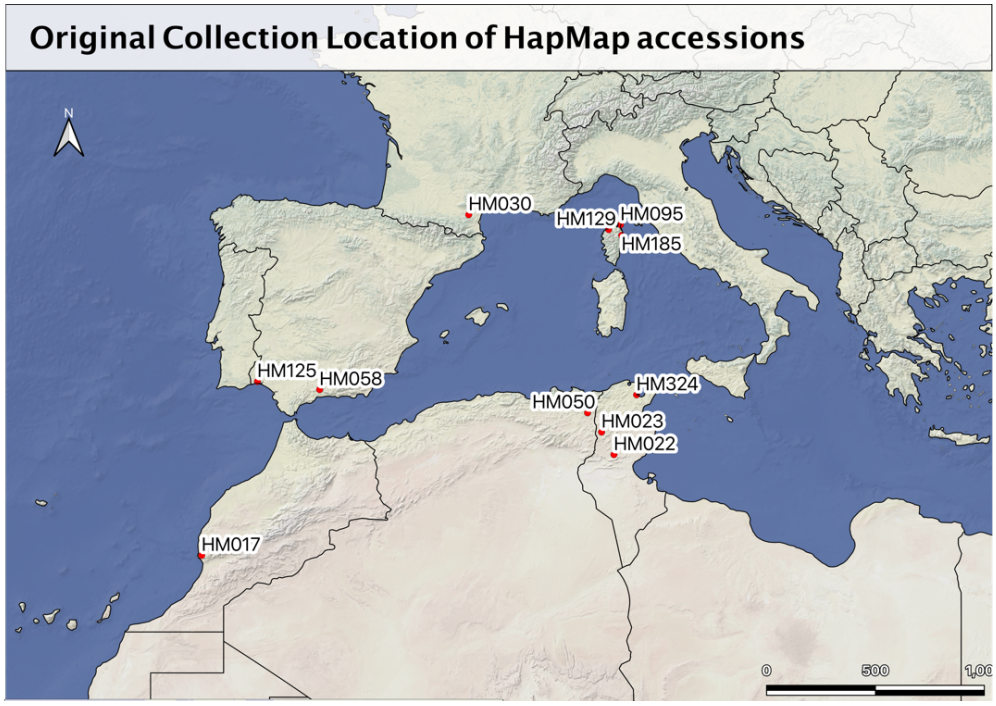
Table 5a. Mean germination proportion, germination rate, germination inhibition, and the number of observations for accessions assigned to each of four clades ('italica', 'littoralis', 'truncatula_1', and 'truncatula_2') and corresponding treatment groups (0 mM, 25 mM, and 50 mM NaCl).

Clade	Treatment (mM NaCl)	Mean Germination Proportion	Mean Germination Rate	Mean Germination Inhibition	Number of Observations
italica	0	0.95	1.88	NA	2
italica	25	0.70	0.83	0.73	2
italica	50	0.90	0.58	1.00	1
littoralis	0	0.91	1.21	NA	8
littoralis	25	0.89	0.58	0.98	8
littoralis	50	0.49	0.30	0.50	7
truncatula_1	0	0.94	1.68	NA	10
truncatula_1	25	0.93	0.90	1.00	10
truncatula_1	50	0.81	0.56	0.88	9
truncatula_2	0	0.97	1.18	NA	6
truncatula_2	25	1.00	0.77	1.02	6
truncatula_2	50	0.95	0.73	0.98	6

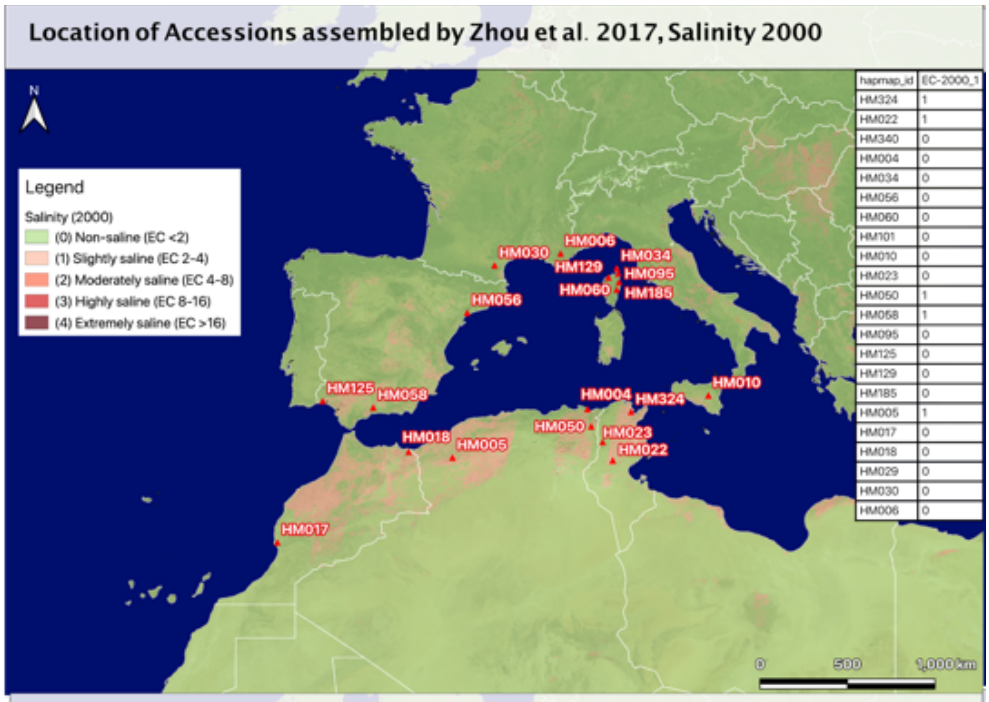
Table 5b. Significance testing for differences in mean germination proportion, rate, and inhibition between accessions assigned to each clade (*p<0.1 and **p<0.05)

Treatment (mM NaCl)	Measure	p-value	Method
littoralis to truncatula_1			
0	Germination Percentage	0.830	Wilcox
0	Germination Rate	0.141	Tukey
0	Germination Inhibition	NA	NA
25	Germination Percentage	0.950	Wilcox
25	Germination Rate	0.071*	Tukey
25	Germination Inhibition	0.716	Wilcox
50	Germination Percentage	0.075*	Tukey
50	Germination Rate	0.153	Tukey
50	Germination Inhibition	0.106	t-test
littoralis to truncatula_2			
0	Germination Percentage	0.830	Wilcox
0	Germination Rate	0.995	Tukey
0	Germination Inhibition	NA	NA
25	Germination Percentage	0.392	Wilcox
25	Germination Rate	0.487	Tukey
25	Germination Inhibition	1.000	Wilcox
50	Germination Percentage	0.018**	Tukey
50	Germination Rate	0.023**	Tukey
50	Germination Inhibition	0.086*	t-test
truncatula_1 to truncatula_2			
0	Germination Percentage	0.830	Wilcox
0	Germination Rate	0.156	Tukey
0	Germination Inhibition	NA	NA
25	Germination Percentage	0.392	Wilcox
25	Germination Rate	0.658	Tukey
25	Germination Inhibition	0.716	Wilcox
50	Germination Percentage	0.607	Tukey
50	Germination Rate	0.462	Tukey
50	Germination Inhibition	0.230	t-test

