RESEARCH

Capacity building in crucifer research—from *BrassiBase* to a taxonomic expert network (*Brassi*TEN)

Markus Kiefer[1] · Dmitry A. German[2] · Christiane Kiefer[1] · Kasper P. Hendriks[3,4] · Andreas Franzke[5] ·
Roswitha Schmickl[6] · Nora Walden[1] · BrassiTEN · Marcus A. Koch[1,5]

Abstract
Networks of scientists working in coordinated efforts to advance plant systematics and taxonomy are becoming increasingly vital. Beyond the necessity of cataloguing and naming plant diversity—which is fundamental for supporting and safeguarding threatened biodiversity—there is also a growing need for robust comparative analyses across all taxonomic levels. Such endeavours rely heavily on a dependable taxonomic framework, ideally embedded within a solid phylogenetic context. *BrassiBase*, a taxonomic knowledge platform, was launched in 2012. Since then, significant progress has been made in phylogenetic research and taxonomic revisions, leading to the availability of comprehensive evolutionary concepts. A newly developed version of *BrassiBase* is presented here, now integrated with World Flora Online (WFO) and underpinned by the most current phylogenetic analyses. This version features an updated species checklist encompassing 61 tribes, 363 accepted genera, 4148 accepted species, 428 subspecies, 58 varieties (excluding nominotypical taxa), and a total of 10,725 synonyms. The phylogenetic placement tool (PPT), which utilises ITS sequence data, has been significantly enhanced and now incorporates off-target ITS sequences obtained through next-generation sequencing. These achievements are the result of collaborative work by a large, well-connected community of scientists dedicated to serving the broader scientific community. We advocate for interdisciplinary collaboration within coordinated networks to advance the taxonomy, systematics, and evolutionary research of crucifers, thereby laying the groundwork for future cross-disciplinary research in plant biology and beyond.

Keywords *BrassiBase* · Brassicaceae · Knowledge database · Phylogenetic placement · Taxonomic expert network · Taxonomy · World flora online

Introduction

The taxonomic history of the Brassicaceae stretches from the early 19th to the mid-twentieth century (De Candolle 1821; Burnett 1835; Prantl 1891; Hayek 1911; Schulz 1936; Janchen 1942), during which various, often

✉ Marcus A. Koch
marcus.koch@cos.uni-heidelberg.de

1   Centre for Organismal Studies Heidelberg (COS), Department of Biodiversity and Plant Systematics, Heidelberg University, 69120 Heidelberg, Germany

2   South-Siberian Botanical Garden, Altai State University, Barnaul, Russia

3   Functional Traits Group, Naturalis Biodiversity Center, Darwinweg 2, 2333 CR Leiden, The Netherlands

4   Department of Biology, Botany, University of Osnabrück, Osnabrück, Germany

5   Heidelberg Botanical Garden und Herbarium (HEID), Heidelberg University, 69120 Heidelberg, Germany

6   Faculty of Science, Charles University, Prague, Czech Republic

Springer

conflicting, classification systems were proposed (Koch et al. 2003a). The emergence of biochemical and molecular markers in the 1990s and early 2000s marked a new era in crucifer systematics, phylogenetics, and taxonomy (Koch et al. 2003a). At that time, the family was thought to comprise approximately 340 genera, around 3350 species, and 19 tribes. Over the following decade, classifications continued to evolve (Al-Shehbaz 2012), reflecting the limitations of early molecular systematics, which frequently relied on single markers or subsets of the nuclear and plastid genomes.

These developments led to significant changes: the number of recognised species rose by 9% to 3660, while the number of accepted genera declined by 9% to 321—highlighting the prevalence of morphological convergence and homoplasy in the family. However, molecular data also enabled the recognition of monophyletic groups that had previously gone unnoticed due to a lack of diagnostic characters, particularly at the generic and tribal levels. This culminated in the formal acceptance of 49 tribes in 2012 (Al-Shehbaz 2012).

Since the early 2000s, researchers from diverse disciplines have collaborated to clarify the evolutionary history of the Brassicaceae and to refine taxonomic frameworks that better reflect its phylogenetic relationships. Among numerous significant contributions, 2 stand out for their influence on taxonomic knowledge dissemination. First, Warwick et al. (2006) published a comprehensive species checklist—available both in print and as a digital Excel file—covering ~14,000 names with taxonomic status, authorship, literature references, and source verification (updated by Francis et al. 2021). Second, a complementary chromosome number database was published in the same journal issue (Warwick and Al-Shehbaz 2006), which has since shaped studies on Brassicaceae genome evolution (Lysak et al. 2009; Lysak and Koch 2011; Kagale et al. 2014; Guo et al. 2021; Mandáková et al. 2020; Walden and Schranz 2023).

These collaborative efforts led to the launch of the German research programme "Adaptomics" in 2011, which concluded in 2018. The initiative aimed to generate predictive insights into how plants adapt to environmental conditions, focusing on the Brassicaceae's ecological diversity. This programme promoted research beyond the model species *Arabidopsis thaliana* and encouraged exploration of evolutionary mechanisms in a broader phylogenetic context. A key infrastructural outcome of Adaptomics was the creation of *BrassiBase*, a knowledge database designed to integrate taxonomic, phylogenetic, genomic, cytological, and trait-based data for the Brassicaceae. Released in 2012 (Koch et al. 2012), *BrassiBase* offered tools for species enumeration, trait comparison, phylogenetic placement, and electronic identification keys. It has since undergone 2 updates (Kiefer et al. 2014; Koch et al. 2018) and continues to serve

as a critical resource for research into systematics, evolution, and biodiversity within the family (e.g. Koch and German 2013; Koch 2019; Koch et al. 2018; Kaya et al. 2022).

Following the last *BrassiBase* update (version 1.2a; Koch et al. 2018), the underlying species checklist and taxonomic backbone were revised in version 1.3 (June 2020). Continued research in taxonomy, systematics, and phylogenomics has since expanded our understanding of the Brassicaceae's evolutionary history, moving firmly into the genomic era. Phylogenetic analyses based on transcriptomes (Huang et al. 2015), plastid genomes (Guo et al. 2017; Walden et al. 2020), and targeted nuclear gene capture (Nikolov et al. 2019) have laid the foundation for current interpretations of lineage relationships. Broad-scale studies, such as the One Thousand Plant Initiative (2019) and Zunitini et al. (2024), have provided critical context at the angiosperm-wide level.

In 2024, shortly after the 20th International Botanical Congress (IBC2024) in Madrid, a dedicated taxonomic expert network—BrassiTEN—was formally established. This initiative brought together scientists with a shared focus on the Brassicaceae (Online Resource 1), many of whom had already been instrumental in shaping the database, phylogenies, and taxonomic insights mentioned above. Building on this expertise a comprehensive genus-level phylogeny based on over 1000 nuclear genes was published (Hendriks et al. 2023), confidently placing 319 Brassicaceae genera from 57 tribes into a well-resolved phylogenetic framework. While species-level resolution was not the primary goal of that study, it complements earlier ITS-based diversification analyses (Huang et al. 2019), which included more than 2000 taxa—all of which have been integrated now into *BrassiBase*. Implementing all this phylogenetic/-genomic knowledge led to a new taxonomic treatment (German et al. 2023), establishing 2 subfamilies—Aethionemoideae and Brassicoideae—and recognising 58 tribes organised within 5 supertribes: Arabodae, Brassicodae, Camelinodae, Heliophilodae, and Hesperodae.

These foundational achievements necessitated a comprehensive reassessment of the entire taxonomic backbone at species level, prompting a major new release of the *BrassiBase* species checklist. Concurrently, at IBC2024, the World Flora Online (WFO) initiative (Loizeau and Jackson 2017; Borsch et al. 2020) invited new collaborators to contribute Taxonomic Expert Networks (TENs). In response, the *BrassiBase* coordination team established an expanded expert consortium under *Brassi*TEN, representing broad expertise in taxonomy, evolutionary biology, bioinformatics, and collection curation.

This collaboration was formally recognised in 2024, and with the WFO December release, *BrassiBase* (version 2.0.1) delivered a major update to WFO's Brassicaceae taxonomy (The World Flora Online Consortium, 2024). The WFO now lists 366 genera and 4544 species for the family. In contrast,

BrassiBase maintains a more conservative count of 363 genera, 4148 accepted species, 428 subspecies, and 58 varieties (excluding nominotypical taxa). The discrepancy is mainly due to unvetted synonyms, often lacking voucher specimens or reliable metadata.

*Brassi*TEN (https://about.worldfloraonline.org/tens/brassicaceae) thus represents a growing, multidisciplinary network with a track record of joint projects and hundreds of publications on Brassicaceae taxonomy, systematics, and evolution over the past 25 years. An earlier, very rough bibliometric analysis on Brassicaceae (Zhou et al. 2024) provided evidence for a growing community and research output since the beginning of the twenty-first century. Obviously, research output of *Brassi*TEN members increased over time, but more importantly, the shared collaborative research output consistently increased as well (Fig. 1).

*Brassi*TEN members have played a leading role in shaping global crucifer research. This is evidenced not only by their extensive publication output, but also by their organisation of Brassicaceae-focused symposia at major scientific events. Dedicated sessions were held at International Botanical Congresses (IBC) in St. Louis (1999), Vienna (2005), Melbourne (2011), Shenzhen (2017), and most recently Madrid (2024). These meetings reflected

scientific progress over successive six-year periods. Notably, the 2005 Vienna symposium led to a thematic Special Issue published in *Plant Systematics and Evolution* (Koch and Mummenhoff 2006).

Beyond taxonomy and phylogenetics (German et al. 2023; Hendriks et al. 2023; Dominicus et al. 2025), *Brassi*TEN has been instrumental in developing new model systems and clades for comparative research (Makenzie et al. 2023). Nearly 15 years ago, several *Brassi*TEN members initiated the *Brassicales Map Alignment Project (*BMAP) under the DOE-JGI (http://bmap.jgi.doe.gov/). This initiative began with a "Top 20" list of Brassicales species—18 of which belonged to the Brassicaceae—and later expanded to a "Top 100" priority list for high-quality genome sequencing.

Central to this effort was access to well-documented, reproducible plant material. Here, *BrassiBase*'s germplasm repository proved vital, housing thousands of genetically defined accessions. This resource supported the publication of several influential genome studies (e.g. Jiao et al. 2017; Kiefer et al. 2019; Hämälä et al. 2024).

