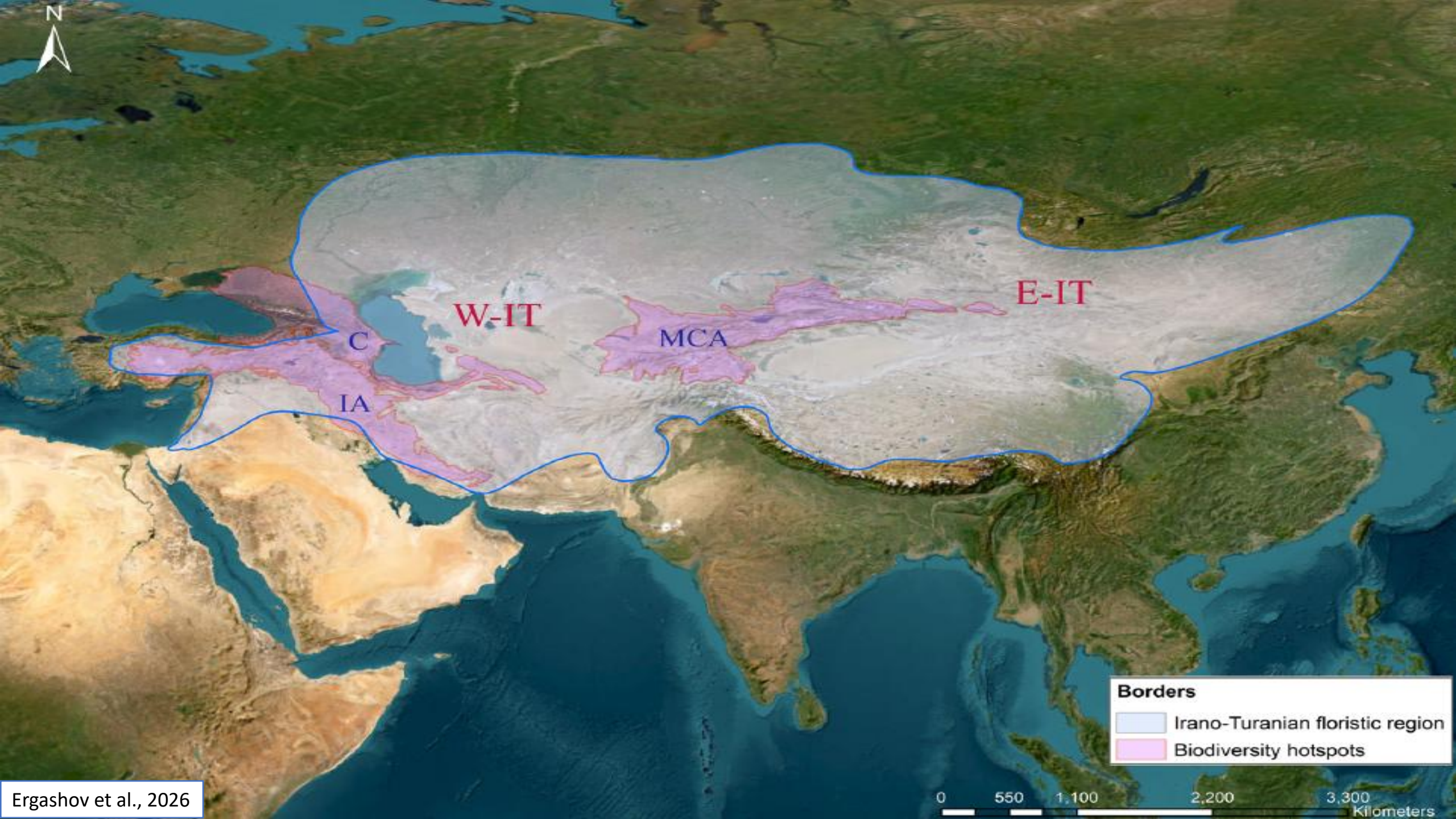
The 4th International Symposium on Mapping Asian Plants
Harbin, China, on August 31, 2026

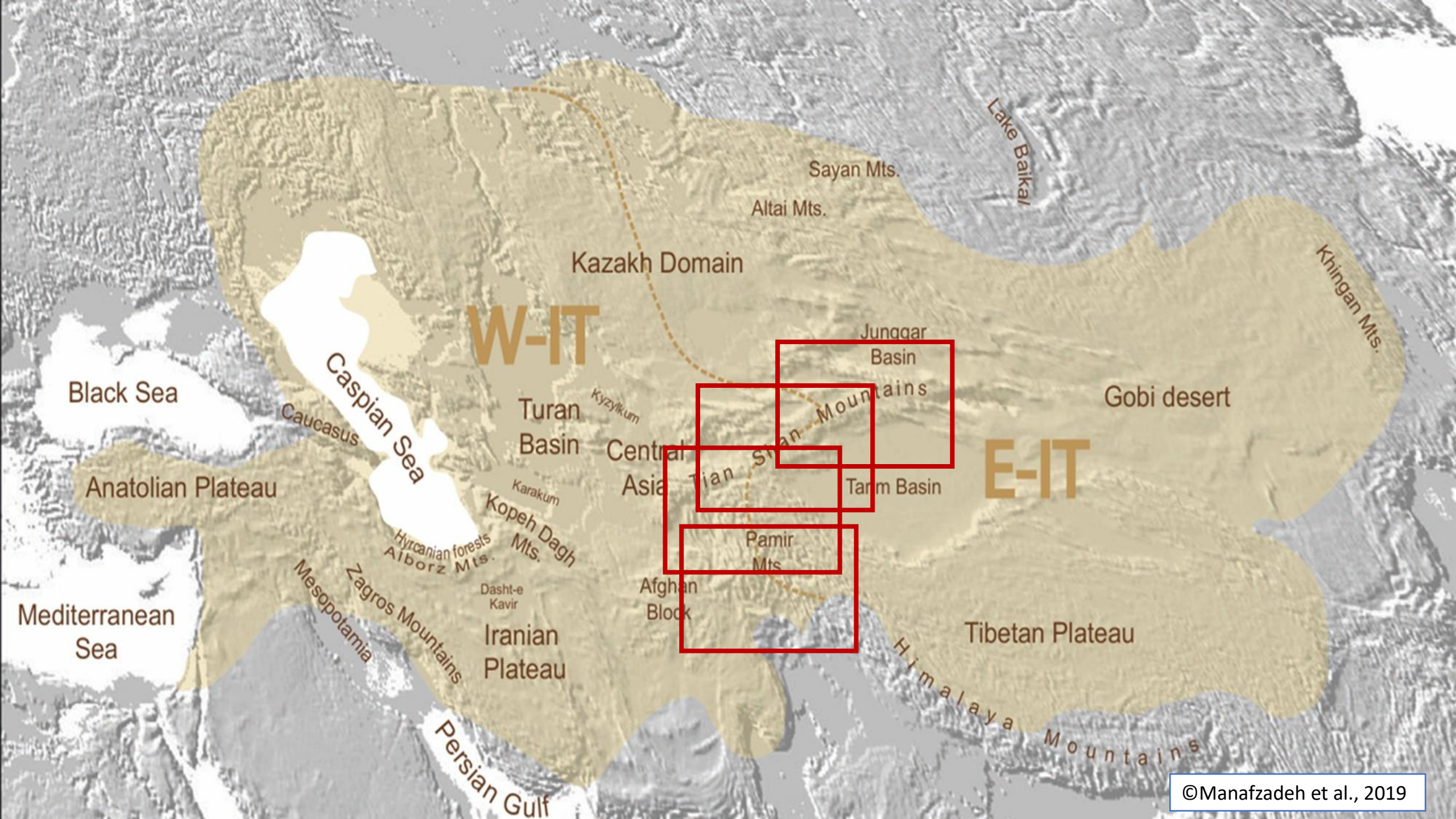
From Biodiversity Shortfalls to Integrated Biodiversity Data in Central Asia

Tojibaev Sh. Komiljon
Institute of Botany, Uzbekistan Academy of Sciences





Ergashov et al., 2026



©Manafzadeh et al., 2019

Central Asia: Biodiversity Hotspot

High species richness and endemism

- ~9,700 vascular plant taxa, including ~3,400 endemics
- The CA mountains support around 7000 species of vascular plants, accounting for more than 75% of the total plant diversity in the region
- Deserts are home to a variety of halophytes and are centres of the origin and differentiation of ephemeral plants, containing more than 400 such species

Unique ecosystems: deserts to alpine zones

- Strong environmental gradients: deserts → mountains → alpine systems
- Major mountain systems: Tian Shan and Pamir-Alay, and some of the world's largest desert landscapes (Kyzylkum, Karakum, and the Ustyurt Plateau)

Biogeographic crossroads

- One of the most diverse yet underexplored regions of Eurasia
- The CA, as part of the Iran-Turan region, plays a key role in understanding how current plant diversity in Eurasia has changed across space and time
- Its biogeographical relationships with adjacent areas remain poorly understood, especially when compared to neighbouring regions (e.g. Mediterranean, South-East Asia)
- The oldest plant domestication sites are located in Central Asia

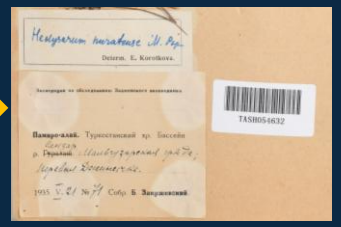
Myers et al., 2000; Manafzadeh et al., 2019; Tojibaev et al., 2020; Ma et al., 2024

Country	Acronym	Number of specimens	Type specimens	Digitized	Georeferenced	Date Founded	Core area
UZB	TASH	1.6M	3241	~ 700 000	~ 400 000	1920	Central Asia
	SAMDU	12 500		~ 8000	Not	1927	Uzbekistan
	KD	5 000		Undigitized	Not	2020	Fergana Valley
	Nukus	~ 10 000		Undigitized	Not	1959	Karakalpakstan
	Namangan	~ 10 000		Undigitized	Not	2025	Uzbekistan
	Termez	~ 7000		Undigitized	Not	2026	S Uzbekistan
KAZ	AA	250 000	~ 500	130 031	No Information	1933	Central Asia
	ABG	30 000		No Information	No Information	1935	Kazakhstan
	PPIU	30 000		No Information	No Information	1930	Kazakhstan
	KG	28 000		No Information	No Information	1946	Kazakhstan
	NUR	25 000		No Information	No Information	2018	Kazakhstan
	KSPI	~ 25 000		No Information	No Information	2004	Kazakhstan
	MANG	~ 15 000		No Information	No Information	1972	Kazakhstan
	QAR	~ 15 000		No Information	No Information	2000	Kazakhstan
	UKSPI	11 300		No Information	No Information	1952	Kazakhstan
KYR	FRU	400 000	~ 750	Undigitized	No Information	1943	Central Asia
TAD	TAD	200 000		Digitized	No Information	1981	Central Asia
	KHOR	30 000		Undigitized	Not	1940	Tajikistan

Proposed AI-Based Workflow

A scalable solution for herbarium digitization, georeferencing and analysis

This slide presents the overall AI roadmap; each component will be explained in the following slides.



INPUT DATA
Digital herbarium sheet

01
AI-based label recognition
Automatically detects and reads text on herbarium labels

02
Structured data extraction
Scientific name, locality, collector, date and ecology

03
AI-assisted georeferencing
Interprets historical localities and predicts coordinates

04
Integrated database and publishing
Ready for Excel, GBIF, GMOCA and other systems



05 BIODIVERSITY ANALYSIS AND SCIENTIFIC USE

- Species distribution
- Ecological analysis
- Geographical analysis
- Mapping
- Monitoring
- Conservation applications

HUMAN-IN-THE-LOOP VALIDATION

AI accelerates the workflow, while experts verify, refine and validate the results.