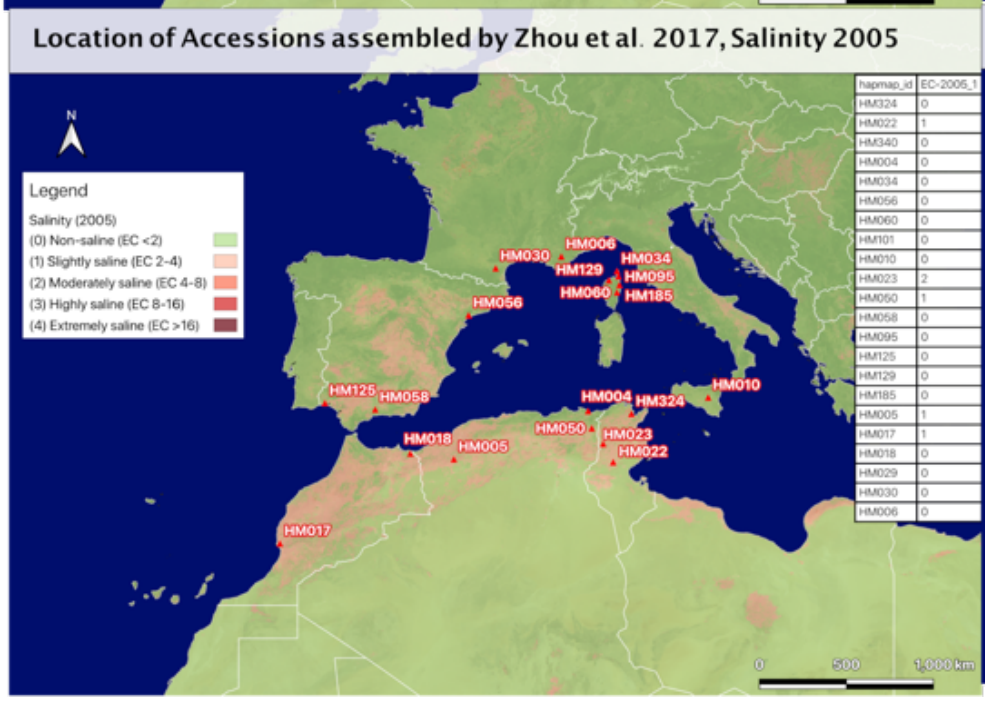
a


Original Collection Location of HapMap accessions
N
HM030
HM129
HM095
HM185
HM125
HM058
HM324
HM050
HM023
HM022
HM017
0
500
1,00

b



hapmap_id	EC-2000_1
HM324	1
HM022	1
HM340	0
HM004	0
HM034	0
HM056	0
HM060	0
HM101	0
HM010	0
HM023	0
HM050	1
HM058	1
HM095	0
HM125	0
HM129	0
HM185	0
HM005	1
HM017	0
HM018	0
HM029	0
HM030	0
HM006	0



hapmap_id	EC-2005_1
HM324	0
HM022	1
HM340	0
HM004	0
HM034	0
HM056	0
HM060	0
HM101	0
HM010	0
HM023	2
HM050	1
HM058	0
HM095	0
HM125	0
HM129	0
HM185	0
HM005	1
HM017	1
HM018	0
HM029	0
HM030	0
HM006	0

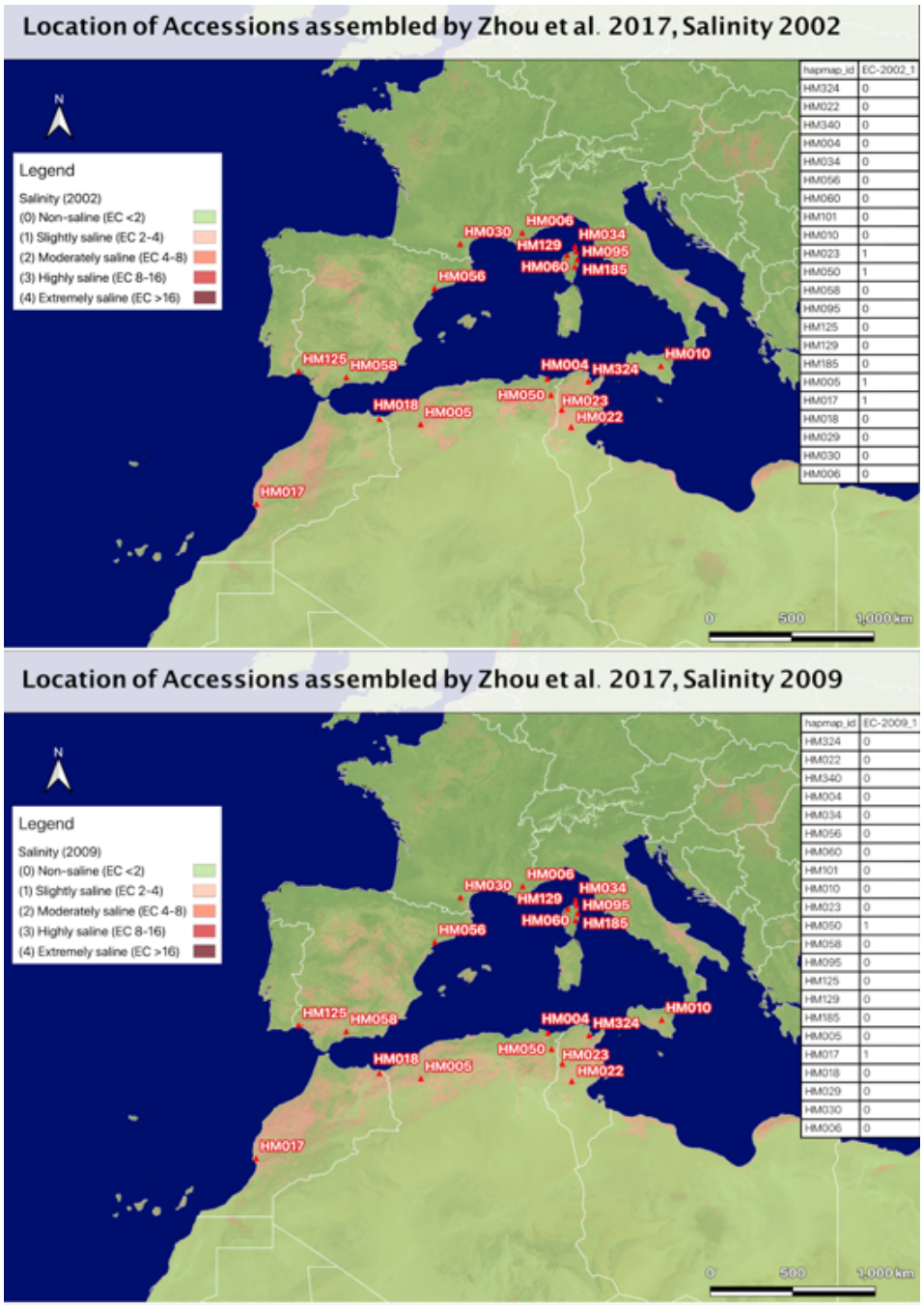


Figure 4. (**a**) Original collection locations for 11 of the 13 accessions sampled in the present study. (**b**) Predicted soil salinity for each collection location in the year 2000, 2002, 2005, and 2009. Soil salinity is categorized as 0 (EC<2 ms·cm^{-1}) or 1 (EC 2-4 ms·cm^{-1}).