These collaborative structures reflect the increasing integration of *Brassi*TEN across disciplines—from classical systematics to cutting-edge genomic research. The
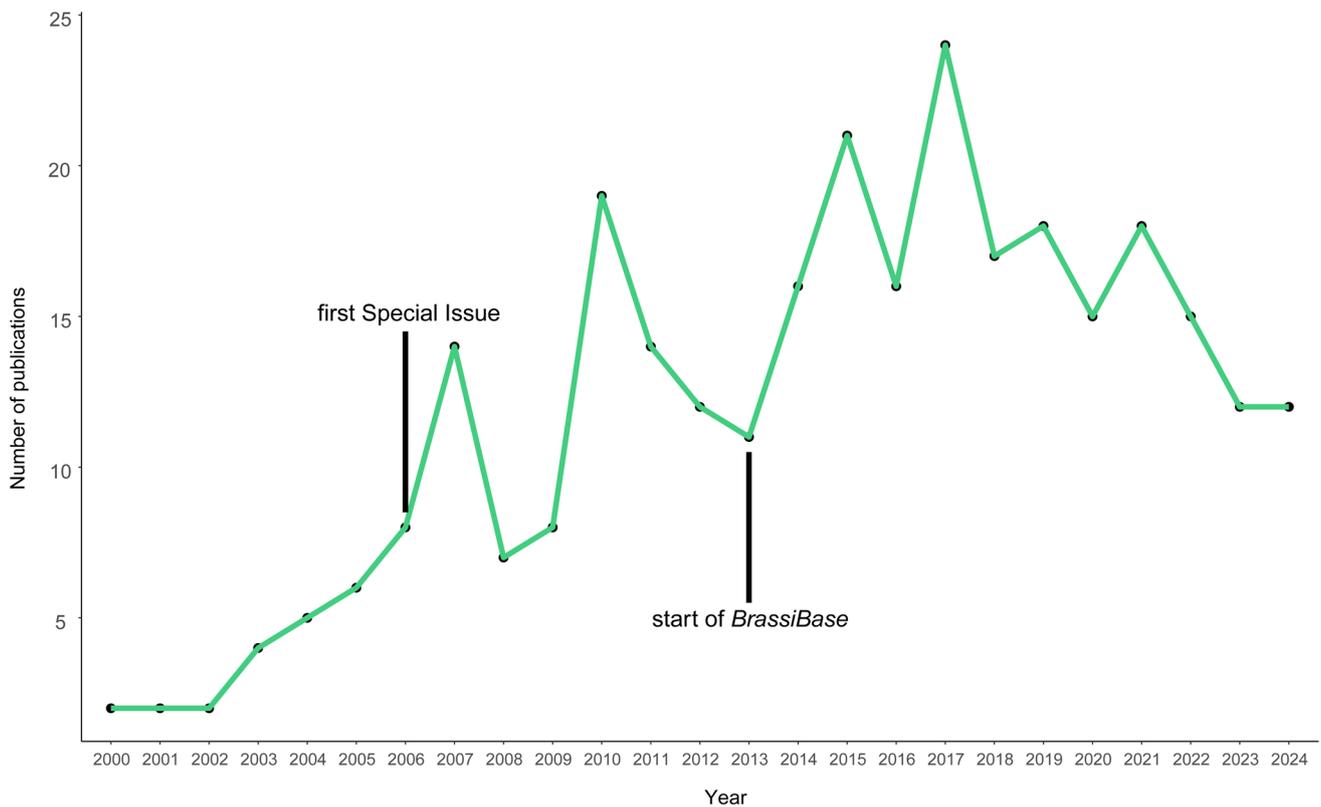

Fig. 1 Summary graph of publications focusing on Brassicaceae (*n* = 301) and authored by at least two *Brassi*TEN members (*n* = 26) throughout the last 25 years. The shared collaborative research is bridging not only scientists but also the various disciplines and key expertise represented by them. Accordingly, interactions displayed for example with joint publications are diverse and "reticulate" (Fig. 2)

group functions not merely as a taxonomic consortium but as a hub for innovation in evolutionary biology and comparative plant science.

The initial version of *BrassiBase* already included a DNA-based phylogenetic placement tool that utilised ITS (internal transcribed spacer regions 1 and 2 of nuclear-encoded ribosomal RNA) sequences from more than 2000 taxa. This tool was developed to address the frequent difficulties in accurately identifying specimens at the genus and species level—issues common in herbarium collections, germplasm banks, and research material, where misidentifications are widespread. *BrassiBase* offered a robust solution by providing a simple but effective Phylogenetic Placement Tool (PPT) based on curated reference alignments. Widely adopted, the PPT has supported both the taxonomic placement of unidentified samples and the formal description of new taxa (e.g. Dönmez et al. 2017, 2023; Koch and Lemmel 2019; Kantor et al. 2023). It has also been incorporated into best-practice guidelines for documenting material in metabolomics research (Đulović et al. 2024). While ITS sequences alone cannot always resolve genetic relationships at the species level, the PPT generally offers reliable placement within closely related groups. This allows users to proceed with more detailed identification using morphological keys, distributional data, and additional traits.

Moreover, the knowledge aggregated in *BrassiBase* has served as a vital resource for testing general hypotheses in plant evolutionary biology. These include studies on genome size variation (Lysak et al. 2009; Hohmann et al. 2015), diploid and polyploid evolution (Román-Palacios et al. 2020), morphological disparity (Walden et al. 2020), comparative genome evolution (Schranz et al. 2006; Murat et al. 2015; Lysak et al. 2016), diversification dynamics across the family (Huang et al. 2020), and mechanisms of concerted evolution (Koch et al. 2003b).

In this paper, we introduce *BrassiBase* v2025, a newly developed and substantially improved version of the database. It incorporates the latest taxonomic species checklist (version 1.7), which forms the dynamic backbone for all linked data within the system. This checklist is connected—via GenBank accession codes—to up-to-date phylogenetic inferences, including ITS1 and ITS2 sequences derived from both traditional Sanger sequencing and a new class of ITS data generated through short-read next-generation sequencing (NGS). In parallel, we highlight the formal launch of *Brassi*TEN at IBC2024 in Madrid—an initiative fully aligned with the goals of the World Flora Online (WFO) and marking a new phase of community-driven scientific collaboration. Through *Brassi*TEN and the innovations of *BrassiBase* v2025, we aim to contribute to the following objectives: a) Advancing the discovery and documentation of global plant biodiversity in coordination with WFO; b) Facilitating integration between scientific communities across both developed and developing regions, using BrassiBase as a shared research platform; c) Accelerating the exploration of Brassicaceae diversity in global biodiversity hotspots, notably the Irano-Turanian region, proposed as a possible centre of origin for the family (Al-Shehbaz et al. 2007; Cheo et al. 2001; Hedge 1976; Karl and Koch 2013; Mohammadin et al. 2017; Kantor et al. 2023); d) Identifying and resolving taxonomic and nomenclatural challenges through transparent scientific scrutiny; and e) Enabling deeper exploration of evolutionary principles within Brassicaceae and beyond.

Material and methods

The *BrassiBase* database comprises multiple interconnected data tables, each containing distinct types of information: the taxonomic checklist, ITS sequence alignments and reference trees (forming the basis of the "Phylogenetic Placement Tool"), a curated bibliography on Brassicaceae-related literature, genome size and chromosome number data, as well as a morphological character matrix. In the original version of *BrassiBase*, these datasets operated independently—changes in one table had no automatic effect on the others. With the current release, we introduce a fundamental conceptual shift: all components are now linked via a dynamic and flexible taxonomic backbone. This ensures that updates to taxonomic names are synchronised across all associated datasets without compromising their integrity or consistency.

From its initial release, *BrassiBase* included high-quality ITS (internal transcribed spacer) sequences—commonly used nuclear markers that separate ribosomal RNA genes—primarily obtained via direct Sanger sequencing using specific PCR primers. These sequences formed the foundation of the original Phylogenetic Placement Tool. However, most of these early ITS accessions lacked additional phylogenomic or multi-locus data. This limitation has been addressed in the current release. We now incorporate ITS sequences assembled de novo from Illumina short-read data, enabling their integration into reference alignments and phylogenetic trees based on large-scale, multi-gene datasets. The sequences were derived from material used in the latest Brassicaceae nuclear phylogeny (Hendriks et al. 2023), all of which was carefully curated by domain experts. As a result, *BrassiBase* now features object-based connectivity between ITS sequence data and the most comprehensive nuclear phylogeny currently available for the family. Moreover, the taxonomic checklist is now actively linked to all other data tables, ensuring that taxon names remain consistent and current across the entire platform.

The following section outlines the updated structure of *BrassiBase* and its core data components in more detail.

The core of *BrassiBase* is a single database table listing Brassicaceae taxonomic names

The central database table in *BrassiBase* stores core information for each taxon, including its taxonomic status, rank, and the corresponding literature reference. Each entry is linked both to the next higher hierarchical level (the "parent" taxon) and, where applicable, to the currently accepted name. As of now, the table includes 16,603 records. These records form the backbone for reconstructing taxonomic placements and relationships among names. All data within the *BrassiBase* knowledgebase are accessible via a web-based user interface and through a stable URL structure that allows individual entries to be bookmarked or shared. Each name is internally represented by a permanent identifier in the format *bb-rank-number*. For example, the species *Brassica oleracea* L. is stored as bb-spc-4454 and can be retrieved via the URLs: /names/get/id/bb-spc-4454 and /names/get/name/Brassica + oleracea. When additional data are available for a given name (e. g. voucher specimens, images, trait data), these are displayed alongside the taxon summary. Supplementary data categories, such as cytogenetic information or bibliographic references, are stored in linked tables. Curated bioinformatics resources—such as tribal reference alignments of ITS sequences and corresponding phylogenetic trees—are maintained separately and linked to the taxonomic database. The *BrassiBase* web interface, accessible at https://brassibase.cos.uni-heidelberg.de/bb2, provides intuitive, menu-driven navigation and search functionality. A dedicated downloads section enables users to retrieve datasets in a variety of formats.