FASTER • MORE SCALABLE • LESS LABOR-INTENSIVE • READY FOR LARGE HERBARIUM COLLECTIONS

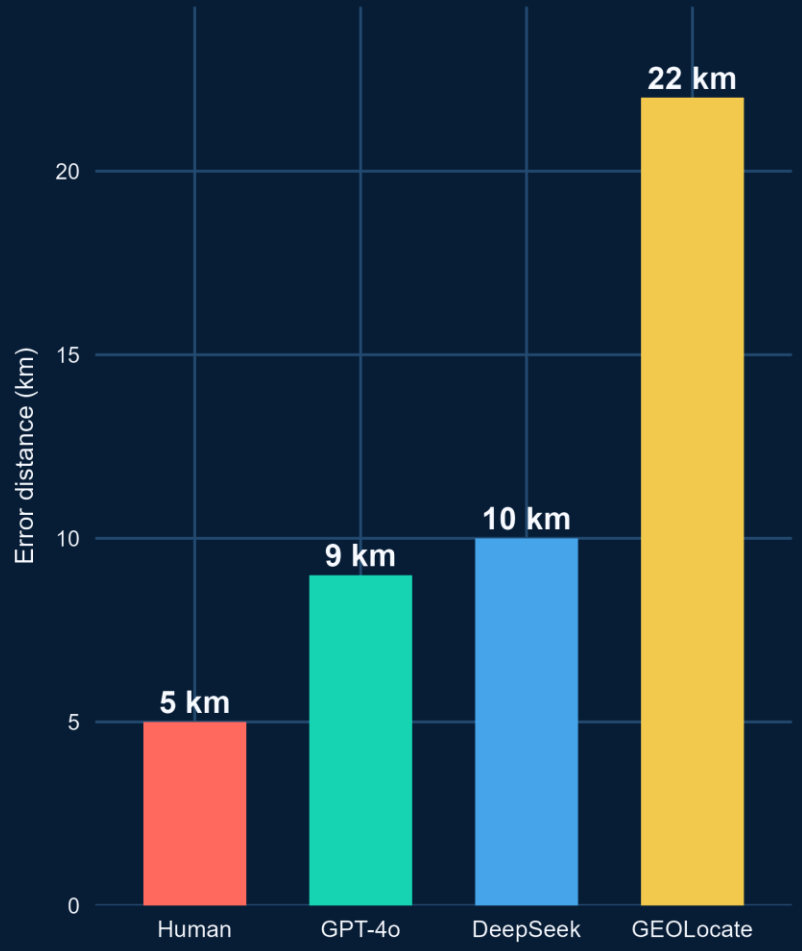
Manual vs AI-Assisted Georeferencing

AI accelerates large-scale georeferencing while experts validate the results

Xie et al. 2025. *Nature Plants*

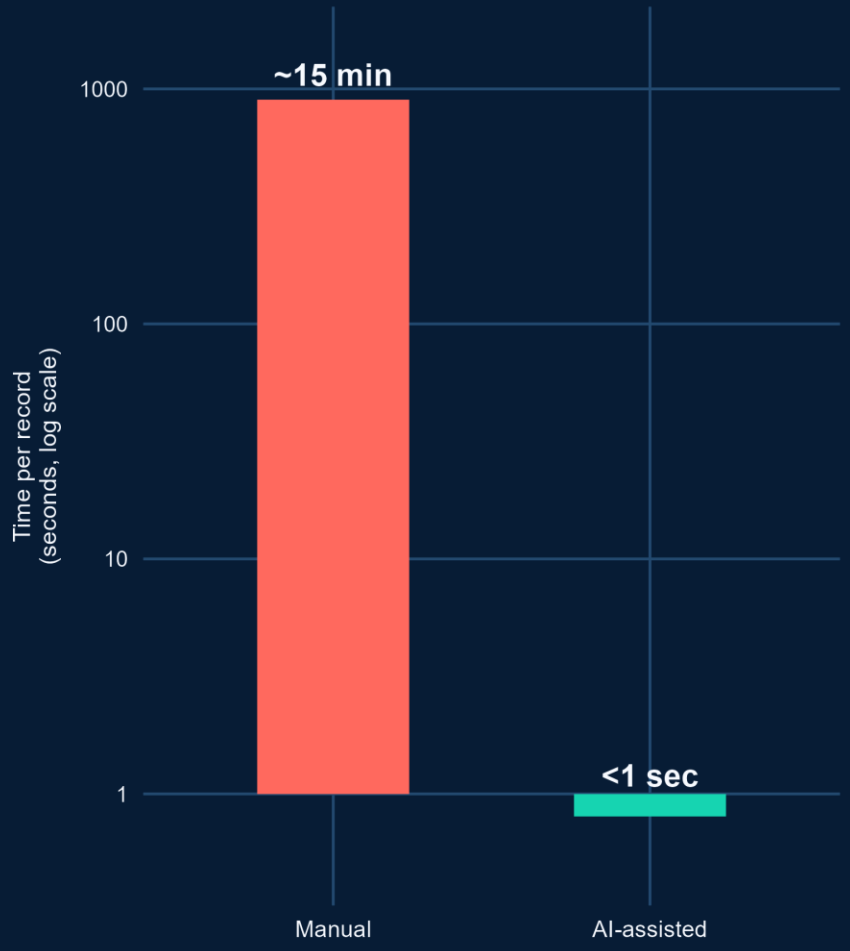
Accuracy

Median error distance



Speed

Time per record



Large-scale scenario

500,000 records

MANUAL		AI-ASSISTED
≈ 3.3 years	VS	≈ 5 days
record-by-record		high-throughput

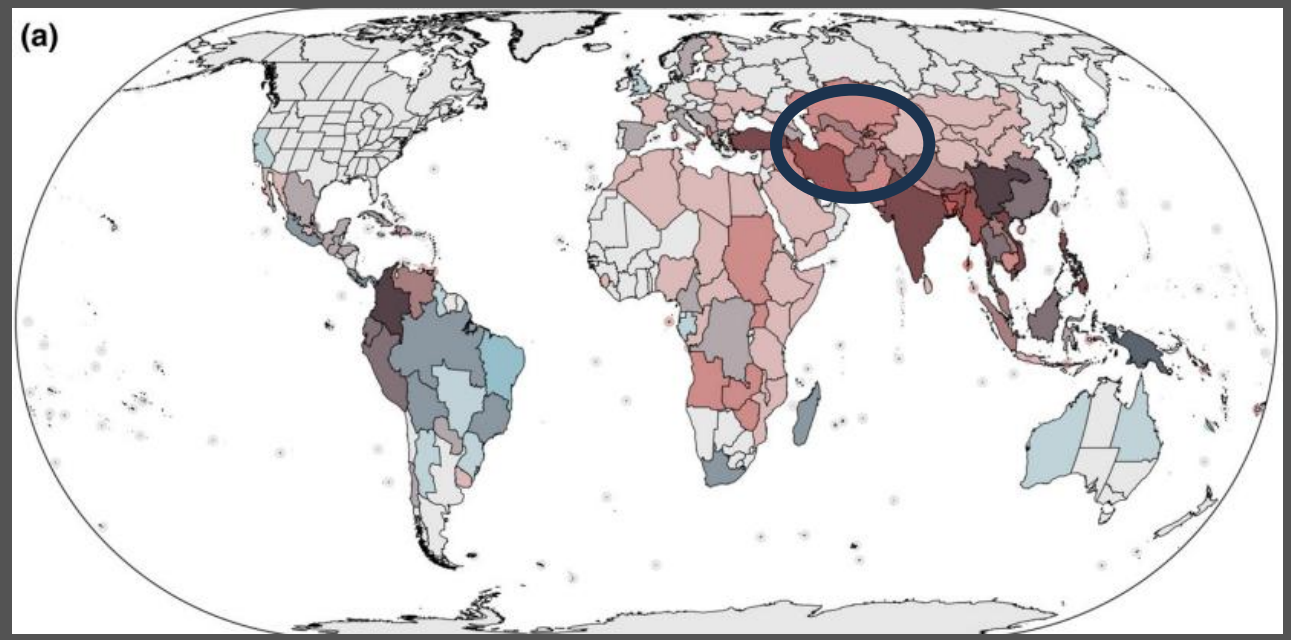
AI performs the first pass; experts verify and correct the results.

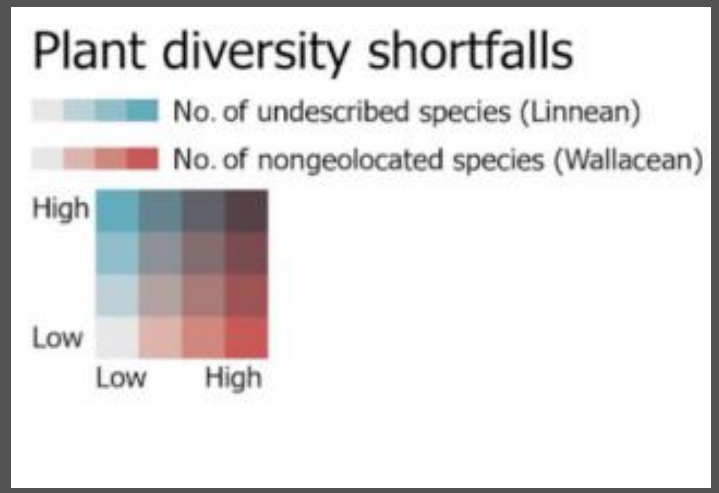
KEY MESSAGE

AI does not replace experts — it automates large-scale first-pass georeferencing, allowing specialists to focus on validation and interpretation.

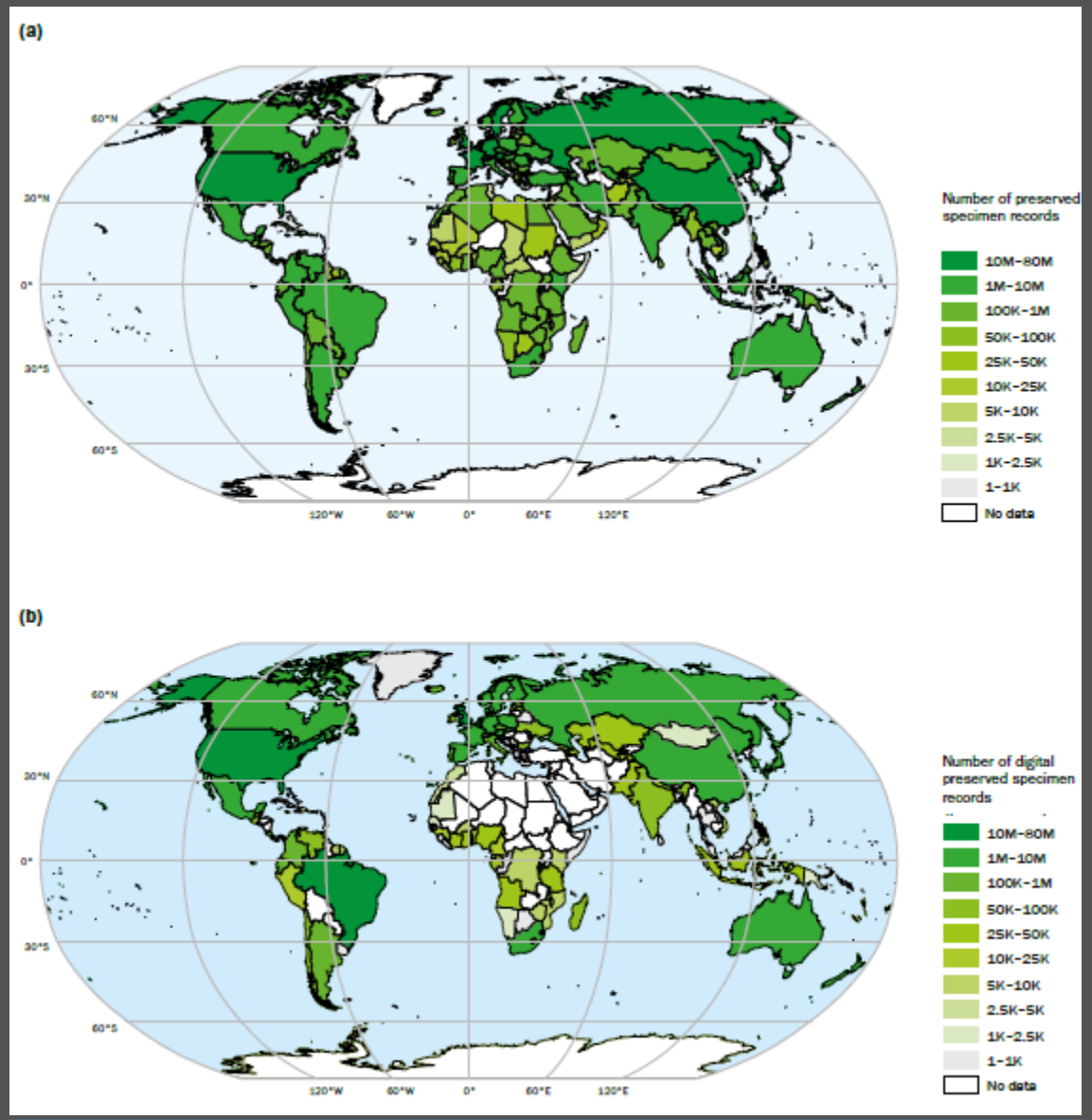
I. Central Asia Biodiversity: Taxonomic and Distributional Knowledge Gaps

