Check for
updates

Pairing omics to decode the diversity of plant specialized metabolism

Felicia C. Wolters[a,b,1], Elena Del Pup[a,1],
Kumar Saurabh Singh[a,c,d,e], Klaas Bouwmeester[b],
M. Eric Schranz[b], Justin J. J. van der Hooft[a] and
Marnix H. Medema[a]

Plants have evolved complex bouquets of specialized natural products that are utilized in medicine, agriculture, and industry. Untargeted natural product discovery has benefitted from growing plant omics data resources. Yet, plant genome complexity limits the identification and curation of biosynthetic pathways via single omics. Pairing multi-omics types within experiments provides multiple layers of evidence for biosynthetic pathway mining. The extraction of paired biological information facilitates connecting genes to transcripts and metabolites, especially when captured across time points, conditions and chemotypes. Experimental design requires specific adaptations to enable effective paired-omics analysis. Ultimately, metadata standards are required to support the integration of paired and unpaired public datasets and to accelerate collaborative efforts for natural product discovery in the plant research community.

Addresses
[a] Bioinformatics Group, Wageningen University & Research, Wageningen, the Netherlands
[b] Biosystematics Group, Wageningen University & Research, Wageningen, the Netherlands
[c] Plant-Microbe Interactions, Institute of Environmental Biology, Utrecht University, the Netherlands
[d] Faculty of Environment, Science and Economy, University of Exeter, TR10 9FE Penryn Cornwall UK
[e] Plant Functional Genomics Group, Brightlands Future Farming Institute, Faculty of Science and Engineering, Maastricht University 5928 SX Venlo, the Netherlands

Corresponding authors: Medema, Marnix H. (marnix.medema@wur.nl); van der Hooft, Justin J. J. (justin.vanderhooft@wur.nl)
✕ (Del Pup E.), ✕ (Singh K.S.), ✕ (Bouwmeester K.), ✕ (van der Hooft J.J.J.), ✕ (Medema M.H.)
[1] Felicia C. Wolters and Elena Del Pup are joint/shared first authors.

Introduction
Plants have evolved an extremely diverse and chemically innovative array of molecules to coordinate environmental and developmental responses. The range of ecological functions encoded in the structural diversity of plant specialized metabolites provides a promising yet underexplored resource for the development of novel drugs, biofuels, agrochemicals, and targets for crop improvement. However, sourcing plant specialized metabolites is constrained by the need to access the source species, cultivate non-crop plants, and purify low-abundant plant extracts that are only conditionally induced, and often compete with primary metabolism [1]. Consequently, the chemical library of the plant specialized metabolome has remained elusive.

Elucidating the biosynthesis of plant specialized metabolites is a promising avenue for discovering plant natural products [2–4] and making them accessible for efficient production [5,6]. Initially, biosynthetic pathways were characterized via targeted metabolite quantification on mass spectrometry platforms to identify pathway intermediates [7,8] or targeted transcriptomics by "bait"-gene co-expression analysis [9] to find additional candidate genes potentially involved in the pathway. Based on this, prioritized lists of candidate genes could then be functionally tested using *in vitro* screens or transient *in planta* expression. Alternatively, untargeted approaches could be used to discover pathways independently of previous knowledge of a target biosynthetic gene or chemical compound of interest.

In bacteria and fungi, untargeted genome mining [10] through the identification of biosynthetic gene clusters (BGCs) with tools such as antiSMASH [11] resulted in hundreds of thousands of biosynthetic pathways being predicted and hundreds being experimentally validated. More than 30 BGCs have now been experimentally

characterized in plants thanks to a combination of omics and genome mining strategies such as plantiSMASH [12−15]. However, gene clustering is not a hallmark feature of plant genomes and the overall structural complexity of the plant genomic landscape across polyploidy levels, sub-genomes, and structural rearrangements [16,17] requires the integration of multiple data resources beyond genomic location to facilitate untargeted pathway discovery.

Recently, pairing multiple omics datasets by matching experimental setups and metadata across samples for different omics types has proven to be a powerful approach for plant natural product discovery by facilitating the detection of patterns that could not be captured by combining single-omics datasets [18−21]. For example, the falcarindiol pathway in tomato was discovered using paired-omics data by identifying metabolite−transcript associations across seven different conditions that were each profiled across three time points [20]. There are several advantages of a paired-omics setup. First, paired dimensions, such as time points, tissue, conditions, and species, allow consistently extracting and associating biologically relevant signals across different omics data types. Second, pairing information across omics layers can clarify causal relationships between genes, transcripts, and metabolites, which can uncover genes underlying the production of certain molecules. Third, pairing omics layers can enhance the mining of public single-omics datasets, by allowing for the propagation of insights to other omics layers, facilitated by gene-transcript-metabolite links identified in the paired data. Ultimately, the integration of species phylogeny enables the exploration of functional relationships between omics data types, strengthening biological signal with evidence from orthology in biosynthetic pathways.

Here, we discuss experimental practices, data standards, and analysis strategies for generating paired-omics data that can support identifying, validating, and prioritizing natural product biosynthetic pathways in plants, and consider how data sharing and metadata standards can expand the utility of these datasets through integration with public datasets.

## Designing experiments for plant omics data integration

Depending on the underlying research question, different omics data types can provide distinct layers of evidence for plant pathway discovery. Additionally, practical feasibility and cost-effectiveness are also fundamental considerations for experimental design. For the discovery of specialized biosynthetic pathways, three molecular omics data types are most frequently considered: genomics, transcriptomics, and metabolomics [14]. For modelling spatial and temporal

metabolic fluxes and metabolic signaling, single-cell multi-omics and spatial omics may provide an additional dimension for data integration to account for tissue- and cell-type specific pathway modules [21].

Unlike genomics data, other omics data types require contextualized interpretation based on experimental conditions. Consequently, integration of plant transcriptomics and metabolomics data is highly sensitive to variation in experimental setups. Regulation of specialized metabolite biosynthesis in plants follows specific patterns in vegetative and generative stage, tissue type, and ontogeny [22,23]. Notably, mimicking complex biotic and abiotic stresses for elicitation of specialized compound biosynthesis demands careful consideration of environmental conditions, e.g., in the greenhouse or climate chamber [24]. To link biosynthetic pathway intermediates to transcript abundance, multiple time points can be included to account for delayed detection of biosynthetic products following transcriptional signaling [14]. Additionally, sampling different tissues and developmental stages can help to distinguish between pathways whose expression is triggered by the same stresses. Further signal can be captured by sampling different genotypes and associated chemotypes of the species of interest. However, combinatorial limits are easily reached, as each additional design parameter multiplies the number of samples required, also considering sufficient replicates (Figure 1a). Yet, we believe that, to be able to capture sufficient variation in gene expression and metabolite production within a standardized setup to facilitate large-scale (semi-) automated pathway discovery, the field should invest in generating more paired-omics datasets that cover all possible dimensions within the same experiment, including diversity in species, tissues, conditions and time points. Designing prototypical paired plant omics experiments covering a carefully selected minimal set of tissues, conditions and time points sufficient for distinguishing between pathways in a cost-effective manner should be explored and standardized to also streamline cross-dataset integration.

