CartograPlant: Bridging genomic, phenotypic, and environmental data to advance plant resilience and eco-evolutionary insight

Brandon M. Lind[1,2*†], Irene Cobo-Simón[1,3†], Meghan Myles[1†], Gabe Barrett[1], Emily Grau[1], Risharde Ramnath[1], Vlad Savitsky[1], Jill L. Wegrzyn[1,2]

[1]Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269, USA
[2]Institute for Systems Genomics, University of Connecticut, Storrs, CT 06269, USA
[3]Departamento de Ecología y Genética Forestal, Instituto de Ciencias Forestales (ICIFOR), Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria - Consejo Superior de Investigaciones Científicas (ICIFOR-INIA, CSIC), Madrid, Spain

26 September 2025

Running title: *Eco-evolutionary insight with CartograPlant*

Keywords: data integration, data interoperability, meta-analysis, biodiversity informatics, plant adaptation, GWAS, GEA, FAIR

[†]These authors contributed equally

***Corresponding Author**

Brandon M. Lind
ORCID: 0000-0002-8560-5417
lind.brandon.m@gmail.com

Abstract

Climate change is threatening plant health and productivity at all spatial scales, and these impacts are further compounded by the rising incidence of invasive pests and pathogens. Effectively addressing these challenges requires a comprehensive understanding of plant demography as well as the mechanisms and drivers of adaptation. Achieving this understanding requires the integration of physiological, ecological, and genetic datasets. However, such integration is often hindered by disconnected data sources, inconsistent metadata standards, and limited tools to link, analyze, and visualize multi-dimensional datasets in a unified framework. Addressing these hurdles is critical to advancing the understanding of species responses to environmental change and developing informed strategies for conservation, restoration, and adaptive management.

CartograPlant (https://cartograplant.org) is a web-based interactive application which facilitates the visualization and analysis of genotypic, phenotypic, and environmental data, as well as associated metadata, from georeferenced individuals. Developed as a Tripal module, CartograPlant addresses a critical gap in biological data integration by enabling users to explore complex eco-

evolutionary patterns across space and time. Recent updates have expanded its data sources, improved interoperability, and introduced NextFlow pipelines alongside new tools for the integration and analysis of these data. CartograPlant offers a scalable, flexible, and continually-updated platform for researchers, conservationists, land managers, and plant breeders to better understand and mitigate the impacts of global change on plant biodiversity, accelerate resilience in breeding programs, and inform data-driven decisions in agriculture and ecosystem management.

1 | Introduction

Globally, plant species face pressures from habitat loss, native and invasive pests and pathogens, and climate change (Urban 2015, 2024; Nadeau et al. 2017). The upcoming century is forecasted to bring increasingly novel climates with no historical analog (Mahony et al. 2018; Smith et al. 2022; Ordonez et al. 2024; Kerr et al. 2025), which is expected to disrupt relationships between locally adaptive genetic variation and environmental optima of many species (Aitken et al. 2008; Wilczek et al. 2014; McGraw et al. 2015; Browne et al. 2019; Fréjaville et al. 2020; Anderson et al. 2025). Such mismatches exacerbate environmental stresses on plants, resulting in increased susceptibility to both native and invasive pests and pathogens (Hellmann et al. 2008; Mainka and Howard 2010; Velásquez et al. 2018; Hartmann et al. 2022) as well as reductions in species' genetic diversity (Exposito-Alonso et al. 2022; Shaw et al. 2025).

Addressing these challenges requires a comprehensive understanding of plant ecology and evolution, as well as the ability to predict responses to global change (Ehrlén and Morris 2015; Díaz et al. 2019; Johnston et al. 2019; Capblancq et al. 2020; Lee-Yaw et al. 2022; Anderson et al. 2025). Recent advancements in genomic sequencing and high-throughput phenotyping offer excellent opportunities to address these knowledge gaps (Mir et al. 2019;

Bernatchez et al. 2023; León et al. 2023; Nguyen et al. 2025). Knowledge of neutral population structure as well as the genetic variation that confers resilience to climate change, pests, and pathogens can be leveraged for tailored management actions to enhance health and productivity at both the species and ecosystem level (Fraser and Bernatchez 2001; Funk et al. 2012; Barbosa et al. 2018; Thorogood et al. 2023). However, to translate these insights into effective management, it is necessary to integrate diverse data types and approaches to capture the complexity of plant responses to environmental change. Such integration can also enhance our understanding of plant adaptations (Sork et al. 2013; Rellstab et al. 2015; Josephs et al. 2017; Carvalho et al. 2021; Volk et al. 2021; Lasky et al. 2022).

Three main data types are used to understand plant adaptations to environmental variation (Sork et al. 2013): genomic data (*i.e.*, DNA nucleotide and structural polymorphism data as well as the intermediate, reference, and pangenome assemblies to which they are mapped), geospatial data (*i.e.*, georeferenced climatic and environmental data), and phenotypic data, which includes morphological, developmental (*e.g.*, phenology and ontogenic traits), physiological, molecular (*e.g.*, transcriptomic, proteomic, metabolomic, and epigenetic traits), and categorical traits (*e.g.*, infection status, survival, and

94 sex). At the genomic level, two main approaches are used to identify the genetic
95 basis of adaptation: top-down and bottom-up approaches. Top-down
96 approaches attempt to link genomic variation with the phenotypic variation
97 thought to underlie adaptation. For example, genome-wide association studies
98 (GWAS) attempt to identify associations between paired phenotypic and
99 genomic polymorphism data (reviewed in Korte and Farlow 2013). Other top-
100 down approaches include quantitative trait locus (QTL) mapping studies
101 (reviewed in Stinchcombe and Hoekstra 2008). Variation within genomic data
102 can also be associated with various forms of molecular variation, such as in
103 expression quantitative trait locus (eQTL) studies (reviewed in Gilad et al.
104 2008). In contrast to top-down approaches, bottom-up approaches can identify
105 putatively adaptive genetic variation without the *a priori* measurement of
106 phenotypes thought to underlie adaptive responses to the environment (Barrett
107 and Hoekstra 2011; Sork et al. 2013; Rellstab et al. 2015). For example, genome-
108 wide scans of selection such as genotype-environment association analysis
109 (GEA; Hedrick et al. 1976), methods that quantify reductions in heterozygosity
110 due to selective sweeps (*sensu* Smith and Haigh 1974), or allele frequency
111 differences among populations (e.g., Whitlock and Lotterhos 2015) can be used
112 to complement top-down approaches by offering independent lines of evidence

for adaptation at the genetic level (but see Barrett and Hoekstra 2011; Rockman 2012; Tiffin and Ross-Ibarra 2014; Booker et al. 2020).

In addition to genomic approaches, phenotypes measured in reciprocal transplant experiments can be used to establish evidence for local adaptation by comparing the fitness of populations in their sympatric environments with their average fitness in allopatry (Kawecki and Ebert 2004; Blanquart et al. 2013). Additionally, phenotype-environment clines (Savolainen et al. 2007; Hut et al. 2013) found in single or reciprocal transplant experiments, as well as Q_{ST}-F_{ST} comparisons (Spitze 1993; Whitlock 2008), which contrast the degree of differentiation among populations for a quantitative trait (Q_{ST}) with neutral expectations from genetic markers (F_{ST}), can also be used to infer spatially heterogeneous selection. Finally, while these datasets can be sampled from a single time point, temporal variation found in time-series data can also be used to identify molecular variation involved in development (Bar-Joseph et al. 2012; Li et al. 2023) or identify genetic signatures underlying ongoing selection (Schraiber et al. 2016; Buffalo and Coop 2019). While the methods discussed above can lead to valuable biological insight, integrating inferences across these data types can lead to a better understanding of the biotic and abiotic

drivers underlying current and ongoing adaptation that may be overlooked when using any single approach.

Although individual genomic and phenotypic datasets have grown in prevalence and volume, there is increasing recognition of the need for well-curated spatiotemporal metadata to enable future data reuse (Toczydlowski et al. 2021; Crandall et al. 2023; Forsdick et al. 2023; Leigh et al. 2024). Well-curated datasets enable a deeper understanding of adaptive and demographic dynamics within species by facilitating reanalysis, meta-analysis, and mega-analysis. Meta-analysis involves the investigation of pooled summary statistics from previous empirical work, while mega-analysis describes the use of pooled data from multiple independent studies (Eisenhauer 2021; Thompson et al. 2025). Indeed, the increased availability of population genetics datasets and associated metadata has facilitated the emergence of macrogenetics (Blanchet et al. 2017), a field of study with the aim of understanding the broad spatial, temporal, and taxonomic patterns and drivers of genetic diversity through the re-analysis of publicly available datasets. This shift towards broad, integrative analyses highlights the value of data reuse as a powerful framework for uncovering the eco-evolutionary processes that shape genetic diversity across space and time (Leigh et al. 2021; Hoban et al. 2022; Schmidt et al. 2023).

Despite the benefits offered by biological data integration, its implementation can be challenging. Storing and integrating data from diverse sources (genotypic, phenotypic, and environmental) can be difficult due to inconsistent reporting standards and decentralized repositories (Deng et al. 2023). The FAIR (Findable, Accessible, Interoperable, and Reusable) principles provide guidelines to address these issues, promoting machine-actionable data through readable metadata, persistent identifiers, standardized data formats, and the enforcement of community-curated ontologies (Wilkinson et al. 2016). However, FAIR implementation for non-model organisms faces additional obstacles due to a combination of biological, technical, and infrastructural challenges. For instance, many non-model organisms lack reference genomes, or consistent versioning where one exists, and rely on intermediate assemblies that are seldom archived (e.g., RADseq contigs or *de novo* assembled transcriptomes). These limitations are compounded in multi-omics contexts, where differences in methods, technology, and data types complicate data integration and machine-actionable use (Jamil et al. 2020; Hao et al. 2025). While ontology-driven frameworks like MIAPPE (Minimum Information About a Plant Phenotyping Experiment) have indeed improved standardization for trait data, their adoption has not been uniform (Papoutsoglou et al. 2023). Despite these efforts, the widespread adoption of FAIR principles has remained

limited, especially for non-model organisms. Moreover, while some web-based tools such as GWAPP (Seren et al. 2012) and easyGWAS (Grimm et al. 2016) offer integrated approaches to genotypic and phenotypic data, most are primarily designed for use with a few model species and do not support the integration of multidimensional, georeferenced datasets that include environmental variables.

In response to these ongoing challenges, CartograPlant was developed as a Tripal module (Falk et al. 2018) to facilitate the integration, visualization, and analysis of genotypic, phenotypic, and environmental data for georeferenced plants. The Tripal toolkit, which supports numerous databases worldwide, provides a common framework for the storage of genomic, genetic, and breeding data, reducing duplication of effort and improving interoperability (Sanderson et al. 2013; Staton et al. 2021). CartograPlant (Figure 1A-D), previously known as CartograTree, provides an interactive web-based geospatial interface along with analytic tools, NextFlow pipelines, and High-Performance Computing (HPC) resources for real-time analysis. It has been developed for the purpose of integrating data resources about any plant taxon, thereby enhancing ease and breadth of research. Here, we review the

functionality of CartograPlant, including data ingestion and integration methodologies, data availability, and supported analyses.

2 | Data ingestion and integration

CartograPlant is designed to create FAIR datasets from existing and ongoing studies of plant populations that integrate genotype, phenotype, and environmental data in a geospatial context. Datasets are sourced from studies of plants on the landscape, in common gardens (including reciprocal transplant experiments), and in controlled environments such as greenhouses and growth chambers. CartograPlant is designed to host FAIR datasets arising from diverse areas of plant research including landscape genomics, quantitative genetics, ecophysiology, phenology, and conservation biology. Closely associated with the TreeGenes database (https://treegenesdb.org/), CartograPlant historically focused on forest tree species but has recently expanded to accept data from all plant species. CartograPlant has a unique role in providing a platform for the ingestion of studies focused on both traditional model- and non-model plant systems.

Data ingestion is organized at the level of a study, with a focus on enhancing the utility of shared data objects across studies (Figure 1A, Figure 2). These

objects can include genotyping assays applied across studies, plant populations monitored over time, traits measured at multiple time points or with identical ontological definitions, climatic variables derived from environmental layers, and markers which are reused or re-discovered over time. This effort is supported by a structured import mechanism that adheres to FAIR data standards and combines community-sourced study submissions (including from the authors themselves), published studies imported directly by trained biocurators, and external sources of trait metrics, including mobile applications (e.g., TreeSnap; Crocker et al. 2020) and independent repositories (e.g., BIEN; Maitner et al. 2018).