Linnaean shortfall (unknown species) and **Wallacean shortfall** (unknown distributions)





CA countries have been identified as part of the world's 33 global diversity dark spots



Ondo et al. New Phytologist, 2024; State of the World's Plants and Fungi, Kew, 2026

The Real Problem

Not a lack of data

Flora of Uzbekistan (1941–1962), 6 vol.; Flora of Uzbekistan (2016–2023), 6 vol. (in progress)
Flora of Central Asia:
(**1**) Conspectus Florae Asiae Mediae (vol. I–X, **8096 speci**es); (**2**) Khassanov, 2015 CFAM (**9341–1300–161**); (**3**) Checklist of Vascular Plants in Central Asia, **9640–1199–139**, Li and Tojibaev, 2025

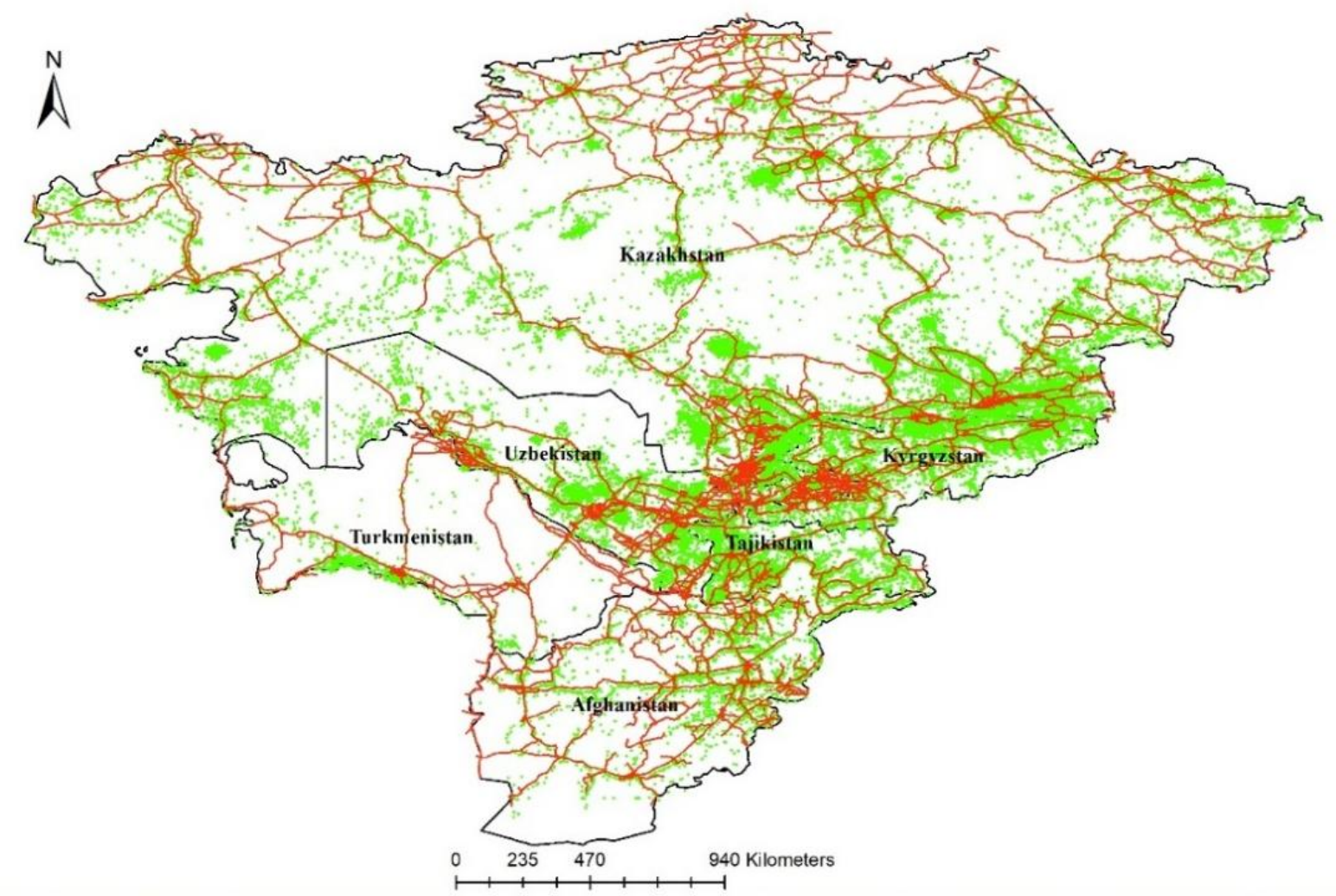
Limited Volume of Accessible & Standardized Data

Millions of specimens remain undigitized; many records lack georeferencing
Limited use of global standards (Darwin Core)
Data not published in global repositories (e.g., GBIF)

Fragmentation of biodiversity knowledge

Biodiversity data are dispersed across institutions
Collections, field data, and publications remain disconnected

Accessibility Bias in Biodiversity Data



Spatial relationship between species occurrence records (green) and the primary road network (red)

© Karimov et al., 2026

JAPB
Journal of Asia-Pacific Biodiversity
Available online 17 March 2026
In Press, Journal Pre-proof What's this?

Original Article

Accessibility bias and the road-map effect in biodiversity data across Central Asia and Afghanistan

Bobur Karimov [a], Sherzod Mamadaliev [b], Diloramxon Kurbanova [b], Ziyoviddin Yusupov [a], Komiljon Sh. Tojibaev [a]

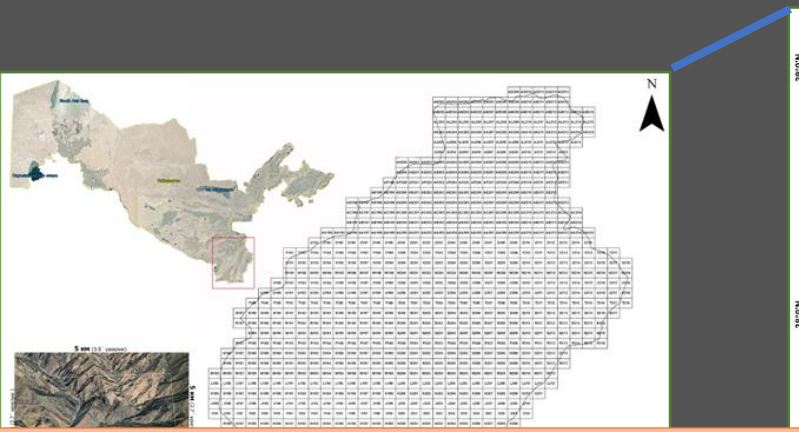
We used 657,299 unique GBIF records to see how distance from roads, elevation, climate, and topography affect sampling intensity

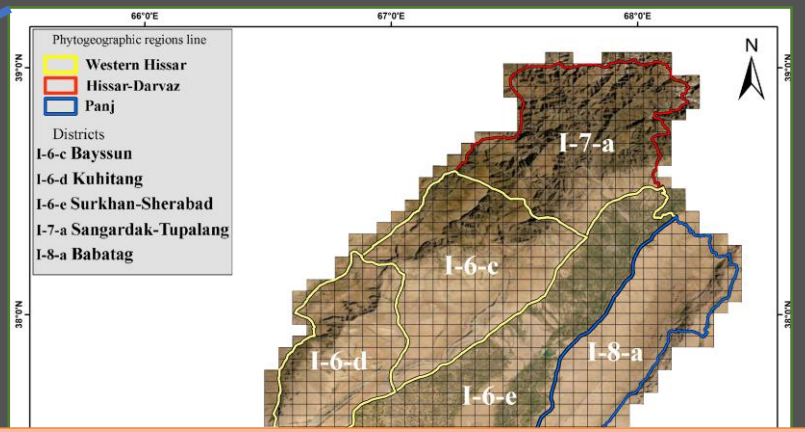
66% of occurrence records are located within 5 km of roads

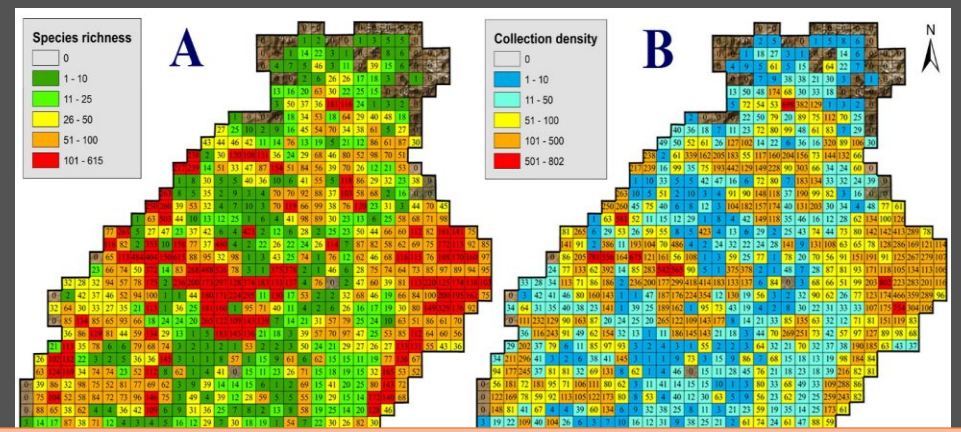
Sampling effort is concentrated in accessible and populated areas

Deserts and high-mountain areas, especially in remote regions, remain under-sampled

Grid-Based Floristic Surveys: Bridging Linnaean and Wallacean Shortfalls in Central Asia