During the transition from *BrassiBase* v1.3 to v2025, the previous version remains temporarily available and includes a link to the new release for continuity of access. From a technical perspective, *BrassiBase* v2025 is built on a conventional LAMP stack, comprising: a Debian Linux virtual machine, an Apache web server, a MariaDB database, and PHP as the primary scripting language. The user interface incorporates standard JavaScript for interactive components. Importantly, the program logic, data files, and user interface are separated from the core database, which is hosted on a separate virtual machine for improved performance and security. The underlying database structure is illustrated in Online Resource 2. For bioinformatic functionality, *BrassiBase* employs a pipeline composed of widely used and peer-reviewed tools, including: MAFFT (Katoh and Frith 2012) for multiple sequence alignment, RAxML (Stamatakis 2014) for phylogenetic inference, components of the EMBOSS suite (Rice et al. 2000), and Newick Utilities (https://github.com/tjunier/newick_utils) for tree processing

and visualisation. All software components were installed in their latest stable versions from official repositories to ensure reproducibility and compatibility.

The taxonomic backbone is key to sharing any information

In *BrassiBase*, the **Taxonomy** section enables users to access comprehensive taxonomic information at species, genus, and tribal levels. The most recent species checklist, along with related datasets, is available for download via the **Downloads** section. The previous version of *BrassiBase* (v1.3, released in 2018) used an earlier checklist that presented several limitations. It was difficult to update, could not easily incorporate external taxonomic content, and lacked programme-friendly access for data exchange with other systems.

As of this writing, the Brassicaceae species checklist has reached version 1.7 (2024), which resolves these issues through the following improvements: (1) Manual correction of inconsistencies: Numerous ambiguous or conflicting taxonomic entries—affecting both synonymy and rank assignments—were identified and corrected. (2) Revised data structure: The checklist is now maintained in a format that is easily updated and can be exported as a single, consistent table suitable for exchange with external systems such as the World Flora Online (WFO). WFO updates its plant species checklist biannually. For the December 2024 release, *BrassiBase* for the first time synchronised its checklist with WFO's Brassicaceae dataset. This involved mirroring the WFO data, identifying taxonomic contradictions and other discrepancies, and resolving these before sending back the corrected dataset for integration into the WFO release.

This update also implemented the latest tribal, supertribal, and subfamily classifications, following the structure proposed by German et al. (2023). The current design of *BrassiBase*'s taxonomy table is fully compatible with WFO's infrastructure and can directly contribute to its global taxonomy releases. Between formal updates, *BrassiBase* uses the WFO API internally to cross-reference its taxonomic content. This allows the system to detect, log, and address discrepancies in near real-time.

Looking ahead, *BrassiBase* intends to support WFO through regular updates, implemented via two complementary approaches: (i) Direct contribution via the Rhakhis Taxonomic Editor: Members of the Brassiaceae Taxonomic Expert Network (TEN) contribute directly to WFO through the Rhakhis Taxonomic Editor (https://plant-list-docs.rbge.info/rhakhis/), a user-friendly online platform for managing taxonomic names and classifications. TEN members log in with their ORCID ID, which tracks individual contributions and generates authorship for WFO outputs (e.g. the WFO Plant List in Zenodo, and records in Catalogue of Life's

ChecklistBank). TEN Leads hold editorial rights at the family level for their respective taxonomic groups. (ii) Synchronised and mirrored updates between BrassiBase and WFO: Regular updates are mirrored between BrassiBase and WFO, enabling cross-checking and alignment of taxonomy. This dual approach supports both individual and collaborative efforts, maximising the impact and accuracy of BrassiTEN contributions.

As a result, all species-level information in *BrassiBase*—such as chromosome counts, genome size data, gene alignments, and GenBank accession numbers—is dynamically linked to the current species checklist, which serves as the central data backbone of the system.

A phylogenetic tool for fast and reliable tests for systematic context

The *Phylogenetic Placement Tool* (PPT) in *BrassiBase*, introduced in the first release (Koch et al. 2012; Kiefer et al. 2014), primarily serves as a DNA identification tool. It places user-supplied ITS sequences (internal transcribed spacer regions) into curated tribal reference alignments and returns a maximum-likelihood tree that includes the query sequence (see Kiefer et al. 2014 for technical details). The ITS region is one of the most widely used genetic markers in Brassicaceae systematics and beyond. Despite certain limitations—such as its high tandem-copy number, paralogous loci, potential chimeric sequences, and patterns of both concerted and non-concerted evolution (Koch et al. 2003b)—it remains highly informative for genus- and tribal-level analyses (e.g. Álvarez and Wendel 2003; Warwick and Sauder 2005; Bailey et al. 2006; Agerbirk et al. 2021). However, ITS alone cannot reliably resolve deeper phylogenetic relationships (e.g. among tribes or subfamilies), nor is a single marker sufficient to reconstruct complex evolutionary histories shaped by polyploidy, hybridisation, and reticulate evolution. Beyond phylogenetic inference, ITS has been widely used in DNA barcoding (e.g. Duan et al. 2019; Banchi et al. 2020), offering a rapid method for identifying unknown or fragmentary material—such as sterile vouchers, seed lots prior to germination, or morphologically incomplete specimens.

Earlier versions of the *BrassiBase* PPT relied on Sanger-sequenced ITS data, including sequences from global diversification studies such as Huang et al. (2020), which drew on high-quality systematic, taxonomic, and phylogeographic research. In the current release, these data have been expanded with newly assembled ITS sequences derived from Illumina sequencing via de novo assembly. These sequences originate from the genus-level Brassicaceae phylogeny by Hendriks et al. (2023; https://treeoflife.naturalis.nl/brassicaceae), which used targeted gene capture of over 1000 nuclear loci. Additional,

previously unpublished data generated with identical laboratory methods were also incorporated. Importantly, Hendriks et al. (2023) enriched their sequencing libraries with whole-genome skimming data, enabling de novo ITS assembly for each sample. Their work also re-analysed earlier data from Nikolov et al. (2019), which we included as well.

To generate these "virtual ITS" sequences, raw reads were trimmed using Trimmomatic v0.40-rc1 (Bolger et al. 2014), applying the following criteria: Removal of low-quality bases using a 4 bp sliding window (average Phred score < 25), removal of leading/trailing bases with Phred < 20, removal of reads shorter than 50 bp. De novo assembly was performed with SPAdes (Prjibelski et al. 2020) using standard parameters. Promising contigs were identified via BLAST (Altschul et al. 1990), aligned with MAFFT v7.520 (Katoh and Standley 2013) against existing *BrassiBase* reference alignments, and manually refined in AliView (Larsson 2014). Sequences under 500 bp or those showing random SNP patterns (likely due to assembly errors) were excluded.

Out of 1385 sequenced libraries (including 137 duplicates), 746 high-quality ITS assemblies remained, representing 1221 accessions from Hendriks et al. (2023), a follow-up study (Hendriks et al. in prep.), and Nikolov et al. (2019) (Online Resource 3). This dataset enables direct object-based linking in BrassiBase between: curated herbarium vouchers, the most comprehensive genus-level Brassicaceae phylogeny to date, and the corresponding ITS sequences.

BrassiBase also integrates a curated selection of ITS sequences from GenBank (NCBI). Although GenBank contains approximately 13,000 ITS sequences from ~ 2000 Brassicaceae taxa (primarily from Sanger sequencing), many entries lack validation or may be based on misidentified specimens. We therefore performed a rigorous review, accepting only sequences from peer-reviewed studies in systematics, phylogenetics, population genetics, or evolutionary biology. Particular attention was given to: verifying taxonomic assignments, correcting outdated or invalid names, and handling chimeric sequences (e.g. joined ITS1 + ITS2 from separate PCR reactions). Sequences under 500 bp were excluded due to the risk of incomplete placement. Following curation, 11,407 high-quality ITS sequences from GenBank were retained. Together with the 746 virtual ITS sequences, BrassiBase now hosts a total of 12,145 ITS sequences (Online Resource 4).

The updated Phylogenetic Placement Tool now follows a two-step process: BLAST Pre-Screening: The user's query sequence is first compared against the complete *BrassiBase* ITS library. Based on top matches, likely tribal affiliations are suggested. Phylogenetic Placement: The selected tribal alignment is then updated with the user's sequence via MAFFT, and the query is placed using RAxML's

evolutionary placement algorithm. The resulting tree is rendered as an SVG image using *NewickUtils*' nw_display. This new approach is significantly faster, more resource-efficient, and supports batch analysis by accepting multi-FASTA uploads.

Phylogenetic placement and reference data sets

The revised ***Phylogenetic Placement Tool*** (PPT) was tested using two independent datasets:

(1) Validation of GenBank Sequences: During the integration of 9286 ITS sequences from GenBank into the *BrassiBase* reference alignments, sequences that appeared suspiciously divergent (based on multiple sequence alignment comparisons) relative to other sequences from the same species were excluded from incorporation. This resulted in the removal of 380 sequences, representing an approximate 4% exclusion rate. These flagged sequences were subsequently analysed via BLAST against the BrassiBase reference ITS library. We compared the genus-level identity provided in GenBank with the genus assignment based on our BLAST results, and cross-checked both against tribal assignments derived from our internal taxonomic reference Table (2) Cross-Validation of de novo ITS assemblies: A second validation involved the de novo assembled ITS sequences derived from Illumina data as part of the recent genus-level Brassicaceae phylogeny (Hendriks et al. 2023; Hendriks et al. in prep.). After excluding highly divergent sequences (likely assembly artefacts), the remaining high-confidence sequences were subjected to BLAST analysis against the curated ITS sequences from GenBank already integrated into BrassiBase. For this analysis, we compared taxonomic delimitation of those accessions being already published (Hendriks et al. 2023) and which had been repeatedly validated by *Brassi*TEN experts and were assumed to be highly reliable (near 100% accuracy), and the unpublished Illumina-derived sequences, which are currently under review for inclusion in a forthcoming phylogenetic study. This comparative approach not only validates the robustness of the updated PPT pipeline but also helps identify a small subset of accessions requiring further curation. Importantly, the likelihood of misidentification or contamination is minimised by the fact that sequencing material was carefully selected and pre-screened by Brassicaceae experts within the *Brassi*TEN network.