Although requirements differ regarding most relevant conditions and time points for the biological question at hand, aligning experimental setups may potentiate interoperability of omics data across studies. Further, standardization of sampling strategies (Figure 1b) and downstream processing workflows (Figure 1c) may enhance the interoperability of extracted features as biological signals (Figure 1d). For integrative omics, material for extraction of metabolites, RNA, or DNA originates either from a pool of the same biological material (split samples), from independent replicates from the same batch, or from independent experiments or batches [25]. Performing transcriptomics and metabolomics on material deriving from exactly the same samples is ideal for omics data integration. However,

Figure 1
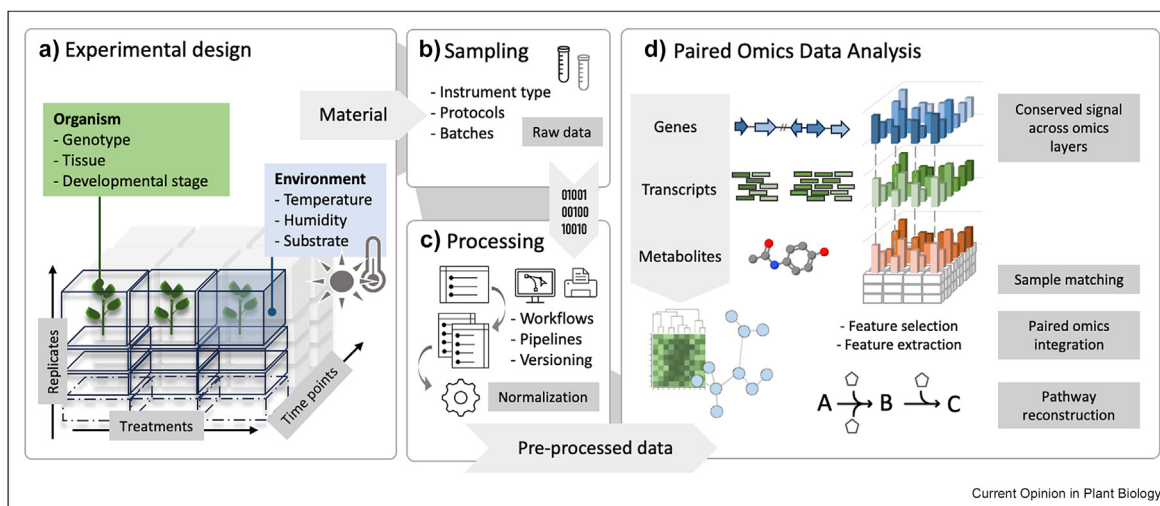
Paired-omics data generation and analysis: starting from experimental design, via sampling and data processing, to paired-omics analysis. **a)** Experimental design for omics data generation. Metadata for environmental conditions and organism, and dimensions of combinatorial setup, including treatments and time points. **b)** Sampling. Raw data acquisition, including sampling, pre-processing protocols, instrument type, and batches in sample analysis. **c)** Data preprocessing protocols, including workflows for processing software, integrative pipelines for large-scale processing, and version control of respective workflows and pipelines for data normalization. **d)** Analysis of paired-omics data. Extraction of biological signal across omics layers via matching conditions in the experimental design. Pathway reconstruction through the integration of genomics, transcriptomics, and metabolomics.

required amounts of sampling material are often unavailable for a combined omics analysis on the same individual plant, depending on plant species, developmental stage, tissue, and conditions under study. For bulk mRNA sequencing, the extraction of plant RNA typically requires fresh input material below 50 mg, while for LC-MS/MS in semi-quantitative analysis, state-of-the-art protocols were optimized for 100 mg fresh weight [26]. The addition of Quality Control samples further increases the total amount of plant tissue material required for combined transcriptomics and metabolomics analysis. For replicate-matched sampling, expected variation between samples limits the integration of distinct omics data types, and covariation across replicates cannot be utilized as signal to connect transcripts to metabolites. Nonetheless, destructive sampling displays an inevitable barrier for omics time-series experiments, since mechanical damage can drastically alter hormone signaling and hence specialized compound biosynthesis [27]. All in all, trade-offs between feasibility and optimality will therefore always need to be navigated.

Analyzing paired-omics data

Paired-omics data analysis aims to combine complementary knowledge from each omics layer, thereby detecting biologically relevant patterns encompassing single omics and experimental dimensions in the paired design. When performing feature selection or extraction in multi-omics studies, early integration by concatenating datasets (followed by, e.g., clustering to identify combined groups of co-abundant metabolites and transcripts) can overcome data multicollinearity and feature redundancy. In contrast, separate processing of single omics (e.g., first identifying coexpression modules and then linking them to metabolites) can better handle signal-to-noise ratio and data imbalances in untargeted datasets [28]. Paired-omics integration performs feature extraction and selection in overlapping conditions, thereby targeting multicollinearity and noise reduction while preserving the most prominent global features and relevant interactions between omics layers.

Correlation of untargeted metabolomics and transcriptomics has been successfully employed as a paired integration strategy that leverages the guilt-by-association principle between metabolite abundance and transcript co-expression across elicitation conditions, combined with evidence from gene clustering [20] or genomics [18]. Similarly, tissue-specific gene expression patterns and single-cell metabolomics have been used to identify candidate genes for the production of tissue-specific metabolites [29], such as monoterpene indole alkaloids in *Catharanthus* [21]. Matching experimental conditions across species could further substantiate predictions thanks to the identification of evolutionarily conserved co-expression modules that indicate likely functional orthology of pathway genes, which led to the identification of biosynthetic modules for alkaloids [30], aliphatic glucosinolates [31], and benzoxazinoids [19]. Such coexpression patterns can be

combined with molecular networking to extend the gene space for tissue and cell-type specific data integration across species, as shown in bacteria [32] and plants [33]. Paired-omics time series profiling can also clarify dynamic regulation and dynamic inference of gene regulatory networks [34], as reported in immune signaling in *Arabidopsis thaliana* [35] and other Brassicaceae [36]. Tools have been developed to capture multiomics patterns across spatial−temporal variation, such as mammalian organogenesis via MEFISTO [37], but their use has not been reported in plants.

