nature portfolio

Corresponding author(s): DAPR NCOMMS-23-32503

Last updated by author(s): Jul 28, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We mapped whole genome resequencing reads onto the reference genome using BWA v.0.7.17. Duplicated reads were removed using the "MarkDuplicates" option in Picard v.2.25.0 (Picard Toolkit. 2019. Broad Institute, GitHubRepository. http://broadinstitute.github.io/picard/). We identified single nucleotide polymorphisms (SNPs) and insertions and deletions (indels) using GATK v.4.1.4.1. For identification of low copy genes, we first performed coverage statistical analysis for each nuclear gene (annotated in the reference genome) in the bam file of each accession using bamstats05 within jvarkit (https://github.com/lindenb/jvarkit), then we used a custom script to estimate the copy number variation (CNV) for each gene, when CNV was between 0.4–1.6, we identified this as a low copy gene. The consensus sequences of these low copy genes were extracted using ANGSD v.0.911 with unique_Only and only_proper_pairs parameters based on the bam files. Coding sequences (CDS) were extracted across all accessions using BEDTools from consensus sequences. We used the NOVOPlasty v.3.8.3 and GetOrganelle pipelines to assemble plastomes of all accessions. We annotated the assembled plastomes using GeSeq with MPI-MP chloroplast references and HMMER profile search. We further confirmed all tRNAs using tRNAscan-SE v.2.0.5. For verification, we compared all annotations with previously published plastomes of these genera available in NCBI, and exon boundaries were corrected in Geneious v.10.2.6. Species distribution information (area) was collected from literature (Small 2011) and online databases (Plants of the World Online, https://powo.science.kew.org/; Go Botany: Native Plant Trust, https://gobotany.nativeplanttrust.org/; Pl@ntNet, https://plantnet.org/en/). We collected extant distributional data (latitude and longitude coordinates) of our studied species from the Global Biodiversity Information Facility (https://www.gbif.org/). We collected environmental variables (Supplementary Data 4) from online databases (http://www.worldclim.com/version2; https://cgiarcsi.community; https://www.fao.org/soils-portal/data-hub/soil-maps-and-databases/harmonized-world-soil-database-v12/en/). We collected three traits information (Supplementary Data 3) from literature (Small 2011). We obtained the global paleo-temperature dataset from literature (Sun 2020).
Data analysis	We used both nuclear genome and plastome datasets for phylogenetic reconstructions. For whole genome single nucleotide polymorphisms (SNPs) dataset from the nuclear genome, we reconstructed an approximately-maximum-likelihood tree using FastTree; for datasets from the

identified low copy nuclear genes, we used both concatenated and coalescent methods. In the concatenation-based analysis, maximum-likelihood (ML) trees of datasets were reconstructed using RAxML. In the coalescent-based analysis, we first reconstructed phylogenetic trees of each low copy nuclear gene using RAxML, and then estimated a coalescent tree using ASTRAL v.5.6.3. For each plastome, we extracted coding sequences (CDS) of protein-coding genes that were shared by all accessions (Supplementary Table 2) and aligned them using MAFFT v7.453. A maximum-likelihood (ML) phylogenetic tree was reconstructed using RAxML based on the concatenated CDS alignments. Divergence times were estimated using the r8s and RelTime methods. We performed ancestral area reconstruction using BioGeoBEARS v.1.1.2. We performed ancestral character reconstruction of three traits using Mesquite v.3.61. We estimated diversification rates, evolutionary rates of three traits, and rates of niche evolution using BAMM v.2.5.0. Then we used BAMMtools v.2.17 to visualize the result. We investigated the effects of global paleo-temperature on diversification using RPANDA v.1.9. We used the "key innovation test" in BayesRates v.1.6 to explore whether the evolution of the two novel pod traits promoted diversification. we reconstructed ancestral niches using RevBayes and used Tracer v.1.7.1 to diagnose and visualize MCMC outputs of RevBayes. We used the "phenogram" function in the phytools package to illustrate the degree of similarity of the selected environmental variables among species with different life-histories. We simulated ILS following the method described in literature (Wang 2018). Specifically, gene trees were simulated using the multispecies coalescent model of Phybase v.1.5. We calculated gene-tree quartet frequencies of incongruent internal nodes in observed and simulated datasets using Twisst. Correlation and Chi-square analyses were carried out in R v.4.2.3. We detected potential hybridization using PhyloNetworks v.0.9.0. We conducted a refined identity by decent (IBD) blocks analysis based on whole-genome single nucleotide polymorphisms (SNPs) using BEAGLE to further detect shared haplotypes between species. We used Admixture v1.3.0 to infer the population genetic structure.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The whole genome resequencing reads newly generated in this study have been deposited in the NCBI Sequence Read Archive under BioProject accession code PRJNA1000112 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1000112). Plastome sequences assembled from whole genome resequencing reads of this study are available in the figshare (DOI:10.6084/m9.figshare.23805507).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Not applicable. This study did not involve human research participants.
Reporting on race, ethnicity, or other socially relevant groupings	Not applicable. This study did not involve human research participants.
Population characteristics	Not applicable. This study did not involve human research participants.
Recruitment	Not applicable. This study did not involve human research participants.
Ethics oversight	Not applicable. This study did not involve human research participants.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Based on phylogenomics, this study infers the spatiotemporal patterns of evolution of Medicago and investigates the evolutionary processes that allow its widespread radiation.
Research sample	We sampled a total of 178 accessions representing 93 species of the tribe Trifolieae (Leguminosae), including 68 Medicago, 19 Trigonella, 4 Melilotus, and 2 Trifolium species. Detailed information, including species, sections, sampling sources, and ploidy levels