Citation analyses and collaboration network graphs

To highlight and quantify collaborative research efforts among Brassicaceae TEN members we collected individual publication lists of TEN members from 2000 onwards. Restricting analyses to peer-reviewed contributions, we counted those publications focusing on Brassicaceae ($n = 301$) and co-authored by at least two out of the 26 *Brassi*TEN members throughout the last 25 years. These data were used to generate a chord diagram illustrating co-authorship links between TEN members with two or more shared publications over the past 25 years. For each publication, pairwise collaborations were calculated and normalised by the total number of TEN co-authors per paper. Data analysis and visualisation were carried out in R 4.4.2 (R Core Team 2024), using the packages: DescTools v0.99.59 (Signorell 2025), dplyr v1.1.4 (Wickham et al. 2023), and circlize v0.4.16 (Gu et al. 2014) for plotting the chord diagram. To place the activities of *Brassi*TEN members in a wider scientific context and assess their interdisciplinary outreach, we conducted a broader bibliometric analysis using Scopus (2000–2025). This search targeted peer-reviewed and indexed publications using the keyword "Brassicaceae", without restricting results to TEN members. To focus on impactful contributors, we applied the following filters: authors must have published at least ten relevant articles, and must be connected by at least one co-authorship link. This yielded 8501 citation records, forming the basis for a researcher association network graph comprising 124 scientists, including 13 *Brassi*TEN members. In this network, citation clusters and co-authorship links were used to visualise collaborative dynamics across disciplines.

All network analyses were conducted using VOSviewer v1.6.18 (Perianes-Rodriguez et al. 2016) in conjunction with the Web of Science citation database. Network generation was based on the following settings: association strength as the similarity measure, default layout parameters, resolution $= 1.00$, minimum cluster size $= 1$, qttraction $= 1$, and repulsion $= -2$. The interactive version of this network is accessible via the online VOSviewer app, allowing users to explore and adjust visualisation parameters dynamically. The full network dataset is published on Zenodo (https://zenodo.org/records/15652856) including all underlying author lists, metadata files. The JSON file can be viewed via a direct link using the vosviewer app (https://app.vosviewer.com/?json=https://zenodo.org/records/15652856/files/20250605_brassicaceae_network.json?/content) (Guerin et al. 2025).

Further improvements

Additional information available in *BrassiBase* from earlier releases includes a substantial cytogenetic dataset comprising DNA content and ploidy level data for numerous Brassicaceae accessions. Since the last version, this dataset has been significantly expanded, critically revised, and made more accessible through the implementation of an improved filtering function, allowing faster and more targeted queries. The literature database has also been a core component of *BrassiBase* since its inception. It has been continuously updated and is now more easily searchable through enhanced

interface features. The current version includes approximately 9100 entries, covering: chromosome data sources, original species descriptions (protologues), and taxonomic, systematic, and phylogenetic studies. A systematic incorporation of literature from the last five years is currently underway. The previous *BrassiBase* release also introduced a morphological database, based on the NaviKey identification system (http://www.navikey.net/) developed and provided by Ihsan Al-Shehbaz (Missouri Botanical Garden). This module enables the categorisation of Brassicaceae genera by morphological traits and supports the exploration of the family's morphospace. While no new data have been added to this module in the current release, the existing dataset remains available and has been linked to the revised taxonomic framework wherever possible. The cytogenetic data module continues to be actively maintained. In this release, an additional 131 chromosome count references were incorporated, integrating the latest available data from the IAPT/IOPB chromosome database (https://www.iaptglobal.org/chromosome-data).

Results and discussion

The new version 2.0.1 of *BrassiBase* is accessible via the same URL as previous releases: https://brassibase.cos.uni-heidelberg.de/. This release consolidates and expands upon all previously introduced functionalities, while introducing several major improvements: (i) A fully revised and updated taxonomic checklist, (ii) the implementation of this checklist (accepted names only) as a dynamic and central data backbone for the entire platform, and (iii) the integration of ITS sequences derived from specimens included in the most recent genus-level Brassicaceae phylogeny (Hendriks et al. 2023), thus linking *BrassiBase* to the most current and robust phylogenetic framework available, and to well-documented specimens reviewed by *Brassi*TEN members. Taxon names in *BrassiBase* remain cross-referenced with external databases such as IPNI (https://www.ipni.org/) and Tropicos (https://tropicos.org/), and are now also mirrored with World Flora Online (WFO) to ensure taxonomic synchronisation across platforms. Over the past two decades, the *Brassi*TEN network has developed into one of the most influential and collaborative research communities in Brassicaceae science (Fig. 3). Accordingly, we view *BrassiBase* and *Brassi*TEN not only as a foundation for advancing taxonomy and systematics, but also as a platform for: bridging disciplines, promoting integrative and collaborative research, and fostering innovation in emerging scientific fields related to Brassicaceae.

The taxon checklist turned into a dynamic backbone taxonomy

As of 25 April 2025, the *BrassiBase* taxonomic checklist comprises 2 subfamilies, 61 tribes (three of which are candidates for further division, based on recent phylogenetic evidence; e.g. Walden et al. 2020; Hendriks et al. 2023), 363 genera, 4148 species, 428 subspecies, and 58 varieties (excluding autonyms). In addition, the database contains 10,725 synonym entries, broken down as follows: 8417 species, 1477 varieties, and 807 subspecies (autonyms excluded in all cases). By comparison, the current WFO (World Flora Online) checklist lists: 4544 species, 366 genera, 56 tribes, and 2 subfamilies. This reveals a discrepancy of 396 species, which, according to the *BrassiBase* checklist, should not be accepted. This discrepancy highlights the need for refinement in the forthcoming WFO December update, particularly concerning species delimitations. Addressing such inconsistencies is central to the mission of *Brassi*TEN, which is actively working to improve WFO data quality through systematic expert contributions. Similar workflows and outcomes have been documented for other plant groups, including: Caryophyllales (Hernández-Ledesma et al. 2015), Cactaceae (Korotkova et al. 2021), and the genus *Erica* L. (Elliott et al. 2024a, b). Since January 2025, WFO has begun sharing monthly updates of new records integrated into its taxonomic backbone via synchronisation with IPNI, streamlining curation and oversight for TENs. This enables a consistent and transparent workflow for updating taxonomic data.

Several other TENs have already established dedicated data portals that serve as knowledge hubs for their respective families, for example Solanaceae (https://solanaceaesource.myspecies.info/), *Begonia* L. (https://padme.rbge.org.uk/Begonia/), Asteraceae (https://www.compositae.org/), Fabaceae (https://www.legumedata.org/working-groups/taxonomy/), or Caryophyllales (https://caryophyllales.org/). These initiatives exemplify the collaborative and multidisciplinary spirit in modern plant systematics. Likewise, *Brassi*TEN is committed to sharing expertise across disciplines, fostering collaboration between taxonomists, phylogeneticists, geneticists, and other researchers in plant sciences.

An analysis of the publication history of *Brassi*TEN members over the past 25 years (Fig. 2) illustrates a strong record of sustained collaboration, providing a solid foundation for maintaining and updating Brassicaceae taxonomic knowledge into the future. The scientific impact and outreach of *Brassi*TEN is further demonstrated by a co-authorship network based on publications between 2000 and 2025 that include "Brassicaceae" as a keyword. This network highlights researchers with at least ten publications, and shows that most *Brassi*TEN members are centrally

Fig. 2 Chord diagram of shared publications (*n* = 301) on Brassicaceae among *Brassi*TEN members (*n* = 26) from 2000 to 2024 indicates highly networked collaboration
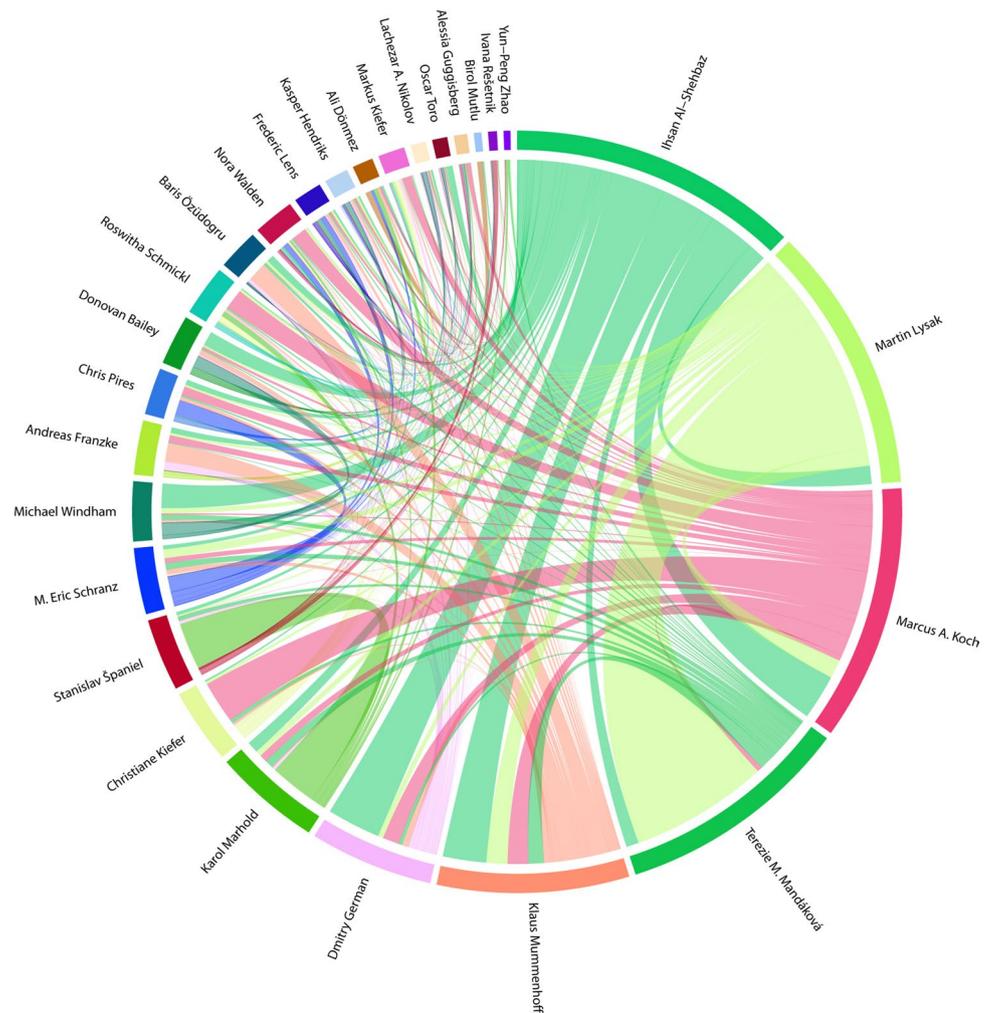
positioned, often acting as key connectors between sub-networks (Fig. 3).