Recently, machine learning approaches beyond linear correlation proved successful to uncover biologically relevant multi-omics trends in human studies [38−41]. Across kingdoms, enrichment analytics have been used in learning thanks to epigenomes [37], regulomes [42], reactomes [43,44], and plant proteomes [45]. In other organisms, interpretable deep learning methods have been developed to drive knowledge-driven feature selection in multi-omics datasets [46,47]. Further, tools for visual analytics across data modalities are available to support the interpretation of complex multi-omics patterns, with examples of applications in human studies [48−50]. The latest approaches for multi-omics analysis have yet to be applied to pilot plant natural product discovery in a paired-omics setting, due to the scarcity of paired-omics studies, linked high-quality metadata, and characterized biosynthetic pathways for validation.
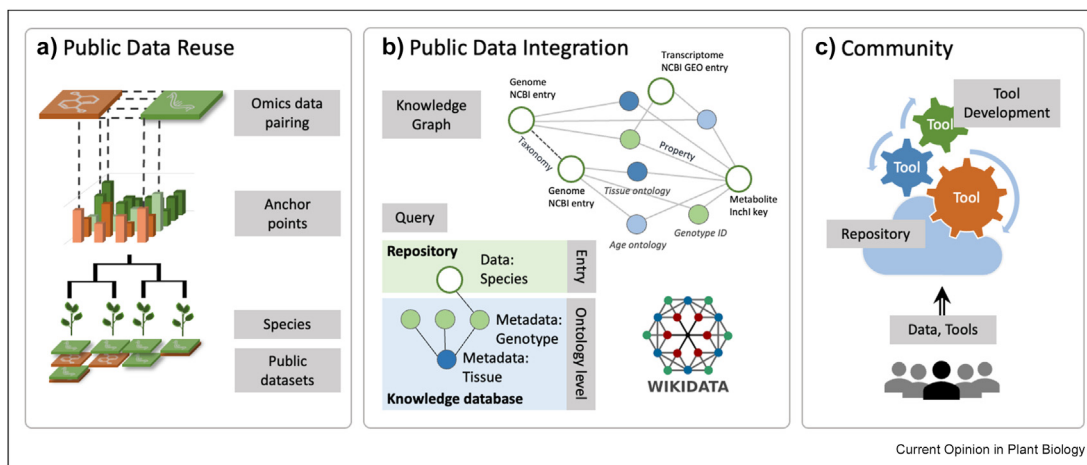
Harnessing the potential of public plant omics data

While the generation of paired-omics datasets is poised to accelerate pathway discovery in plants in the near future, there exists significant potential to leverage the vast body of publicly available single-omics data. Indeed, single and paired-omics datasets have the potential to mutually inform and enhance each other. For instance, single omics can be used to enhance omics data pairing. Alternatively, large-scale single-omics datasets can identify expression modules or families of chemically or ecologically related molecules, thereby increasing the statistical power for linking genes to molecules in paired datasets. Conversely, predicted links between genes and metabolites from paired studies could enhance the functional interpretation of single-omics datasets. Prototypical paired-omics studies can thus act as anchor points for routinely pairing further layers of multi or single omics (Figure 2a).

Standardization and harmonization of public datasets facilitate creating communities around data generation to pool insights, as successfully practiced in the GNPS metabolomics community platform [51]. Moreover, reusing public datasets also enables building or benchmarking tools without having to generate original datasets. The newly developed tool plantMASST [52], implemented within GNPS [51], automates spectral library searches and direct metabolite annotation including their taxonomic distribution. Re-use and re-analysis of existing datasets on the MassIVE platform via REDU [53] enables implementation and comparison of computational analysis software, and integration into modular workflows. At the same time, integration of independent datasets requires alignment of metadata, which currently still poses a major challenge for molecular omics. Recently, a pan-repository scale approach based on harmonization of metadata and standardization of identifiers across platforms showcased the potential

Figure 2

a) Public Data Reuse — Omics data pairing, Anchor points, Species, Public datasets

b) Public Data Integration — Knowledge Graph, Query, Repository: Data: Species, Metadata: Genotype, Metadata: Tissue, Knowledge database; Transcriptome NCBI GEO entry, Genome NCBI entry, Taxonomy, Property, Genome NCBI entry, Tissue ontology, Metabolite InchI key, Age ontology, Genotype ID, WIKIDATA

c) Community — Tool Development, Tool, Repository, Data, Tools

Current Opinion in Plant Biology

Biocuration for data integration via paired-omics data supports a collaborative effort for plant natural product discovery. a) Paired-omics datasets act as anchor points for the alignment of additional public data resources of multi and single-omics datasets, including phylogenetically related species. **b)** Data integration of genomics, transcriptomics, and metabolomics data via queryable knowledge graphs, connecting repository data entries and properties according to levels of metadata ontologies. **c)** Creating a community for developing multi-omics tools, building on and linking single-omics tools, and supporting a collaborative effort for discovering and annotating plant natural products.

of leveraging existing data across multiple repositories, including GNPS [51], MetaboLights [54], and Metabolomics Workbench [55]. Specific metadata structures of single omics include gene, transcript, and protein IDs, and metabolite representations (SMILES or InchIkeys), while common metadata defines categories such as sample type, conditions, extraction method and tissue type (Figure 2b). Although availability of metadata across repositories plays a pivotal role in dimensionality reduction for omics integration [56], improvements in metadata structure standardization are mostly limited to individual omics data types [57−59]. Employing existing minimum information standards for common metadata across omics types may provide a common structure to bridge repositories and facilitate effective cross-omics analyses.

Suitable experimental metadata ontologies are provided by the recently updated Planteome knowledgebase as Plant Ontology (PO) and Plant Experimental Conditions Ontology (PECO) terms [60,61]. PECO terms are integrated into omics databases, such as The Arabidopsis Information Resource (TAIR), with options for querying by PO and PECO hierarchies [60]. Furthermore, the online tool Annotare (https://www.ebi.ac.uk/fg/annotare) provides standardized metadata submission and tagging with PO terms [62]. However, PO and PECO are not yet implemented as ontologies in popular omics data repositories, such as the NCBI Gene Expression Omnibus (GEO) or Plant Expression Omnibus (PEO) [63]. To mine biosynthetic pathways in microbes, links between metabolomics repositories such as GNPS-MassIVE [51] and MetaboLights [54] and (meta)genome assemblies have been successfully established via metadata ontologies on the Paired omics Data Platform (PoDP) [64], facilitating automated gene−metabolite association analysis through tools such as NPLinker [65]. Altogether, linking identifiers across omics-type-specific repositories and adopting shared metadata standards would greatly simplify and facilitate combined analysis both within and beyond the study of plant metabolism.

Biocuration for plant natural product discovery

Collection, curation, and integration of information by linking properties via biocuration allows to connect data across omics-specific databases and perform semantic searches. The fundamental objective to facilitate machine-readable database structures was already underlined in the first proposal of the FAIR principles [66]. For NCBI, E-utilities provide an infrastructure to effectively fetch data with corresponding crosslinks and metadata via Entrezpy [67,68]. Bridging omics repositories via third level instances has been introduced recently for metabolomics data [69] for the curation of queryable knowledge graphs. Linking taxonomic

information with generated genomics, transcriptomics, and metabolomics data could provide a generic structure for data mining approaches (Figure 2b). Wikidata has been proposed earlier as linking knowledgebase for life science data, as it provides entry points for SPARQL querying, and an infrastructure of metadata ontologies based on properties [70]. The spectral database LOTUS [71] is queryable via Wikidata, and integrates taxonomy with IDs of experimentally validated compounds in plants [69,71]. For genomics data, WikiGenomes has been introduced as an integration of data repositories into Wikidata, but is limited to prokaryotic organisms [72]. Other integrative knowledge databases such as RAMP-DB [73] and the *Brassica napus* multi-omics database BnIR [74] are curated from multiple external data sources, such as KEGG and WikiPathways.