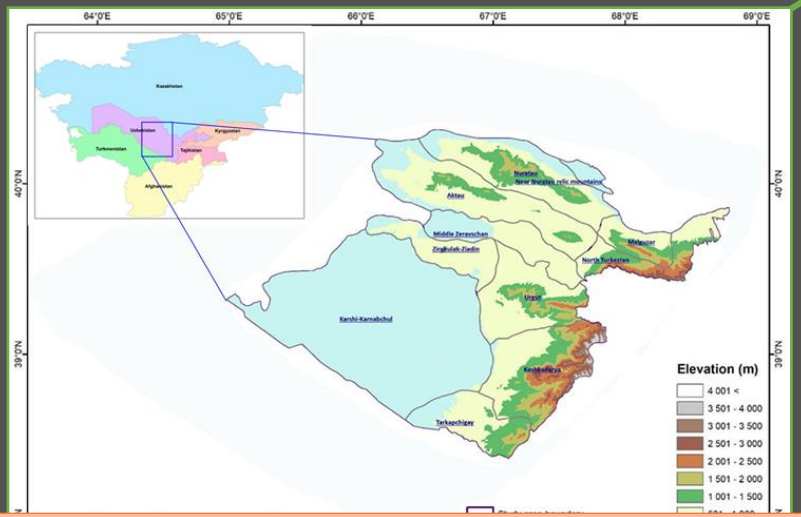


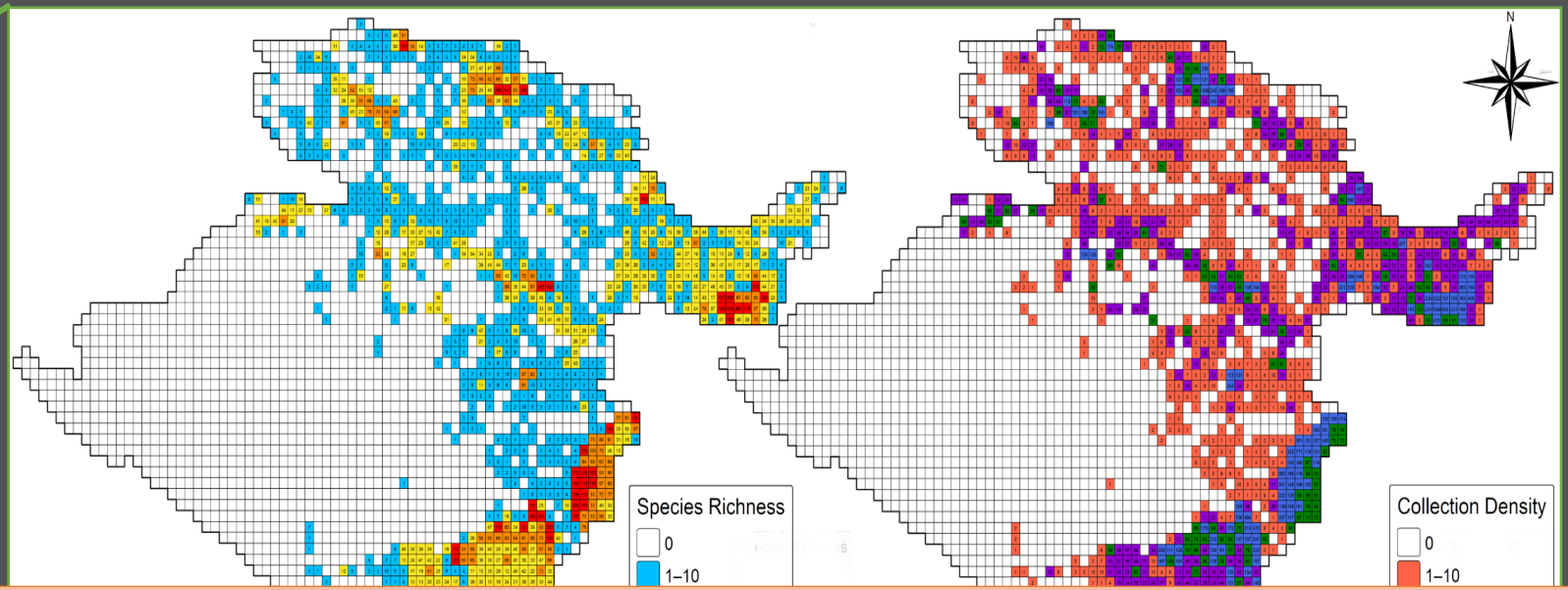




Previous phase: Towards a Data Ecosystem (2020–2024)

70,000+ records integrated; 64,500+ curated for analysis; Taxonomic scope: 2,202 species, 615 genera, 96 families

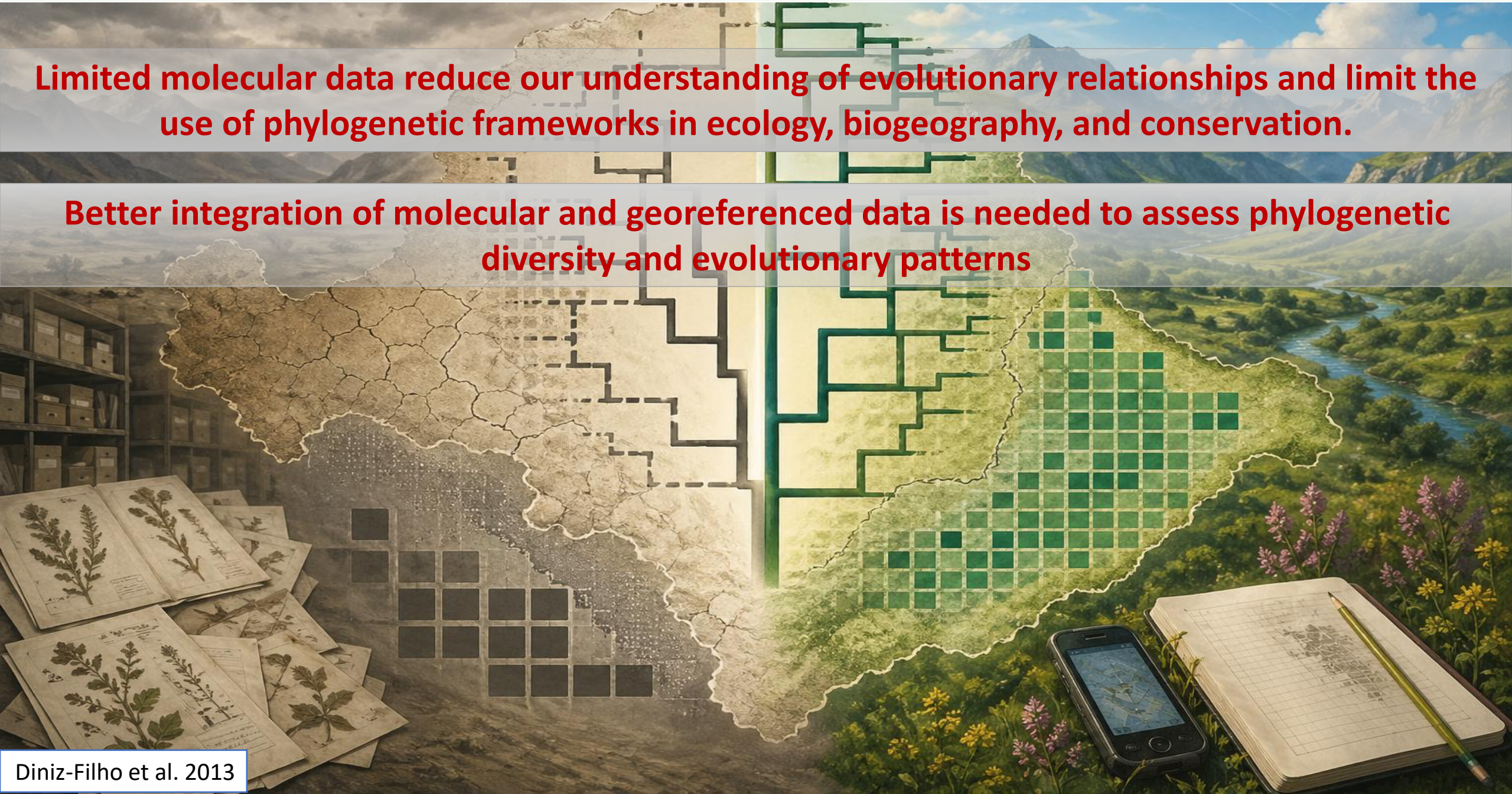




Current phase (2025–2029): Expanding the Data Ecosystem

<u>2025 baseline</u>: 2,557 species documented; 45,000+ georeferenced records; 43.3% spatial coverage (2,878 grid cells)

II. Central Asia Biodiversity: Phylogenetic Knowledge Gaps (The Darwinian Shortfall)



Limited molecular data reduce our understanding of evolutionary relationships and limit the use of phylogenetic frameworks in ecology, biogeography, and conservation.

Better integration of molecular and georeferenced data is needed to assess phylogenetic diversity and evolutionary patterns

Diniz-Filho et al. 2013

The current state of knowledge of many small genera remains at a basic level

Allardia Decne. (= *Waldheimia* Kar. & Kir.), 12 species



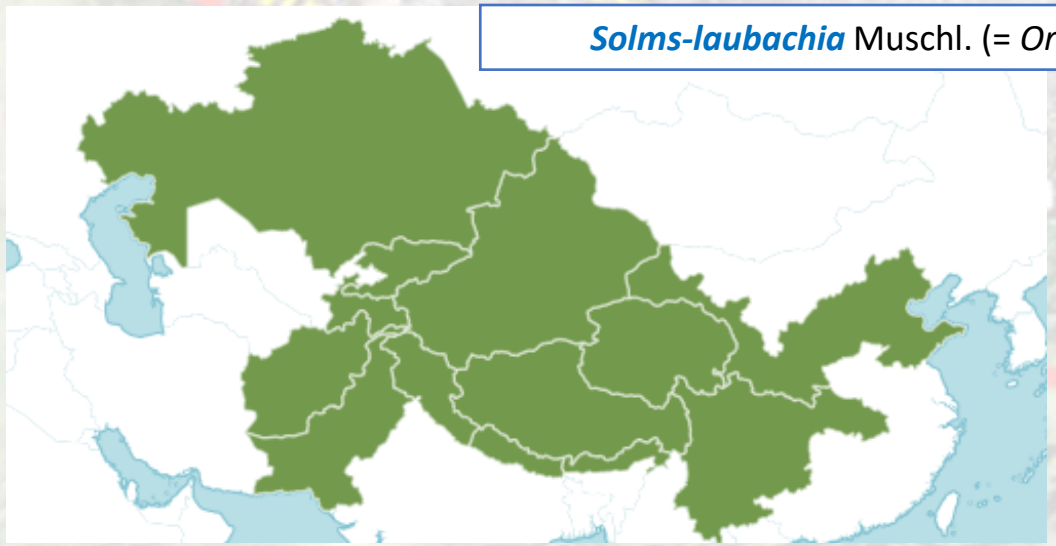
Eriophyton Benth. (= *Alajja* Ikonn.), 12 species



Thylacospermum Fenzl (2 species)



Solms-laubachia Muschl. (= *Oreoblastus* Suslova, *Phaeonychium* O.E.Schulz), 34 species







34 species distributed in Central Asia, Western Himalayas, Central Xizang (Tibet), and the Hengduan Mountains and all species occur at altitudes of (3200) 4000–6200 m (Yue et al., *JSE*, 2009**)**

II. Central Asia Biodiversity: Phylogenetic Knowledge Gaps Gaps

The Darwinian Shortfall

Allium L.: Filling the evolutionary dimension of biodiversity data

Knowledge gap

Species lists and distribution records alone do not show show where unique evolutionary history persists.

Institute of Botany response

Integrating georeferenced records, plastome/ITS phylogenies, phylogenies, and spatial evolutionary metrics.