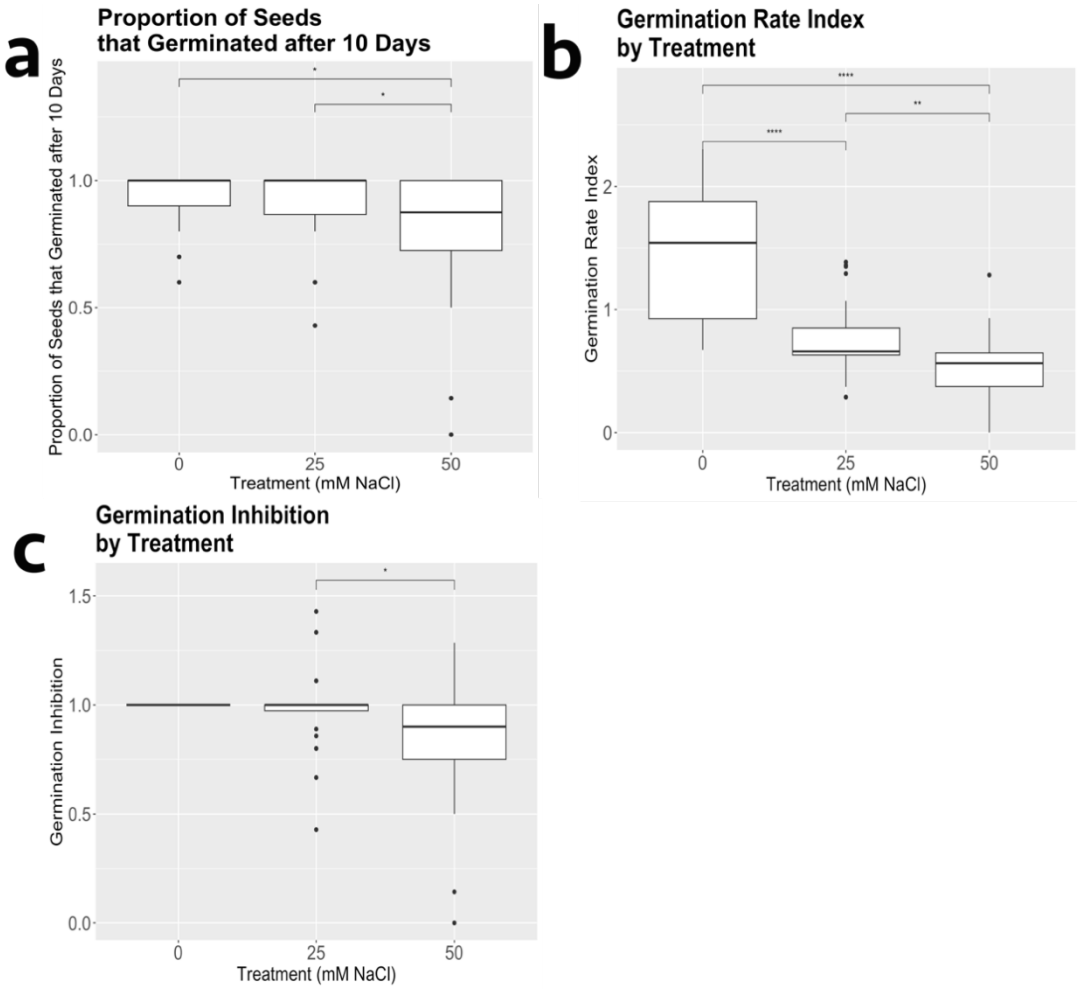


Figure 5. (**a**) Final germination proportion plotted by treatment group, (**b**) germination rate index plotted by treatment group, and (**c**) Germination inhibition plotted by treatment group (inhibition is not meaningful for the 0 mM NaCl control group). P-values indicating significance, or lack thereof, are shown above brackets between treatment groups.

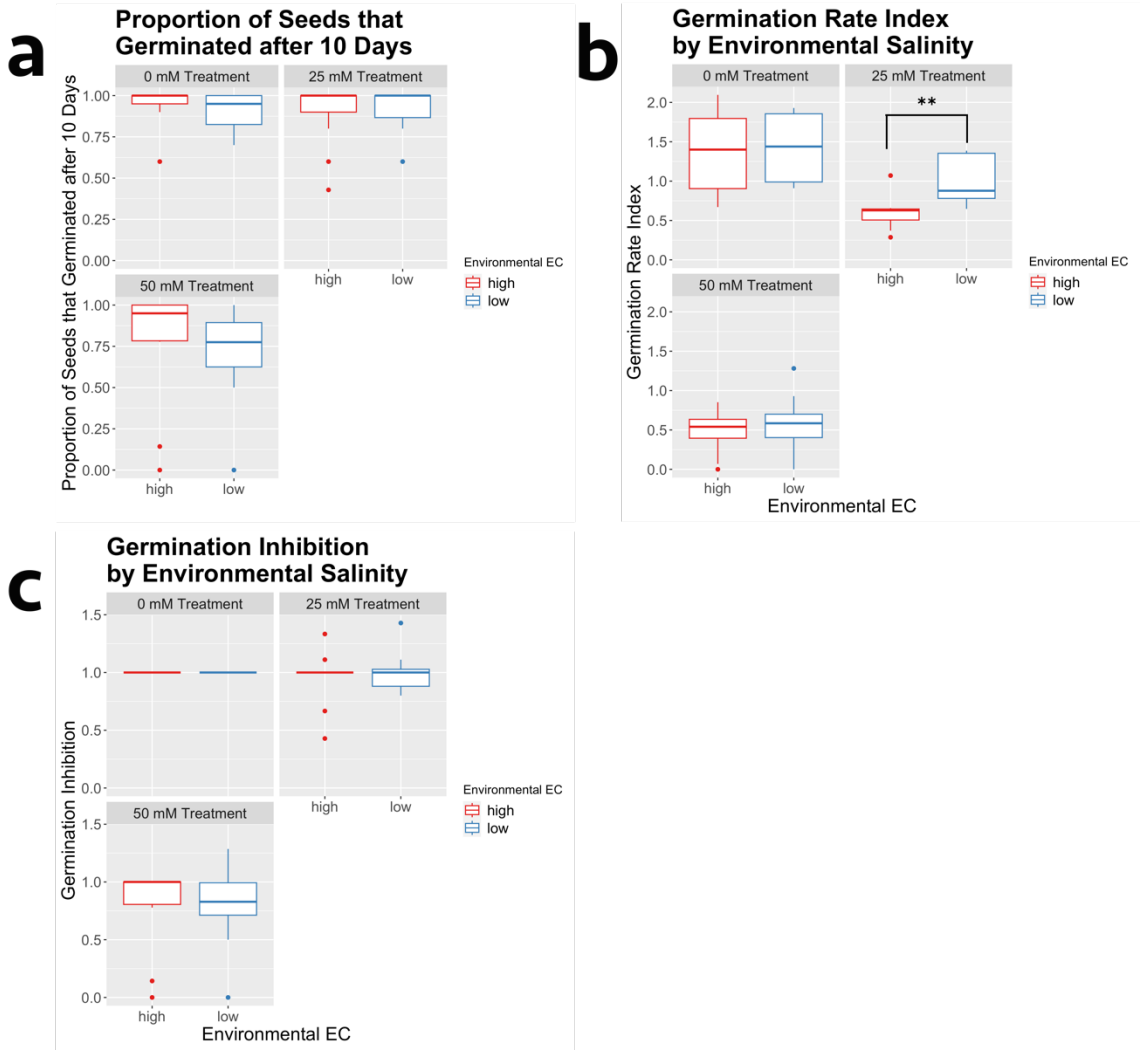
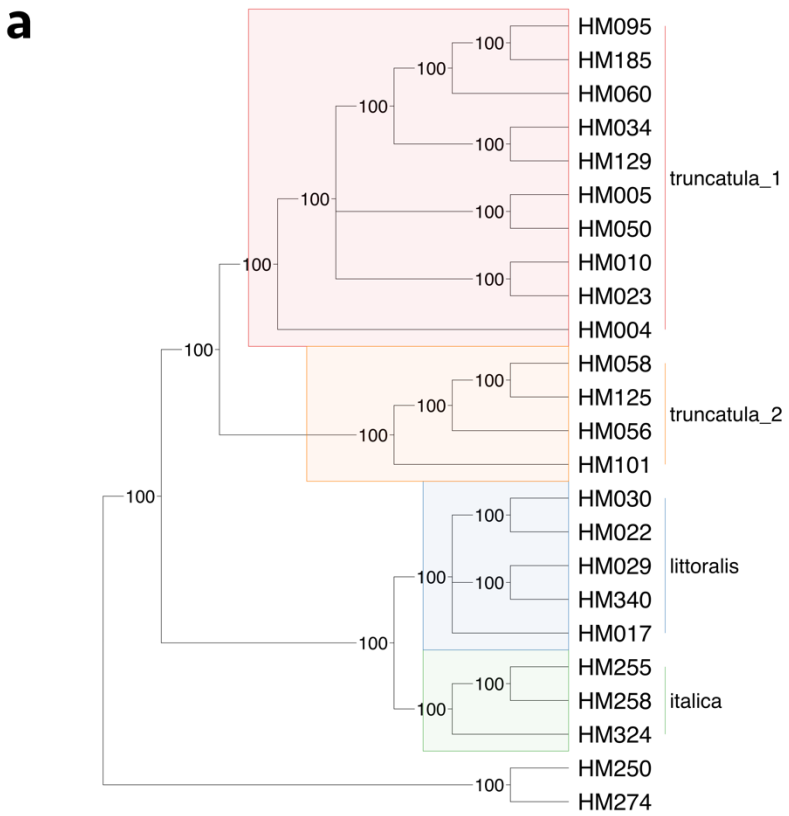


Figure 6. (**a**) Final germination proportions (**b**) germination rates and (**c**) germination inhibition plotted by accessions collected from environments with low (blue) vs high (red) predicted soil salinity. Significant differences are indicated with brackets between boxplots (** $p<0.05$).


a
HM095
HM185
HM060
HM034
HM129
HM005
HM050
HM010
HM023
HM004
truncatula_1
HM058
HM125
HM056
HM101
truncatula_2
HM030
HM022
HM029
HM340
HM017
littoralis
HM255
HM258
HM324
italica
HM250
HM274
100