Building up on repository-scale integration, omics data generation can drive insights across the tree of life. Paired-omics data communities could potentially enable transfer learning approaches in new species or new omics datatypes by employing biocuration for linking properties such as taxonomic information (Figure 2b) [75]. In recent years, publicly available omics data has expanded beyond model species to encompass broad taxonomic clades with increasing resolution up to the level of plant chemotypes. Hence, phylogenetic information allows the propagation of orthologous relationships from different species in an automated way, as recently showcased for transcriptomics [33] and metabolomics data [32,76]. Besides conservation of coexpression (as mentioned above), the localization of genes within syntenic blocks can be a powerful method to predict orthologues and, by proxy, orthologous pathways [77,78]. However, resolving the overarching species taxonomy is an ongoing challenge in plant systematics. Recently refined taxonomies such as in the Brassicaceae plant family [79] are not yet implemented in Wikidata or NCBI taxonomy, and thus require manual curation. Cross-species enrichment tests can rely on orthology where data for a species is not available [80]; however, such orthology and paralogy relationships across plant genomes could be complex to establish correctly. Nevertheless, integrative omics could provide substantial insights into evolutionary trajectories of biosynthetic pathway modules across species phylogenies, with the potential to propagate annotations and to predict pathway assembly and stability [30,81]. To automate the generation of insights across species, plant-specific databases and in-house solutions with individual server architectures, data standards, preferred genome versions, and manual genome annotation standards should be discouraged in favor of harmonized standards.

Conclusions & future perspectives
Pairing omics data for the same sample material presents a powerful approach for targeted as well as untargeted discovery of specialized biosynthetic pathways in plants.

Further, paired datasets can be used as anchors for the alignment of additional public data resources and single-omics datasets to automate hypothesis generation and support common efforts by the plant research community to annotate the plant biosynthetic space. Integrating metadata annotations, creating consistent ontologies, and propagating their use will be key to creating living resources that can be powered at repository-scale to tap into the plant biosynthetic and biochemical space. The generated hypotheses could be automatically ranked and prioritized for experimental validation based on new or prior evidence, chemical novelty, or relevant chemical properties, thereby accelerating collaborative efforts in the plant natural product discovery community.

Authorship contribution
Felicia C. Wolters: Writing − original draft, Writing − review & editing, Visualization, Conceptualization. **Elena Del Pup:** Writing − original draft, Writing − review & editing, Visualization, Conceptualization. **Kumar Saurabh Singh:** Writing − review & editing. **Klaas Bouwmeester:** Writing − review & editing, Supervision, Conceptualization. **M. Eric Schranz:** Writing − review & editing, Supervision, Conceptualization. **Justin J.J. van der Hooft:** Writing − review & editing, Writing − original draft, Supervision, Conceptualization. **Marnix H. Medema:** Writing − review & editing, Writing − original draft, Supervision, Conceptualization, Funding acquisition.

Declaration of competing interest
The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: JJJvdH is member of the scientific advisory board of NAICONS Srl., Milano, Italy, and consults for Corteva Agriscience, Indianapolis, IN, USA. MHM is a member of the scientific advisory board of Hexagon Bio.

Acknowledgments

Data availability
No data was used for the research described in the article.

References
Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Owen C, Patron NJ, Huang A, Osbourn A: **Harnessing plant metabolic diversity**. *Curr Opin Chem Biol* 2017, **40**:24−30, https://doi.org/10.1016/j.cbpa.2017.04.015.

2. Kim SS, Wengier DL, Ragland CJ, Sattely ES: **Transcriptional reactivation of lignin biosynthesis for the heterologous production of etoposide aglycone in *Nicotiana benthamiana***. *ACS Synth Biol* 2022, **11**:3379−3387, https://doi.org/10.1021/acssynbio.2c00289.

3. Liu JC-T, De La Peña R, Tocol C, Sattely ES: **Reconstitution of early paclitaxel biosynthetic network**. *Nat Commun* 2024, **15**: 1419, https://doi.org/10.1038/s41467-024-45574-8.

4. Ricardo DLP, Sattely ES: **Rerouting plant terpene biosynthesis enables momilactone pathway elucidation**. *Nat Chem Biol* 2021, **17**:205−212, https://doi.org/10.1038/s41589-020-00669-3.

5. Cravens A, Payne J, Smolke CD: **Synthetic biology strategies for microbial biosynthesis of plant natural products**. *Nat Commun* 2019, **10**:2142, https://doi.org/10.1038/s41467-019-09848-w.

6. Ranawaka B, An J, Lorenc MT, Jung H, Sulli M, Aprea G, Roden S, Llaca V, Hayashi S, Asadyar L, LeBlanc Z, Ahmed Z, Naim F, De Campos SB, Cooper T, De Felippes FF, Dong P, Zhong S, Garcia-Carpintero V, Orzaez D, Dudley KJ, Bombarely A, Bally J, Winefield C, Giuliano G, Waterhouse PM: **A multi-omic *Nicotiana benthamiana* resource for fundamental research and biotechnology**. *Nat Plants* 2023, https://doi.org/10.1038/s41477-023-01489-8.

7. Cohen JD, Tang Q, Hegeman AD: **Chapter Nine - using targeted metabolomics to elucidate the indole auxin network in plants**. In *Methods enzymol*. Edited by Jez J, Academic Press; 2022:239−278, https://doi.org/10.1016/bs.mie.2022.07.038.

8. Frey M, Chomet P, Glawischnig E, Stettner C, Grün S, Winklmair A, Eisenreich W, Bacher A, Meeley RB, Briggs SP, Simcox K, Gierl A: **Analysis of a chemical plant defense mechanism in grasses**. *Science* 1997, **277**:696−699.

9. Lau W, Sattely ES: **Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone**. *Science* 2015, **349**:1224−1228, https://doi.org/10.1126/science.aac7202.

10. Medema MH, de Rond T, Moore BS: **Mining genomes to illuminate the specialized chemistry of life**. *Nat Rev Genet* 2021, **22**:553−571, https://doi.org/10.1038/s41576-021-00363-7.

11. Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, Kautsar SA, Suarez Duran HG, de los Santos ELC, Kim HU, Nave M, Dickschat JS, Mitchell DA, Shelest E, Breitling R, Takano E, Lee SY, Weber T, Medema MH: **antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification**. *Nucleic Acids Res* 2017, **45**:W36−W41, https://doi.org/10.1093/nar/gkx319.

12. Medema MH, Osbourn A: **Computational genomic identification and functional reconstitution of plant natural product biosynthetic pathways**. *Nat Prod Rep* 2016, **33**:951−962, https://doi.org/10.1039/C6NP00035E.

13. Polturak G, Liu Z, Osbourn A: **New and emerging concepts in the evolution and function of plant biosynthetic gene clusters**. *Curr Opin Green Sustainable Chem* 2022, **33**, 100568, https://doi.org/10.1016/j.cogsc.2021.100568.

14. Singh KS, van der Hooft JJJ, van Wees SCM, Medema MH: **Integrative omics approaches for biosynthetic pathway discovery in plants**. *Nat Prod Rep* 2022, **39**:1876−1896, https://doi.org/10.1039/D2NP00032F.

15. Kautsar SA, Suarez Duran HG, Blin K, Osbourn A, Medema MH: **plantiSMASH: automated identification, annotation and expression analysis of plant biosynthetic gene clusters**. *Nucleic Acids Res* 2017, **45**:W55−W63, https://doi.org/10.1093/nar/gkx305.