DATA
4,642 records
records
170 taxa

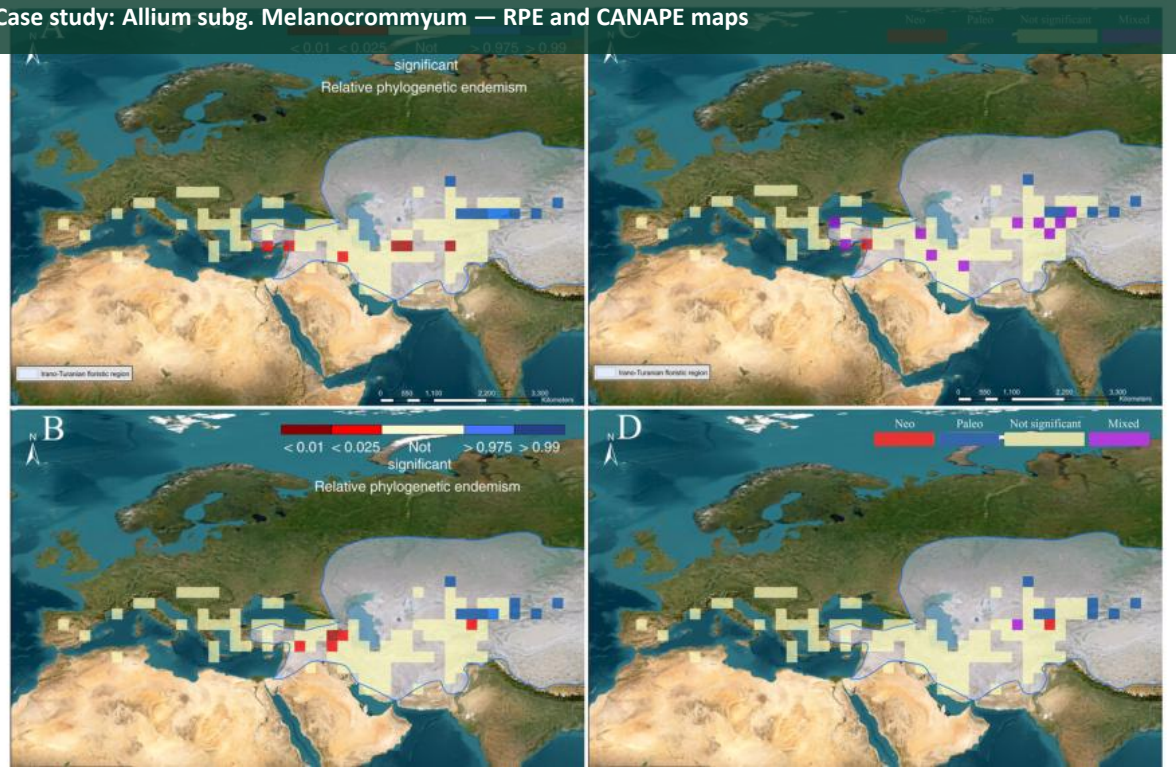
→

TREE
plastome
+ ITS

→

MAPS
PD · PE · RPE
CANAPE

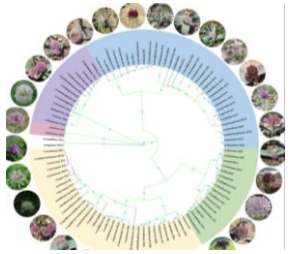
Case study: Allium subg. Melanocrommyum — RPE and CANAPE maps



Key conservation signal

Tian Shan and Pamir–Alay emerge as reservoirs of irreplaceable evolutionary diversity.

Shift in question: "Which species occur?" → "Where does evolutionary history persist?"



Phylogenies make evolutionary history spatially explicit

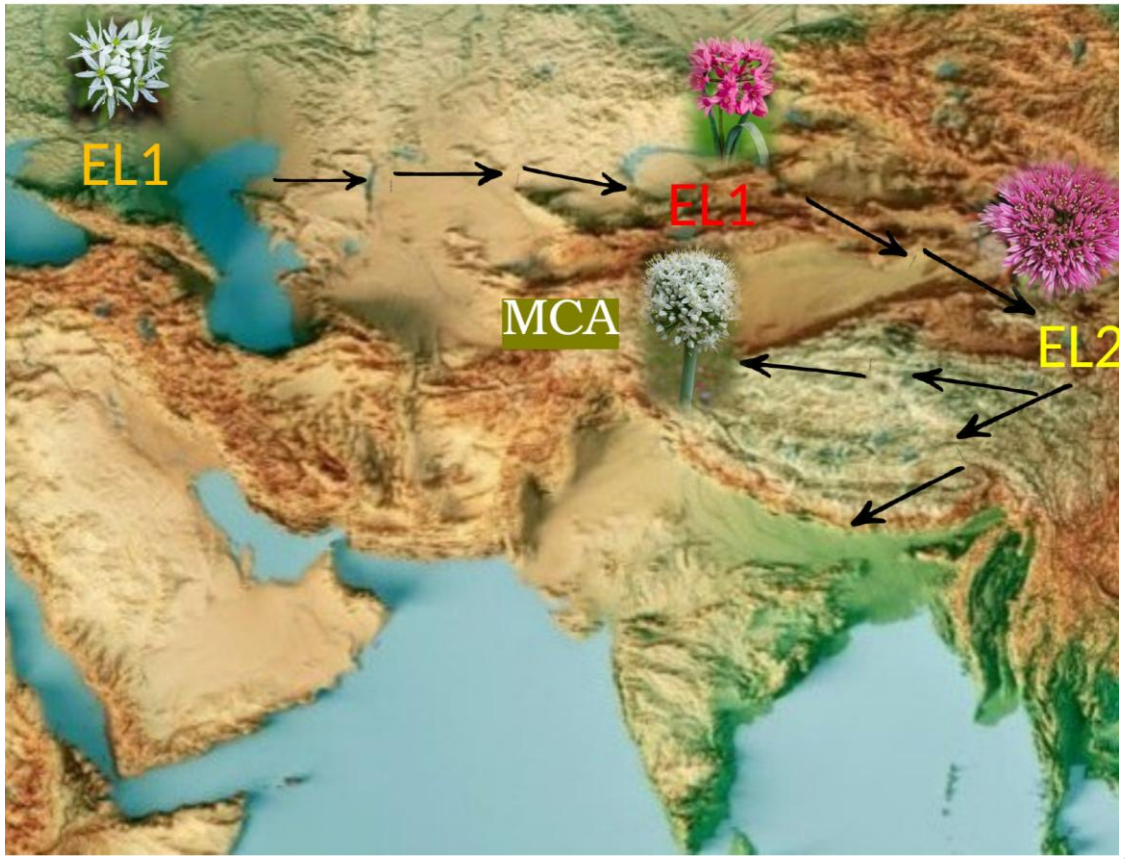
Institute of Botany studies convert occurrence data into evolution-based conservation priorities for Central Asia.

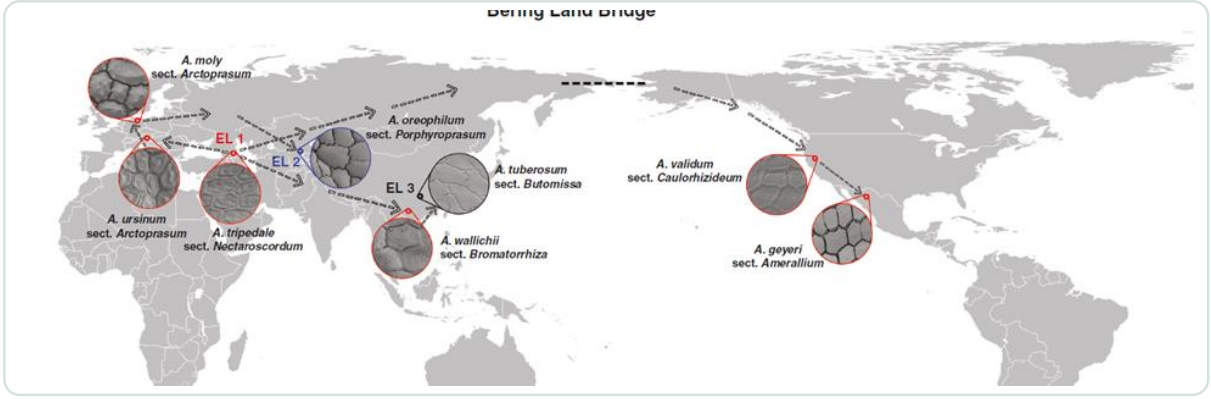
Institute of Botany · Digital Nature 2025–2029

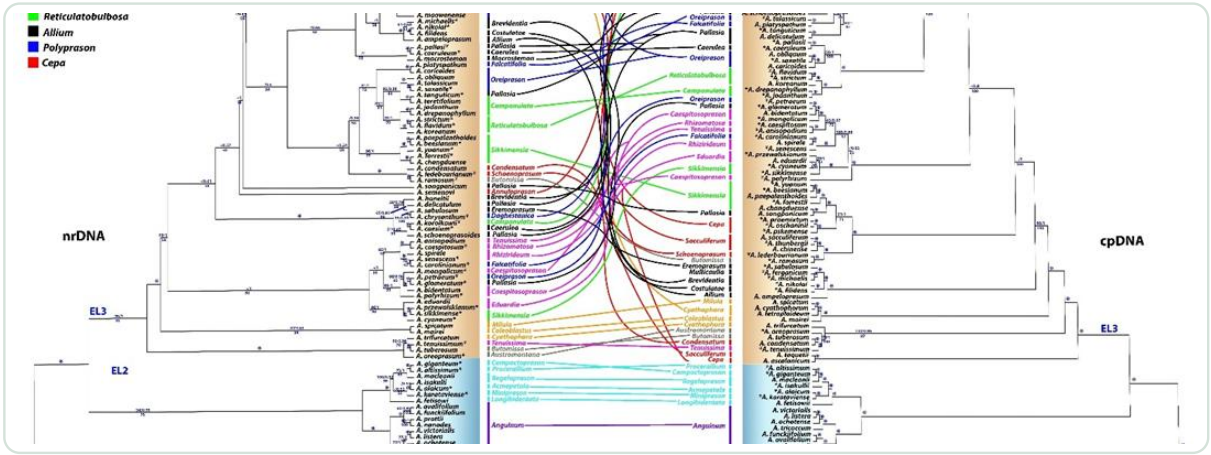
Ergashov et al. 2026, Botanical Journal of the Linnean Society

Allium L.: Evolutionary Lineages Across the Central Asiatic Highland Corridor

A phylogenetic framework links origin, lineage shifts and highland dispersal pathways.