Both direct submissions and internal study biocuration are facilitated by the Tripal Plant Population Submit pipeline (TPPS; http://treegenesdb.org/tpps, Figure 2). TPPS provides users with a questionnaire that adapts dynamically to user responses and permits the direct upload of genotypic, phenotypic, and environmental data (Table 1). Metadata describing the publication and the experimental design is also collected and is consistent with the MIAPPE *v1.1* standard, which supports the description of studies involving both perennial plants and traditional crop models (Papoutsoglou et al. 2020). Reporting standards in MIAPPE are realized through connections to existing biological

ontologies. CartograPlant hosts and actively curates several ontology mappings that describe aspects of genotype, phenotype, and environment (Table 1). Ingested trait and field-measured environmental data include metadata that describes instrumentation, units, scales, timing, and the associated plant anatomical structure(s) (Figure 1A). This detailed annotation facilitates informed mapping to ontologies. Alternatively, environmental data may be incorporated into a study through the selection of any of CartograPlant's 998 global and regional environmental layers, with metrics from chosen layers systematically associated with the georeferenced coordinates of all studied plants (Figure 1B). Whether ingested directly or derived from spatial layers, these datasets are standardized for integration. Once a study has been submitted and approved by CartograPlant biocurators, it can be assigned a Digital Object Identifier (DOI) through Zenodo (European Organization for Nuclear Research 2013). If the dataset already has a DOI (e.g., through Dryad) or accession (e.g., through NCBI), CartograPlant will also reference those existing identifiers, ensuring the linkage between the dataset and any associated publication or previously cited data resources. CartograPlant now hosts geospatially contextualized biological and environmental data from more than 400 studies representing over 25.6 million individuals across 846 species, cultivars, and hybrid complexes georeferenced in more than 40 countries.

These records also span a wide range of eco-evolutionary contexts, and the structured metadata supports integrative cross-study comparisons and synthesis across the breadth of plant science disciplines.

Because CartograPlant aggregates studies collected across many years, the imported datasets reflect a wide range of genomic technologies used to identify genetic variants. In addition, advances in genome sequencing and assembly tools now allow researchers to generate increasingly complete and accurate genomic and transcriptomic reference assemblies. When variants are anchored to a sufficiently accurate and contiguous genomic or transcriptomic reference, it is possible to realign them to new reference assemblies over time. Realigning and storing previously known markers to new assemblies enables the rapid access of marker datasets across all known reference targets, preserves the utility of earlier discovery efforts, and generates consistent identifiers for re-used markers that map onto published names. CartograPlant updates variant datasets by performing automated flank-based remapping of markers to each stored conspecific assembly (implementation described in Supplementary File S1). In addition, while primary sequence repositories like the NCBI's GenBank sequence database (Benson et al. 2013) support the upload of intermediate assemblies including *de novo* transcriptome assemblies, RAD-Seq assemblies,

and resequenced amplicons, in practice they are rarely deposited. Recent developments in CartograPlant have emphasized the submission of intermediate assemblies, along with systematic tracking of reference genomes.

CartograPlant's rich metadata availability, enforcement of standards and ontologies, and automated marker remapping facilitate the integration of both summary statistics and raw data. This ensures that data are prepared for downstream analyses, including meta- and mega-analyses in integration with all (or a subset) of other relevant datasets (see Data analysis).

3 | Data visualization and collection

CartograPlant is designed to facilitate data discovery in a geospatial context. All plant accessions are visualized on a world map (Figure 3A, Figure 3B). Users can subset displayed data by provenance, study descriptors (title, authorship, accession), taxon (family, genus, species), phenotype (structure, description), and marker type (microsatellites, SNPs, indels) using the Plant Data Source (Figure 3C) and Filters (Figure 3D) panels at the top of the page. On the Browse page, accessible from the site's landing page, built-in logical operators allow filtering by any combination of these options, and specific queries can be saved

to the user's session. Users may also toggle on any of 998 environmental layers, whose values at accession coordinates can be viewed directly. Collections of plants within studies or across studies can be transferred into the Analysis Panel for refined queries, download, or HPC-supported analysis. The Coordinate Search panel (Figure 3E) allows users to move to any point on the map.

When a user selects an individual plant on the Browse page (map interface), CartograPlant displays an Accession Summary with available data relevant to that individual (see inset of Figure 3B). These data include the plant's latitude and longitude (and their measurement accuracy), elevation, any included photographs, and metadata about the study or studies with which the plant is associated. In addition, if a user has toggled on any of CartograPlant's environmental layers, the values from that layer at the coordinates of the selected individual are displayed (Figure 3B). The associated study link available from the Accession Summary displays data collected from all individuals described within the source dataset. When a user selects the Information icon associated with an individual, summaries of measured genotypes and phenotypes are displayed. This also allows the user to download flat files associated with these data types to facilitate offline analysis.

Registered users can download or analyze data directly within CartograPlant's Graphic User Interface (GUI), which provides HPC-supported access to several custom bioinformatic workflows (Figure 3E; Figure 4; Table 2; see Data analysis). There are several ways to collect accessions for download or analysis. A Collection is a set of plants and their associated data and metadata. These data can be analyzed or downloaded later (described in the following section). The Accession Summary display, which is activated when a user clicks on a plant, can be used to collect an individual. The Accession Summary display also allows for the Collection of all plants associated with an individual's source study or studies. Alternatively, a user can use the polygon select tool, located in the Map Summary panel, to collect accessions for data download or analysis. After identifying a Collection of plants, users can view information about the studies with which they are associated including the number of plants, the number of measured genotypes, and the number of measured phenotypes per study (Figure 3F).

4 | Data analysis

CartograPlant's Analysis Panel provides an interactive GUI-mediated environment for subsetting, integrating, downloading, and conducting

bioinformatic analyses of stored datasets. These capabilities enable scalable, customizable, and reproducible analyses of plant datasets across diverse taxa and study designs. This flexibility offers unprecedented support for the discovery and description of the genetic and ecological factors that contribute to plant resilience and adaptation to environmental pressures.

Once a user has a Collection of plants of interest from CartograPlant's Browse page (see previous section), the associated data can be downloaded or used to perform various forms of Analysis on CartograPlant (Figure 1C, Figure 4). An Analysis is a specific computational workflow which has been applied to a set of input data. HPC-supported workflows, implemented with NextFlow *v25.04.6* (Di Tommaso et al. 2017), are available within the Analysis Panel for the Collection of plants and/or studies. Registered users can create multiple Workspaces to store data associated with various projects. A CartograPlant Workspace consists of a directory holding raw data associated with a Collection of individuals and, following an Analysis, any generated outputs. Additional raw data and other input files necessary to run an analytic workflow of interest can be uploaded directly to a Workspace from a user's local machine. Each Workspace can be associated with one or more Analyses. Both Analyses and

Workspaces (and their input, intermediate, and resultant files) are private to users and can be saved for future use.

Current supported Analysis options are described in the following sections and include selection and correlation analysis of traits, marker overlap analysis, marker filtering, population structure inference, environmental layer data selection, GWAS, and GEA (Table 2; Figure 4). To meet the needs of bioinformaticians, ecologists, land managers, and breeders, CartograPlant is continually updated with new analytic workflows as data, technologies, and methods evolve.

4.1 | Study-level filtering

Within the Analysis Panel's Select Studies tab, a list of study titles associated with a Collection of plants is displayed. For each study, counts of individuals within the Collection, genotype calls, and phenotype measures are displayed. Users can remove entire studies from their Collection before proceeding with Analysis. The Select Studies tab draws from CartograPlant's standardized plant sample names such that if data about an individual is present in multiple studies within the Collection, these data can be integrated and compared.

4.2 | Phenotypes

Because CartograPlant enforces the mapping of phenotype measures to existing ontologies, trait data from multiple sources can be analyzed together. To facilitate trait-level exploration, the Analysis Panel's Filter Traits tab enables the selection of specific phenotypic measures from input data. A plot, generated with the R *v3.6.0* (R Core Team 2025) libraries *ggplot2* (Wickham 2011) and *ggbiplot* (Vu et al. 2024), is displayed including a two-dimensional trait Principal Component Analysis (PCA; Menozzi et al. 1978; Patterson et al. 2006) and an overlaid biplot with loading values for each of the selected traits. Users can view the displayed plot to determine which, if any, of their selected traits are correlated. Final user trait selection is saved for future use (see Marker association analysis).

4.3 | Leveraging Genotypes across Studies

Identifying shared genetic markers across studies is essential to synthesizing information across studies, expanding the scope of eco-evolutionary insight, and uncovering broad patterns of genetic variation across space and time. Because CartograPlant automatically remaps submitted markers to all available conspecific reference assemblies, including newly released

assemblies as they are added, it is possible to merge all (or any subset of) marker data associated with a species into a single data structure. When studies represent markers mapped to disparate conspecific assemblies, CartograPlant's Select Genotypes tab allows users to leverage *bcftools* (Danecek et al. 2021) to merge them into a combined VCF mapped to a species' *most recent* reference assembly. This merged file is added to the current Workspace and can be used in downstream analysis. In the future, users will be able to visualize marker remapping statistics across *all* available assemblies before choosing a reference for Analysis (remapping implementation described in Supplemental File S1).

Various approaches for the identification of variant loci in a species can sometimes detect the same marker or utilize the same genotype panel. Within the Analysis Panel's Select Genotypes tab, statistics are displayed describing the intersection of marker sets among studies within a Collection. Two studies are considered to share a marker if (i) they both report a marker with the same reference assembly and locus, or if (ii) they report markers on different assemblies that remap to the same locus on the most recent conspecific assembly. The accession of each study with marker data in the Collection is shown along with the reference assembly (or reference assemblies) to which

they were originally mapped. An UpSet plot illustrating overlaps among study marker sets within the Collection is shown in Figure 4.

The accurate detection of signals within genetic data that underlie plant adaptation and resilience depends on well-curated genotype datasets that minimize noise and bias. Rigorous filtering is essential to ensure that downstream GWAS and GEA analyses reflect true biological patterns rather than confounding effects such as population genetic structure (Rellstab et al. 2015; Sul et al. 2018). In the Analysis Panel's Filter Markers and Genotypes tab, genotype data quality indicators including percentage of missing data (both per marker and per individual) and minor allele frequency per marker can be displayed. Users can set thresholds for these statistics to filter markers at both the marker and individual level. The objective of SNP filtering is to remove markers, individuals, and genotype calls likely containing errors or insufficient data due to bioinformatic or wet lab processes (including errant sample labeling). Modular logical operators allow for any combination of these filters to be applied. The Filter Markers and Genotypes tab provides interactive histograms that allow users to select the best threshold values for the SNP quality filtering parameters.

Standard filtering processes include enforcement of a minimum genotype quality score to remove genotypes with poor instrumental confidence, a minimum allele count for each marker (*i.e.*, half of the targeted coverage depth) to ensure well-supported genotype calls, a minimum percentage of missing genotypes per individual to remove individuals that were poorly genotyped, a threshold for missing genotype calls per marker across samples to retain only informative markers, and a minor allele frequency threshold to exclude alleles with insufficient statistical power (Pavan et al. 2020). In the Filter Markers and Genotypes tab, CartograPlant provides a GUI for constructing customizable modular *bcftools view* commands with logical operators to filter a VCF. When a user selects a statistic to begin defining a filter, a histogram of that statistic's distribution is generated to guide threshold choice. If the input VCF already contains pre-computed filter statistics in its header or INFO fields, those values can also be used directly for filtering. The underlying marker data distributions are retrieved with *bcftools query*, and the plots are produced with D3.js (Bostock et al. 2011).

4.4 | Population structure inference

Inferring population genetic structure is a critical step in understanding how demographic history and genetic differentiation shape adaptive potential and

resilience across landscapes. Population genetic structure refers to the presence of subgroups within a population exhibiting differences in allele frequencies shaped by factors including demographic history, isolation by distance, genetic drift, and kinship (Wright 1931, 1943, 1951). The results of many analyses of population-level data, including GWAS, can be skewed or biased if this structure is not considered (Pavan et al. 2020). To allow the inference of population structure from genetic datasets, the Analysis Panel's Assess Population Structure tab allows for the use of fastSTRUCTURE (Raj et al. 2014). STRUCTURE-like analyses are Bayesian approaches that calculate a posterior probability to estimate genetic structure (Pritchard et al. 2000; Falush et al. 2003, 2007; Hubisz et al. 2009). Prior to running fastSTRUCTURE, the genotypes are pruned based on linkage disequilibrium using PLINK (Purcell et al. 2007; Slifer 2018). Inferences from STRUCTURE-like approaches are highly dependent on the chosen number of subpopulations, k (Pritchard et al. 2000). CartograPlant allows users to estimate the degree of population structure across their selected accessions across a range of k-values. For each k, the workflow computes a Q-matrix and a set of marginal likelihoods and chooses a maximizing k. A global ancestry barplot, in which each individual is represented by a stacked bar reflecting the ancestry proportions in the best fastSTRUCTURE Q-matrix, is generated (Wickham 2011).

4.5 | Environmental layers for georeferenced populations

Environmental variation is a major driver of plant distribution, ecology, evolution, and resilience. To enable analyses of genotypic variation within an environmental context, CartograPlant integrates a curated library of environmental layers with relevance to plant ecology, adaptation, and conservation. Environmental data from the location of each selected individual can be retrieved for analysis from CartograPlant's stored environmental layers using the Analysis Panel's Select Environmental Metrics tab. If environmental data is added to an Analysis that accompanies genotypic data, a user can perform GEA (see Marker association analysis).