ITS reference sequence data in *BrassiBase* demonstrate taxonomic and technical improvement

From its inception, *BrassiBase* has aimed not only to provide robust taxonomic content, but also to offer a reliable tool for identifying Brassicaceae material, including herbarium vouchers, field samples, and germplasm accessions. Traditional identification is often hampered by missing diagnostic characters and the extensive homoplasy observed in nearly all morphological traits (Walden et al. 2020). As a result, reliable identification at the species, genus, or even tribal level can be challenging or impossible using morphology alone. To address this, *BrassiBase* implemented a phylogenetic placement approach using the Internal Transcribed Spacer (ITS) region as phylogenetic marker. ITS is widely used in Brassicaceae systematics and is relatively easy to obtain—even from aged herbarium specimens or seeds. The pipeline from DNA extraction to sequencing and analysis is rapid and cost-effective, making ITS a practical marker for large-scale applications. Over the years, the Phylogenetic Placement Tool (PPT) in *BrassiBase* has undergone continuous refinement.

Prior to the current release, *BrassiBase* included 2113 ITS sequences derived from Sanger sequencing, many of which were contributed by Huang et al. (2020). With version v2025, the ITS dataset has expanded considerably to include: 746 ITS sequences assembled from next-generation sequencing (NGS) data, and 11,407 Sanger-derived ITS sequences (including both legacy and newly curated sequences), resulting in a total of 12,145 ITS sequences currently housed in BrassiBase. In terms of taxonomic coverage, *BrassiBase* v2025 now includes: ITS data from all 61 tribes, 359 genera (previously 324), and 2578 species (previously 2014), representing an increase of 22% in tribes, 11% in genera, and 28% in species compared to earlier releases. Some tribal alignments were further subdivided due to evidence of non-monophyly, specifically in Iberideae, Camelineae, and Brassiceae, as demonstrated by Walden et al. (2020) and Hendriks et al. (2023). For
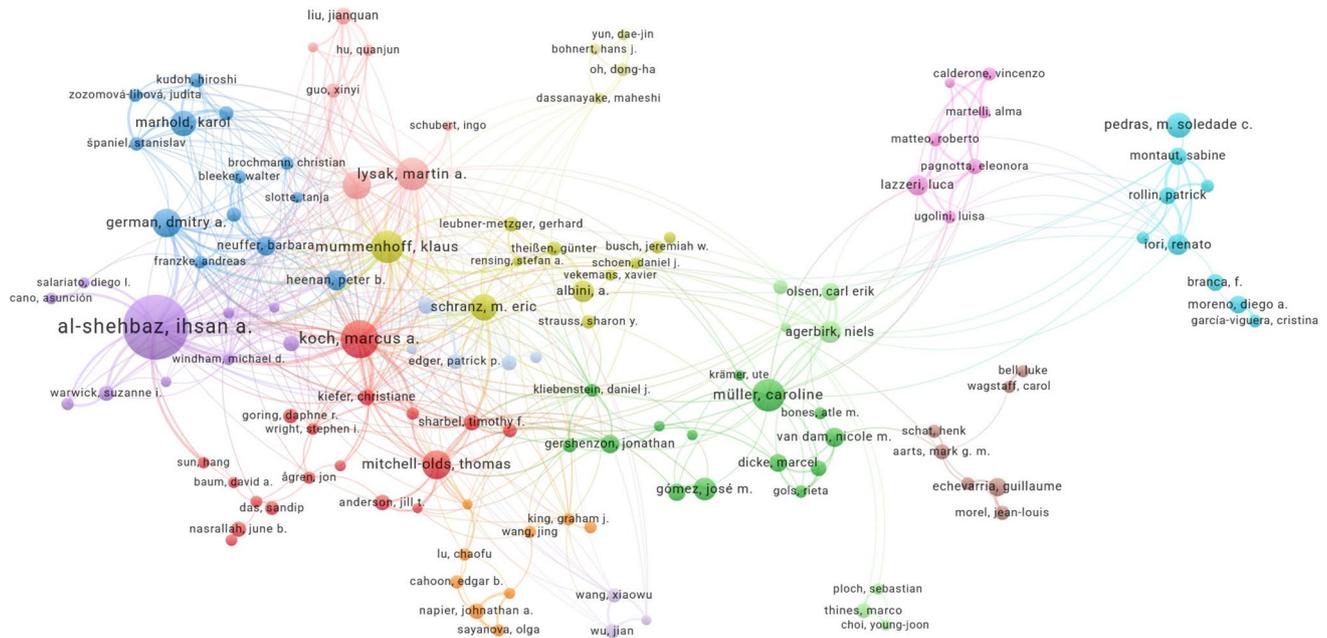
Fig. 3 Network visualization of shared publications (2000–2025) among researchers (keyword "Brassicaceae"; >10 documents). Out of the total 124 authors, 13 are members of *Brassi*TEN. The colour determines different clusters of researchers. Lines between researchers represent shared publications. In general, the closer 2 researchers are presented to each other, the stronger their relatedness. The interactive online copy of this analysis is available at [https://app.vosviewer.com/?json=https://zenodo.org/records/15652856/files/20250605_brassicaceae_network.json?/content]

example, the genus *Bivonaea* was affected by the restructuring of Brassiceae. A breakdown of the current ITS dataset shows: The 746 NGS-derived sequences cover 56 tribes, 240 genera, and 690 species. The GenBank-derived sequences represent 59 tribes (including the three split tribes), 355 genera, and 2427 species. In total, *BrassiBase* v2025 provides ITS representation for: 100% of recognised tribes, 99% of genera, and 62% of known species in the family.

This high level of coverage translates into a strong probability of identifying unknown samples at the genus or tribe level, and frequently even at the species level. Nevertheless, accurate placement may still depend on the evolutionary dynamics of ITS loci, such as concerted vs. non-concerted evolution and the effects of polyploidy (see Koch et al. 2003b for details). The future expansion of ITS coverage is expected to be gradual. Many remaining species are either rare, geographically restricted, or difficult to access, due to: lack of suitable voucher material, poorly documented distributions, or logistical challenges in fieldwork.

To mitigate these limitations, *BrassiBase* now incorporates "virtual ITS" sequences, assembled from short-read genome-wide Illumina datasets, even when ITS was not an explicit target. Such data often arise in studies aiming to capture single-copy genes or plastid genomes (e.g. Hendriks et al. 2023), and can yield sufficiently complete ITS sequences for analysis. Importantly, the ongoing Brassicaceae Tree of Life (*Brassi*ToL) initiative (Hendriks et al. in prep.) is expected to further enhance taxonomic resolution. The project will allow for comprehensive cross-linking of all available genomic and ITS data for each accession, thereby significantly advancing the power and precision of *BrassiBase*'s placement tools.

Test cases of ITS datasets highlight the general usability and future reliability of the "phylogenetic placement tool"

To evaluate the reliability of the Phylogenetic Placement Tool (PPT) and to gain insights into the expected error rates in taxon identification, we conducted 2 test cases: (i) Voucher-Based NGS Data (Hendriks et al. 2023 and in prep.): We first assessed a dataset derived from carefully curated herbarium vouchers used in the genus-level phylogeny by Hendriks et al. (2023). In total, 1387 ITS sequences were initially assembled from short-read Illumina data. Following quality filtering (assembly metrics, sequence length, alignment behaviour), 641 samples were excluded, leaving 746 sequences for downstream analysis. A full overview

of the filtering pipeline, quality control steps, and alignment statistics is provided in Online Resource 3. Of the 746 retained ITS sequences: 255 were already included in Hendriks et al. (2023), 61 of those were originally published in Nikolov et al. (2019), and 491 are to be included in the forthcoming Brassicaceae Tree of Life (*Brassi*ToL) project (Hendriks et al. in prep.). We then performed BLAST searches against the *BrassiBase* reference ITS database and compared the tribal and generic identity of each sample (as assigned by the original authors) with the top BLAST hit in *BrassiBase*.

Results: Among the published sequences, 34 did not return the correct genus as the top BLAST hit. Among the unpublished sequences, 40 produced mismatches at the genus level.

These mismatches may be due to one or more of the following factors: Incorrect taxonomic identification of the specimen, erroneous sequence assembly (e.g. assembly of a paralogue or pseudogene), BLAST limitations, as it ranks similarity based on percentage identity rather than positional mutation patterns, low phylogenetic resolution in younger or reticulate lineages, where ITS does not sufficiently distinguish genera or species. To resolve these inconsistencies, all 74 mismatched sequences were re-analysed using the BrassiBase PPT. This approach showed: Only 4 of the published and 12 of the unpublished sequences still clustered with different genera, suggesting they were potentially misidentified. Thus, the true error rates were: 1.2% for published material, and 2.4% for unpublished material. In most other cases, mismatches were attributable to either: poor sequence quality (resulting in ambiguous placement near the tree base), or unresolved phylogenetic relationships (e.g. within Alysseae). Importantly, these estimates likely represent a lower bound, as divergent sequences had already been excluded during earlier quality control (step 3 of our pipeline; see Suppl. Fig. S2). Nevertheless, among the excluded sequences: 10 of the published, and 66 of the unpublished still returned the correct genus in the top BLAST hit—indicating that low sequence coverage (causing apparent SNP inflation) was a common reason for exclusion rather than true misidentification. Thus, only sequences with both a poor BLAST match and strong assembly metrics can be confidently flagged as misidentified. In total, this yields an estimated error rate of 2.1% for the curated Illumina-derived dataset. This underscores the value of using expert-curated voucher material and highlights the benefit of integrating *BrassiBase* with the most comprehensive phylogenies available. It also validates the continued use of the ITS region for accurate phylogenetic placement when combined with proper quality control and expert oversight.