	are listed in Supplementary Data 2. The 68 Medicago species accounted for about ~76% of the 90 recognized species within this genus and covered 11 of the total 14 sections. We could not obtain the materials of the remaining three sections that contain only one species for each. As outgroups, we used Cicer arietinum (tribe Cicereae) and three species in the tribe Fabeae (Pisum sativum, Lathyrus sativus, and Vicia sativa) based on their close relationships with Trifolieae. We generated whole genome resequencing data for 175/178 accessions for this study, and downloaded the remaining three accessions and four outgroups from NCBI (https://www.ncbi.nlm.nih.gov/). The newly generated data were deposited in the Sequence Read Archive of NCBI under BioProject accession code PRJNA1000112 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1000112).
Sampling strategy	Our goal was to create a taxon sampling across Medicago as comprehensive as possible. The 68 Medicago species accounted for about ~76% of the 90 recognized species within this genus and covered 11 of the total 14 sections.
Data collection	Z.L. provided the seeds and planted them in a greenhouse. G.R. collected samples from the wild. Z.L. carried out DNA extractions. G.R. and S.W. performed phylogenetic and evolutionary analyses.
Timing and spatial scale	The sequence data for this study were collected between May 2020 and September 2021. This time frame was delimited by the timing of available grant funding for the study. Sample sources of the taxa included in this study stem from U.S. National Plant Germplasm System, CHN. National Animal Husbandry Station, China wild, and NCBI (https://www.ncbi.nlm.nih.gov/).
Data exclusions	Sequence data were cleaned and trimmed according to standard protocols to remove contamination and sequencing error. After this step, no data were excluded.
Reproducibility	We have provided all data and code used in the study in figshare and GitHub, respectively. Further, we provided detailed methodology for analyses in the Methods. We therefore believe that the study is fully reproducible.
Randomization	We performed phylogenetic and evolutionary analyses based on the taxon sampling outlined. Their methods are fundamentally different from a standard statistical experimental design that requires randomization.
Blinding	The type of questions and phylogenetic analyses carried out in this study do not require blinding experiment design, because there are no participants who may be influenced by the treatments.

Did the study involve field work? ☐ Yes ☒ No

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☐	☒ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes
☒	☐ Public health
☒	☐ National security
☒	☐ Crops and/or livestock
☒	☐ Ecosystems
☒	☐ Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☒	☐	Demonstrate how to render a vaccine ineffective
☒	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☒	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☒	☐	Increase transmissibility of a pathogen
☒	☐	Alter the host range of a pathogen
☒	☐	Enable evasion of diagnostic/detection modalities
☒	☐	Enable the weaponization of a biological agent or toxin
☒	☐	Any other potentially harmful combination of experiments and agents