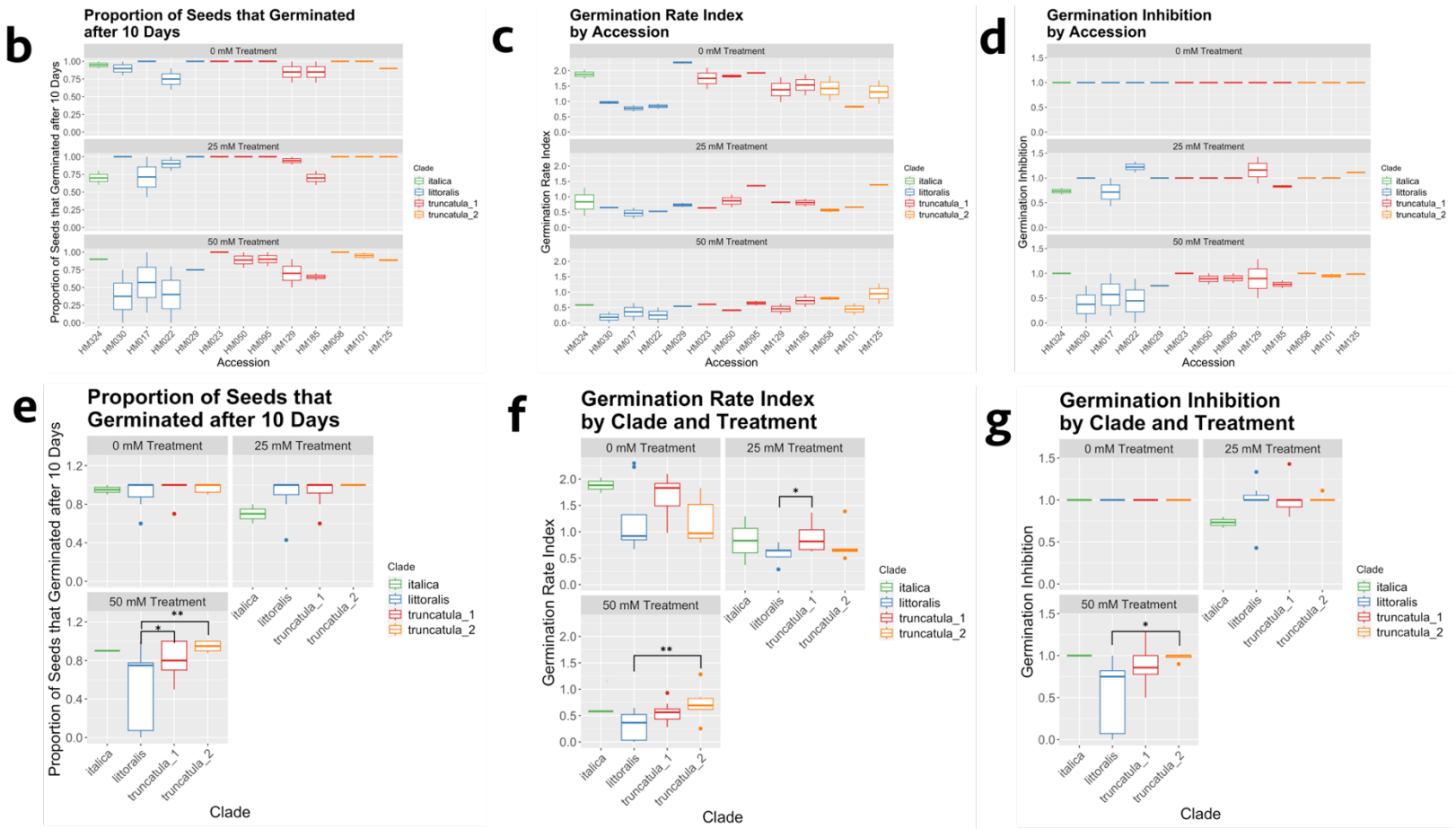
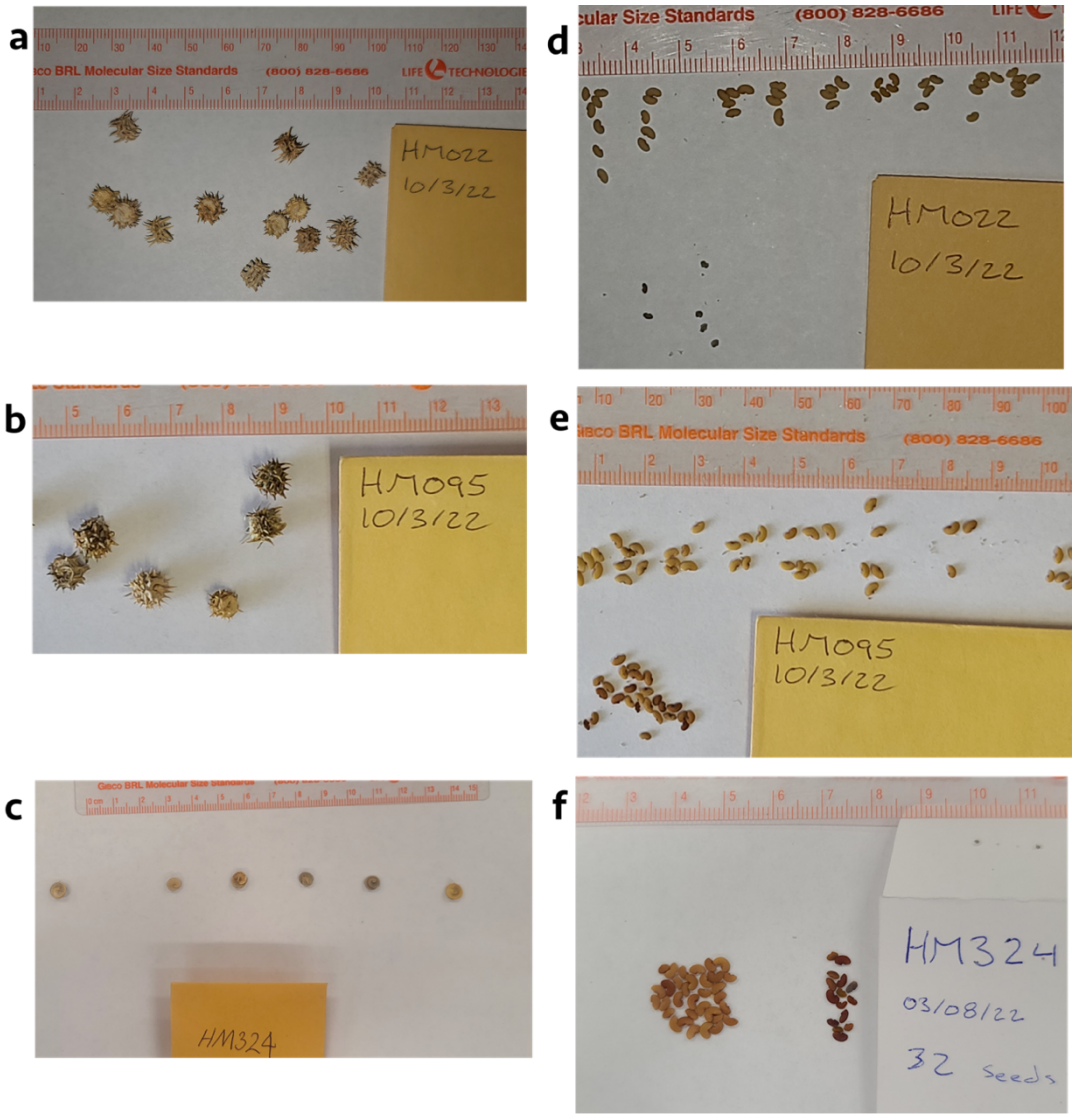
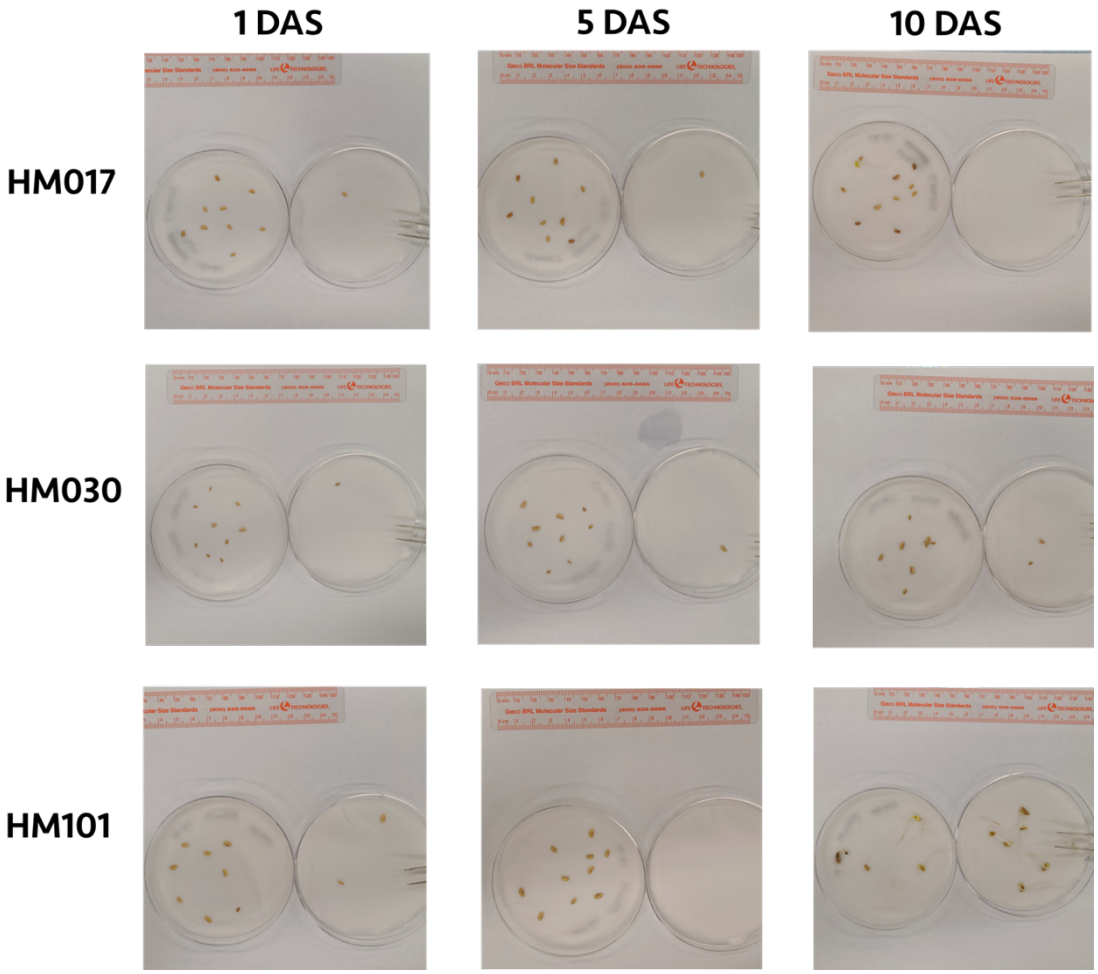