(ii) Publicly Available GenBank Sequences: In the second test, we assessed the accuracy of 9286 ITS sequences recently downloaded from GenBank. These were manually screened, and 380 sequences were excluded from the BrassiBase reference alignments based on visual inspection due to excessive divergence. For these 380 divergent sequences, we compared: the tribe and genus assignments in GenBank, with the top BLAST result from *BrassiBase*. We found that 82 sequences had an incorrect genus assignment in GenBank. Given the full dataset size, this corresponds to a misidentification rate of 0.9%. However, several caveats must be considered: Only sequences > 500 bp were included in the analysis. Sequences missing the 5.8S rRNA gene (between ITS1 and ITS2) or containing excessive missing data were excluded. Only sequences associated with peer-reviewed publications in taxonomic, phylogenetic, or evolutionary contexts were retained. This likely underestimates the true error rate in the broader GenBank ITS dataset for Brassicaceae. Therefore, although the observed error rate is similar between *BrassiBase* and GenBank (2.1% vs. 0.9%), the curation level and sequence quality in *BrassiBase*—alongside its integration with phylogenetic trees, expert-verified taxonomies, and voucher records—make it the preferred reference source for phylogenetic placement and taxonomic verification.

Conclusion and future perspectives of *BrassiBase*

The updated version of *BrassiBase* currently lacks one important feature: the integration of a character matrix linked to the dynamic taxonomic backbone. Such a module would significantly enhance the platform's utility for morphological and trait-based analyses of evolutionary patterns and processes (e.g. Walden et al. 2020; Gómez et al. 2022; Maccagni and Willi 2022; Baczyński et al. 2025). We aim to pilot this functionality using a comprehensive morphological dataset for the tribe Arabideae, at species level. This dataset will cover approximately 550 species, representing around 13% of the current 4148 species in *BrassiBase*, and will be linked directly to the taxonomic backbone. A further major priority is the integration of data reflecting reticulate evolutionary processes. Our recent work on Arabideae (Walden et al. 2024) demonstrates that traditional bifurcating phylogenetic trees are insufficient for reconstructing the complex evolutionary history of this tribe. The main cause is the prevalence of reticulate evolution, hybridisation, and introgression, with phylogenetic signatures spanning the entire ~ 15 million year history of the group (Huang et al. 2020). An indirect proxy for this complexity is the high frequency of neopolyploidy—estimated at 63% for Arabideae and around 43% across the Brassicaceae (Hohmann et al. 2015). To better capture this complexity, we developed a novel analytical pipeline to evaluate the placement of paralogous gene copies from polyploid taxa on a diploid backbone phylogeny. This approach enables quantification of reticulate patterns, even in deep evolutionary timescales. In Arabideae, for example, polyploid taxa were distributed

across 43 distinct, often reticulate clusters within the diploid backbone (Walden et al. 2024). The same method has since been applied successfully to the genus *Murbeckiella* and its allies (Walden et al. 2025).

In parallel, plastid genome data, as a maternally inherited marker system, offer an independent phylogenetic signal that can enrich taxonomic insights and clarify maternal lineages (Walden et al. 2020; Hendriks et al. 2023; Thomas et al. 2025). A well-resolved plastid phylogeny, in turn, facilitates the placement of commonly used plastid markers, such as the *trn*L intron/*trn*LF intergenic spacer, in a broader evolutionary context. These regions have been widely employed in Brassicaceae phylogenetics (e.g. Bailey et al. 2006; Koch et al. 2007, 2017; Dobes et al. 2007; Schmickl et al. 2008). As of May 2025, GenBank includes approximately 4950 *trn*L intron and 4800 *trn*LF spacer sequences from Brassicaceae.

Together, the integration of taxonomy, molecular phylogenies, and plastid and nuclear genomic data will further enable the reconstruction of the complex and reticulate evolutionary history of the family. These efforts complement current activities, including the development of the Brassicaceae Tree of Life (*Brassi*ToL) (Hendriks and Lens 2024), and may contribute to our understanding of parallel and convergent evolution in the family.

BrassiBase also aligns with Target 1 of the Global Strategy for Plant Conservation, aiming to provide a comprehensive global overview of Brassicaceae species diversity. This functionality is intended to support conservation planners, policymakers, and practitioners by serving as a centralised, curated resource. Numerous Brassicaceae species are endangered or vulnerable and require targeted, evidence-based conservation strategies (e.g. Castillo-Lorenzo et al. 2024). Future updates to *BrassiBase* will therefore aim to include: Curated distribution maps, assignments of species to WWF ecoregions (Olson et al. 2001), and further integration with conservation-related frameworks such as World Flora Online (WFO). This not only enhances the platform's scientific utility but also its contribution to global biodiversity conservation efforts.

Information on electronic supplementary material

Online Resource 1. *Brassi*TEN members, affiliations and ORCID (if available).
Online Resource 2. Schematic overview of the database structure. The database consists of several interlinked tables referring to the content. Respective data fields are shown. The table tb_names (taxonomy) is core to the database. Alignments, sequences and treefiles are stored in the server's file system.

Online Resource 3. Workflow for determining high quality ITS sequences generated from NGS data for integration into *BrassiBase*.
Online Resource 4. Distribution of ITS sequence data hold in *BrassiBase* sorted according to tribes.

Declarations

References

Agerbirk N, Hansen CC, Kiefer C, Hauser TP, Ørgaard M, Bruun Asmussen Lange C, Cipollini D, Koch MA (2021) Comparison of glucosinolate diversity in the crucifer tribe Cardamineae and the remaining order Brassicales highlights repetitive evolutionary loss and gain of biosynthetic steps. Phytochemistry 185:e112668. https://doi.org/10.1016/j.phytochem.2021.112668

Al-Shehbaz IA (2012) A generic and tribal synopsis of the Brassicaceae (Cruciferae). Taxon 61:931–954. https://doi.org/10.1002/tax.615002

Al-Shehbaz IA, Mutlu B, Dönmez AA (2007) The Brassicaceae (Cruciferae) of Turkey, updated. Turk J Bot 31:327–336

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Molec Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Álvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. Molec Phylogen Evol 29:417–434. https://doi.org/10.1016/S1055-7903(03)00208-2

Baczyński J, Oskolski AA, Winter PJD, Manuel R, Lyner T, Magee AR, Muasya AM, Frankiewicz KE (2025) Lifespan outperforms climate as a predictor of wood functional traits, but secondary woodiness shows no clear climatic pattern in *Heliophila*, a diverse clade from the Cape Floristic Region. Ann Bot (Oxford) (in press): mcaf046. https://doi.org/10.1093/aob/mcaf046

Bailey CD, Koch MA, Mayer M, Mummenhoff K, O'Kane SL, Warwick SI, Windham MD, Al-Shehbaz IA (2006) Toward a global phylogeny of the Brassicaceae. Molec Biol Evol 23:2142–2160. https://doi.org/10.1093/molbev/msl087

Banchi E, Ametrano CG, Greco S, Stanković D, Muggia L, Pallavicini A (2020) PLANiTS: a curated sequence reference dataset for plant ITS DNA metabarcoding. Database 2020:baz155. https://doi.org/10.1093/database/baz155

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinform 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Borsch T, Berendsohn W, Dalcin E, Delmas M, Demissew S, Elliott A, Fritsch P, Fuchs A, Geltman D, Güner A, Haevermans T, Knapp S, le Roux MM, Loizeau PA, Miller C, Miller J, Miller JT, Palese R, Paton A, Parnell J, Pendry C, Qin HN, Sosa V, Sosef M, von Raab-Straube E, Ranwashe F, Raz L, Salimov R, Smets E, Thiers B, Thomas W, Tulig M, Ulate W, Ung V, Watson M, Jackson PW, Zamora N (2020) World Flora Online: Placing taxonomists at the heart of a definitive and comprehensive global resource on the world's plants. TAXON 69:1311–1341. https://doi.org/10.1002/tax.12373

Burnett GT (1835) (3847–3933) Brassicaceae or Cruciferae. Outlines of Botany, vol. 2. John Churchill, London

Castillo-Lorenzo E, Breman E, Gómez Barreiro P, Viruel J (2024) Current status of global conservation and characterisation of wild and cultivated Brassicaceae genetic resources. Gigascience 13:giae050. https://doi.org/10.1093/gigascience/giae050

Cheo TY, Lu L, Yang G, Al-Shehbaz I, Dorofeev V (2001) Flora of China, vol. 8. Science Press, Beijing, and Missouri Botanical Garden Press, St. Louis. http://www.efloras.org/florataxon.aspx?flora_id=2&taxon_id=10120

De Candolle AP (1821) Ordo Undecimus Cruciferae. Systema Naturale, vol. 2. Treuttel et Würtz, Parisiis, pp 139–700

Dobes C, Kiefer C, Kiefer M, Koch M (2007) *Trn*f pseudogenes in North American genus *Boechera* (Brassicaceae): mechanistic aspects of evolution. Pl Biol 9:1–14. https://doi.org/10.1055/s-2006-955978

Dominicus LJMA, Al-Shehbaz IA, German DA, Mummenhoff K, Hay NM, Lysak MA, Koch MA, Lens F, Hendriks KP (2025) Mito-plastomic discordance in Brassicaceae phylogenomics confirms the complex evolutionary history of the family. Ann Bot (Oxford) (in press): mcaf084. https://doi.org/10.1093/aob/mcaf084

Dönmez AA, Uğurlu Aydın Z, Koch MA (2017) *Aubrieta al-shehbazii* (Brassicaceae), a new species from Central Turkey. PhytoTaxa 299:103–110. https://doi.org/10.11646/phytotaxa.299.1.8

Dönmez AA, Uğurlu Aydın Z, Kaya Y, Yüzbaşıoğlu S (2023) *Aubrieta birolmutlui* (Brassicaceae), a new species from Eastern Turkey with molecular phylogenetic support. Phytotaxa 579:278–288. https://doi.org/10.11646/phytotaxa.579.4.5

Duan H, Wang W, Zeng Y, Mengmeng G, Zhou Y (2019) The screening and identification of DNA barcode sequences for Rehmannia. Sci Rep 9:17295. https://doi.org/10.1038/s41598-019-53752-8

Đulović A, Koch MA, Thongyoo P, Pattison DI, Blažević I, Rollin P, Agerbirk N (2024) Glucosinolates in non-Brassicales plant species: critical literature evaluation and testing of two high chemical quality reports. Biochem Syst Ecol 116:104864. https://doi.org/10.1016/j.bse.2024.104864