CartograPlant currently provides access to 998 North American and global environmental layers. These environmental layers fall into four broad categories: 1) biological, environmental, and climate data such as Normalized Difference Vegetation Indices for phenology estimation, delineations of areas of global human influence or biotic damage, aridity indices, and climate data from widely used sources such as ClimateWNA and WorldClim; 2) forest and vegetation data such as tree cover, canopy height, forest fragmentation indices, national forest boundaries, land use, and seed zones for the Eastern United

States; 3) biodiversity and species data including species ranges, biodiversity hotspots, protected areas, ecoregions, NEON ecoclimatic domains, and tree species population data; and 4) agriculturally cultivated and protected areas as well as soil data (Table 3). To add environmental data to an Analysis, users can select specific layers and then click “Gather and Upload to Workspace” (Figure 4). CartograPlant uses the R package *psych (Revelle 2025)* to generate a trellis plot containing several statistics for visualizing relationships between selected environmental variables. These trellis plots allow users to visually assess relationships among environmental variables and identify sets of variables that are relatively uncorrelated, which can be especially useful when preparing data for downstream analyses such as GEA modeling.

4.6 | Genotype-environment and genotype-trait association analysis

Marker association studies provide a powerful framework for the identification of the genetic variation underlying variability in phenotypic traits as well as adaptive responses to the environment, especially within analytical workflows that support the integration of diverse data types and user configurations. In genotype-phenotype analyses, genetic variation at individual loci or windows

is associated with trait measures using models in which genotypes act as predictor variables and phenotypes as response variables. Within the Analysis Panel's Conduct Analysis tab, CartograPlant offers GUI-mediated access to two models from GEMMA (Genome-wide Efficient Mixed Model Association; Zhou and Stephens 2012, 2014): a basic linear model (LM) and a univariate linear mixed model (LMM). The LM tests associations between SNPs and traits without accounting for confounding effects, while the univariate LMM extends this framework by incorporating both fixed effects and random effects that model sample relatedness and population structure. The LMM can also estimate the proportion of variance explained (PVE) by SNP heritability in a phenotype. By supporting both models, CartograPlant enables users to perform GWAS that range from simple association tests to more robust analyses that account for population structure and relatedness.

To complement genotype-phenotype analyses, Cartograplant implements genotype-environment analysis using LFMM2 from the R package LEA (Caye et al. 2019). LFMM2 implements a latent factor mixed model (LFMM) to perform GEA analysis. This LFMM implementation allows for the control of confounding effects, such as population structure, in the dataset. Specifically, LFMM2 uses a least-squares estimation approach for confounder estimation

that provides a unique framework for several categories of genomic data. Upon completion, *p*-value results of a GWAS or GEA analyses are displayed within a Manhattan plot generated with the R package *ggplot2*. This plot groups markers by chromosome and allows for the visualization of markers with high association signals.

Statistical methods which involve the testing of many hypotheses, including GWAS and GEA, often require a multiple testing correction of the resultant *p*-values. As the number of association tests performed grows larger, so too does the probability that one or more results are representative of a Type 1 error (Moskvina and Schmidt 2008; Rellstab et al. 2015). To account for this, *p*-values generated under multiple hypothesis testing can be adjusted using a variety of techniques.

5 | Data-driven approaches to plant adaptation and conservation

5.1. Enabling GWAS and GEA across studies though meta-analysis in model systems

Populus trichocarpa, a widely recognized model for bioenergy research, exhibits extensive latitudinal variation across western North America with substantial evidence for local adaptation to the climatic conditions of its native habitat (Evans et al. 2014; McKown et al. 2014a; Zhang et al. 2019). As a result, *P. trichocarpa* has emerged as a model species for genetics and plant biology research (Tuskan et al. 2006; Taylor et al. 2019), and has been the focus of numerous GWAS and GEA studies. These investigations have explored natural variation in phenology, wood properties, and traits related to lignin biosynthesis. Since the publication of the first reference genome in 2006, four major reference assemblies, along with several intermediate versions and updated annotations, have been released through the Phytozome data portal (Goodstein et al. 2012), each supporting new studies (Tuskan et al. 2006; Sreedasyam et al. 2023). Across these investigations, researchers have applied a range of genotyping and sequencing approaches, including genotyping assays (Geraldes et al. 2013, 2014; McKown et al. 2014b,c), reduced representation methods such as exome capture (Zhou et al. 2014; Guerra et al. 2019), and whole genome resequencing (Slavov et al. 2012).

Many of these studies have relied upon a single, well-characterized common garden experiment. Despite this shared foundation, differences in marker systems, reference genome versions, and annotation frameworks make it challenging to integrate results across studies. Integrating these datasets allows users to identify consistent genotype-phenotype-environment relationships across studies, even when experimental designs and genomic resolutions differ. As a result, the cumulative impact of nearly two decades of research has not been fully realized. CartograPlant seeks to address this by enabling the biocuration of a subset of over 400 studies, systematically mapping variants across reference versions, standardizing trait metadata, and linking findings with spatial and temporal environmental data layers. This integrated framework enhances the reuse of legacy data and supports discovery of trait and environment associations.

5.2 Mobile applications to connect traits, environment, and genotypes for plant resilience

Citizen science has emerged as a powerful resource for plant biologists, and enables the collection of large-scale ecological and trait data that would

577 otherwise be logistically and financially infeasible. The growing participation
578 of the public underscores the need for platforms to translate such observations
579 into actionable resources, particularly for non-model plant species. This is
580 especially true for species of conservation concern. For example, all *Fraxinus*
581 (ash tree) species native to North America are under exceptional threat from
582 the invasive Emerald Ash Borer (EAB, *Agrilus planipennis*; Cappaert et al. 2005;
583 Herms & McCullough 2013). The conservation and restoration of ash species is
584 a coordinated effort involving federal and state agencies, non-profit
585 organizations, academic partners, and citizen scientists. These efforts focus on
586 identifying lingering individuals that may exhibit resistance, understanding
587 stand level dynamics, and documenting ecological pressures (Koch et al. 2012).
588 This work integrates data from federal, state, and county level EAB surveys,
589 citizen science contributions, and landowner observations, alongside the
590 collection of plant material for seed banking, ex situ preservation, and the
591 establishment of long-term monitoring plots.
592
593 CartograPlant provides a framework for these efforts by serving as the
594 integration hub between mobile applications and genomic resources.
595 Currently, TreeSnap (Crocker et al. 2020), a mobile citizen science application
596 designed for species of forest health concern, enables custom surveys that

capture both individual and stand-level data. In the case of ash, TreeSnap hosts survey modules tailored for both general users and experts who can provide more detailed information on EAB progression. In addition, CartograPlant has begun to incorporate data from MaMA (Monitoring and Managing Ash; https://www.monitoringash.org/), an independent assessment program that collects both qualitative and quantitative data from regional monitoring plots. Traits collected through TreeSnap and MaMA, can be connected to genotype data on a subset of the phenotyped accessions through TPPS.

5.3. Eco-evolutionary dynamics in rapidly changing climates

Climate change is rapidly reshaping species distributions, altering community dynamics and species interactions, and exerting strong selection pressure on organisms that can cascade to ecosystem-level effects (Bailey et al. 2009). To understand and ameliorate these impacts, researchers increasingly emphasize the integration of ecological and evolutionary perspectives (Bolnick et al. 2011; Müller 2017; Brady et al. 2019; Fronhofer et al. 2023; Fouqueau and Polechová 2024).

For instance, the Arctic, which is warming four times faster than global norms (ACIA 2004; Hobbie et al. 2017), offers a uniquely tractable system to study

evolutionary responses across multiple levels of biological organization under this eco-evolutionary framework. With relatively low species diversity and a strong coupling between biotic and abiotic processes, even subtle shifts could result in cascading effects on Arctic ecosystem structure and function (Wookey et al. 2009). Research on Arctic stream-riparian systems - particularly those dominated by keystone willow (*Salix*) species - are shedding light on how adaptive evolution influences ecosystem function. These stream-riparian systems are sensitive to ongoing climate change (Myers-Smith et al. 2011; Frost and Epstein 2014; Criado et al. 2020; Anderson et al. 2024), with potential cascading effects across trophic levels (Hollister et al. 2015; Zhou et al. 2020; Mekonnen et al. 2021). To support such investigations, CartograPlant aids in linking field observations, *in situ* organismal traits, common gardens, genetic sampling, and pertinent environmental data layers. For example, layers associated with permafrost, ground temperature, and hydrology have been incorporated to support this integration. Alongside, standardized, long-term environmental and ecological data from the National Ecological Observatory Network (NEON), these layers together provide the foundation for assessing ecological and evolutionary processes in *Salix* and other Arctic species, including rates of hybridization, environmental clines in trait and genetic diversity, as well as shifts in ploidy.

6 | Summary and Future Directions

CartograPlant streamlines the integration and analysis of diverse biological data by adhering to FAIR principles and standardized ontologies. It offers flexible, quality-controlled workflows for data filtering, analysis and visualization, facilitating reproducible insights into plant resilience, adaptation, and demography. Designed to evolve alongside emerging technologies and data sources, CartograPlant supports future integration of structural variation, pangenomes, high-throughput phenotyping data, as well as connections with primary repositories, ensuring its continued relevance in the rapidly advancing fields of ecology and evolution. For example, CartograPlant is actively developing two-way data integration with the European Variation Archive (Cezard et al. 2021), enabling automatic synchronization of variant datasets across platforms. This interoperability will further streamline data discovery, ensure that dataset identifiers are mappable between systems, and reduce duplication of biocuration effort. Because CartograPlant has the ability to ingest environmental and phenotypic data as well as detailed study metadata in addition to variant data, this collaboration

has the potential to increase the amount of information available about a given variant dataset.

By integrating genotypic, phenotypic, and environmental data alongside a user-friendly interface, CartograPlant advances our understanding of biodiversity, ecological responses, and the traits that shape plant evolution in the face of environmental change and biotic stressors.

7 | Author Contributions

Brandon M. Lind: Conceptualization; Investigation; Methodology; Writing - original draft; Writing - review & editing; Visualization

Irene Cobo-Simón: Conceptualization; Investigation; Methodology; Formal analysis, Writing - original draft; Writing - review & editing; Visualization

Meghan Myles: Conceptualization; Data Curation; Formal analysis; Methodology; Software; Writing - original draft; Writing - review & editing; Visualization

Gabe Barrett: Data Curation; Formal analysis; Methodology; Software; Writing - review & editing

Emily Grau: Data Curation; Formal analysis; Software; Writing - review & editing

Risharde Ramnath: Data Curation; Formal analysis; Software; Writing - review & editing

Vlad Savitsky: Data Curation; Formal analysis; Software; Writing - review & editing

Jill L. Wegrzyn: Conceptualization; Funding acquisition; Project administration; Resources; Software; Writing - review & editing

8 | Acknowledgments

This research was supported by the intramural research program of the U.S. Department of Agriculture (USDA), National Institute of Food and Agriculture (NIFA), Agricultural Genome to Phenome Initiative, #2022-70412-38454, #2021-70412-35233, and #2020-70412-32615 (JLW, MM, ICS); USDA NIFA FACT #2019-67021-29920 (JLW); USDA NIFA #2024-51181-43233 (JLW); NSF BII 2320675 (JLW); The Trees In Peril Project funded by the Nature Conservancy (JLW). We would also like to thank the Institute for Systems Genetics and the Computational Biology Core at the University of Connecticut for access to high-performance computing infrastructure and technical support. In addition, we would like to thank Radka Wildova and Jonathan Rosenthal of the Monitoring and Managing Ash (MaMA) program for access to collected data. We would like to thank the undergraduate biocuration team, including Isabella Harding, Trang Nguyen, Phoebe Zhou, for their contributions. Finally, we would like to thank the developers of TreeSnap, Meg Staton and Florence Caldwell of the University of Tennessee, for app-based support.

9 | Data Availability

All data used in this study are publicly available through the CartograPlant online database (https://cartograplant.org/). Specific datasets referenced in this publication can be accessed by navigating to the relevant study pages within CartograPlant. Additional information about CartograPlant is available at https://cartograplant-tpps.readthedocs.io/en/latest/. CartograPlant accession TGDR 1892 (Geraldes et al., 2013) and accession TGDR 2269 (McKown et al., 2014c) were used to create Figure 4.

References

ACIA. 2004. Impacts of a Warming Arctic, Arctic Climate Impact Assessment. Cambridge University Press.

Aitken, S. N., S. Yeaman, J. A. Holliday, T. Wang, and S. Curtis-McLane. 2008. Adaptation, migration or extirpation: climate change outcomes for tree populations. Evol. Appl. 1:95–111.

Anderson, J. T., M. L. DeMarche, D. A. Denney, I. Breckheimer, J. Santangelo, and S. M. Wadgymar. 2025. Adaptation and gene flow are insufficient to rescue a montane plant under climate change. Science 388:525–531.

Anderson, M., I. Myers-Smith, E. Zaja, H. Thomas, M. G. Criado, G. Daskalova, E. Gallois, J. J. Šubrt, and M. Vellend. 2024. Earlier and increased growth of tundra willows after a decade of growth in a warmer common garden environment. EcoEvoRxiv.