Figure 7. (**a**) Phylogenetic relationships within *Medicago* subsect. *Pachyspireae* based on maximum likelihood analysis of nuclear SNP data. The *M. truncatula* subclade formed by A17 (HM101) and three additional accessions of *M. truncatula* is highlighted as 'truncatula_2'. The clade formed by *M. littoralis* (HM030), R108 (HM029), and two likely misidentified *M. truncatula* (HM022 and HM017) is highlighted as "littoralis". (**b**, **e**) Germination proportion plotted by accession and clade assignment. (**c**, **f**) Germination rate plotted by accession and clade assignment. (**d**, **g**) Germination inhibition plotted by accession and clade assignment. Significant differences are indicated by brackets above boxplots (*p<0.10, **p<0.05)



Supplemental Figure 1. Images taken of pods (**a**-**c**) and seeds (**d**-**f**) for three representative accessions sampled in the present study: HM022 (**a**,**d**), HM095 (**b**,**e**), and HM324 (**c**,**f**). Malformed and discolored seeds were removed from analysis and can be seen in the bottom of (**d**), the bottom of (**e**), and in the right of (**f**).



Supplemental Figure 2. Germination progress 1 day after sowing (DAS), 5 DAS, and 10 DAS for three representative accessions (HM017, HM030, and HM101).

REFERENCES

Arraouadi, S., Badri, M., Taamalli, W., Huguet, T., & Aouani, M. E. (2011). Variability salt stress response analysis of Tunisian natural populations of *Medicago truncatula* (Fabaceae) using salt response index (SRI) ratio. *African Journal of Biotechnology*, 10(52), 10636–10647. https://doi.org/10.5897/ajb10.1784

Aloui, M., Mahjoub, A., Cheikh, B. N., Ludidi, N., Abdelly, C., & Badri, M. (2022). Genetic Variation in Responses to Salt Stress in Tunisian Populations of *Medicago ciliaris*. *Agronomy*, 12(8), 1781–. https://doi.org/10.3390/agronomy12081781

Branca, A., Paape, T. D., Zhou, P., Briskine, R., Farmer, A. D., Mudge, J., Bharti, A. K., Woodward, J. E., May, G. D., Gentzbittel, L., Ben, C., Denny, R., Sadowsky, M. J., Ronfort, J., Bataillon, T., Young, N. D., & Tiffin, P. (2011). Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. *Proceedings of the National Academy of Sciences - PNAS*, 108(42), E864–E870. https://doi.org/10.1073/pnas.1104032108

Bena, M. (2001). Molecular phylogeny supports the morphologically based taxonomic transfer of the "medicagoid" *Trigonella* species to the genus *Medicago* L. *Plant Systematics and Evolution*, 229(3/4), 217–236. https://doi.org/10.1007/s006060170012

Branca, A., Paape, T. D., Zhou, P., Briskine, R., Farmer, A. D., Mudge, J., Bharti, A. K., Woodward, J. E., May, G. D., Gentzbittel, L., Ben, C., Denny, R., Sadowsky, M. J., Ronfort, J., Bataillon, T., Young, N. D., & Tiffin, P. (2011). Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. *Proceedings of the National Academy of Sciences - PNAS*, 108(42), E864–E870. https://doi.org/10.1073/pnas.1104032108

Castro, B. M., Moriuchi, K. S., Friesen, M. L., Badri, M., Nuzhdin, S. V., Strauss, S. Y., Cook, D. R., & von Wettberg, E. (2013). Parental environments and interactions with conspecifics alter salinity tolerance of offspring in the annual *Medicago truncatula*. *The Journal of Ecology*, 101(5), 1281–1287. https://doi.org/10.1111/1365-2745.12125

Cerbah, M., Kevei, Z., Siljak-Yakovlev, S., Kondorosi, E., Kondorosi, A., & Trinh, T. H. (1999). FISH chromosome mapping allowing karyotype analysis in *Medicago truncatula* lines jemalong J5 and R-108-1. *Molecular Plant-Microbe Interactions*, 12(11), 947–950. https://doi.org/10.1094/MPMI.1999.12.11.947

Chen, H., Zeng, Y., Yang, Y., Huang, L., Tang, B., Zhang, H., Hao, F., Liu, W., Li, Y., Liu, Y., Zhang, X., Zhang, R., Zhang, Y., Li, Y., Wang, K., He, H., Wang, Z., Fan, G., Yang, H., Bao, A., Shang, Z., Chen, J., Wang, W., & Qiu, Q. (2020). Allele-aware chromosome-level genome assembly and efficient transgene-free genome editing for the autotetraploid cultivated alfalfa. *Nature Communications*, 11(1), 2494–2494. https://doi.org/10.1038/s41467-020-16338-x

Chen, Z., Vu, J. L., Vu, B. L., Buitink, J., Leprince, O., & Verdier, J. (2021). Genome-Wide Association Studies of Seed Performance Traits in Response to Heat Stress in *Medicago truncatula* Uncover MIEL1 as a Regulator of Seed Germination Plasticity. *Frontiers in Plant Science*, 12, 673072–673072. https://doi.org/10.3389/fpls.2021.673072

Choi, I.-S., Wojciechowski, M. F., Steele, K. P., Hopkins, A., Ruhlman, T. A., & Jansen, R. K. (2022). Plastid phylogenomics uncovers multiple species in *Medicago truncatula* (Fabaceae) germplasm accessions. *Scientific Reports*, 12(1), 21172–21172. https://doi.org/10.1038/s41598-022-25381-1

Choi, H.-K., Kim, D., Uhm, T., Limpens, E., Lim, H., Mun, J.-H., Kalo, P., Penmetsa, R. ., Seres, A., Kulikova, O., Roe, B. A., Bisseling, T., Kiss, G. B., & Cook, D. R. (2004). sequence-based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. sativa*. *Genetics* (Austin), 166(3), 1463–1502. https://doi.org/10.1534/genetics.166.3.1463

Cook, D. R. (1999). *Medicago truncatula* - A model in the making. *Current Opinion in Plant Biology*, 2(4), 301–304. https://doi.org/10.1016/S1369-5266(99)80053-3