Elliott AC, Bester SP, Klopper RR, Nelson EC, Pirie MD (2024a) Curating an online checklist for *Erica* L. (Ericaceae): contributing to and supporting global conservation through the World Flora Online. PhytoKeys 243:121–135. https://doi.org/10.3897/phytokeys.243.121555

Elliott A, Hyam R, Watson M, Wrankmore E, Hartley H, Krieger J, Gandhi K, Almeida RFD, Anderson G, Andrella GC, Anguiano M, Antonio-Domingues H, Ardi WH, Atkins H, Atwood JJ, Aubriot X, Baker W, Balan AP, Zizka G, The World Flora Online Consortium (2024b) World Flora Online Plant List December 2024 (2024–12). Zenodo. https://doi.org/10.5281/zenodo.14538251

Francis A, Lujan-Toro BE, Warwick SI, Macklin JA, Martin SL (2021) Update on the Brassicaceae species checklist. Biodivers Data J 9:e58773. https://doi.org/10.3897/BDJ.9.e58773

German DA, Hendriks KP, Koch MA, Lens F, Lysak MA, Bailey CD, Mummenhoff K, Al-Shehbaz IA (2023) An updated classification of the Brassicaceae (Cruciferae). PhytoKeys 220:127–144. https://doi.org/10.3897/phytokeys.220.97724

Gómez JM, González-Megías A, Narbona E, Navarro L, Perfectti F, Armas C (2022) Phenotypic plasticity guides Moricandia arvensis divergence and convergence across the Brassicaceae floral morphospace. New Phytol 233:1479–1493. https://doi.org/10.1111/nph.17807

Gu Z, Gu L, Eils R, Schlesner M, Brors B (2014) Circlize implements and enhances circular visualization in R. Bioinform 30:2811–2812. https://doi.org/10.1093/bioinformatics/btu393

Guerin O, Schranz ME, Koch MA (2025) A publication network of research in Brassicaceae from the past 25 years (2000–2025). Zenodo Available at: https://doi.org/10.5281/zenodo.15652856

Guo X, Liu J, Hao G, Zhang L, Mao K, Wang X, Zhang D, Ma T, Hu Q, Al-Shehbaz IA, Koch MA (2017) Plastome phylogeny and early diversification of Brassicaceae. BMC Genomics 18:e176. https://doi.org/10.1186/s12864-017-3555-3

Guo X, Mandáková T, Trachtová K, Özüdoğru B, Liu J, Lysak MA (2021) Linked by ancestral bonds: multiple whole-genome duplications and reticulate evolution in a Brassicaceae tribe. Mol Biol Evol 38:1695–1714. https://doi.org/10.1093/molbev/msaa327

Hämälä T, Moore C, Cowan L, Carlile M, Gopaulchan D, Brandrud MK, Birkeland S, Loose M, Kolar F, Koch MA, Yant L (2024) Impact of whole-genome duplications on structural variant evolution in the plant genus *Cochlearia*. Nat Comm 15:5377. https://doi.org/10.1038/s41467-024-49679-y

Hayek A (1911) Entwurf Eines Cruciferen Systemes auf phylogenetischer Grundlage. Beiheifte Bot Centralbl 27:127–335

Hedge IC (1976) A systematic and geographicalsurvey of the Old World Cruciferae. In: MacLeod AJ, Jones BMG (eds) The biology

and chemistry of the Cruciferae. Academic Press, London, New York, San Francisco, pp 1–45

Hendriks KP, Kiefer C, Al-Shehbaz IA, Bailey D, Hooft van Huysduynen A, Nikolov LA, Nauheimer L, Zuntini AR, German DA, Franzke A, Koch MA, Lysak MA, Toro-Nunez O, Özüdoğru B, Invernón VR, Walden N, Maurin O, Hay NM, Shushkov P, Mandakova T, Schranz ME, Thulin M, Windham MD, Resetnik I, Spaniel S, Ly E, Pires JC, Harkess A, Neuffer B, Vogt R, Bräuchler C, Rainer H, Janssens SB, Schmull M, Forrest A, Guggisberg A, Zmarzty S, Lepschi BJ, Scarlett N, Stauffer FW, Schönberger I, Heenan P, Baker WJ, Forest F, Mummenhoff K, Lens F (2023) Global Brassicaceae phylogeny based on filtering of 1,000-gene dataset. Curr Biol 33:1–17. https://doi.org/10.1016/j.cub.2023.08.026

Hendriks KP, Lens F (2024) Brassicaceae tree of life app. Available at: https://brassitol.vercel.app/

Hernández-Ledesma P, Berendsohn WG, Borsch T, Von Mering S, Akhani H, Arias S, Castañeda-Noa I, Eggli U, Eriksson R, Flores-Olvera H, Fuentes-Bazán S, Kadereit G, Klak C, Korotkova N, Nyffeler R, Ocampo G, Ochoterena H, Oxelman B, Rabeler RK, Sanchez A, Schlumpberger BO, Uotila P (2015) A taxonomic backbone for the global synthesis of species diversity in the angiosperm order *Caryophyllales*. Willldenowia 45:281–383. https://doi.org/10.3372/wi.45.45301

Hohmann N, Wolf E, Lysak MA, Koch MA (2015) A time-calibrated road map of Brassicaceae species radiation and evolutionary history. Pl Cell 27:2770–2784. https://doi.org/10.1105/tpc.15.00482

Huang CH, Sun R, Zeng YH, Zhang N, Cai L, Zhang Q, Koch MA, Al-Shehbaz IA, Edger PP, Pires JC, Tan DY, Zhong Y, Ma H (2015) Resolution of Brassicaceae phylogeny using nuclear genes uncovers nested radiations and supports convergent morphological evolution. Molec Biol Evol 33:394–412. https://doi.org/10.1093/molbev/msv226

Huang XC, German D, Koch MA (2019) Temporal patterns of diversification in Brassicaceae demonstrate decoupling of rate shifts and mesopolyploidization events. Ann Bot 125:29–47. https://doi.org/10.1093/aob/mcz123

Huang XC, German D, Koch MA (2020) Temporal patterns of diversification in Brassicaceae demonstrate decoupling of rate shifts and mesopolyploidization events. Ann Bot (Oxford) 125:29–47. https://doi.org/10.1093/aob/mcz123

Janchen E (1942) Das system der Cruciferen. Oesterr Bot Z 91:1–28

Jiao WB, Accinelli GG, Hartwig B, Kiefer C, Baker D, Severing E, Willing EM, Piednoel M, Woetzel S, Madrid-Herrero E, Huettel B, Hümann U, Reinhard R, Koch MA, Swan D, Clavijo B, Coupland G, Schneeberger K (2017) Improving and correcting the contiguity of long-read genome assemblies of three plant species using optical mapping and chromosome conformation capture data. Genome Res 27:778–786. https://doi.org/10.1111/mec.15648

Kagale S, Robinson SJ, Nixon J, Xiao R, Huebert T, Condie J, Kessler D, Clarke WE, Edger PP, Links MG, Sharpe AG, Parkin IA (2014) Polyploid evolution of the Brassicaceae during the Cenozoic era. Pl Cell 26:2777–2791. https://doi.org/10.1105/tpc.114.126391

Kantor A, Šlenker M, Kučera J, Breidy J, Marhold K, Dönmez AA, Yüzbaşıoğlu S, Zozomova-Lihova J (2023) Balkan-Anatolian biogeographic links and the evolutionary significance of Anatolian mountains, as evidenced by *Cardamine* (Brassicaceae). Taxon 72:1303–1320. https://doi.org/10.1002/tax.13098

Karl R, Koch MA (2013) A world-wide perspective on crucifer speciation and evolution: phylogenetics, biogeography and trait evolution in tribe Arabideae. Ann Bot (Oxford) 112:983–1001. https://doi.org/10.1093/aob/mct165

Katoh K, Frith MC (2012) Adding unaligned sequences into an existing alignment using MAFFT and LAST. Bioinform 28:3144–3146. https://doi.org/10.1093/bioinformatics/bts578

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molec Biol Evol 30:772–780. https://doi.org/10.1093/molbev/mst010

Kaya Y, Uğurlu Aydın Z, Cai Z, Wang X, Dönmez AA (2022) Genome-wide characterization of two Aubrieta taxa: Aubrieta canescens subsp. canescens and Au. macrostyla (Brassicaceae). AoB PLANTS 14:plac035. https://doi.org/10.1093/aobpla/plac03

Kiefer M, Schmickl R, German DA, Mandáková T, Lysak MA, Al-Shehbaz IA, Franzke A, Mummenhoff K, Stamatakis A, Koch MA (2014) BrassiBase: introduction to a novel knowledge database on Brassicaceae evolution. Pl Cell Physiol 55:e3. https://doi.org/10.1093/pcp/pct158

Kiefer C, Willing EM, Jiao WB, Sun H, Piednoel M, Hümann U, Hartwig B, Koch MA, Schneeberger K (2019) Inter-species association mapping links reduced CG-TG substitution rates to the loss of gene body methylation. Nat Pl 5:846–855. https://doi.org/10.1038/s41477-019-0486-9

Koch MA (2019) The plant model system *Arabidopsis* set in an evolutionary, systematic, and spatio-temporal context. J Exp Bot 70:55–67. https://doi.org/10.1093/jxb/ery340

Koch MA, German DA (2013) Taxonomy and systematics are key to biological information: *Arabidopsis, Eutrema (Thellungiella), Noccaea* and *Schrenkiella* (Brassicaceae) as examples. Frontiers Pl Sci 4:267. https://doi.org/10.3389/fpls.2013.00267

Koch MA, Lemmel C (2019) *Zahora*, a new monotypic genus from tribe Brassiceae (Brassicaceae) endemic to the Moroccan Sahara. PhytoKeys 135:119–131. https://doi.org/10.3897/phytokeys.135.46946

Koch M, Mummenhoff K (2006) Editorial: Evolution and phylogeny of the Brassicaceae. Pl Syst Evol 259:81–83. https://doi.org/10.1007/s00606-006-0433-x

Koch MA, Al-Shehbaz IA, Mummenhoff K (2003) Molecular systematics, evolution, and population biology in the mustard family (Brassicaceae). Ann Missouri Bot Gard 90:151–171. https://doi.org/10.2307/3298580