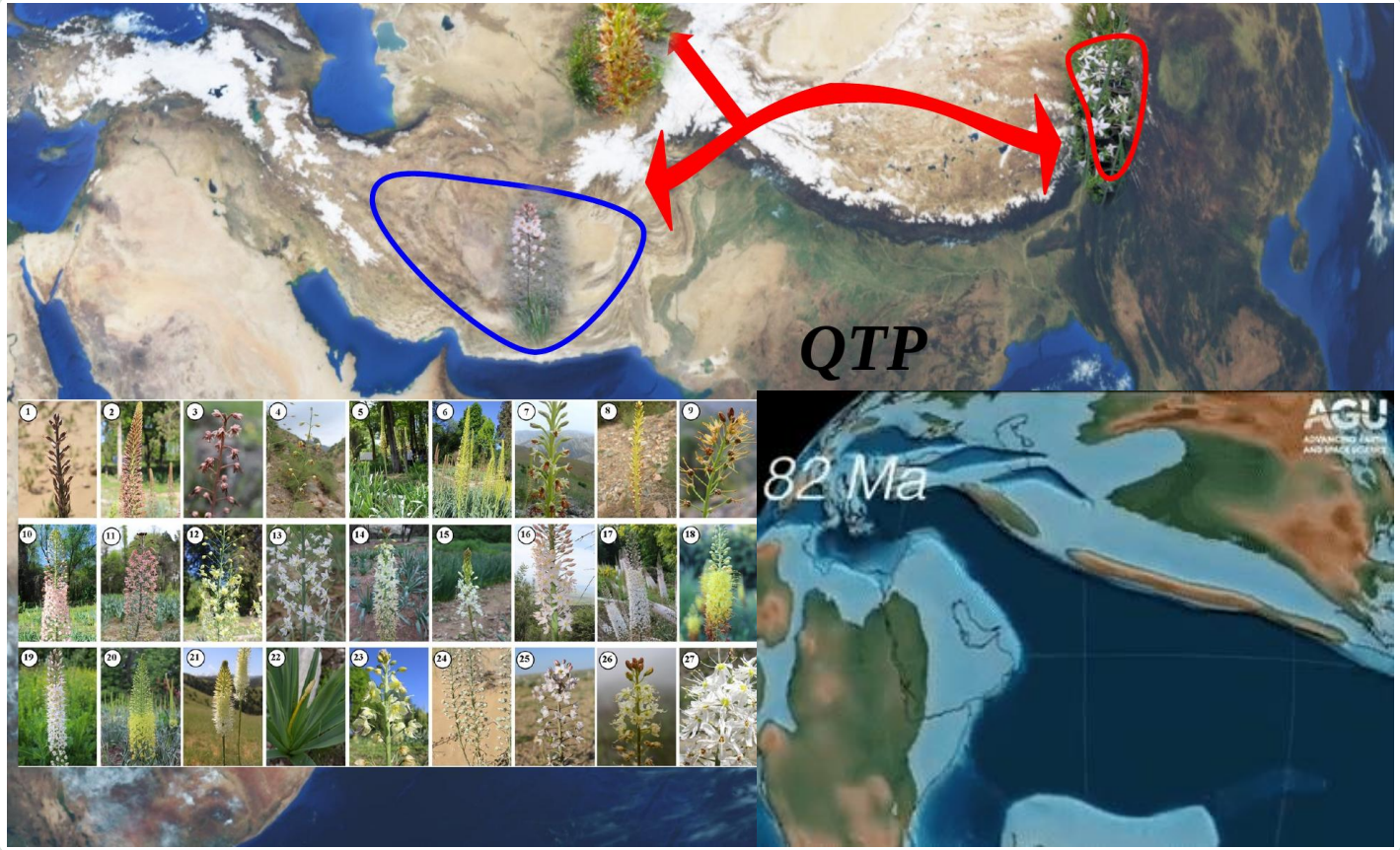
Key message: The Darwinian shortfall is narrowed by connecting floristic records with evolutionary lineages and geographic dispersal routes.

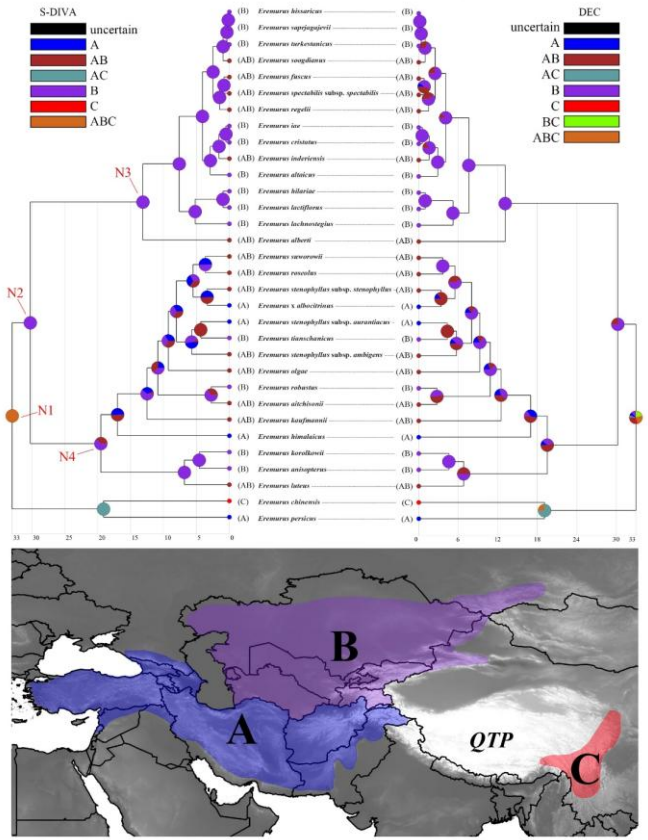
phylogeny → biogeography

Yusupov et al., Annals of Botany, 2022 | QTP = Qinghai–Tibet Plateau; MCA = Mountains of Central Asia

Eremurus M. Bieb.: Tethyan Origin and QTP-Driven Aridification

Time-calibrated phylogeny and ancestral-range reconstruction connect lineage evolution with Central Asian environmental change.





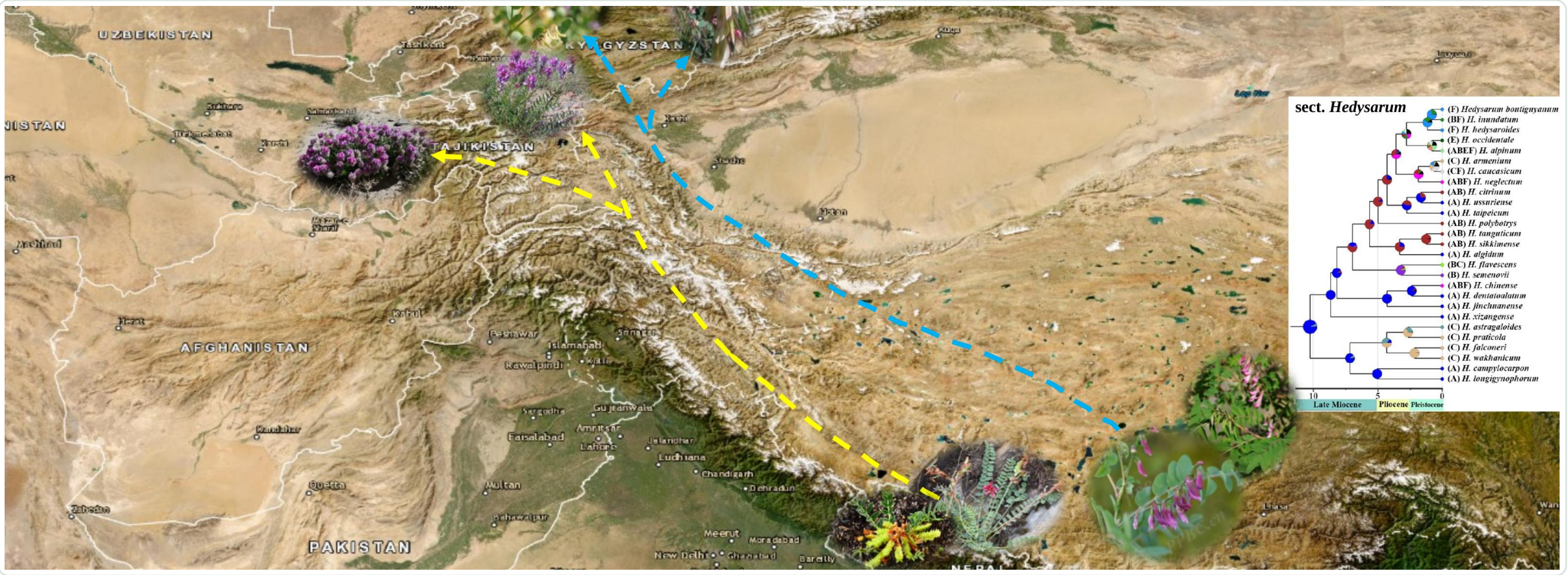
Key message: Institute of Botany research links plant diversification with the uplift of the Himalaya–QTP and aridification processes in Central Asia.

evolution + environment

Makhmudjanov et al. (2024), Plant Diversity 46(3):333–343

Hedysarum L.: QTP Origins and Central Asian Expansion

Phylogenetic reconstruction traces highland Fabaceae lineages from origin centres to regional dispersal pathways.



Key message: The Darwinian shortfall is reduced by integrating phylogenetic trees with spatial distribution patterns and inferred dispersal routes.

trees + maps

Juramurodov et al., Taxon, 2023, 72(6):1262–1284

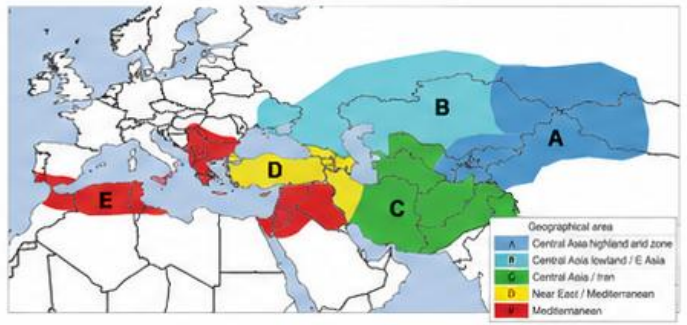
Tulipa L.: Central Asian Origin and Eurasian Dispersal

BioGeoBEARS ancestral-range models and dated phylogeny identify Central Asia as the evolutionary core of the genus.

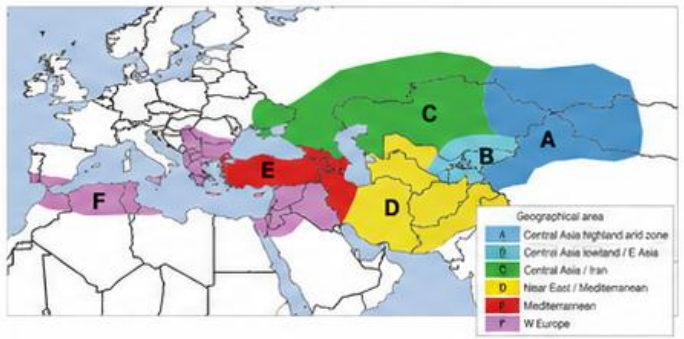
BioGeoBEARS models

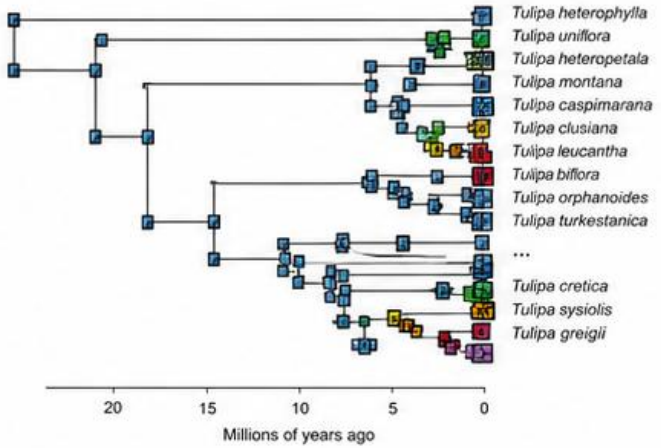
~86% of accepted species | 254 plastomes