16. Bayer PE, Scheben A, Golicz AA, Yuan Y, Faure S, Lee H, Chawla HS, Anderson R, Bancroft I, Raman H, Lim YP, Robbens S, Jiang L, Liu S, Barker MS, Schranz ME, Wang X, King GJ, Pires JC, Chalhoub B, Snowdon RJ, Batley J, Edwards D: **Modelling of gene loss propensity in the pangenomes of three *Brassica* species suggests different mechanisms between polyploids and diploids**. *Plant Biotechnol J* 2021, **19**:2488−2500, https://doi.org/10.1111/pbi.13674.

17. Liu Z, Suarez Duran HG, Harnvanichvech Y, Stephenson MJ, Schranz ME, Nelson D, Medema MH, Osbourn A: **Drivers of**

metabolic diversification: how dynamic genomic neighbour-hoods generate new biosynthetic pathways in the Brassica-ceae. *New Phytol* 2020, **227**:1109−1123, https://doi.org/10.1111/nph.16338.

18. Bai Y, Yang C, Halitschke R, Paetz C, Kessler D, Burkard K,
** Gaquerel E, Baldwin IT, Li D: **Natural history−guided omics reveals plant defensive chemistry against leafhopper pests**. *Science* 2022, **375**, eabm2948, https://doi.org/10.1126/science.abm2948.
This study uses genomics, transcriptomics, and metabolomics to characterize jasmonate-mediated defense signaling in plants.

19. Florean M, Luck K, Hong B, Nakamura Y, O'Connor SE,
** Köllner TG: **Reinventing metabolic pathways: independent evolution of benzoxazinoids in flowering plants**. *Proc Natl Acad Sci USA* 2023, **120**, e2307981120, https://doi.org/10.1073/pnas.2307981120.
This study uses untargeted metabolomics and transcriptomics to characterize benzoxazinoids biosynthesis and shed light on its evolution in phylogenetically related species.

20. Jeon JE, Kim J-G, Fischer CR, Mehta N, Dufour-Schroif C,
** Wemmer K, Mudgett MB, Sattely E: **A pathogen-responsive gene cluster for highly modified fatty acids in tomato**. *Cell* 2020, **180**:176−187.e19, https://doi.org/10.1016/j.cell.2019.11.037.
This study integrates untargeted transcriptomics and metabolomics to characterize the falcarindiol biosynthetic pathway in tomato.

21. Li C, Wood JC, Vu AH, Hamilton JP, Rodriguez Lopez CE,
** Payne RME, Serna Guerrero DA, Gase K, Yamamoto K, Vaillancourt B, Caputi L, O'Connor SE, Robin Buell C: **Single-cell multi-omics in the medicinal plant Catharanthus roseus**. *Nat Chem Biol* 2023, **19**:1031−1041, https://doi.org/10.1038/s41589-023-01327-0.
This study uses single-cell trascriptomics and metabolomics to characterize the monoterpene indole alkaloid pathway in *Catharanthus roseus*.

22. Doan A-T, Ervin G, Felton G: **Temporal effects on jasmonate induction of anti-herbivore defense in Physalis angulata: seasonal and ontogenetic gradients**. *Biochem Systemat Ecol* 2004, **32**:117−126, https://doi.org/10.1016/S0305-1978(03)00142-X.

23. Kagale S, Nixon J, Khedikar Y, Pasha A, Provart NJ, Clarke WE, Bollina V, Robinson SJ, Coutu C, Hegedus DD, Sharpe AG, Parkin IAP: **The developmental transcriptome atlas of the biofuel crop Camelina sativa**. *Plant J* 2016, **88**:879−894, https://doi.org/10.1111/tpj.13302.

24. Houshyani B, Kabouw P, Muth D, De Vos RCH, Bino RJ, Bouwmeester HJ: **Characterization of the natural variation in Arabidopsis thaliana metabolome by the analysis of meta-bolic distance**. *Metabolomics* 2012, **8**:131−145, https://doi.org/10.1007/s11306-011-0375-3.

25. Cavill R, Jennen D, Kleinjans J, Briedé JJ: **Transcriptomic and metabolomic data integration**. *Briefings Bioinf* 2016, **17**:891−901, https://doi.org/10.1091/bib/bbv090.

26. De Vos RC, Moco S, Lommen A, Keurentjes JJ, Bino RJ, Hall RD: **Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry**. *Nat Protoc* 2007, **2**:778−791, https://doi.org/10.1038/nprot.2007.95.

27. Sun J-Q, Jiang H-L, Li C-Y: **Systemin/jasmonate-mediated systemic defense signaling in tomato**. *Mol Plant* 2011, **4**:607−615, https://doi.org/10.1093/mp/ssr008.

28. Picard M, Scott-Boyer M-P, Bodein A, Périn O, Droit A: **Integra-tion strategies of multi-omics data for machine learning analysis**. *Comput Struct Biotechnol J* 2021, **19**:3735−3746, https://doi.org/10.1016/j.csbj.2021.06.030.

29. Li D, Heiling S, Baldwin IT, Gaquerel E: **Illuminating a plant's tissue-specific metabolic diversity using computational metabolomics and information theory**. *Proc Natl Acad Sci USA* 2016, **113**:E7610−E7618, https://doi.org/10.1073/pnas.1610218113.

30. Itkin M, Heinig U, Tzfadia O, Bhide AJ, Shinde B, Cardenas PD, Bocobza SE, Unger T, Malitsky S, Finkers R, Tikunov Y, Bovy A, Chikate Y, Singh P, Rogachev I, Beekwilder J, Giri AP, Aharoni A: **Biosynthesis of antinutritional alkaloids in solanaceous**

crops is mediated by clustered genes. *Science* 2013, **341**:175−179, https://doi.org/10.1126/science.1240230.

31. Wisecaver JH, Borowsky AT, Tzin V, Jander G, Kliebenstein DJ, Rokas A: **A global coexpression network approach for connecting genes to specialized metabolic pathways in plants**. *Plant Cell* 2017, **29**:944−959, https://doi.org/10.1105/tpc.17.00009.

32. Ramon C, Stelling J: **Functional comparison of metabolic networks across species**. *Nat Commun* 2023, **14**:1699, https://doi.org/10.1038/s41467-023-37429-5.

33. Passalacqua MJ, Gillis J: **Coexpression enhances cross-species integration of single-cell RNA sequencing across diverse plant species**. *Nat Plants* 2024, **10**:1075−1080, https://doi.org/10.1038/s41477-024-01738-4.

34. Schulz MH, Devanny WE, Gitter A, Zhong S, Ernst J, Bar-Joseph Z: **Drem 2.0: improved reconstruction of dynamic regulatory networks from time-series expression data**. *BMC Syst Biol* 2012, **6**:104, https://doi.org/10.1186/1752-0509-6-104.

35. Hillmer RA, Igarashi D, Stoddard T, Lu Y, Liu X, Tsuda K, Katagiri F: **The kinetics and basal levels of the transcriptome response during Effector-Triggered Immunity in Arabidopsis are mainly controlled by four immune signaling sectors**. *Syst Biol* 2023, https://doi.org/10.1101/2023.05.10.540266.