Koch MA, Dobeš C, Mitchell-Olds T (2003) Multiple hybrid formation in natural populations: concerted evolution of the internal transcribed spacer of nuclear ribosomal DNA (ITS) in North American *Arabis divaricarpa* (Brassicaceae). Molec Biol Evol 20:338–350. https://doi.org/10.1093/molbev/msg046

Koch MA, Dobes C, Kiefer C, Schmickl R, Klimes L, Lysak MA (2007) Supernetwork identifies multiple events of plastid *trnF*(GAA) pseudogene evolution in the Brassicaceae. Molec Biol Evol 24:63–73. https://doi.org/10.1093/molbev/msl130

Koch MA, Kiefer M, German DA, Al-Shehbaz IA, Franzke A, Mummenhoff K, Schmickl R (2012) BrassiBase: tools and biological resources to study characters and traits in the Brassicaceae—version 1.1. Taxon 61:1001–1009. https://doi.org/10.1002/tax.615007

Koch MA, Karl R, German GA (2017) Underexplored biodiversity of Eastern Mediterranean biota: Systematics and evolutionary history of the genus *Aubrieta* (Brassicaceae). Ann Bot (Oxford) 119:39–57. https://doi.org/10.1093/aob/mcw204

Koch MA, German DA, Kiefer M, Franzke A (2018) Database taxonomics as key to modern plant biology. Trends Pl Sci 23:4–6. https://doi.org/10.1016/j.tplants.2017.10.005

Korotkova N, Aquino D, Arias S, Eggli U, Franck A, Gómez-Hinostrosa C, Guerrero PC, Hernández HM, Kohlbecker A, Köhler M, Luther K, Majure LC, Müller A, Metzing D, Nyffeler R, Sánchez D, Schlumpberger B, Berendsohn WG (2021) *Cactaceae* at Caryophyllales.org—a dynamic online species-level taxonomic

backbone for the family. Willdenowia 51:251–270. https://doi.org/10.3372/wi.51.51208

Larsson A (2014) AliView: a fast and lightweight alignment viewer and editor for large data sets. Bioinform 30:3276–3278. https://doi.org/10.1093/bioinformatics/btu531

Loizeau PA, Jackson PW (2017) World Flora Online mid-term update. Ann Missouri Bot Gard 102:341–346. https://doi.org/10.3417/D-16-00008A

Lysak MA, Koch MA (2011) Phylogeny, genome and karyotype evolution of crucifers (Brassicaceae). In: Schmidt R, Bancroft I (eds) Genetics and genomics of the Brassicaceae. Springer Verlag, New York

Lysak MA, Koch MA, Leitch IJ, Beaulieau JM, Meister A (2009) The dynamic ups and downs of genome size evolution in Brassicaceae. Molec Biol Evol 26:85–98. https://doi.org/10.1093/molbev/msn223

Lysak MA, Mandáková T, Schranz ME (2016) Comparative paleogenomics of crucifers: ancestral genomic blocks revisited. Curr Opin Pl Biol 30:108–115. https://doi.org/10.1016/j.pbi.2016.02.001

Maccagni A, Willi Y (2022) Trait divergence and trade-offs among Brassicaceae species differing in elevational distribution. Evolution 76:1986–2003. https://doi.org/10.1111/evo.14554

Makenzie EM, Abrahams RS, Al-Shehbaz IA, Baker WJ, Barak S, BarkerMS BRL, Beric A, Bhattacharya S, Carey SB, Conant GC, Conran JG, Dassanayake M, Edger PP, Hall JC, Hao Y, Hendriks KP, Hibberd JM, King GJ, Kliebenstein DJ, Koch MA, Leitch IJ, Lens F, Lysak MA, McAlvay AC, McKibben MTW, Mercati F, Moore RC, Mummenhoff K, Murphy DJ, Nikolov LA, Pisias M, Roalson EH, Schranz ME, Thomas SK, Yu Q, Yocca A, Pires JC, Harkess AE (2023) Complementing model species with model clades. Pl Cell 36:1205–1226. https://doi.org/10.1093/plcell/koad260

Mandáková T, Hloušková P, Koch MA, Lysak MA (2020) Genome evolution in Arabideae was marked by frequent centromere repositioning. Pl Cell 32:650–665. https://doi.org/10.1105/tpc.19.00557

Mohammadin S, Peterse K, van de Kerke SJ, Chatrou LW, Dönmez AA, Mummenhoff K, Pires JC, Edger PP, Al-Shehbaz IA, Schranz ME (2017) Anatolian origins and diversification of *Aethionema*, the sister lineage of the core Brassicaceae. Amer J Bot 104:1042–1054. https://doi.org/10.3732/ajb.1700091

Murat F, Louis A, Maumus F, Armero A, Cooke R, Quesneville H, Crollius HR, Salse J (2015) Understanding Brassicaceae evolution through ancestral genome reconstruction. Genome Biol 16:262. https://doi.org/10.1186/s13059-015-0814-y

Nikolov LA, Shushkov P, Nevado B, Gan X, Al-Shehbaz IA, Filatov D, Bailey CD, Tsiantis M (2019) Resolving the backbone of the Brassicaceae phylogeny for investigating trait diversity. New Phytol 222:1638–1651. https://doi.org/10.1111/nph.15732

Olson DM, Dinerstein E, Wikramanayake ED, Burgess ND, Powell GVN, Underwood EC, D'Amico JA, Itoua I, Strand HE, Morrison JC, Loucks CJ, Allnutt TF, Ricketts TH, Kura Y, Lamoreux JF, Wettengel WW, Hedao P, Kassem KR (2001) Terrestrial ecoregions of the world: a new map of life on earth. Bioscience 51:933–938. https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2

Perianes-Rodriguez A, Waltman L, Van Eck NJ (2016) Constructing bibliometric networks: a comparison between full and fractional counting. J Informetr 10:1178–1195. https://doi.org/10.1016/j.joi.2016.10.006

Prantl K (1891) Cruciferae. In: Engler A, Prantl K (eds) Die natürlichen Pflanzenfamilien Teil. III (2). Wilhelm Engelmann, Leipzig, pp 145–206

Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A (2020) Using SPAdes de novo assembler. Curr Protoc Bioinform 70:e102. https://doi.org/10.1002/cpbi.102

R Core Team (2024) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Available at: https://www.R-project.org/

Rice P, Longden I, Bleasby A (2000) EMBOSS: The European molecular biology open software suite. Trends Genet 16:276–277. https://doi.org/10.1016/s0168-9525(00)02024-2

Román-Palacios C, Molina-Henao YF, Barker MS (2020) Polyploids increase overall diversity despite higher turnover than diploids in the Brassicaceae. Proc Biol Sci 287:20200962. https://doi.org/10.1098/rspb.2020.0962

Schmickl R, Kiefer M, Dobes C, Koch MA (2008) Evolution of *trnF*(GAA) pseudogenes in cruciferous plants. Pl Syst Evol 282:229–240. https://doi.org/10.1007/s00606-008-0030-2

Schranz ME, Lysak MA, Mitchell-Olds T (2006) The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. Trends Pl Sci 11:535–542. https://doi.org/10.1016/j.tplants.2006.09.002

Schulz OE (1936) Cruciferae. In: Engler A, Harms H (eds) Die natürlichen Pflanzenfamilien 17B. Wilhelm Engelmann, Leipzig, pp 227–658

Signorell A (2025) DescTools: tools for descriptive statistics. R package version 0.99.59. Available at: https://CRAN.R-project.org/package=DescTools

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

Thomas SK, Abrahams RS, Kick DR, Walden N, Conant G, McKain MR, An H, Arias T, Edger PP, Harkess A, Henriks KP, Koch MA, Lens F, Lysak MA, McAlvay AC, Mummenhoff K, Al-Shehbaz IA, Washburn JD, Pires JC (2025) Mesoolyploidy as a taxonomic clade marker for *Brassica* and relatives (tribe Brassiceae). Ann Bot (Oxford) (in press):mcaf170. https://doi.org/10.1093/aob/mcaf170

Walden N, Schranz ME (2023) Synteny identifies reliable orthologs for phylogenomics and comparative genomics of the Brassicaceae. Genome Biol Evol 15:evad034. https://doi.org/10.1093/gbe/evad034

Walden N, German DA, Wolf EM, Kiefer M, Rigault P, Huang XC, Kiefer C, Schmickl R, Franzke A, Neuffer B, Mummenhoff K, Koch MA (2020) Nested whole-genome duplications coincide with diversification and high morphological disparity in Brassicaceae. Nat Comm 11:3795. https://doi.org/10.1038/s41467-020-17605-7

Walden N, Kiefer C, Koch MA (2024) Unravelling complex hybrid and polyploid evolutionary relationships using phylogenetic placement of paralogs from target enrichment data. bioRxiv (preprint). https://doi.org/10.1101/2024.06.28.601132

Walden N, Mansion G, Schnorf H, Koch MA, Guggisberg A (2025) *Murbeckiella* is dead, long live *Oreophyton*—origin and systematics of tribe Oreophytoneae (Brassicaceae). Ann Bot (Oxford) (in press):mcaf115. https://doi.org/10.1093/aob/mcaf115

Warwick SI, Al-Shehbaz IA (2006) Chromosome number index and database on CD-rom. Pl Syst Evol 259:237–248. https://doi.org/10.1007/s00606-006-0421-1

Warwick SI, Sauder CA (2005) Phylogeny of tribe Brassiceae (Brassicaceae) based on chloroplast restriction site polymorphisms and nuclear ribosomal internal transcribed spacer and chloroplast *trnL* intron sequences. Canad J Bot 83:467–483. https://doi.org/10.1139/b05-02

Warwick SI, Francis A, Al-Shehbaz IA (2006) Brassicaceae: species checklist and database on CD-rom. Pl Syst Evol 259:249–258. https://doi.org/10.1007/s00606-006-0422-0

Wickham H, François R, Henry L, Müller K, Vaughan D (2023) dplyr: A Grammar of Data Manipulation. R package version 1.1.4. Available at: https://CRAN.R-project.org/package=dplyr

Zhou R, Qin X, Hou J, Liu Y (2024) Research progress on Brassicaceae plants: a bibliometrics analysis. Frontiers Pl Sci 15:1285050. https://doi.org/10.3389/fpls.2024.1285050

Zunitini AR, Carruthers T, Maurin O et al (2024) Phylogenomics and the rise of angiosperms. Nature 629:843–850. https://doi.org/10.1038/s41586-024-07324-0