Bailey, J. K., J. A. Schweitzer, F. beda, J. Koricheva, C. J. LeRoy, M. D. Madritch, B. J. Rehill, R. K. Bangert, D. G. Fischer, G. J. Allan, and T. G. Whitham. 2009. From genes to ecosystems: a synthesis of the effects of plant genetic factors across levels of organization. Philos. Trans. R. Soc. B: Biol. Sci. 364:1607–1616.

Barbosa, S., F. Mestre, T. A. White, J. Paupério, P. C. Alves, and J. B. Searle. 2018. Integrative approaches to guide conservation decisions: Using genomics to define conservation units and functional corridors. Mol. Ecol. 27:3452–3465.

Bar-Joseph, Z., A. Gitter, and I. Simon. 2012. Studying and modelling dynamic biological processes using time-series gene expression data. Nat. Rev. Genet. 13:552–564.

Barrett, R. D. H., and H. E. Hoekstra. 2011. Molecular spandrels: tests of adaptation at the genetic level. Nat. Rev. Genet. 12:767–780.

Benson, D. A., M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, J. Ostell, K. D. Pruitt, and E. W. Sayers. 2013. GenBank. Nucleic Acids Res. 46:gkx1094-.

Bernatchez, L., A.-L. Ferchaud, C. S. Berger, C. J. Venney, and A. Xuereb. 2023. Genomics for monitoring and understanding species responses to global climate change. Nat. Rev. Genet. 1–19.

Blanchet, S., J. G. Prunier, and H. D. Kort. 2017. Time to Go Bigger: Emerging Patterns in Macrogenetics. Trends Genet. 33:579–580.

Blanquart, F., O. Kaltz, S. L. Nuismer, and S. Gandon. 2013. A practical guide to measuring local adaptation. Ecol. Lett. 16:1195–1205.

Bolnick, D. I., P. Amarasekare, M. S. Araújo, R. Bürger, J. M. Levine, M. Novak, V. H. W. Rudolf, S. J. Schreiber, M. C. Urban, and D. A. Vasseur. 2011. Why intraspecific trait variation matters in community ecology. Trends Ecol. Evol. 26:183–192.

739 Booker, T. R., S. Yeaman, and M. C. Whitlock. 2020. Variation in recombination
740 rate affects detection of outliers in genome scans under neutrality. Mol. Ecol.
741 29:4274–4279.
742 Bostock, M., V. Ogievetsky, and J. Heer. 2011. D³ Data-Driven Documents.
743 IEEE Trans. Vis. Comput. Graph. 17:2301–2309.
744 Brady, S. P., D. I. Bolnick, R. D. H. Barrett, L. Chapman, E. Crispo, A. M. Derry, C.
745 G. Eckert, D. J. Fraser, G. F. Fussmann, A. Gonzalez, F. Guichard, T. Lamy, J.
746 Lane, A. G. McAdam, A. E. M. Newman, A. Paccard, B. Robertson, G.
747 Rolshausen, P. M. Schulte, A. M. Simons, M. Vellend, and A. Hendry. 2019.
748 Understanding Maladaptation by Uniting Ecological and Evolutionary
749 Perspectives. Am. Nat. 194:495–515.
750 Browne, L., J. W. Wright, S. Fitz-Gibbon, P. F. Gugger, and V. L. Sork. 2019.
751 Adaptational lag to temperature in valley oak (Quercus lobata) can be
752 mitigated by genome-informed assisted gene flow. Proc. Natl. Acad. Sci.
753 116:25179–25185.
754 Buffalo, V., and G. Coop. 2019. The Linked Selection Signature of Rapid
755 Adaptation in Temporal Genomic Data. Genetics 213:1007–1045.
756 Capblancq, T., M. C. Fitzpatrick, R. A. Bay, M. Exposito-Alonso, and S. R. Keller.
757 2020. Genomic Prediction of (Mal)Adaptation Across Current and Future
758 Climatic Landscapes. Annu. Rev. Ecol. Evol. Syst. 51:245–269.
759 Cappaert, D., D. G. McCullough, T. M. Poland, and N. W. Siegert. 2005. Emerald
760 ash borer in North America: a research and regulatory challenge. *Am.*
761 *Entomol.* 51(3):152–165.
762 Carvalho, C. S., B. R. Forester, S. K. Mitre, R. Alves, V. L. Imperatriz-Fonseca, S.
763 J. Ramos, L. C. Resende-Moreira, J. O. Siqueira, L. C. Trevelin, C. F. Caldeira,
764 M. Gastauer, and R. Jaffé. 2021. Combining genotype, phenotype, and
765 environmental data to delineate site-adjusted provenance strategies for
766 ecological restoration. Mol. Ecol. Resour. 21:44–58.
767 Caye, K., B. Jumentier, J. Lepeule, and O. François. 2019. LFMM 2: Fast and
768 Accurate Inference of Gene-Environment Associations in Genome-Wide
769 Studies. Mol. Biol. Evol. 36:852–860.
770 Cezard, T., F. Cunningham, S. E. Hunt, B. Koylass, N. Kumar, G. Saunders, A.
771 Shen, A. F. Silva, K. Tsukanov, S. Venkataraman, P. Flicek, H. Parkinson, and
772 T. M. Keane. 2021. The European Variation Archive: a FAIR resource of
773 genomic variation for all species. Nucleic Acids Res. 50:D1216–D1220.
774 Crandall, E. D., R. H. Toczydlowski, L. Liggins, A. E. Holmes, M. Ghoojaei, M. R.
775 Gaither, B. E. Wham, A. L. Pritt, C. Noble, T. J. Anderson, R. L. Barton, J. T.
776 Berg, S. G. Beskid, A. Delgado, E. Farrell, N. Himmelsbach, S. R. Queeno, T.
777 Trinh, C. Weyand, A. Bentley, J. Deck, C. Riginos, G. S. Bradburd, and R. J.

778 Toonen. 2023. Importance of timely metadata curation to the global
779 surveillance of genetic diversity. Conserv. Biol. 37:e14061.
780 Criado, M. G., I. H. Myers-Smith, A. D. Bjorkman, C. E. R. Lehmann, and N.
781 Stevens. 2020. Woody plant encroachment intensifies under climate change
782 across tundra and savanna biomes. Glob. Ecol. Biogeogr. 29:925–943.
783 Crocker, E., B. Condon, A. Almsaeed, B. Jarret, C. D. Nelson, A. G. Abbott, D.
784 Main, and M. Staton. 2020. TreeSnap: A citizen science app connecting tree
785 enthusiasts and forest scientists. Plants, People, Planet 2:47–52.
786 Danecek, P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A.
787 Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, and H. Li. 2021. Twelve
788 years of SAMtools and BCFtools. GigaScience 10:giab008.
789 Deng, C. H., S. Naithani, S. Kumari, I. Cobo-Simón, E. H. Quezada-Rodríguez, M.
790 Skrabisova, N. Gladman, M. J. Correll, A. B. Sikiru, O. O. Afuwape, A.
791 Marrano, I. Rebollo, W. Zhang, and S. Jung. 2023. Genotype and phenotype
792 data standardization, utilization and integration in the big data era for
793 agricultural sciences. Database 2023:baad088.
794 Di Tommaso, P. D., M. Chatzou, E. W. Floden, P. P. Barja, E. Palumbo, and C.
795 Notredame. 2017. Nextflow enables reproducible computational workflows.
796 Nat. Biotechnol. 35:316–319.
797 Díaz, S., J. Settele, E. S. Brondízio, H. T. Ngo, J. Agard, A. Arneth, P. Balvanera, K.
798 A. Brauman, S. H. M. Butchart, K. M. A. Chan, L. A. Garibaldi, K. Ichii, J. Liu,
799 S. M. Subramanian, G. F. Midgley, P. Miloslavich, Z. Molnár, D. Obura, A.
800 Pfaff, S. Polasky, A. Purvis, J. Razzaque, B. Reyers, R. R. Chowdhury, Y.-J.
801 Shin, I. Visseren-Hamakers, K. J. Willis, and C. N. Zayas. 2019. Pervasive
802 human-driven decline of life on Earth points to the need for transformative
803 change. Science 366.
804 Ehrlén, J., and W. F. Morris. 2015. Predicting changes in the distribution and
805 abundance of species under environmental change. Ecol. Lett. 18:303–314.
806 Eisenhauer, J. G. 2021. Meta-analysis and mega-analysis: A simple introduction.
807 Teach. Stat. 43:21–27.
808 European Organization for Nuclear Research (CERN) and OpenAIRE, eds.
809 Zenodo.org. Geneva: CERN, 2013. https://zenodo.org.
810 Evans, L. M., G. T. Slavov, E. Rodgers-Melnick, J. Martin, P. Ranjan, W. Muchero,
811 A. M. Brunner, W. Schackwitz, L. Gunter, J.-G. Chen, G. A. Tuskan, and S. P.
812 DiFazio. 2014. Population genomics of Populus trichocarpa identifies
813 signatures of selection and adaptive trait associations. Nat. Genet. 46:1089–
814 1096.
815 Exposito-Alonso, M., T. R. Booker, L. Czech, L. Gillespie, S. Hateley, C. C.
816 Kyriazis, P. L. M. Lang, L. Leventhal, D. Nogues-Bravo, V. Pagowski, M.

817 Ruffley, J. P. Spence, S. E. T. Arana, C. L. Weiß, and E. Zess. 2022. Genetic
818 diversity loss in the Anthropocene. Science 377:1431–1435.
819 Falk, T., N. Herndon, E. Grau, S. Buehler, P. Richter, S. Zaman, E. M. Baker, R.
820 Ramnath, S. Ficklin, M. Staton, F. A. Feltus, S. Jung, D. Main, and J. L.
821 Wegrzyn. 2018. Growing and cultivating the forest genomics database,
822 TreeGenes. Database 2018:bay084.
823 Falush, D., M. Stephens, and J. K. Pritchard. 2007. Inference of population
824 structure using multilocus genotype data: dominant markers and null
825 alleles. Mol. Ecol. Notes 7:574–578.
826 Falush, D., M. Stephens, and J. K. Pritchard. 2003. Inference of Population
827 Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele
828 Frequencies. Genetics 164:1567–1587.
829 Fick, S. E., and R. J. Hijmans. 2017. WorldClim 2: new 1 km spatial resolution
830 climate surfaces for global land areas. Int. J. Climat. 37:4302–4315.
831 Forsdick, N. J., J. Wold, A. Angelo, F. Bissey, J. Hart, M. Head, L. Liggins, D.
832 Senanayake, and T. E. Steeves. 2023. Journeying towards best practice data
833 management in biodiversity genomics. Mol. Ecol. Res. 00:1-13.
834 Fouqueau, L., and J. Polechová. 2024. Eco-evolutionary dynamics in changing
835 environments: integrating theory with data. J. Evol. Biol. 37:579–587.
836 Fraser, D. J., and L. Bernatchez. 2001. Adaptive evolutionary conservation:
837 towards a unified concept for defining conservation units. Mol. Ecol.
838 10:2741–2752.
839 Fréjaville, T., N. Vizcaíno-Palomar, B. Fady, A. Kremer, and M. B. Garzón. 2020.
840 Range margin populations show high climate adaptation lags in European
841 trees. Glob. Chang. Biol. 26:484–495.
842 Fronhofer, E. A., D. Corenblit, J. N. Deshpande, L. Govaert, P. Huneman, F. Viard,
843 P. Jarne, and S. Puijalon. 2023. Eco-evolution from deep time to
844 contemporary dynamics: The role of timescales and rate modulators. Ecol.
845 Lett. 26:S91–S108.
846 Frost, G. V., and H. E. Epstein. 2014. Tall shrub and tree expansion in Siberian
847 tundra ecotones since the 1960s. Glob. Chang. Biol. 20:1264–1277.
848 Funk, W. C., J. K. McKay, P. A. Hohenlohe, and F. W. Allendorf. 2012. Harnessing
849 genomics for delineating conservation units. Trends Ecol. Evol. 27:489–496.
850 Geraldes, A., S. P. DiFazio, G. T. Slavov, P. Ranjan, W. Muchero, J. Hannemann,
851 L. E. Gunter, A. M. Wymore, C. J. Grassa, N. Farzaneh, I. Porth, A. D.
852 McKown, O. Skyba, E. Li, M. Fujita, J. Klápště, J. Martin, W. Schackwitz, C.
853 Pennacchio, D. Rokhsar, M. C. Friedmann, G. O. Wasteneys, R. D. Guy, Y. A.
854 El-Kassaby, S. D. Mansfield, Q. C. B. Cronk, J. Ehlting, C. J. Douglas, and G. A.
855 Tuskan. 2013. A 34K SNP genotyping array for Populus trichocarpa: Design,