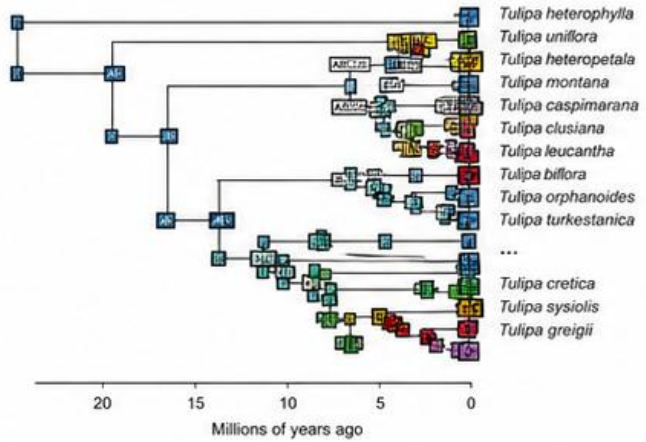
DEC model



DEC+J model

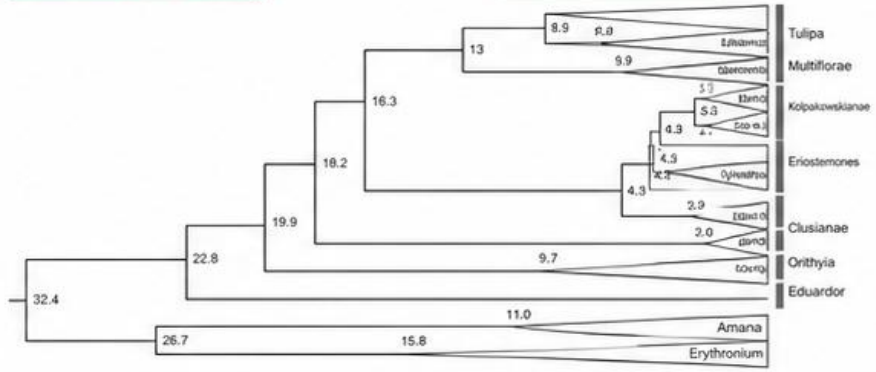




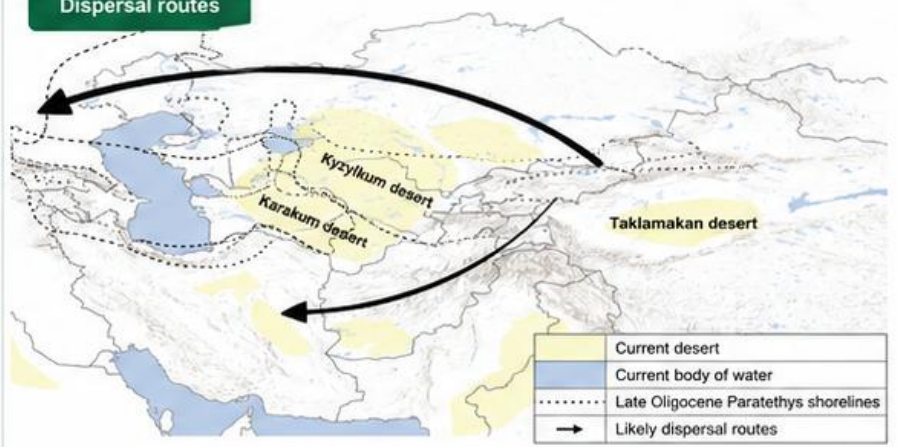


Dated phylogeny

MRCA ≈22.8 Mya



Dispersal routes



Key message: The Darwinian shortfall is reduced by linking ancestral-range modelling, dated phylogeny and dispersal pathways, revealing Central Asia as the evolutionary source and major expansion centre of *Tulipa*.

phylogenomics + biogeography

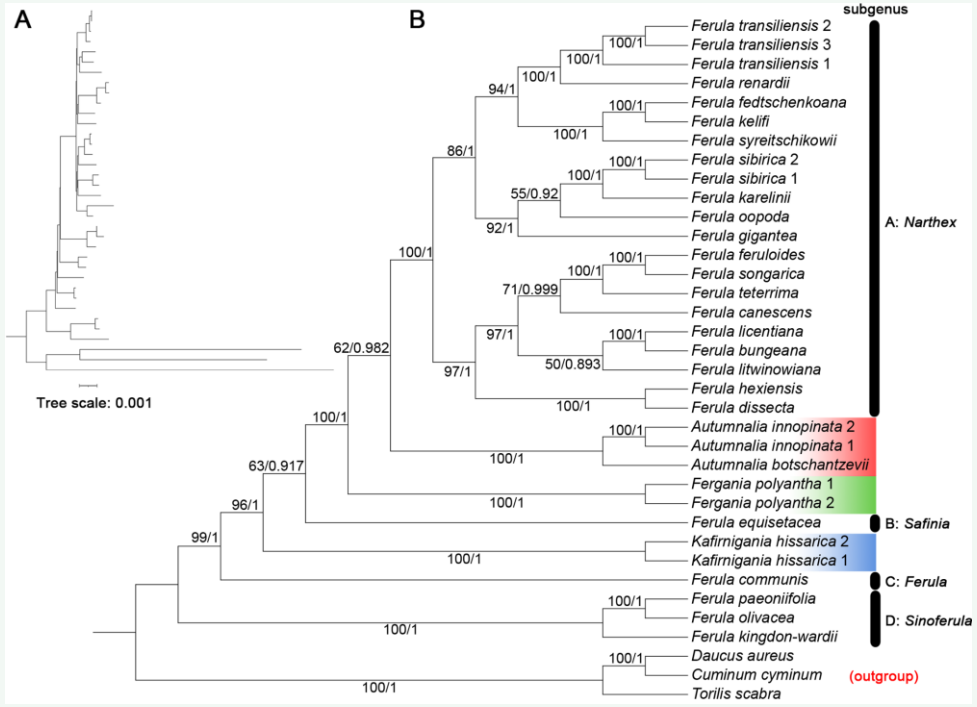
Wilson et al., American Journal of Botany, 2026

II. Central Asia Biodiversity: Phylogenetic Knowledge Gaps

Ferula L.: Resolving Generic Boundaries in Central Asian *Ferulinae*

Plastome, nrDNA and mericarp evidence place *Autumnalia*, *Fergania* and *Kafirnigania* within *Ferula*

Plastome phylogeny



Autumnalia, *Fergania,* and *Kafirnigania* are not isolated branches; they are nested within nested within the *Ferula* complex

Phylogenetic gap addressed

Three Central Asian endemic genera had uncertain placement and placement and unstable taxonomic boundaries.

7 new plastomes

36 plastomes in tree

55 nrDNA accessions



morphology also supports the revision

Outcome: *Autumnalia* + *Fergania* + + *Kafirnigania* → *Ferula*

This moves poorly resolved endemic genera into a tested evolutionary framework

Key message: The Darwinian shortfall is reduced by combining phylogenomics and morphology to clarify the evolutionary placement of Central Asian endemic genera

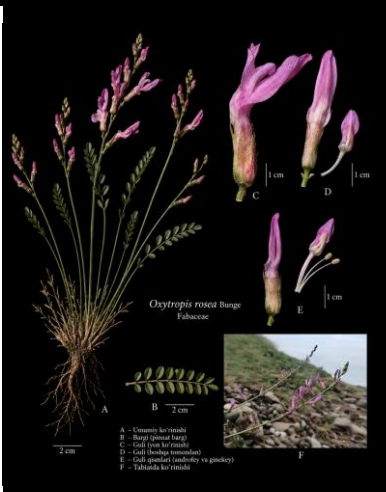
phylogenomics + taxonomy

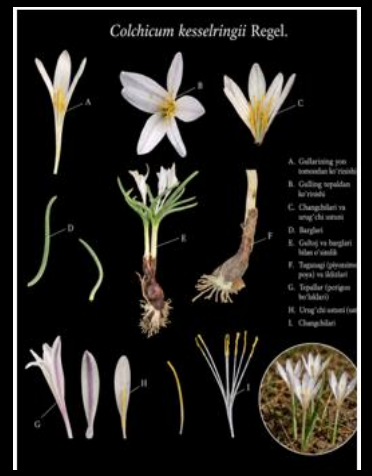
Yang et al., Taxon, 2026, 75(4): e70194 | Institute of Botany · Digital Nature 2025–2029

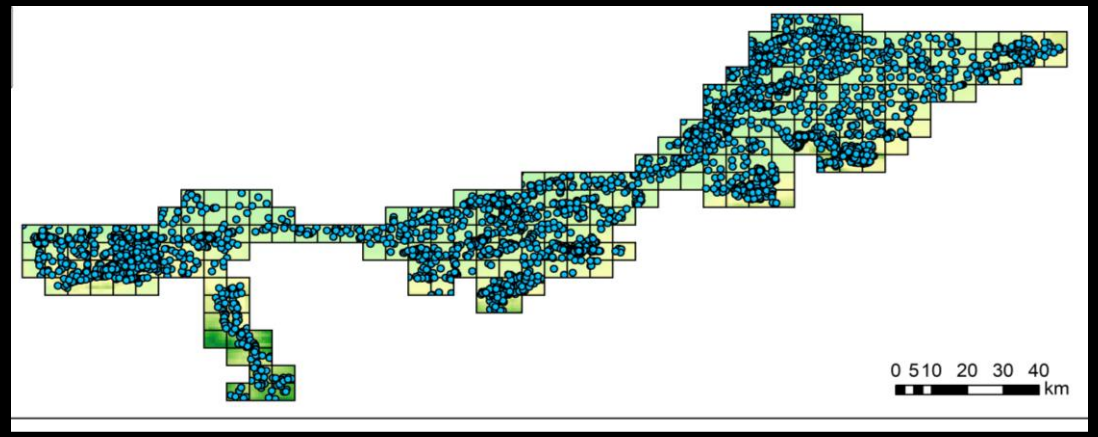
Classical Floristics Revisited: High Priority in Modern Research