36. Winkelmüller TM, Entila F, Anver S, Piasecka A, Song B, Dahms E, Sakakibara H, Gan X, Kułak K, Sawikowska A, Krajewski P, Tsiantis M, Garrido-Oter R, Fukushima K, Schulze-Lefert P, Laurent S, Bednarek P, Tsuda K: **Gene expression evolution in pattern-triggered immunity within Arabidopsis thaliana and across Brassicaceae species**. *Plant Cell* 2021, **33**:1863−1887, https://doi.org/10.1093/plcell/koab073.

37. Velten B, Braunger JM, Argelaguet R, Arnol D, Wirbel J,
* Bredikhin D, Zeller G, Stegle O: **Identifying temporal and spatial patterns of variation from multimodal data using MEFISTO**. *Nat Methods* 2022, **19**:179−186, https://doi.org/10.1038/s41592-021-01343-9.
This paper presents a factor analytic approach to identify spatial and temporal patterns across multi-omics layers.

38. Argelaguet R, Velten B, Arnol D, Dietrich S, Zenz T, Marioni JC, Buettner F, Huber W, Stegle O: **Multi-Omics Factor Anal-ysis—a framework for unsupervised integration of multi-omics data sets**. *Mol Syst Biol* 2018, **14**:e8124, https://doi.org/10.15252/msb.20178124.

39. Rahimikollu J, Xiao H, Rosengart A, Rosen ABI, Tabib T, Zdinak PM, He K, Bing X, Bunea F, Wegkamp M, Poholek AC, Joglekar AV, Lafyatis RA, Das J: **SLIDE: significant latent factor interaction discovery and exploration across biological do-mains**. *Nat Methods* 2024:1−11, https://doi.org/10.1038/s41592-024-02175-z.

40. Rohart F, Gautier B, Singh A, Cao K-AL: **mixOmics: an R package for 'omics feature selection and multiple data inte-gration**. *PLoS Comput Biol* 2017, **13**, e1005752, https://doi.org/10.1371/journal.pcbi.1005752.

41. Singh A, Shannon CP, Gautier B, Rohart F, Vacher M, Tebbutt SJ, Lê Cao K-A: **DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays**. *Bioinforma. Oxf. Engl.* 2019, **35**:3055−3062, https://doi.org/10.1093/bioinformatics/bty1054.

42. Sheffield NC, Bock C: **LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor**. *Bioinforma. Oxf. Engl.* 2016, **32**:587−589, https://doi.org/10.1093/bioinformatics/btv612.

43. Joe H, Kim H-G: **Multi-label classification with XGBoost for metabolic pathway prediction**. *BMC Bioinf* 2024, **25**:1−15, https://doi.org/10.1186/s12859-024-05666-0.

44. Karp PD, Latendresse M, Caspi R: **The pathway tools pathway prediction algorithm**. *Stand. Genomic Sci.* 2011, **5**:424−429, https://doi.org/10.4056/sigs.1794338.

45. McWhite CD, Papoulas O, Drew K, Cox RM, June V, Dong OX, Kwon T, Wan C, Salmi ML, Roux SJ, Browning KS, Chen ZJ, Ronald PC, Marcotte EM: **A pan-plant protein complex map**

reveals deep conservation and novel assemblies. *Cell* 2020, **181**:460−474.e14, https://doi.org/10.1016/j.cell.2020.02.049.

46. Wang H, Lu H, Sun J, Safo SE: **Interpretable deep learning methods for multiview learning**. *BMC Bioinf* 2024, **25**:1−30, https://doi.org/10.1186/s12859-024-05679-9.

47. Zhao L, Dong Q, Luo C, Wu Y, Bu D, Qi X, Luo Y, Zhao Y: **DeepOmix: a scalable and interpretable multi-omics deep learning framework and application in cancer survival analysis**. *Comput Struct Biotechnol J* 2021, **19**:2719−2725, https://doi.org/10.1016/j.csbj.2021.04.067.

48. Ghosh S, Datta A, Choi H: **multiSLIDE is a web server for exploring connected elements of biological pathways in multi-omics data**. *Nat Commun* 2021, **12**:2279, https://doi.org/10.1038/s41467-021-22650-x.

49. Nishida K, Maruyama J, Kaizu K, Takahashi K, Yugi K: **Trans-omics2cytoscape: an automated software for interpretable 2.5-dimensional visualization of trans-omic networks**. *Npj Syst. Biol. Appl.* 2024, **10**:1−9, https://doi.org/10.1038/s41540-024-00342-8.

50. Zhou G, Ewald J, Xia J: **OmicsAnalyst: a comprehensive web-based platform for visual analytics of multi-omics data**. *Nucleic Acids Res* 2021, **49**:W476−W482, https://doi.org/10.1093/nar/gkab394.

51. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu W-T, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu C-C, Floros DJ, Gavilan RG, Kleigrewe K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw C-C, Yang Y-L, Humpf H-U, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, Klitgaard A, Larson CB, P. CAB, Torres-Mendoza D, Gonzalez DJ, Silva DB, Marques LM, Demarque DP, Pociute E, O'Neill EC, Briand E, Helfrich EJN, Granatosky EA, Glukhov E, Ryffel F, Houson H, Mohimani H, Kharbush JJ, Zeng Y, Vorholt JA, Kurita KL, Charusanti P, McPhail KL, Nielsen KF, Vuong L, Elfeki M, Traxler MF, Engene N, Koyama N, Vining OB, Baric R, Silva RR, Mascuch SJ, Tomasi S, Jenkins S, Macherla V, Hoffman T, Agarwal V, Williams PG, Dai J, Neupane R, Gurr J, Rodríguez AMC, Lamsa A, Zhang C, Dorrestein K, Duggan BM, Almaliti J, Allard P-M, Phapale P, Nothias L-F, Alexandrov T, Litaudon M, Wolfender J-L, Kyle JE, Metz TO, Peryea T, Nguyen D-T, VanLeer D, Shinn P, Jadhav A, Müller R, Waters KM, Shi W, Liu X, Zhang L, Knight R, Jensen PR, Palsson BO, Pogliano K, Linington RG, Gutiérrez M, Lopes NP, Gerwick WH, Moore BS, Dorrestein PC, Bandeira N: **Sharing and community curation of mass spectrometry data with GNPS**. *Nat Biotechnol* 2016, **34**:828−837, https://doi.org/10.1038/nbt.3597.