856 application to the study of natural populations and transferability to other
857 Populus species. Mol. Ecol. Resour. 13:306–323.
858 Geraldes, A., N. Farzaneh, C. J. Grassa, A. D. McKown, R. D. Guy, S. D. Mansfield,
859 C. J. Douglas, and Q. C. B. Cronk. 2014. LANDSCAPE GENOMICS OF POPULUS
860 TRICHOCARPA: THE ROLE OF HYBRIDIZATION, LIMITED GENE FLOW,
861 AND NATURAL SELECTION IN SHAPING PATTERNS OF POPULATION
862 STRUCTURE. Evolution 68:3260–3280.
863 Gilad, Y., S. A. Rifkin, and J. K. Pritchard. 2008. Revealing the architecture of gene
864 regulation: the promise of eQTL studies. Trends Genet. 24:408–415.
865 Goodstein, D. M., S. Shu, R. Howson, R. Neupane, R. D. Hayes, J. Fazo, T. Mitros,
866 W. Dirks, U. Hellsten, N. Putnam, and D. S. Rokhsar. 2012. Phytozome: a
867 comparative platform for green plant genomics. Nucleic Acids Res.
868 40:D1178–D1186.
869 Grimm, D. G., D. Roqueiro, P. A. Salomé, S. Kleeberger, B. Greshake, W. Zhu, C.
870 Liu, C. Lippert, O. Stegle, B. Schölkopf, D. Weigel, and K. M. Borgwardt. 2016.
871 easyGWAS: A Cloud-Based Platform for Comparing the Results of Genome-
872 Wide Association Studies. Plant Cell 29:5–19.
873 Guerra, F. P., H. Suren, J. Holliday, J. H. Richards, O. Fiehn, R. Famula, B. J.
874 Stanton, R. Shuren, R. Sykes, M. F. Davis, and D. B. Neale. 2019. Exome
875 resequencing and GWAS for growth, ecophysiology, and chemical and
876 metabolomic composition of wood of Populus trichocarpa. BMC Genom.
877 20:875.
878 Hao, Y., Z. Zhang, E. Luo, J. Yang, and S. Wang. 2025. Plant metabolomics:
879 applications and challenges in the era of multi-omics big data. aBIOTECH
880 6:116–132.
881 Hartmann, H., A. Bastos, A. J. Das, A. Esquivel-Muelbert, W. M. Hammond, J.
882 Martínez-Vilalta, N. G. McDowell, J. S. Powers, T. A. M. Pugh, K. X. Ruthrof,
883 and C. D. Allen. 2022. Climate Change Risks to Global Forest Health:
884 Emergence of Unexpected Events of Elevated Tree Mortality Worldwide.
885 Annu. Rev. Plant Biol. 73:673–702.
886 Hedrick, P. W., M. E. Ginevan, and E. P. Ewing. 1976. Genetic Polymorphism in
887 Heterogeneous Environments. Annu. Rev. Ecol. Syst. 7:1–32.
888 Hellmann, J. J., J. E. Byers, B. G. Bierwagen, and J. S. Dukes. 2008. Five Potential
889 Consequences of Climate Change for Invasive Species. Conserv. Biol. 22:534–
890 543.
891 Herms, D. A., and D. G. McCullough. 2014. Emerald ash borer invasion of North
892 America: history, biology, ecology, impacts, and management. Annu. Rev.
893 Entomol. 59:13–30.

894 Hoban, S., F. I. Archer, L. D. Bertola, J. G. Bragg, M. F. Breed, M. W. Bruford, M.
895 A. Coleman, R. Ekblom, W. C. Funk, C. E. Grueber, B. K. Hand, R. Jaffé, E.
896 Jensen, J. S. Johnson, F. Kershaw, L. Liggins, A. J. MacDonald, J. Mergeay, J.
897 M. Miller, F. Muller-Karger, D. O'Brien, I. Paz-Vinas, K. M. Potter, O. Razgour,
898 C. Vernesi, and M. E. Hunter. 2022. Global genetic diversity status and
899 trends: towards a suite of Essential Biodiversity Variables (EBVs) for genetic
900 composition. Biol. Rev. 97:1511–1538.
901 Hobbie, J. E., G. R. Shaver, E. B. Rastetter, J. E. Cherry, S. J. Goetz, K. C. Guay, W.
902 A. Gould, and G. W. Kling. 2017. Ecosystem responses to climate change at a
903 Low Arctic and a High Arctic long-term research site. Ambio 46:160–173.
904 Hollister, R. D., J. L. May, K. S. Kremers, C. E. Tweedie, S. F. Oberbauer, J. A.
905 Liebig, T. F. Botting, R. T. Barrett, and J. L. Gregory. 2015. Warming
906 experiments elucidate the drivers of observed directional changes in tundra
907 vegetation. Ecol. Evol. 5:1881–1895.
908 Hubisz, M. J., D. Falush, M. Stephens, and J. K. Pritchard. 2009. Inferring weak
909 population structure with the assistance of sample group information. Mol.
910 Ecol. Resour. 9:1322–1332.
911 Hut, R. A., S. Paolucci, R. Dor, C. P. Kyriacou, and S. Daan. 2013. Latitudinal
912 clines: an evolutionary view on biological rhythms. Proc. R. Soc. B: Biol. Sci.
913 280:20130433.
914 Jamil, I. N., J. Remali, K. A. Azizan, N. A. N. Muhammad, M. Arita, H.-H. Goh, and
915 W. M. Aizat. 2020. Systematic Multi-Omics Integration (MOI) Approach in
916 Plant Systems Biology. Front. Plant Sci. 11:944.
917 Johnston, A. S. A., R. J. Boyd, J. W. Watson, A. Paul, L. C. Evans, E. L. Gardner,
918 and V. L. Boult. 2019. Predicting population responses to environmental
919 change from individual-level mechanisms: towards a standardized
920 mechanistic approach. Proc. R. Soc. B 286:20191916.
921 Josephs, E. B., J. R. Stinchcombe, and S. I. Wright. 2017. What can genome-wide
922 association studies tell us about the evolutionary forces maintaining genetic
923 variation for quantitative traits? New Phytologist 214:21–33.
924 Kawecki, T. J., and D. Ebert. 2004. Conceptual issues in local adaptation. Ecol.
925 Lett. 7:1225–1241.
926 Kerr, M. R., A. Ordonez, F. Riede, J. Atkinson, E. A. Pearce, M. Sykut, J. Trepel,
927 and J.-C. Svenning. 2025. Widespread ecological novelty across the
928 terrestrial biosphere. Nat. Ecol. Evol. 9:589–598.
929 Koch, J. L., D. W. Carey, K. S. Knight, T. Poland, D. A. Herms, and M. E. Mason.
930 2012. Breeding strategies for the development of emerald ash borer-resistant
931 North American ash. Pp. 235–239 in R. A. Sniezko, A. D. Yanchuk, J. T.
932 Kliejunas, K. M. Palmieri, J. M. Alexander, and S. J. Frankel, tech.

coords. Proceedings of the fourth international workshop on the genetics of host-parasite interactions in forestry: Disease and insect resistance in forest trees. Gen. Tech. Rep. PSW-GTR-240. Pacific Southwest Research Station, Forest Service, U.S. Department of Agriculture, Albany, CA.

Korte, A., and A. Farlow. 2013. The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9:29.

Lasky, J. R., E. B. Josephs, and G. P. Morris. 2022. Genotype–environment associations to reveal the molecular basis of environmental adaptation. Plant Cell 35:125–138.

Lee-Yaw, J. A., J. L. McCune, S. Pironon, and S. N. Sheth. 2022. Species distribution models rarely predict the biology of real populations. Ecography 2022.

Leigh, D. M., C. B. van Rees, K. L. Millette, M. F. Breed, C. Schmidt, L. D. Bertola, B. K. Hand, M. E. Hunter, E. L. Jensen, F. Kershaw, L. Liggins, G. Luikart, S. Manel, J. Mergeay, J. M. Miller, G. Segelbacher, S. Hoban, and I. Paz-Vinas. 2021. Opportunities and challenges of macrogenetic studies. Nat. Rev. Genet. 22:791–807.

Leigh, D. M., A. G. Vandergast, M. E. Hunter, E. D. Crandall, W. C. Funk, C. J. Garroway, S. Hoban, S. J. Oyler-McCance, C. Rellstab, G. Segelbacher, C. Schmidt, E. Vázquez-Domínguez, and I. Paz-Vinas. 2024. Best practices for genetic and genomic data archiving. Nat. Ecol. Evol. 8:1224–1232.

León, L. F. D., B. Silva, K. J. Avilés-Rodríguez, and D. Buitrago-Rosas. 2023. Harnessing the omics revolution to address the global biodiversity crisis. 80:102901.

Li, D., D. Bai, Y. Tian, Y. Li, C. Zhao, Q. Wang, S. Guo, Y. Gu, X. Luan, R. Wang, J. Yang, M. J. Hawkesford, J. C. Schnable, X. Jin, and L. Qiu. 2023. Time series canopy phenotyping enables the identification of genetic variants controlling dynamic phenotypes in soybean. J. Integr. Plant Biol. 65:117–132.

Mahony, C. R., W. H. MacKenzie, and S. Aitken. 2018. Novel climates: Trajectories of climate change beyond the boundaries of British Columbia's forest management knowledge system. For. Ecol. Manag. 410:35–47.

Mainka, S. A., and G. W. Howard. 2010. Climate change and invasive species: double jeopardy. Integr. Zoöl. 5:102–111.

Maitner, B. S., B. Boyle, N. Casler, R. Condit, J. Donoghue, S. M. Durán, D. Guaderrama, C. E. Hinchliff, P. M. Jørgensen, N. J. B. Kraft, B. McGill, C. Merow, N. Morueta-Holme, R. K. Peet, B. Sandel, M. Schildhauer, S. A. Smith, J. Svenning, B. Thiers, C. Violle, S. Wiser, and B. J. Enquist. 2018. The bien r package: A tool to access the Botanical Information and Ecology Network (BIEN) database. Methods Ecol. Evol. 9:373–379.

McGraw, J. B., J. B. Turner, S. Souther, C. C. Bennington, M. C. Vavrek, G. R. Shaver, and N. Fetcher. 2015. Northward displacement of optimal climate conditions for ecotypes of Eriophorum vaginatum L. across a latitudinal gradient in Alaska. Glob. Chang. Biol. 21:3827–3835.

McKown, A. D., R. D. Guy, J. Klápště, A. Geraldes, M. Friedmann, Q. C. B. Cronk, Y. A. El-Kassaby, S. D. Mansfield, and C. J. Douglas. 2014a. Geographical and environmental gradients shape phenotypic trait variation and genetic structure in Populus trichocarpa. N. Phytol. 201:1263–1276.

McKown, A. D., R. D. Guy, L. Quamme, J. Klápště, J. L. Mantia, C. P. Constabel, Y. A. El-Kassaby, R. C. Hamelin, M. Zifkin, and M. S. Azam. 2014b. Association genetics, geography and ecophysiology link stomatal patterning in Populus trichocarpa with carbon gain and disease resistance trade-offs. Mol. Ecol. 23:5771–5790.

McKown, A. D., J. Klápště, R. D. Guy, A. Geraldes, I. Porth, J. Hannemann, M. Friedmann, W. Muchero, G. A. Tuskan, J. Ehlting, Q. C. B. Cronk, Y. A. El-Kassaby, S. D. Mansfield, and C. J. Douglas. 2014c. Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of Populus trichocarpa. N. Phytol. 203:535–553.

Mekonnen, Z. A., W. J. Riley, L. T. Berner, N. J. Bouskill, M. S. Torn, G. Iwahana, A. L. Breen, I. H. Myers-Smith, M. G. Criado, Y. Liu, E. S. Euskirchen, S. J. Goetz, M. C. Mack, and R. F. Grant. 2021. Arctic tundra shrubification: a review of mechanisms and impacts on ecosystem carbon balance. Environ. Res. Lett. 16:053001.

Menozzi, P., A. Piazza, and L. Cavalli-Sforza. 1978. Synthetic Maps of Human Gene Frequencies in Europeans. Science 201:786–792.

Mir, R. R., M. Reynolds, F. Pinto, M. A. Khan, and M. A. Bhat. 2019. High-throughput phenotyping for crop improvement in the genomics era. The 4th International Plant Phenotyping Symposium 282:60–72.

Moskvina, V., and K. M. Schmidt. 2008. On multiple-testing correction in genome-wide association studies. Genet. Epidemiol. 32:567–73.

Müller, G. B. 2017. Why an extended evolutionary synthesis is necessary. Interface Focus 7:20170015.