Cordeiro, M. A., Moriuchi, K. S., Fotinos, T. D., Miller, K. E., Nuzhdin, S. V., von Wettberg, E. J., & Cook, D. R. (2014). Population differentiation for germination and early seedling root growth traits under saline conditions in the annual legume *Medicago truncatula* (Fabaceae). *American Journal of Botany*, 101(3), 488–498. https://doi.org/10.3732/ajb.1300285

Cui, J., Lu, Z., Wang, T., Chen, G., Mostafa, S., Ren, H., Liu, S., Fu, C., Wang, L., Zhu, Y., Lu, J., Chen, X., Wei, Z., & Jin, B. (2021). The genome of *Medicago polymorpha* provides insights into its edibility and nutritional value as a vegetable and forage legume. *Horticulture Research*, 8(1). https://doi.org/10.1038/s41438-021-00483-5

Downie, S. R., Katz-Downie, D. S., Rogers, E. J., Zujewski, H. L., & Small, E. (1998). Multiple independent losses of the plastid rpoC1 intron in *Medicago* (*Fabaceae*) as inferred from phylogenetic analyses of nuclear ribosomal DNA internal transcribed spacer sequences. *Canadian Journal of Botany*, 76(5). https://doi.org/10.1139/cjb-76-5-791

Eswar, D., Karuppusamy, R., & Chellamuthu, S. (2021). Drivers of soil salinity and their correlation with climate change. *Current Opinion in Environmental Sustainability*, 50, 310–318. https://doi.org/10.1016/j.cosust.2020.10.015

FAO-ITPS-GSP. (2015). Status of the world's soil resources. *FAO- ITPS-GSP Main Report*. Rome, Italy. 124-127. https://hdl.handle.net/10669/78011

Friesen, M. L., von Wettberg, E. J. B., Badri, M., Moriuchi, K. S., Barhoumi, F., Chang, P. L., Cuellar-Ortiz, S., Cordeiro, M. A., Vu, W. T., Arraouadi, S., Djébali, N., Zribi, K., Badri, Y., Porter, S. S., Aouani, M. E., Cook, D. R., Strauss, S. Y., & Nuzhdin, S. V. (2014). The ecological genomic basis of salinity adaptation in Tunisian Medicago truncatula. *BMC Genomics, 15*(1), 1160–1160. https://doi.org/10.1186/1471-2164-15-1160

Garcia, J., Barker D. G., & Journet E.-P (2006) Seed Storage and Germination. *In* The *Medicago truncatula* Handbook. The Samuel Roberts Noble Foundation, Ardmore, OK, (pp 1–9). Available online at (https://www.cabdirect.org/cabdirect/abstract/20073080185)

Guan, B., Zhou, D., Zhang, H., Tian, Y., Japhet, W., & Wang, P. (2009). Germination responses of *Medicago ruthenica* seeds to salinity, alkalinity, and temperature. *Journal of Arid Environments*, 73(1), 135–138. https://doi.org/10.1016/j.jaridenv.2008.08.009

Gurdon C., & Pal, M. (2014). Two distinct plastid genome configurations and unprecedented intraspecies length variation in the accD coding region in *Medicago truncatula*. *DNA Research*, 21(4), 417–427. https://doi.org/10.1093/dnares/dsu007

Hassani, A., Azapagic, A., & Shokri, N. (2021). Global predictions of primary soil salinization under changing climate in the 21st century. *Nature Communications*, 12(1), 6663–6663. https://doi.org/10.1038/s41467-021-26907-3

Hoffmann, B., Trinh, T. H., Leung, J., Kondorosi, A., & Kondorosi, E. (1997). A new *Medicago truncatula* line with superior in vitro regeneration, transformation, and symbiotic properties isolated through cell culture selection. *Molecular Plant-Microbe Interactions*, 10(3), 307–315. https://doi.org/10.1094/MPMI.1997.10.3.307

Ivushkin, K., Bartholomeus, H., Bregt, A. K., Pulatov, A., Kempen, B., & de Sousa, L. (2019). Global mapping of soil salinity change. *Remote Sensing of Environment*, 231, 111260–. https://doi.org/10.1016/j.rse.2019.111260

Jiao, Y.-X., He, X.-F., Song, R., Wang, X.-M., Zhang, H., Aili, R., Chao, Y.-H., Shen, Y.-H., Yu, L.-X., Zhang, T.-J., & Jia, S.-G. (2022). Recent structural variations in the *Medicago* chloroplast genomes and their horizontal transfer into nuclear chromosomes. *Journal of Systematics and Evolution : JSE*. https://doi.org/10.1111/jse.12900

Kamphuis, L. G., Williams, A. H., D'Souza, N. K., Pfaff, T., Ellwood, S. R., Groves, E. J., Singh, K. B., Oliver, R. P., & Lichtenzveig, J. (2007). *Medicago truncatula* reference accession A17 has an aberrant chromosomal configuration. *The New Phytologist*, 174(2), 299–303. https://doi.org/10.1111/j.1469-8137.2007.02039.x

Kaur, P., Lui, C., Dudchenko, O., Nandety, R. S., Hurgobin, B., Pham, M., Aiden, E. L., Wen, J., & Mysore, K. S. (2021). Delineating the Tnt1 Insertion Landscape of the Model Legume *Medicago truncatula* cv. R108 at the Hi-C Resolution Using a Chromosome-Length Genome Assembly. *International Journal of Molecular Sciences*, 22(9), 4326–. https://doi.org/10.3390/ijms22094326

Khalil, C., Boufous E., Mousadik. A. E. (2011). Diversity of Salt Tolerance During Germination in *Medicago ciliaris* (L.) and *Medicago polymorpha* (L.). *Atlas Journal of Plant Biology* 1 (1): 6–12, 2011.https://doi.org/10.5147/ajpb.v1i1.83

Lazrek, F., Roussel, V., Ronfort, J., Cardinet, G., Chardon, F., Aouani, M. E., & Huguet, T. (2009). use of neutral and non-neutral SSRs to analyse the genetic structure of a Tunisian collection of *Medicago truncatula* lines and to reveal associations with eco-environmental variables. *Genetica*, 135(3), 391–402. https://doi.org/10.1007/s10709-008-9285-3

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, A., Liu, A., Wu, S., Qu, K., Hu, H., Yang, J., Shrestha, N., Liu, J., & Ren, G. (2022). Comparison of structural variants in the whole genome sequences of two *Medicago truncatula* ecotypes: Jemalong A17 and R108. *BMC Plant Biology*, 22(1), 77–77. https://doi.org/10.1186/s12870-022-03469-0

Liang, W., Ma, X., Wan, P., & Liu, L. (2018). Plant salt-tolerance mechanism: A review. *Biochemical and Biophysical Research Communications*, 495(1), 286–291. https://doi.org/10.1016/j.bbrc.2017.11.043

Maureira-Butler, I. J., Pfeil, B. E., Muangprom, A., Osborn, T. C., & Doyle, J. J. (2008). The Reticulate History of *Medicago* (*Fabaceae*). *Systematic Biology*, 57(3), 466–482. https://doi.org/10.1080/10635150802172168

Mbarki, S., Skalicky, M., Vachova, P., Hajihashemi, S., Jouini, L., Zivcak, M., Tlustos, P., Brestic, M., Hejnak, V., & Khelil, A. Z. (2020). Comparing salt tolerance at seedling and germination stages in local populations of *Medicago ciliaris* L. To *Medicago intertexta* L. and *Medicago scutellata* L. *Plants* (Basel), 9(4), 526–. https://doi.org/10.3390/plants9040526

Moll, K. M., Zhou, P., Ramaraj, T., Fajardo, D., Devitt, N. P., Sadowsky, M. J., Stupar, R. M., Tiffin, P., Miller, J. R., Young, N. D., Silverstein, K. A. T., & Mudge, J. (2017). Strategies for optimizing BioNano and Dovetail explored through a second reference quality assembly for the legume model, *Medicago truncatula*. *BMC Genomics*, 18(1), 578–578. https://doi.org/10.1186/s12864-017-3971-4

Monirifar, H., & Barghi, M. (2009). Identification and Selection for Salt Tolerance in Alfalfa (*Medicago sativa* L.) Ecotypes via Physiological Traits. Notulae Scientia Biologicae, 1(1), 63–66. https://doi.org/10.15835/nsb113498