52. Gomes PWP, Mannochio-Russo H, Schmid R, Zuffa S, Damiani T, Quiros-Guerrero L-M, Caraballo-Rodríguez AM, Zhao HN, Yang H, Xing S, Charron-Lamoureux V, Chigumba DN, Sedio BE, Myers JA, Allard P-M, Harwood TV, Tamayo-Castillo G, Kang KB, Defossez E, Koolen HHF, Da Silva MN, E Silva CYY, Rasmann S, Walker TWN, Glauser G, Chaves-Fallas JM, David B, Kim H, Lee KH, Kim MJ, Choi WJ, Keum Y-S, De Lima EJSP, De Medeiros LS, Bataglion GA, Costa EV, Da Silva FMA, Carvalho ARV, Reis JDE, Pamplona S, Jeong E, Lee K, Kim GJ, Kil Y-S, Nam J-W, Choi H, Han YK, Park SY, Lee KY, Hu C, Dong Y, Sang S, Morrison CR, Borges RM, Teixeira AM, Lee SY, Lee BS, Jeong SY, Kim KH, Rutz A, Gaudry A, Bruelhart E, Kappers IF, Karlova R, Meisenburg M, Berdaguer R, Tello JS, Henderson D, Cayola L, Wright SJ, Allen DN, Anderson-Teixeira KJ, Baltzer JL, Lutz JA, McMahon SM, Parker GG, Parker JD, Northen TR, Bowen BP, Pluskal T, Van Der Hooft JJJ, Carver JJ, Bandeira N, Pullman BS, Wolfender J-L, Kersten RD, Wang M, Dorrestein PC: ***plant-MASST - community-driven chemotaxonomic digitization of plants***. https://doi.org/10.1101/2024.05.13.593988.

53. Jarmusch AK, Wang M, Aceves CM, Advani RS, Aguirre S, Aksenov AA, Aleti G, Aron AT, Bauermeister A, Bolleddu S, Bouslimani A, Caraballo Rodriguez AM, Chaar R, Coras R, Elijah EO, Ernst M, Gauglitz JM, Gentry EC, Husband M, Jarmusch SA, Jones KL, Kamenik Z, Le Gouellec A, Lu A, McCall L-I, McPhail KL, Meehan MJ, Melnik AV, Menezes RC, Montoya Giraldo YA, Nguyen NH, Nothias LF, Nothias-Esposito M, Panitchpakdi M, Petras D, Quinn RA, Sikora N, Van

Der Hooft JJJ, Vargas F, Vrbanac A, Weldon KC, Knight R, Bandeira N, Dorrestein PC: **ReDU: a framework to find and reanalyze public mass spectrometry data**. *Nat Methods* 2020, **17**:901−904, https://doi.org/10.1038/s41592-020-0916-7.

54. Yurekten O, Payne T, Tejera N, Amaladoss FX, Martin C, Williams M, O'Donovan C: **MetaboLights: open data repository for metabolomics**. *Nucleic Acids Res* 2024, **52**:D640−D646, https://doi.org/10.1093/nar/gkad1045.

55. Sud M, Fahy E, Cotter D, Azam K, Vadivelu I, Burant C, Edison A, Fiehn O, Higashi R, Nair KS, Sumner S, Subramaniam S: **Metabolomics Workbench: an international repository for metabolomics data and metadata, metabolite standards, protocols, tutorials and training, and analysis tools**. *Nucleic Acids Res* 2016, **44**:D463−D470, https://doi.org/10.1093/nar/gkv1042.

56. Ewald JD, Zhou G, Lu Y, Kolic J, Ellis C, Johnson JD,
    * Macdonald PE, Xia J: **Web-based multi-omics integration using the Analyst software suite**. *Nat Protoc* 2024, https://doi.org/10.1038/s41596-023-00950-4.
This recently published protocol provides a comprehensive overview of important considerations for multi-omics experimental design, data integration and analysis.

57. Canakoglu A, Bernasconi A, Colombo A, Masseroli M, Ceri S: **GenoSurf: metadata driven semantic search system for integrated genomic datasets**. *Database* 2019, **2019**, baz132, https://doi.org/10.1093/database/baz132.

58. Bremer PL, Fiehn O: **SMetaS: a sample metadata standardizer for metabolomics**. *Metabolites* 2023, **13**:941, https://doi.org/10.3390/metabo13080941.

59. Jackson KC, Pachter L: **A standard for sharing spatial transcriptomics data**. *Cell Genomics* 2023, **3**, 100374, https://doi.org/10.1016/j.xgen.2023.100374.

60. Cooper L, Meier A, Laporte M-A, Elser JL, Mungall C, Sinn BT, Cavaliere D, Carbon S, Dunn NA, Smith B, Qu B, Preece J, Zhang E, Todorovic S, Gkoutos G, Doonan JH, Stevenson DW, Arnaud E, Jaiswal P: **The Planteome database: an integrated resource for reference ontologies, plant genomics and phenomics**. *Nucleic Acids Res* 2018, **46**:D1168−D1180, https://doi.org/10.1093/nar/gkx1152.

61. Cooper L, Elser J, Laporte M-A, Arnaud E, Jaiswal P: **Planteome
    * 2024 update: reference ontologies and knowledgebase for plant biology**. *Nucleic Acids Res* 2024, **52**:D1548−D1555, https://doi.org/10.1093/nar/gkad1028.
In this publication, a recent update is presented on the curation of metadata ontologies specific for experimental plant sciences.

62. Athar A, Füllgrabe A, George N, Iqbal H, Huerta L, Ali A, Snow C, Fonseca NA, Petryszak R, Papatheodorou I, Sarkans U, Brazma A: **ArrayExpress update − from bulk to single-cell expression data**. *Nucleic Acids Res* 2019, **47**:D711−D715, https://doi.org/10.1093/nar/gky964.

63. Koh E, Goh W, Julca I, Villanueva E, Mutwil M: **Peo : plant expression Omnibus − a comparative transcriptomic database for 103 archaeplastida**. *Plant J* 2024, **117**:1592−1603, https://doi.org/10.1111/tpj.16566.

64. Schorn MA, Verhoeven S, Ridder L, Huber F, Acharya DD, Aksenov AA, Aleti G, Moghaddam JA, Aron AT, Aziz S, Bauermeister A, Bauman KD, Baunach M, Beemelmanns C, Beman JM, Berlanga-Clavero MV, Blacutt AA, Bode HB, Boullie A, Brejnrod A, Bugni TS, Calteau A, Cao L, Carrión VJ, Castelo-Branco R, Chanana S, Chase AB, Chevrette MG, Costa-Lotufo LV, Crawford JM, Currie CR, Cuypers B, Dang T, De Rond T, Demko AM, Dittmann E, Du C, Drozd C, Dujardin J-C, Dutton RJ, Edlund A, Fewer DP, Garg N, Gauglitz JM, Gentry EC, Gerwick L, Glukhov E, Gross H, Gugger M, Guillén Matus DG, Helfrich EJN, Hempel B-F, Hur J-S, Iorio M, Jensen PR, Kang KB, Kaysser L, Kelleher NL, Kim CS, Kim KH, Koester I, König GM, Leao T, Lee SR, Lee Y-Y, Li X, Little JC, Maloney KN, Männle D, Martin H. C, McAvoy AC, Metcalf WW, Mohimani H, Molina-Santiago C, Moore BS, Mullowney MW, Muskat M, Nothias L-F, O'Neill EC, Parkinson EI, Petras D, Piel J, Pierce EC, Pires K, Reher R, Romero D, Roper MC, Rust M, Saad H, Saenz C, Sanchez LM, Sørensen SJ, Sosio M, Süssmuth RD, Sweeney D, Tahlan K, Thomson RJ, Tobias NJ, Trindade-Silva AE, Van Wezel GP, Wang M, Weldon KC, Zhang F, Ziemert N, Duncan KR, Crüsemann M, Rogers S, Dorrestein PC,

Medema MH, Van Der Hooft JJJ: **A community resource for paired genomic and metabolomic data mining**. *Nat Chem Biol* 2021, **17**:363−368, https://doi.org/10.1038/s41589-020-00724-z.