Myers-Smith, I. H., B. C. Forbes, M. Wilmking, M. Hallinger, T. Lantz, D. Blok, K. D. Tape, M. Macias-Fauria, U. Sass-Klaassen, E. Lévesque, S. Boudreau, P. Ropars, L. Hermanutz, A. Trant, L. S. Collier, S. Weijers, J. Rozema, S. A. Rayback, N. M. Schmidt, G. Schaepman-Strub, S. Wipf, C. Rixen, C. B. Ménard, S. Venn, S. Goetz, L. Andreu-Hayles, S. Elmendorf, V. Ravolainen, J. Welker, P. Grogan, H. E. Epstein, and D. S. Hik. 2011. Shrub expansion in

tundra ecosystems: dynamics, impacts and research priorities. Environ. Res. Lett. 6:045509.
Nadeau, C. P., M. C. Urban, and J. R. Bridle. 2017. Climates Past, Present, and Yet-to-Come Shape Climate Change Vulnerabilities. Trends Ecol. Evol. 32:786–800.
Nguyen, H. T., M. A. R. Khan, T. T. Nguyen, N. T. Pham, T. T. B. Nguyen, T. R. Anik, M. D. Nguyen, M. Li, K. H. Nguyen, U. K. Ghosh, L.-S. P. Tran, and C. V. Ha. 2025. Advancing Crop Resilience Through High-Throughput Phenotyping for Crop Improvement in the Face of Climate Change. Plants 14:907.
Ordonez, A., F. Riede, S. Normand, and J.-C. Svenning. 2024. Towards a novel biosphere in 2300: rapid and extensive global and biome-wide climatic novelty in the Anthropocene. Philos. Trans. R. Soc. B 379:20230022.
Papoutsoglou, E. A., I. N. Athanasiadis, R. G. F. Visser, and R. Finkers. 2023. The benefits and struggles of FAIR data: the case of reusing plant phenotyping data. Sci. Data 10:457.
Papoutsoglou, E. A., D. Faria, D. Arend, E. Arnaud, I. N. Athanasiadis, I. Chaves, F. Coppens, G. Cornut, B. V. Costa, H. Ćwiek-Kupczyńska, B. Droesbeke, R. Finkers, K. Gruden, A. Junker, G. J. King, P. Krajewski, M. Lange, M. Laporte, C. Michotey, M. Oppermann, R. Ostler, H. Poorter, R. Ramírez-Gonzalez, Ž. Ramšak, J. C. Reif, P. Rocca-Serra, S. Sansone, U. Scholz, F. Tardieu, C. Uauy, B. Usadel, R. G. F. Visser, S. Weise, P. J. Kersey, C. M. Miguel, A. Adam-Blondon, and C. Pommier. 2020. Enabling reusability of plant phenomic datasets with MIAPPE 1.1. N. Phytol. 227:260–273.
Patterson, N., A. L. Price, and D. Reich. 2006. Population structure and eigenanalysis. PLoS Genet. 2:e190.
Pavan, S., C. Delvento, L. Ricciardi, C. Lotti, E. Ciani, and N. D'Agostino. 2020. Recommendations for Choosing the Genotyping Method and Best Practices for Quality Control in Crop Genome-Wide Association Studies. Front. Genet. 11:447.
Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of Population Structure Using Multilocus Genotype Data. Genetics 155:945–959.
Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–75.
Raj, A., M. Stephens, and J. K. Pritchard. 2014. fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. Genetics 197:573–589.

1049 Rellstab, C., F. Gugerli, A. J. Eckert, A. M. Hancock, and R. Holderegger. 2015. A
1050 practical guide to environmental association analysis in landscape
1051 genomics. Mol. Ecol. 24:4348–4370.
1052 Revelle, W. 2025. psych: Procedures for Psychological, Psychometric, and
1053 Personality Research. https://CRAN.R-project.org/package=psych
1054 Rockman, M. V. 2012. The QTN program and the alleles that matter for evolution:
1055 all that's gold does not glitter. Evolution 66:1–17.
1056 Sanderson, L.-A., S. P. Ficklin, C.-H. Cheng, S. Jung, F. A. Feltus, K. E. Bett, and
1057 D. Main. 2013. Tripal v1.1: a standards-based toolkit for construction of
1058 online genetic and genomic databases. Database 2013:bat075.
1059 Savolainen, O., T. Pyhäjärvi, and T. Knürr. 2007. Gene Flow and Local Adaptation
1060 in Trees. Ann. Rev. Ecol. Evol. Syst. 38:595–619.
1061 Schmidt, C., S. Hoban, and W. Jetz. 2023. Conservation macrogenetics:
1062 harnessing genetic data to meet conservation commitments. Trends Genet.
1063 39:816–829.
1064 Schraiber, J. G., S. N. Evans, and M. Slatkin. 2016. Bayesian Inference of Natural
1065 Selection from Allele Frequency Time Series. Genetics 203:493–511.
1066 Seren, Ü., B. J. Vilhjálmsson, M. W. Horton, D. Meng, P. Forai, Y. S. Huang, Q.
1067 Long, V. Segura, and M. Nordborg. 2012. GWAPP: A Web Application for
1068 Genome-Wide Association Mapping in Arabidopsis. Plant Cell 24:4793–4805.
1069 Shaw, R. E., K. A. Farquharson, M. W. Bruford, D. J. Coates, C. P. Elliott, J.
1070 Mergeay, K. M. Ottewell, G. Segelbacher, S. Hoban, C. Hvilsom, S. Pérez-
1071 Espona, D. Ruņģis, F. Aravanopoulos, L. D. Bertola, H. Cotrim, K. Cox, V.
1072 Cubric-Curik, R. Ekblom, J. A. Godoy, M. K. Konopiński, L. Laikre, I.-R. M.
1073 Russo, N. Veličković, P. Vergeer, C. Vilà, V. Brajkovic, D. L. Field, W. P.
1074 Goodall-Copestake, F. Hailer, T. Hopley, F. E. Zachos, P. C. Alves, A.
1075 Biedrzycka, R. M. Binks, J. Buiteveld, E. Buzan, M. Byrne, B. Huntley, L.
1076 Iacolina, N. L. P. Keehnen, P. Klinga, A. Kopatz, S. Kurland, J. A. Leonard, C.
1077 Manfrin, A. Marchesini, M. A. Millar, P. Orozco-terWengel, J. Ottenburghs,
1078 D. Posledovich, P. B. Spencer, N. Tourvas, T. U. Nahberger, P. van Hooft, R.
1079 Verbylaite, C. Vernesi, and C. E. Grueber. 2025. Global meta-analysis shows
1080 action is needed to halt genetic diversity loss. Nature 638:704–710.
1081 Slavov, G. T., S. P. DiFazio, J. Martin, W. Schackwitz, W. Muchero, E. Rodgers-
1082 Melnick, M. F. Lipphardt, C. P. Pennacchio, U. Hellsten, L. A. Pennacchio, L.
1083 E. Gunter, P. Ranjan, K. Vining, K. R. Pomraning, L. J. Wilhelm, M. Pellegrini,
1084 T. C. Mockler, M. Freitag, A. Geraldes, Y. A. El-Kassaby, S. D. Mansfield, Q. C.
1085 B. Cronk, C. J. Douglas, S. H. Strauss, D. Rokhsar, and G. A. Tuskan. 2012.
1086 Genome resequencing reveals multiscale geographic structure and

extensive linkage disequilibrium in the forest tree Populus trichocarpa. N. Phytol. 196:713–725.
Slifer, S. H. 2018. PLINK: Key Functions for Data Analysis. Curr. Protoc. Hum. Genet. 97:e59.
Smith, J. A., M. P. Buil, J. Fiechter, D. Tommasi, and M. G. Jacox. 2022. Projected novelty in the climate envelope of the California Current at multiple spatial-temporal scales. PLOS Clim. 1:e0000022.
Smith, J. M., and J. Haigh. 1974. The hitch-hiking effect of a favourable gene. Genetical Research 23:23–35.
Sork, V. L., S. N. Aitken, R. J. Dyer, A. J. Eckert, P. Legendre, and D. B. Neale. 2013. Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. Tree Genet. Genomes 9:901–911.
Spitze, K. 1993. Population structure in Daphnia obtusa: quantitative genetic and allozymic variation. Genetics 135:367–374.
Sreedasyam, A., C. Plott, M. S. Hossain, J. T. Lovell, J. Grimwood, J. W. Jenkins, C. Daum, K. Barry, J. Carlson, S. Shu, J. Phillips, M. Amirebrahimi, M. Zane, M. Wang, D. Goodstein, F. B. Haas, M. Hiss, P.-F. Perroud, S. S. Jawdy, Y. Yang, R. Hu, J. Johnson, J. Kropat, S. D. Gallaher, A. Lipzen, E. V. Shakirov, X. Weng, I. Torres-Jerez, B. Weers, D. Conde, M. R. Pappas, L. Liu, A. Muchlinski, H. Jiang, C. Shyu, P. Huang, J. Sebastian, C. Laiben, A. Medlin, S. Carey, A. A. Carrell, J.-G. Chen, M. Perales, K. Swaminathan, I. Allona, D. Grattapaglia, E. A. Cooper, D. Tholl, J. P. Vogel, D. J. Weston, X. Yang, T. P. Brutnell, E. A. Kellogg, I. Baxter, M. Udvardi, Y. Tang, T. C. Mockler, T. E. Juenger, J. Mullet, S. A. Rensing, G. A. Tuskan, S. S. Merchant, G. Stacey, and J. Schmutz. 2023. JGI Plant Gene Atlas: an updateable transcriptome resource to improve functional gene descriptions across the plant kingdom. Nucleic Acids Res. 51:8383–8401.
Staton, M., E. Cannon, L.-A. Sanderson, J. Wegrzyn, T. Anderson, S. Buehler, I. Cobo-Simón, K. Faaberg, E. Grau, V. Guignon, J. Gunoskey, B. Inderski, S. Jung, K. Lager, D. Main, M. Poelchau, R. Ramnath, P. Richter, J. West, and S. Ficklin. 2021. Tripal, a community update after 10 years of supporting open source, standards-based genetic, genomic and breeding databases. Brief. Bioinform. 22:bbab238.
Stinchcombe, J. R., and H. E. Hoekstra. 2008. Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity 100:158–170.
Sul, J. H., L. S. Martin, and E. Eskin. 2018. Population structure in genetic studies: Confounding factors and mixed models. PLoS Genet. 14:e1007309.

Taylor, G., I. S. Donnison, D. Murphy-Bokern, M. Morgante, M.-B. Bogeat-Triboulot, R. Bhalerao, M. Hertzberg, A. Polle, A. Harfouche, F. Alasia, V. Petoussi, D. Trebbi, K. Schwarz, J. J. B. Keurentjes, M. Centritto, B. Genty, J. Flexas, E. Grill, S. Salvi, and W. J. Davies. 2019. Sustainable bioenergy for climate mitigation: developing drought-tolerant trees and grasses. Ann. Bot. 124:513–520.

R Core Team. 2025. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Thompson, M. J., J. G. A. Martin, C. Biard, J. Bleu, C. J. Branston, P. Capilla-Lasheras, N. J. Dingemanse, D. M. Dominoni, M. Eens, T. Eeva, K. L. Evans, C. Isaksson, A. Liker, S. Massemin, E. Matthysen, A. Mouchet, S. Perret, J. C. Senar, G. Seress, M. Szulkin, E. Vincze, H. Watson, D. Réale, and A. Charmantier. 2025. Continental Patterns of Phenotypic Variation Along Replicated Urban Gradients: A Mega-Analysis. Ecol. Lett. 28:e70180.

Thorogood, R., V. Mustonen, A. Aleixo, P. J. Aphalo, F. O. Asiegbu, M. Cabeza, J. Cairns, U. Candolin, P. Cardoso, J. T. Eronen, M. Hällfors, I. Hovatta, A. Juslén, A. Kovalchuk, J. Kulmuni, L. Kuula, R. Mäkipää, O. Ovaskainen, A.-K. Pesonen, C. R. Primmer, M. Saastamoinen, A. H. Schulman, L. Schulman, G. Strona, and J. Vanhatalo. 2023. Understanding and applying biological resilience, from genes to ecosystems. npj Biodivers. 2:16.

Tiffin, P., and J. Ross-Ibarra. 2014. Advances and limits of using population genetics to understand local adaptation. Trends Ecol. Evol. 29:673–680.

Toczydlowski, R. H., L. Liggins, M. R. Gaither, T. J. Anderson, R. L. Barton, J. T. Berg, S. G. Beskid, B. Davis, A. Delgado, E. Farrell, M. Ghoojaei, N. Himmelsbach, A. E. Holmes, S. R. Queeno, T. Trinh, C. A. Weyand, G. S. Bradburd, C. Riginos, R. J. Toonen, and E. D. Crandall. 2021. Poor data stewardship will hinder global genetic diversity surveillance. Proc. Natl. Acad. Sci. 118:e2107934118.