Office of Chemical Safety and Pollution Prevention. (2018) ZerolTol 2.0 (EPA Reg. No. 70299-12) [Master Label] . United States Environmental Protection Agency. https://www3.epa.gov/pesticides/chem_search/ppls/070299-00012-20180712.pdf

Paape, T., Heiniger, B., Domingo, M. S., Clear, M. R., Lucas, M. M., & Pueyo, J. J. (2021). Genome-Wide Association Study Reveals Complex Genetic Architecture of Cadmium and Mercury Accumulation and Tolerance Traits in *Medicago truncatula*. *Frontiers in Plant Science*, 12, 806949–806949. https://doi.org/10.3389/fpls.2021.806949

Pecrix, Y., Staton, S. E., Sallet, E., Lelandais-Brière, C., Moreau, S., Carrère, S., Blein, T., Jardinaud, M.-F., Latrasse, D., Zouine, M., Zahm, M., Kreplak, J., Mayjonade, B., Satgé, C., Perez, M., Cauet, S., Marande, W., Chantry-Darmon, C., Lopez-Roques, C., Bouchez, O., Bérard, A., Debellé, F., Muños, S., Bendahmane, A., Bergès, H., Niebel, A., Buitink, J., Frugier, F., Benhamed, M., Crespi, M., Gouzy, J., Gamas, P. (2018). Whole-genome landscape of *Medicago truncatula* symbiotic genes. In *Nature plants* (Vol. 4, Issue 12, pp. 1017–1025). *Springer Nature*. https://doi.org/10.1038/s41477-018-0286-7

Pokhrel, Y., Felfelani, F., Satoh, Y., Boulange, J., Burek, P., Gädeke, A., Gerten, D., Gosling, S. N., Grillakis, M., Gudmundsson, L., Hanasaki, N., Kim, H., Koutroulis, A., Liu, J., Papadimitriou, L., Schewe, J., Müller Schmied, H., Stacke, T., Telteu, C. E., Thiery, W., Veldkamp, T., Zhao, F., & Wada, Y. (2021). Global terrestrial water storage and drought severity under climate change. *Nature Climate Change*, 11(3), 226–233. https://doi.org/10.1038/s41558-020-00972-w

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics*, 81(3), 559–575. https://doi.org/10.1086/519795

Qadir, M., Quillérou, E., Nangia, V., Murtaza, G., Singh, M., Thomas, R. J., Drechsel, P., & Noble, A. D. (2014). Economics of salt-induced land degradation and restoration. *Natural Resources Forum*, 38(4), 282–295. https://doi.org/10.1111/1477-8947.12054

QGIS.org, (2023). QGIS Geographic Information System. QGIS Association. http://www.qgis.org

Rengasamy, P. (2010). Soil processes affecting crop production in salt-affected soils. *Functional Plant Biology : FPB*, 37(7), 613–620. https://doi.org/10.1071/FP09249

Ronfort, J., Bataillon, T., Santoni, S., Delalande, M., David, J. L., & Prosperi, J.-M. (2006). Microsatellite diversity and broad scale geographic structure in a model legume: Building a set of nested core collection for studying naturally occurring variation in *Medicago truncatula*. *BMC Plant Biology*, 6(1), 28–28. https://doi.org/10.1186/1471-2229-6-28

Rumbaugh M. D. (1991) *Salt Tolerance of Germinating Alfalfa Seeds* [conference presentation]. The North American Alfalfa Improvement Conference. https://www.naaic.org/stdtests/saltseeds.pdf

Sarri, E., Termentzi, A., Abraham, E. M., Papadopoulos, G. K., Baira, E., Machera, K., Loukas, V., Komaitis, F., & Tani, E. (2021). Salinity stress alters the secondary metabolic profile of *M. Sativa*, *M. Arborea* and their hybrid (alborea). *International Journal of Molecular Sciences*, 22(9), 4882–. https://doi.org/10.3390/ijms22094882

Scasta, J., Trostle, C. L., & Foster, M. A. (2012). Evaluating Alfalfa (*Medicago sativa* L.) Cultivars for Salt Tolerance Using Laboratory, Greenhouse and Field Methods. *Journal of Agricultural Science*, 4(6). https://doi.org/10.5539/jas.v4n6p90

Shen, C., Du, H., Chen, Z., Lu, H., Zhu, F., Chen, H., Meng, X., Liu, Q., Liu, P., Zheng, L., Li, X., Dong, J., Liang, C., & Wang, T. (2020). The Chromosome-Level Genome Sequence of the Autotetraploid Alfalfa and Resequencing of Core Germplasms Provide Genomic Resources for Alfalfa Research. *Molecular Plant*, 13(9), 1250–1261. https://doi.org/10.1016/j.molp.2020.07.003

Small, E. (2011). Alfalfa and relatives: evolution and classification of *Medicago*. *NRC research press*.

Stanton-Geddes, J., Paape, T., Epstein, B., Briskine, R., Yoder, J., Mudge, J., Bharti, A. K., Farmer, A. D., Zhou, P., Denny, R., May, G. D., Erlandson, S., Yakub, M., Sugawara, M., Sadowsky, M. J., Young, N. D., & Tiffin, P. (2013). Candidate Genes and Genetic Architecture of Symbiotic and Agronomic Traits Revealed by Whole-Genome, Sequence-Based Association Genetics in *Medicago truncatula*. *PloS One*, 8(5), e65688–e65688. https://doi.org/10.1371/journal.pone.0065688

Swofford, D. L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts. https://paup.phylosolutions.com/

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Steele, K. P., Ickert-Bond, S. M., Zarre, S., & Wojciechowski, M. F. (2010). Phylogeny and character evolution in Medicago (Leguminosae): Evidence from analyses of plastid *trnK*/*matK* and nuclear *GA3ox1* sequences. *American Journal of Botany*, 97(7), 1142–1155. https://doi.org/10.3732/ajb.1000009

Tadege, M., Ratet, P., & Mysore, K. S. (2005). Insertional mutagenesis: a Swiss Army knife for functional genomics of *Medicago truncatula*. *Trends in Plant Science*, 10(5), 229–235. https://doi.org/10.1016/j.tplants.2005.03.009

Tadege, M., Wen, J., He, J., Tu, H., Kwak, Y., Eschstruth, A., Cayrel, A., Endre, G., Zhao, P. X., Chabaud, M., Ratet, P., & Mysore, K. S. (2008). Large-scale insertional mutagenesis using the Tnt1 retrotransposon in the model legume *Medicago truncatula*. *The Plant Journal* : for Cell and Molecular Biology, 54(2), 335–347. https://doi.org/10.1111/j.1365-313X.2008.03418.x

Tang, H., Krishnakumar, V., Bidwell, S., Rosen, B., Chan, A., Zhou, S., Gentzbittel, L., Childs, K. L., Yandell, M., Gundlach, H., Mayer, K. F. X., Schwartz, D. C., & Town, C. D. (2014). An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*. *BMC Genomics*, 15(1), 312–312. https://doi.org/10.1186/1471-2164-15-312

Wang, T.-Z., Tian, Q.-Y., Wang, B.-L., Zhao, M.-G., & Zhang, W.-H. (2014). Genome variations account for different response to three mineral elements between *Medicago truncatula* ecotypes Jemalong A17 and R108. *BMC Plant Biology*, 14(1), 122–122. https://doi.org/10.1186/1471-2229-14-122

Wang, T., Ren, L., Li, C., Zhang, D., Zhang, X., Zhou, G., Gao, D., Chen, R., Chen, Y., Wang, Z., Shi, F., Farmer, A. D., Li, Y., Zhou, M., Young, N. D., & Zhang, W.-H. (2021). The genome of a wild *Medicago* species provides insights into the tolerant mechanisms of legume forage to environmental stress. *BMC Biology*, 19(1), 96–96. https://doi.org/10.1186/s12915-021-01033-0

Yaish, M. W., Al-Lawati, A., Jana, G. A., Patankar, H. V., & Glick, B. R. (2016). Impact of soil salinity on the structure of the bacterial endophytic community identified from the roots of Caliph Medic (*Medicago truncatula*). *PloS One*, 11(7), e0159007–e0159007. https://doi.org/10.1371/journal.pone.0159007

Yaish, M. W., Al-Lawati, A., Al-Harrasi, I., & Patankar, H. V. (2018). Genome-wide DNA Methylation analysis in response to salinity in the model plant caliph medic (*Medicago truncatula*). *BMC Genomics*, 19(1), 78–78. https://doi.org/10.1186/s12864-018-4484-5