65. Hjörleifsson Eldjárn G, Ramsay A, Van Der Hooft JJJ, Duncan KR, Soldatou S, Rousu J, Daly R, Wandy J, Rogers S: **Ranking microbial metabolomic and genomic links in the NPLinker framework using complementary scoring functions**. *PLoS Comput Biol* 2021, **17**, e1008920, https://doi.org/10.1371/journal.pcbi.1008920.

66. Wilkinson MD, Dumontier M, Aalbersberg IjJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, Da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, 'T Hoen PAC, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, Van Schaik R, Sansone S-A, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, Van Der Lei J, Van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B: **The FAIR Guiding Principles for scientific data management and stewardship**. *Sci Data* 2016, **3**, 160018, https://doi.org/10.1038/sdata.2016.18.

67. Buchmann JP, Holmes EC: **Entrezpy: a Python library to dynamically interact with the NCBI Entrez databases**. *Bioinformatics* 2019, **35**:4511−4514, https://doi.org/10.1093/bioinformatics/btz385.

68. Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, Madej T, Marchler-Bauer A, Lanczycki C, Lathrop S, Lu Z, Thibaud-Nissen F, Murphy T, Phan L, Skripchenko Y, Tse T, Wang J, Williams R, Trawick BW, Pruitt KD, Sherry ST: **Database resources of the national center for biotechnology information**. *Nucleic Acids Res* 2022, **50**:D20−D26, https://doi.org/10.1093/nar/gkab1112.

69. Gaudry A, Pagni M, Mehl F, Moretti S, Quiros-Guerrero L-M,
** Cappelletti L, Rutz A, Kaiser M, Marcourt L, Queiroz EF, Ioset J-R, Grondin A, David B, Wolfender J-L, Allard P-M: **A sample-centric and knowledge-driven computational framework for natural products drug discovery**. *ACS Cent Sci* 2024, acscentsci.3c00800, https://doi.org/10.1021/acscentsci.3c00800.
The authors introduce the concept of linking metabolite features with species taxonomy for curation of queryable knowledge graphs. Further, the integration of public data resources via harmonization and standardization of metadata is exemplified for the metabolomics community.

70. Waagmeester A, Stupp G, Burgstaller-Muehlbacher S, Good BM,
* Griffith M, Griffith OL, Hanspers K, Hermjakob H, Hudson TS, Hybiske K, Keating SM, Manske M, Mayers M, Mietchen D, Mitraka E, Pico AR, Putman T, Riutta A, Queralt-Rosinach N, Schriml LM, Shafee T, Slenter D, Stephan R, Thornton K, Tsueng G, Tu R, Ul-Hasan S, Willighagen E, Wu C, Su AI: **Wikidata as a knowledge graph for the life sciences**. *Elife* 2020, **9**, e52614, https://doi.org/10.7554/eLife.52614.
The authors employ Wikidata for the curation of queryable knowledge graphs in the broader context of life sciences. In particular, the interoperability of Wikidata as third level instance for bridging repositories is exemplified for integration of genomics, proteomics, and metabolomics data.

71. Rutz A, Sorokina M, Galgonek J, Mietchen D, Willighagen E,
* Gaudry A, Graham JG, Stephan R, Page R, Vondrášek J, Steinbeck C, Pauli GF, Wolfender J-L, Bisson J, Allard P-M: **The LOTUS initiative for open knowledge management in natural products research**. *Elife* 2022, **11**, e70780, https://doi.org/10.7554/eLife.70780.
The authors introduce the recently curated spectral database LOTUS, which is integrated with Wikidata, and includes a SPARQL query endpoint for integration with taxonomic information.

72. Putman TE, Lelong S, Burgstaller-Muehlbacher S, Waagmeester A, Diesh C, Dunn N, Munoz-Torres M, Stupp GS, Wu C, Su AI, Good BM: **WikiGenomes: an open web application for community consumption and curation of gene annotation data in Wikidata, Database 2017** https://doi.org/10.1093/database/bax025.

73. Braisted J, Patt A, Tindall C, Sheils T, Neyra J, Spencer K, Eicher T, Mathé EA: **RaMP-DB 2.0: a renovated knowledge-base for deriving biological and chemical insight from metabolites, proteins, and genes**. *Bioinformatics* 2023, **39**, btac726, https://doi.org/10.1093/bioinformatics/btac726.

74. Yang Z, Wang S, Wei L, Huang Y, Liu D, Jia Y, Luo C, Lin Y, Liang C, Hu Y, Dai C, Guo L, Zhou Y, Yang Q-Y: **BnIR: a multi-omics database with various tools for *Brassica napus* research and breeding**. *Mol Plant* 2023, **16**:775−789, https://doi.org/10.1016/j.molp.2023.03.007.

75. Jin J, Zhang H, Kong L, Gao G, Luo J: **PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors**. *Nucleic Acids Res* 2014, **42**:D1182−D1187, https://doi.org/10.1093/nar/gkt1016.

76. Busta L, Hall D, Johnson B, Schaut M, Hanson CM, Gupta A, Gundrum M, Wang Y, Maeda HA: **Mapping of specialized metabolite terms onto a plant phylogeny using text mining and large language models**. *Plant J* 2024, tpj.16906, https://doi.org/10.1111/tpj.16906.

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

78. Kerstens MHL, Schranz ME, Bouwmeester K: **Phylogenomic analysis of the APETALA2 transcription factor subfamily across angiosperms reveals both deep conservation and lineage-specific patterns**. *Plant J* 2020, **103**:1516−1524, https://doi.org/10.1111/tpj.14843.

79. Hendriks KP, Kiefer C, Al-Shehbaz IA, Bailey CD, Hooft Van Huysduynen A, Nikolov LA, Nauheimer L, Zuntini AR, German DA, Franzke A, Koch MA, Lysak MA, Toro-Núñez Ó, Özüdoğru B, Invernón VR, Walden N, Maurin O, Hay NM, Shushkov P, Mandáková T, Schranz ME, Thulin M, Windham MD, Rešetnik I, Španiel 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: **Global Brassicaceae phylogeny based on filtering of 1,000-gene dataset**. *Curr Biol* 2023, **33**: 4052−4068.e6, https://doi.org/10.1016/j.cub.2023.08.026.

80. Faust K, Dupont P, Callut J, van Helden J: **Pathway discovery in metabolic networks by subgraph extraction**. *Bioinformatics* 2010, **26**:1211−1218, https://doi.org/10.1093/bioinformatics/btq105.

81. Tzfadia O, Diels T, De Meyer S, Vandepoele K, Aharoni A, Van de Peer Y: **CoExpNetViz: comparative Co-expression networks construction and visualization tool**. *Front Plant Sci* 2016, **6**, https://doi.org/10.3389/fpls.2015.01194.