Tuskan, G. A., S. DiFazio, S. Jansson, J. Bohlmann, I. Grigoriev, U. Hellsten, N. Putnam, S. Ralph, S. Rombauts, A. Salamov, J. Schein, L. Sterck, A. Aerts, R. R. Bhalerao, R. P. Bhalerao, D. Blaudez, W. Boerjan, A. Brun, A. Brunner, V. Busov, M. Campbell, J. Carlson, M. Chalot, J. Chapman, G.-L. Chen, D. Cooper, P. M. Coutinho, J. Couturier, S. Covert, Q. Cronk, R. Cunningham, J. Davis, S. Degroeve, A. Déjardin, C. dePamphilis, J. Detter, B. Dirks, I. Dubchak, S. Duplessis, J. Ehlting, B. Ellis, K. Gendler, D. Goodstein, M. Gribskov, J. Grimwood, A. Groover, L. Gunter, B. Hamberger, B. Heinze, Y. Helariutta, B. Henrissat, D. Holligan, R. Holt, W. Huang, N. Islam-Faridi, S. Jones, M. Jones-Rhoades, R. Jorgensen, C. Joshi, J. Kangasjärvi, J. Karlsson,

C. Kelleher, R. Kirkpatrick, M. Kirst, A. Kohler, U. Kalluri, F. Larimer, J. Leebens-Mack, J.-C. Leplé, P. Locascio, Y. Lou, S. Lucas, F. Martin, B. Montanini, C. Napoli, D. R. Nelson, C. Nelson, K. Nieminen, O. Nilsson, V. Pereda, G. Peter, R. Philippe, G. Pilate, A. Poliakov, J. Razumovskaya, P. Richardson, C. Rinaldi, K. Ritland, P. Rouzé, D. Ryaboy, J. Schmutz, J. Schrader, B. Segerman, H. Shin, A. Siddiqui, F. Sterky, A. Terry, C.-J. Tsai, E. Uberbacher, P. Unneberg, J. Vahala, K. Wall, S. Wessler, G. Yang, T. Yin, C. Douglas, M. Marra, G. Sandberg, Y. V. de Peer, and D. Rokhsar. 2006. The Genome of Black Cottonwood, Populus trichocarpa (Torr. & Gray). Science 313:1596–1604.

Urban, M. C. 2015. Accelerating extinction risk from climate change. Science 348:571–573.

Urban, M. C. 2024. Climate change extinctions. Science 386:1123–1128.

Velásquez, A. C., C. D. M. Castroverde, and S. Y. He. 2018. Plant–Pathogen Warfare under Changing Climate Conditions. Curr. Biol. 28:R619–R634.

Volk, G. M., P. F. Byrne, C. J. Coyne, S. Flint-Garcia, P. A. Reeves, and C. Richards. 2021. Integrating Genomic and Phenomic Approaches to Support Plant Genetic Resources Conservation and Use. Plants 10:2260.

Vu, V. Q., M. Friendly, and A. Tavadyan. 2024. ggbiplot: A Grammar of Graphics Implementation of Biplots.

Whitlock, M. C. 2008. Evolutionary inference from QST. Molecular Ecology 17:1885–1896.

Whitlock, M. C., and K. E. Lotterhos. 2015. Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of FST*. The American Naturalist 186:S24–S36.

Wickham, H. 2011. ggplot2. Wiley Interdiscip. Rev.: Comput. Stat. 3:180–185.

Wilczek, A. M., M. D. Cooper, T. M. Korves, and J. Schmitt. 2014. Lagging adaptation to warming climate in Arabidopsis thaliana. Proc. Natl. Acad. of Sci. 111:7906–7913.

Wilkinson, M. D., M. Dumontier, Ij. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da S. Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons. 2016.

The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018.

Wookey, P. A., R. Aerts, R. D. Bardgett, F. Baptist, K. A. Bråthen, J. H. C. Cornelissen, L. Gough, I. P. Hartley, D. W. Hopkins, S. Lavorel, and G. R. Shaver. 2009. Ecosystem feedbacks and cascade processes: understanding their role in the responses of Arctic and alpine ecosystems to environmental change. Glob. Chang. Biol. 15:1153–1172.

Wright, S. 1931. Evolution in Mendelian populations. Genetics 16:97–159.

Wright, S. 1943. Isolation by distance. Genetics 28:114–138.

Wright, S. 1951. The genetical structure of populations. Ann. Appl. Stat. 15:323–354.

Zhang, M., H. Suren, and J. A. Holliday. 2019. Phenotypic and Genomic Local Adaptation across Latitude and Altitude in Populus trichocarpa. Genome Biol. Evol. 11:2256–2272.

Zhou, J., K. D. Tape, L. Prugh, G. Kofinas, G. Carroll, and K. Kielland. 2020. Enhanced shrub growth in the Arctic increases habitat connectivity for browsing herbivores. Glob. Chang. Biol. 26:3809–3820.

Zhou, L., R. Bawa, and J. A. Holliday. 2014. Exome resequencing reveals signatures of demographic and adaptive processes across the genome and range of black cottonwood (Populus trichocarpa). Mol. Ecol. 23:2486–2499.

Zhou, X., and M. Stephens. 2014. Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nature Methods 11:407–409.

Zhou, X., and M. Stephens. 2012. Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. 44:821–824.

Tables and Figures for

CartograPlant: Bridging genomic, phenotypic, and environmental data to advance plant resilience and eco-evolutionary insight

Brandon M. Lind[1,2*†], Irene Cobo-Simón[1,3†], Meghan Myles[1†], Gabe Barrett[1], Emily Grau[1], Risharde Ramnath[1], Vlad Savitsky[1], Jill L. Wegrzyn[1,2]

[1]Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269, USA
[2]Institute for Systems Genomics, University of Connecticut, Storrs, CT 06269, USA
[3]Departamento de Ecología y Genética Forestal, Instituto de Ciencias Forestales (ICIFOR), Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria - Consejo Superior de Investigaciones Científicas (ICIFOR-INIA, CSIC), Madrid, Spain

26 September 2025

Running title: *Eco-evolutionary insight with CartograPlant*

Keywords: data integration, data interoperability, meta-analysis, biodiversity informatics, plant adaptation

[†]These authors contributed equally

***Corresponding Author**

Brandon M. Lind
ORCID: 0000-0002-8560-5417

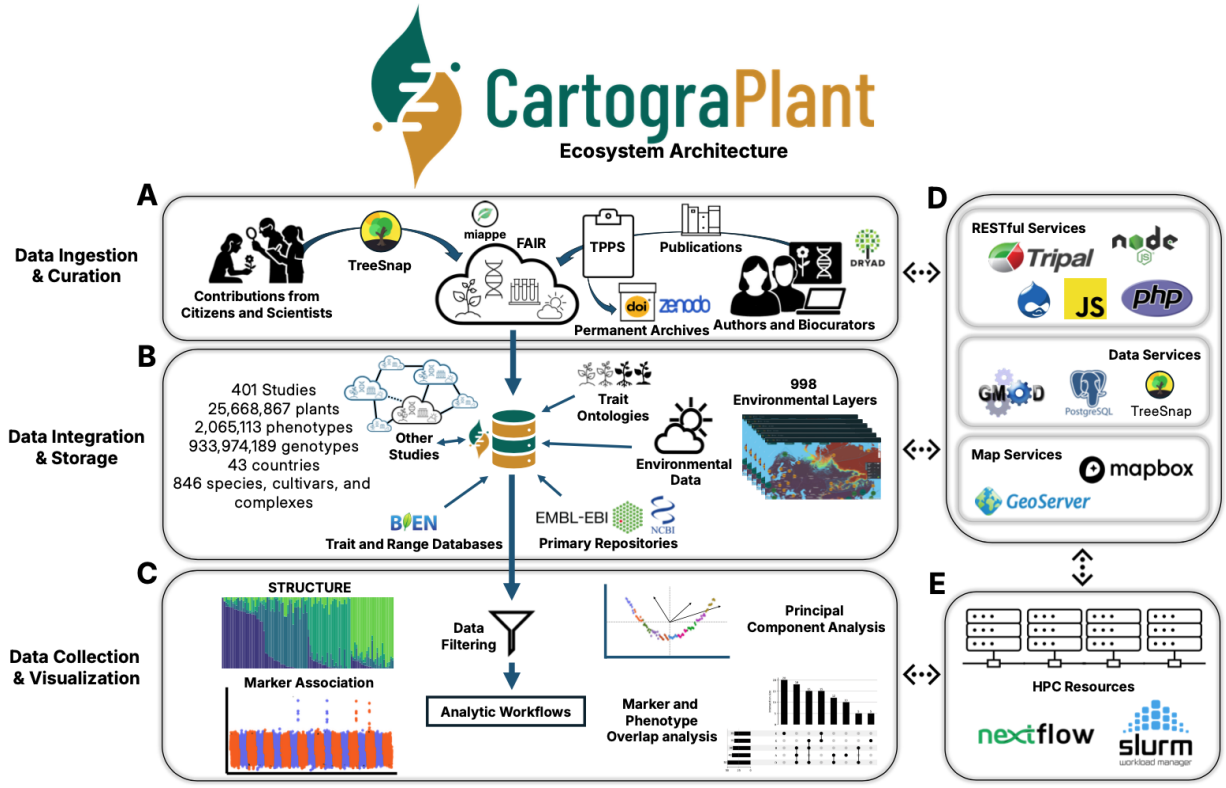


Figure 1 The CartograPlant ecosystem architecture facilitates eco-evolutionary insight. CartograPlant is web-based interactive application composed of interconnected modules that support integrative, reproducible analyses though scalable, standards-based infrastructure and services: A) *Data Ingestion & Curation* - Phenotypic, genetic, and environmental data, and the associated metadata and experimental design information, are collected from both citizen science platforms such as TreeSnap as well as trained biocurators through the TPPS pipeline (Tripal Plant Population Submit - see Figure 2). These data are curated using community standards such as the MIAPPE (Minimum Information About a Plant Phenotyping Experiment) and FAIR principles (Findable, Accessible, Interoperable, and Reusable). When data has not been previously archived, CartograPlant offers the option to create a Zenodo archive. B) *Data Integration & Storage* - Ingested data are incorporated into a Chado database schema that integrates current study information and metadata with other ingested studies, databases, repositories, ontologies, and environmental layers (see Table 1). C) *Data Collection & Visualization* - Registered users of CartograPlant can create saved Workspaces that retain raw data linked to a Collection of individuals as well as any output files and figures generated through subsequent analysis. D-E) *Services* - CartograPlant's ecosystem architecture is enabled by RESTful, data, and map services alongside high-performance computing (HPC) resources and NextFlow pipelines to support reproducible and scalable data processing. CartograPlant is available online at https://cartograplant.org.

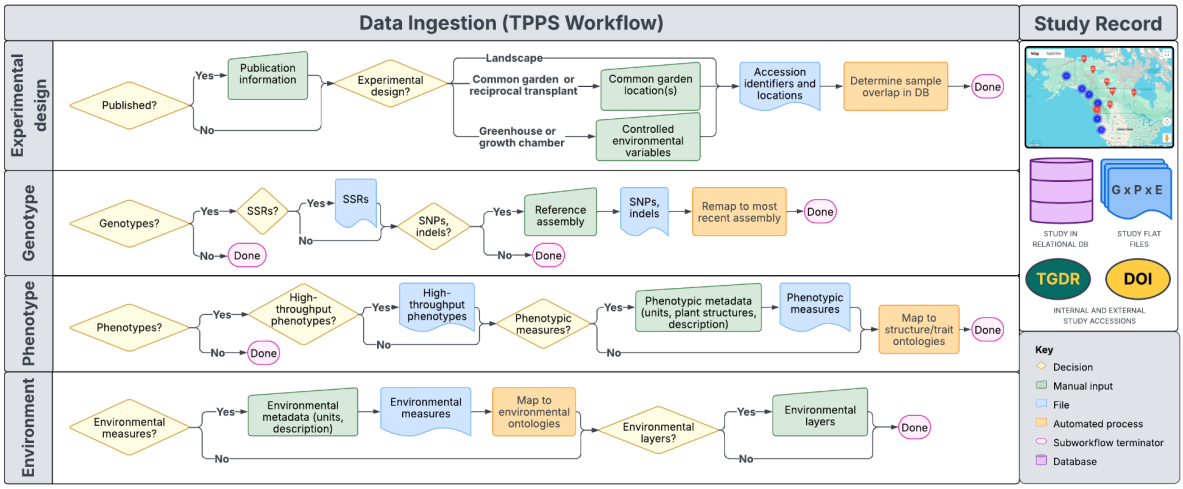


Figure 2. Data ingestion workflows for integrating genotypic, phenotypic, and environmental data through the Tripal Plant Population Submit pipeline (TPPS; http://treegenesdb.org/tpps). This flowchart outlines the structured processes on CartograPlant for ingesting diverse biological datasets, faceted by four primary information sources: Experimental Design, Genotypic Data, Phenotypic Data, and Environmental Data. Each workflow includes decision points (yellow diamonds), manual data input steps (green rectangles), file upload actions (blue rectangles), automated processing steps (orange rectangles), and subworkflow endpoints (purple ovals). This flexible and modular framework ensures consistent and accurate data ingestion and integration across diverse study types. Upon data ingestion, a study record is created which holds study flat files, is assigned a permanent TreeGenes Database Record (TGDR) number, and can be assigned a DOI. Files are held in association with ontologies within a relational database (DB).