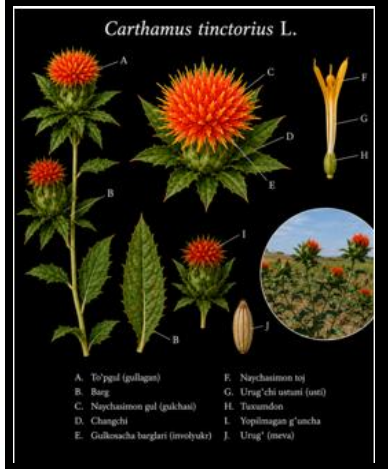


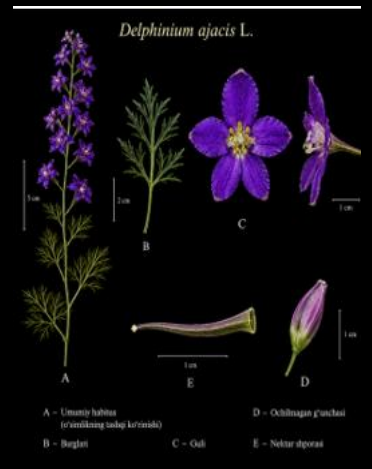






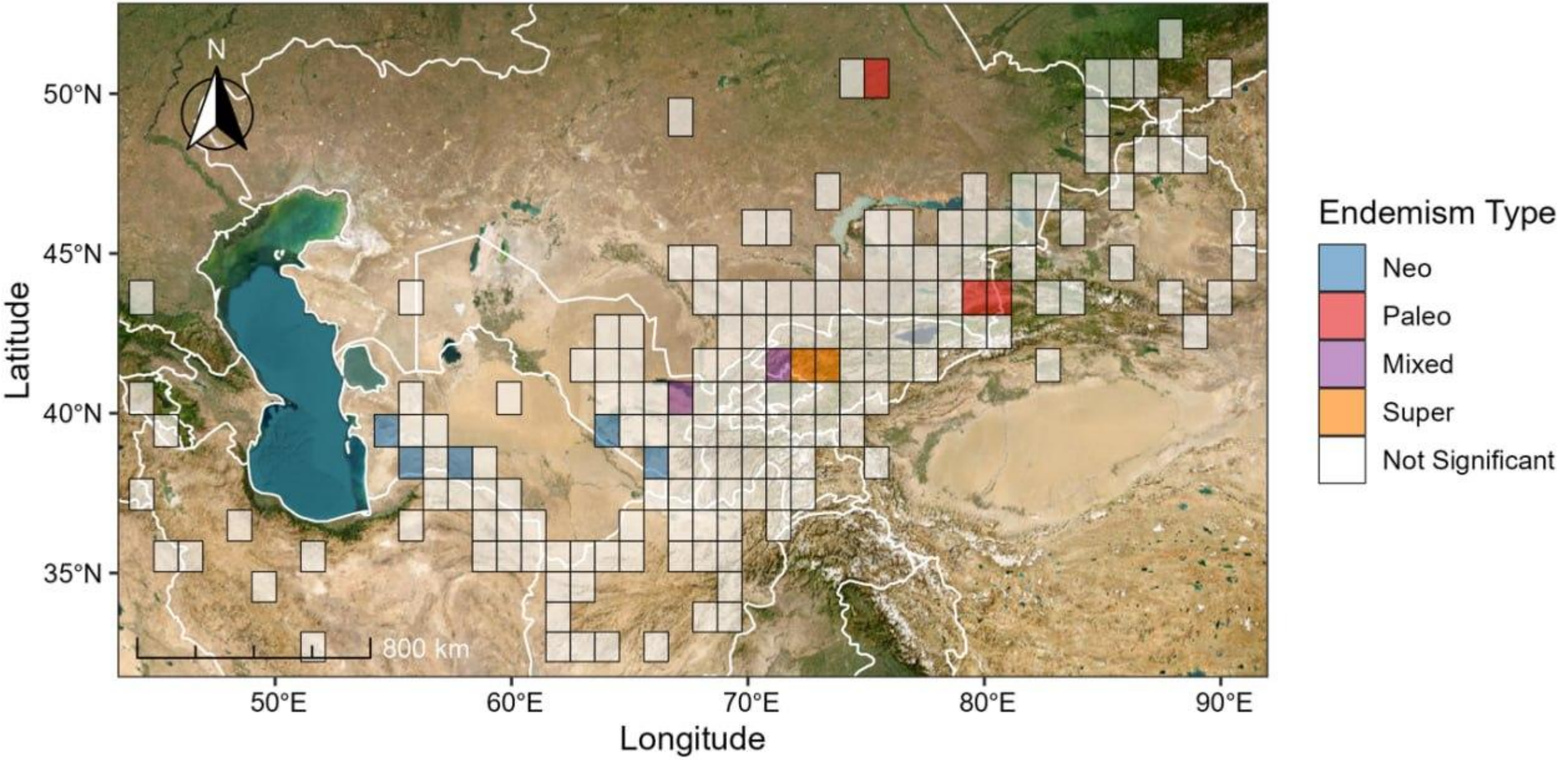






Early 2000s — 300–500 specimens per year: annual collections were limited and only occasionally approached 1,000 specimens.
No georeferencing: specimen records lacked precise coordinates, restricting spatial analysis and distribution mapping.
2025–2026 — up to 25,000 specimens per year: field collecting and herbarium accessioning increased dramatically.
Fully georeferenced collections: each specimen is now linked to accurate locality data and suitable for spatial research.
Open digital access: biodiversity records are published through **iNaturalist, GBIF, and GMOCA**.
Key message: classical floristics provides the verified data foundation for modern biodiversity research and conservation

Genus-Level Spatial Analysis of Regional Biogeographic and Evolutionary Patterns



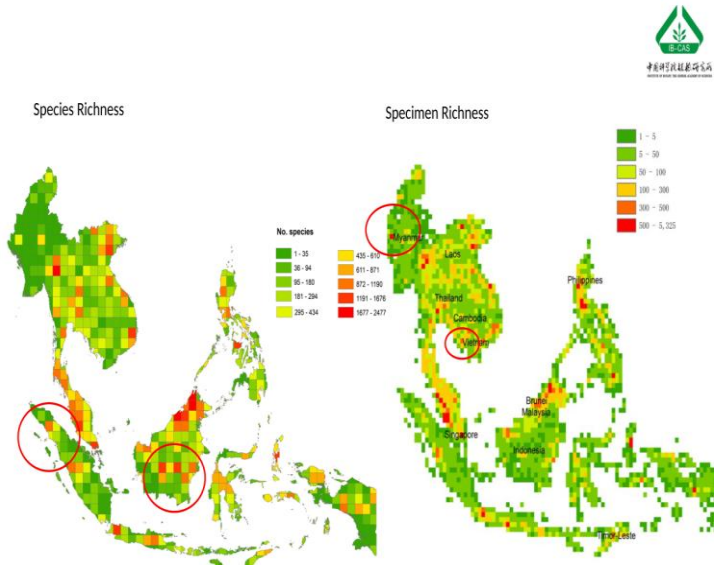
©Plant Species of Uzbekistan Group

III. Digital Nature Uzbekistan 2030: Building the Data Foundation

Rapid Acceleration of Biodiversity Data in Uzbekistan and Central Asia

Coverage is still more dynamic and uneven than in Southeast Asia, but data generation is now growing rapidly

Regional benchmark



Southeast Asia shows what an integrated grid-based assessment can look like.
Uzbekistan and Central Asia are not there yet — but the growth trajectory is now very fast.

Uzbekistan / Central Asia

Annual herbarium
>25,000

Field records
>25,000

2035 target
>4 million

Grid mapping

Expanding step by step from local studies to a regional platform

South of Uzbekistan
Pilot study

❯

Central Uzbekistan
Current study

❯

Uzbekistan
National-scale integration

❯

Central Asia
Regional floristic platform

Key message: **rapid growth of field collecting and data accumulation can bring Central Asia to a much higher, global significant level**

© Minzu University of China, MAP Group; ©Institute of Botany · Digital Nature 2030 Strategy

A critical expertise gap in flora studies and plant taxonomy

Tashkent as a historical center of outstanding botanical expertise in Central Asia



V.P. Drobov
(1885–1956)



R.I. Abolin
(1886–1938)



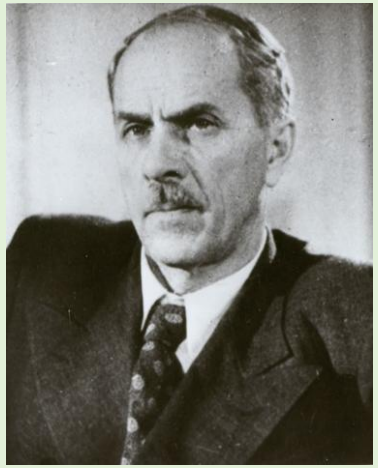
M.V. Kultiasov
(1891–1968)



M.G. Popov
(1893–1955)



A.I. Vvedensky
(1898-1972)



E.P. Korovin
(1891–1963)



K.Z. Zakirov
(1906–1992)



M.M. Arifxanova
(1911–1974)



S.S. Kovalevskaya
(1929–1981)



T.A. Adilov
(1920–1998)

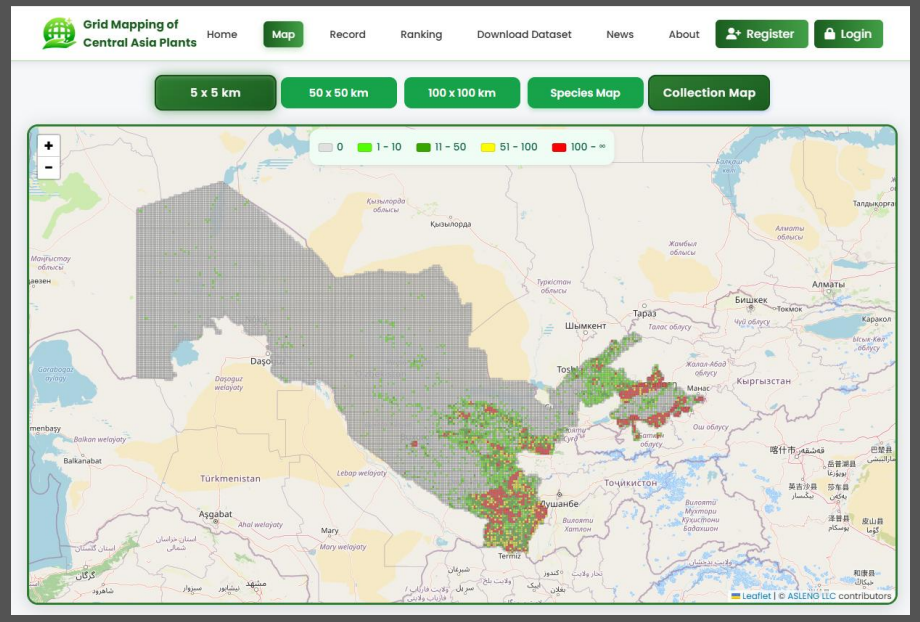


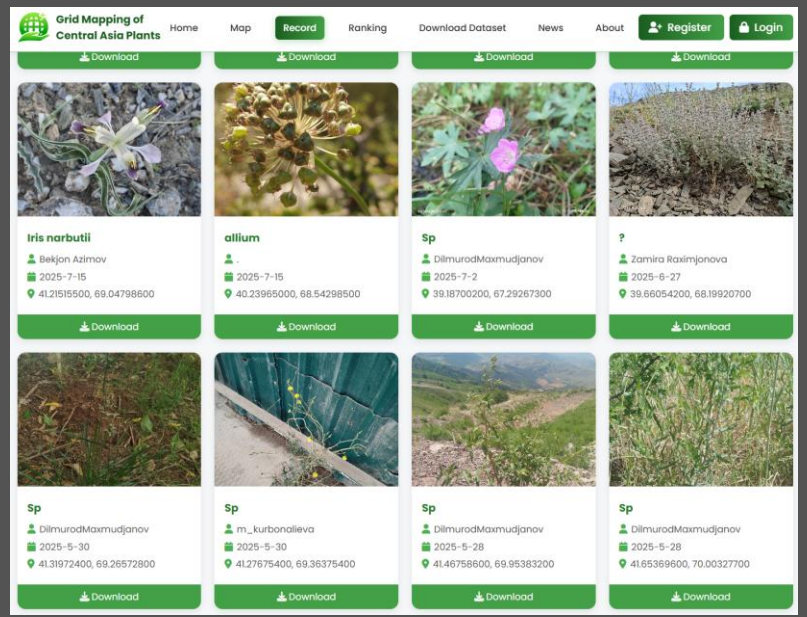
U.P. Pratov
(1934–2017)

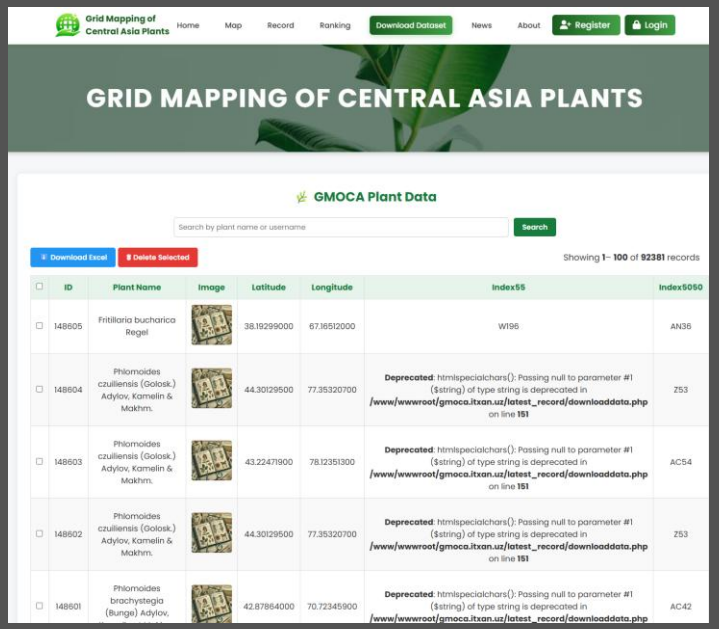


F.O. Khassanov
(1959–2024)

GRID MAPPING OF CENTRAL ASIA PLANTS (GMOCA.org)
serves as regional platform for collecting and accumulating plant data







https://gmoca.org/

Grid-based floristic mapping across Central Asia
Integration of herbarium, field and digital data
Standardized georeferenced biodiversity database
Supports large-scale spatial and phylogenetic analyses

Field data + Herbarium + Literature
↓
GMOCA (regional integration)
↓
GBIF (global integration)

GMOCA translates GBIF and or MAP principles into a regional, scalable biodiversity system, bridging local data with global biodiversity science

Thank you for your attention!

ACADEMY OF SCIENCES OF THE REPUBLIC OF UZBEKISTAN