Yoder, J. B., Briskine, R., Mudge, J., Farmer, A., Paape, T., Steele, K., Weiblen, G. D., Bharti, A. K., Zhou, P., May, G. D., Young, N. D., & Tiffin, P. (2013). Phylogenetic Signal Variation in the Genomes of *Medicago* (Fabaceae). *Systematic Biology*, 62(3), 424–438. https://doi.org/10.1093/sysbio/syt009

Young, N., Debellé, F., Oldroyd, G. et al. The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* 480, 520–524 (2011). https://doi.org/10.1038/nature10625

Yu, G., Smith, D. K., Zhu, H., Guan, Y., Lam, T. T., & McInerny, G. (2017). ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8(1), 28–36. https://doi.org/10.1111/2041-210X.12628

Zhou, P., Silverstein, K. A. T., Ramaraj, T., Guhlin, J., Denny, R., Liu, J., Farmer, A. D., Steele, K. P., Stupar, R. M., Miller, J. R., Tiffin, P., Mudge, J., & Young, N. D. (2017). Exploring structural variation and gene family architecture with De Novo assemblies of 15 *Medicago* genomes. *BMC Genomics*, 18(1), 261–261. https://doi.org/10.1186/s12864-017-3654-1

APPENDIX A

HAPMAP ACCESSION INFORMATION AND SNP DATASET STATISTICS

Hapmap Accession Information and SNP Dataset Statistics. HapMap ID of the accessions sampled, species identification, mean coverage in the SNP dataset (number of reads per site), standard deviation of coverage in the SNP dataset, and proportion of sites missing data in the SNP dataset. Accessions in bold have been previously categorized as highly divergent, or misidentified, *M. truncatula*.

Sample	Species Identification	Original Collection Location	Coverage Mean (number of reads per site)	Coverage Standard Deviation	Proportion of Sites Missing Data
HM004	*M. truncatula* [a]	Algeria [e]	84.21	15.96	0.0008
HM005	*M. truncatula* [a]	Algeria [e]	28.88	14.18	0.0010
HM010	*M. truncatula* [a]	Italy [e]	76.82	15.64	0.0006
HM017	***M. truncatula*** [b]	N/A	32.80	14.38	0.0013
HM022	***M. truncatula*** [b]	N/A	91.87	19.86	0.0038
HM023	*M. truncatula* [a]	N/A	11.04	4.73	0.0197
HM029	*M. tricycla* [c]	N/A	13.75	6.79	0.0146
HM030	*M. littoralis* [c]	N/A	31.52	20.50	0.0035
HM034	*M. truncatula* [a]	France [e]	69.39	13.99	0.0005
HM050	*M. truncatula* [a]	Algeria [e]	81.94	19.73	0.0009
HM056	*M. truncatula* [a]	Spain [e]	154.24	18.87	0.0074
HM058	*M. truncatula* [a]	Spain [e]	21.65	7.44	0.0005
HM060	*M. truncatula* [a]	France [e]	60.83	12.52	0.0005
HM095	*M. truncatula* [a]	France [e]	73.53	14.74	0.0004
HM101	*M. truncatula* [a]	N/A	39.72	16.20	0.0001
HM125	*M. truncatula* [a]	Spain [e]	66.77	13.18	0.0006
HM129	*M. truncatula* [a]	France [e]	6.83	2.14	0.3165

Sample	Species Identification	Original Collection Location	Coverage Mean (number of reads per site)	Coverage Standard Deviation	Proportion of Sites Missing Data
HM185	*M. truncatula* [a]	France [e]	11.05	5.15	0.0217
HM250	***M. truncatula*** [a] ***M. murex*** [c,d]	Spain [c]	6.58	3.06	0.4920
HM255	***M. truncatula*** [a]	Spain [d]	6.26	1.78	0.4564
HM258	***M. truncatula*** [a]	Morocco [d]	6.19	1.67	0.5157
HM274	***M. truncatula*** [a] ***M. murex*** [c] ***M. doliata*** [d]	Algeria [c]	6.70	2.45	0.2798
HM324	*M. italica* [c]	Tunisia [d]	5.93	1.83	0.6572
HM340	*M. truncatula* [d]	N/A	90.59	19.80	0.0028

([a]) Stanton-Geddes et al., 2013, ([b]) Yoder et al., 2013, ([c]) Medicago Analysis Portal, ([d]) Germplasm Resource Information Network – GRIN, and ([e]) Ronfort et al., 2006.

APPENDIX B

HAPMAP ACCESSION INFORMATION AND GERMINATION STATISTICS

Hapmap Accession Information And Germination Statistics. HapMap ID of accessions sampled, treatment group (mM NaCl applied), predicted soil salinity at the original collection location, categorization of predicted soil salinity at the original collection location, phylogenetic clade assignment according to nuclear SNP data, mean germination percentage, mean germination rate, mean germination inhibition, and the number of observations (replications) for each sample and treatment group

Accession	Treatment	Env. Soil Salinity	Env. Soil Salinity (Categorized)	Clade	Mean Germination Proportion	Mean Germination Rate	Mean Germination Inhibition	Number of Obs.
HM017	0	0.75	high	littoralis	1.00	0.776	1.00	2
HM017	25	0.75	high	littoralis	0.71	0.466	0.71	2
HM017	50	0.75	high	littoralis	0.57	0.357	0.57	2
HM022	0	0.5	high	littoralis	0.75	0.840	1.00	2
HM022	25	0.5	high	littoralis	0.90	0.524	1.22	2
HM022	50	0.5	high	littoralis	0.40	0.250	0.44	2
HM023	0	0.75	high	truncatula_1	1.00	1.748	1.00	2
HM023	25	0.75	high	truncatula_1	1.00	0.637	1.00	2
HM023	50	0.75	high	truncatula_1	1.00	0.602	1.00	1
HM029	0	NA	NA	littoralis	1.00	2.265	1.00	2
HM029	25	NA	NA	littoralis	1.00	0.731	1.00	2
HM029	50	NA	NA	littoralis	0.75	0.540	0.75	1
HM030	0	0	low	littoralis	0.90	0.965	1.00	2
HM030	25	0	low	littoralis	1.00	0.647	1.00	2

Accession	Treatment	Env. Soil Salinity	Env. Soil Salinity (Categorized)	Clade	Mean Germination Proportion	Mean Germination Rate	Mean Germination Inhibition	Number of Obs.
HM030	50	0	low	littoralis	0.38	0.183	0.38	2
HM050	0	1	high	truncatula_1	1.00	1.821	1.00	2
HM050	25	1	high	truncatula_1	1.00	0.864	1.00	2
HM050	50	1	high	truncatula_1	0.89	0.407	0.89	2
HM058	0	0.25	high	truncatula_2	1.00	1.422	1.00	2
HM058	25	0.25	high	truncatula_2	1.00	0.564	1.00	2
HM058	50	0.25	high	truncatula_2	1.00	0.796	1.00	2
HM095	0	0	low	truncatula_1	1.00	1.929	1.00	2
HM095	25	0	low	truncatula_1	1.00	1.356	1.00	2
HM095	50	0	low	truncatula_1	0.90	0.644	0.90	2
HM101	0	NA	NA	truncatula_2	1.00	0.828	1.00	2
HM101	25	NA	NA	truncatula_2	1.00	0.660	1.00	2
HM101	50	NA	NA	truncatula_2	0.95	0.449	0.95	2
HM125	0	0	low	truncatula_2	0.90	1.304	1.00	2
HM125	25	0	low	truncatula_2	1.00	1.387	1.11	2
HM125	50	0	low	truncatula_2	0.89	0.944	0.99	2

Accession	Treatment	Env. Soil Salinity	Env. Soil Salinity (Categorized)	Clade	Mean Germination Proportion	Mean Germination Rate	Mean Germination Inhibition	Number of Obs.
HM129	0	0	low	truncatula_1	0.85	1.381	1.00	2
HM129	25	0	low	truncatula_1	0.94	0.818	1.16	2
HM129	50	0	low	truncatula_1	0.70	0.454	0.89	2
HM185	0	0	low	truncatula_1	0.85	1.535	1.00	2
HM185	25	0	low	truncatula_1	0.70	0.807	0.83	2
HM185	50	0	low	truncatula_1	0.65	0.723	0.78	2
HM324	0	0.25	high	italica	0.95	1.883	1.00	2
HM324	25	0.25	high	italica	0.70	0.832	0.73	2
HM324	50	0.25	high	italica	0.90	0.581	1.00	1