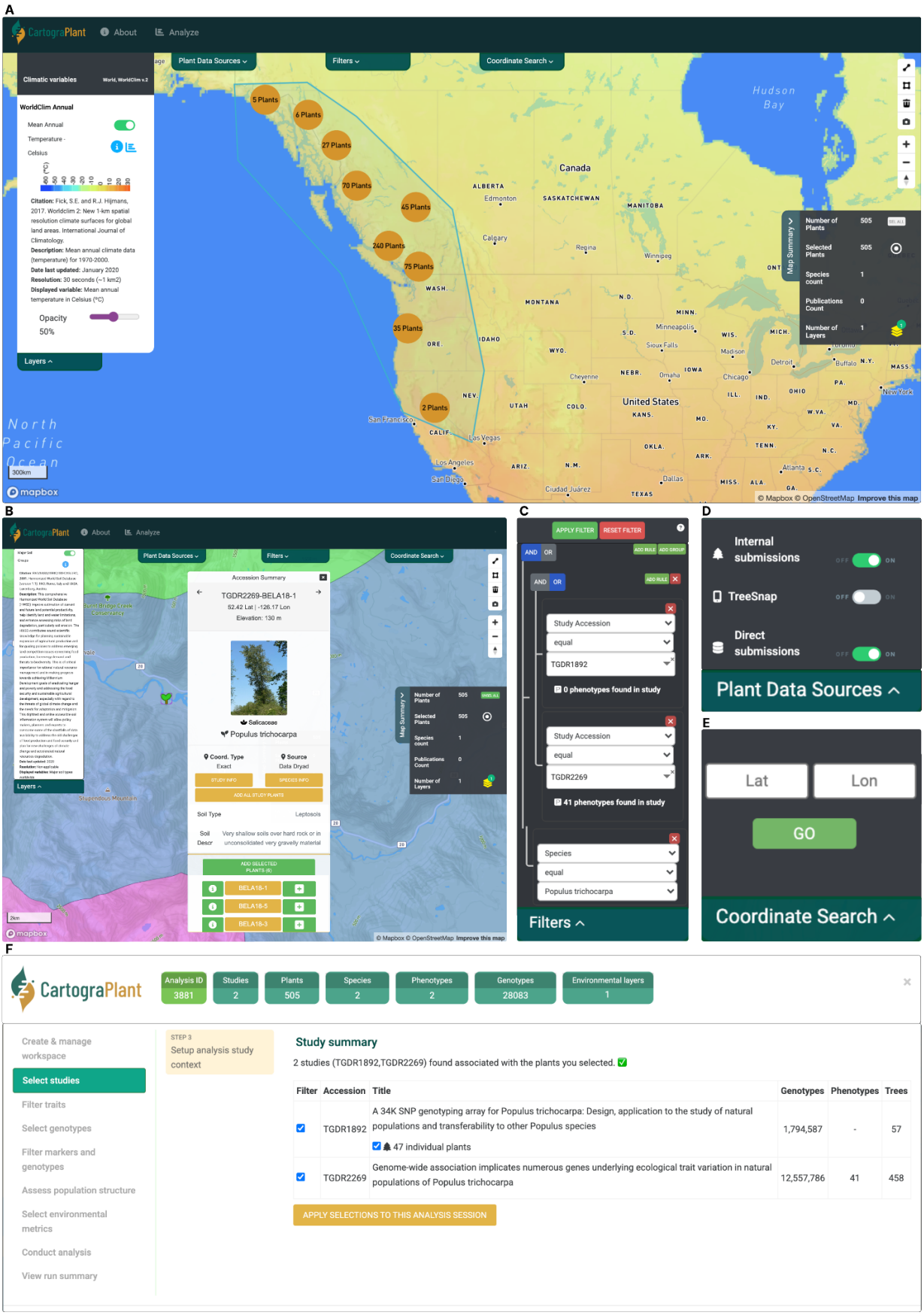

A
CartograPlant
About
Analyze
Plant Data Sources
Filters
Coordinate Search
Climatic variables
WorldClim Annual
Mean Annual Temperature - Celsius
Opacity 50%
Layers
5 Plants
6 Plants
27 Plants
70 Plants
45 Plants
240 Plants
75 Plants
35 Plants
2 Plants
Canada
United States
Hudson Bay
North Pacific Ocean
Map Summary
Number of Plants 505
Selected Plants 505
Species count 1
Publications Count 0
Number of Layers 1
© Mapbox © OpenStreetMap Improve this map
B
Accession Summary
TGDR2269-BELA18-1
52.42 Lat | -126.17 Lon
Elevation: 130 m
Salicaceae
Populus trichocarpa
Coord. Type Exact
Source Data Dryad
STUDY INFO
SPECIES INFO
ADD ALL STUDY PLANTS
Soil Type Leptosols
Soil Descr Very shallow soils over hard rock or in unconsolidated very gravelly material
ADD SELECTED PLANTS (6)
BELA18-1
BELA18-5
BELA18-3
C
APPLY FILTER
RESET FILTER
AND
OR
ADD RULE
ADD GROUP
Study Accession
equal
TGDR1892
0 phenotypes found in study
TGDR2269
41 phenotypes found in study
Species
Populus trichocarpa
Filters
D
Internal submissions
TreeSnap
Direct submissions
OFF
ON
Plant Data Sources
E
Lat
Lon
GO
Coordinate Search
F
CartograPlant
Analysis ID 3881
Studies 2
Plants 505
Species 2
Phenotypes 2
Genotypes 28083
Environmental layers 1
Create & manage workspace
Select studies
Filter traits
Select genotypes
Filter markers and genotypes
Assess population structure
Select environmental metrics
Conduct analysis
View run summary
STEP 3
Setup analysis study context
Study summary
2 studies (TGDR1892,TGDR2269) found associated with the plants you selected.
Filter
Accession
Title
Genotypes
Phenotypes
Trees
TGDR1892
A 34K SNP genotyping array for Populus trichocarpa: Design, application to the study of natural populations and transferability to other Populus species
1,794,587
-
57
47 individual plants
TGDR2269
Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of Populus trichocarpa
12,557,786
41
458
APPLY SELECTIONS TO THIS ANALYSIS SESSION

Figure 3. CartograPlant's Browse page enables data filtering and visualization in a number of ways. A) The CartograPlant Browse page, with the Worldclim2 (Fick and Hijmans 2017) layer Mean Annual Temperature climate normal enabled. The Map Summary displays both the number of plants available after filtering and the number of plants within a Collection for analysis. The polygon tool has been used to create a Collection (blue bounding box). B) The Browse page with the Harmonized World Soil Database Major Soil Groups layer enabled. A single plant has been selected, enabling the display of that plant's Accession Summary. The Accession Summary displays soil variables from the coordinates of the selected plant, and allows users to add this plant (or all plants associated with this study) to their Collection for further use. C) Users can filter displayed data by a number of options with interactive logical operators. Here, filters have been applied such that all displayed plants are members of the species *Populus trichocarpa* and are associated with one of two specific CartograPlant study accessions (TGDR = TreeGenes Database Reference). D) Users can subset displayed data by source. Internal submissions are those made by CartograPlant biocurators, and external submissions are those made directly by authors; TreeSnap = TreeSnap.org. E) Individual coordinates can be searched to zoom to them on the map. F) Data within a Collection is described in the Analysis Panel's Select Studies tab.

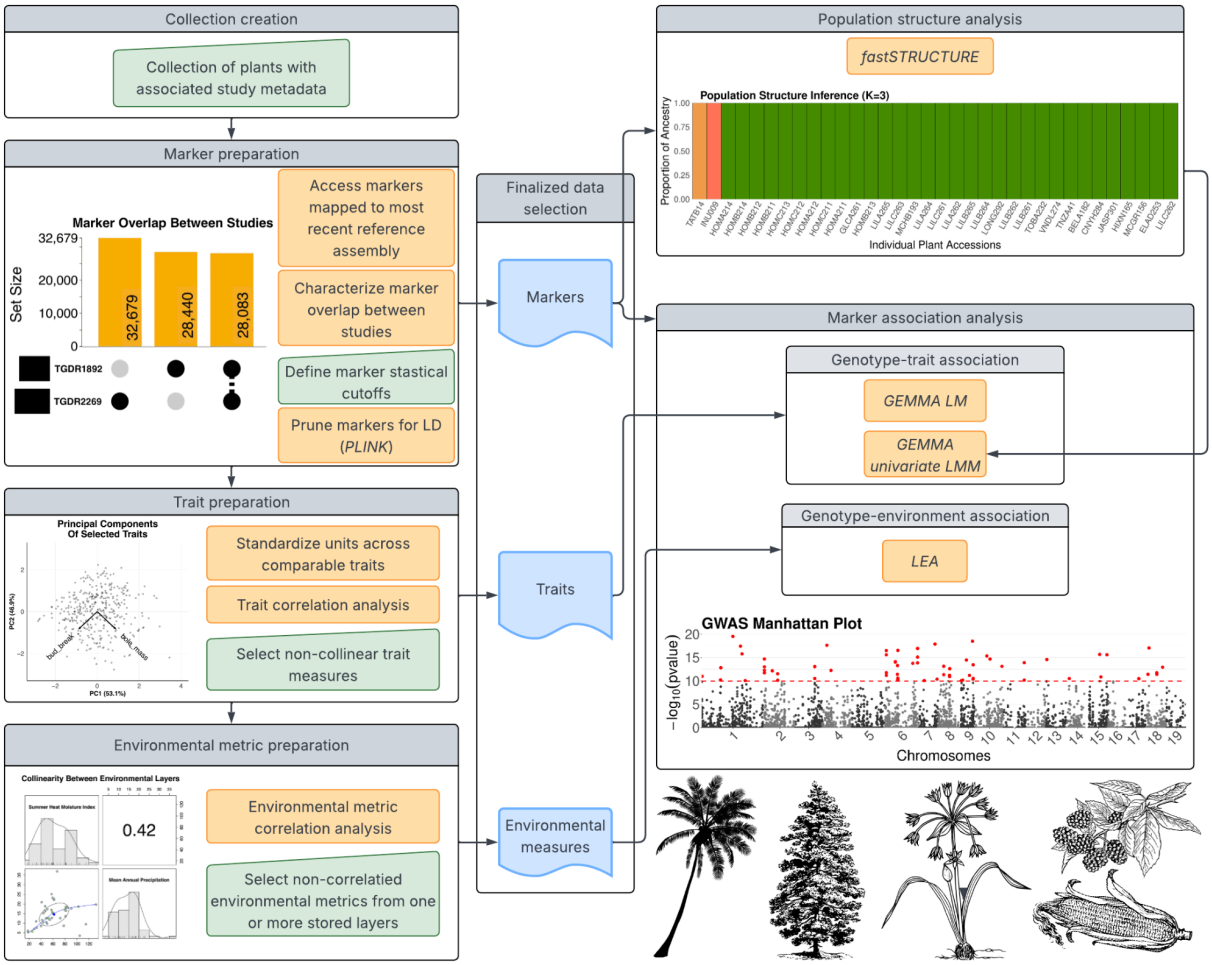


Figure 4. Availability of data filtering, visualization, and analysis methods within CartograPlant's Analysis Panel. After creating a Collection of plants (see Figure 3), a user can prepare the data for analysis. Marker overlap between studies is characterized in an UpSet plot. Users can define statistical cutoffs for marker quality and prune their markers for linkage disequilibrium (LD). For phenotypic data preparation, users are aided in selecting non-collinear measures with a PCA biplot. Here, the measures *bud_break* and *bole_mass* have been selected. For environmental layer data preparation, users are aided in selecting non-correlated measures with a trellis plot. The upper triangular panels display the pairwise Pearson correlation between environmental variables. Panels on the diagonal display histograms of the distribution of each environmental variable. The lower triangular panels display pairwise scatterplots between each selected variable. Once data selection has been finalized, users can perform analyses of population structure, GWAS, and GEA. File outputs at any step can be downloaded from a user's Workspace. All data in this plot are associated with CartograPlant accession TGDR 1892 (Geraldes et al., 2013) or with accession TGDR 2269 (McKown et al., 2014c).

Table 1. Data types and subtypes held within CartograPlant as well as the ontologies with which they are described. File extensions accepted for upload of these data are given as well as the extensions with which data can be retrieved. CSV = Comma-Separated Values; TSV = Tab-Separated Values; VCF = Variant Call Format.

Data type	Associated ontologies	Data subtype	Accepted ingestion formats	Retrieval formats
Genotypic	Sequence Ontology[1] (SO), Gene Ontology[2,3] (GO)	Assembled or unassembled DNA or cDNA sequence	FASTA	FASTA
		Microsatellite	CSV, TSV	CSV
		SNP	VCF, CSV, TSV	VCF
		Indel	VCF, CSV, TSV	VCF
Phenotypic	Plant Ontology[4-5] (PO), Crop Ontology[6] (CO), Chemical Entities of Biological Interest[7] (ChEBI), Phenotype And Trait Ontology[8] (PATO), Plant Trait Ontology[5] (TO)	Morphological	CSV, TSV	CSV
		Physiological	CSV, TSV	CSV
		Phenological	CSV, TSV	CSV
		Metabolomic	CSV, TSV	CSV
		Expression	CSV, TSV	CSV
Environmental	The Environment Ontology[9] (ENVO), Plant Experimental Conditions Ontology[5] (PECO)	Point measure	CSV, TSV	CSV
		Layer	Shapefile, geoTIFF, MBTiles	CSV

1. Eilbeck, Karen, Suzanna E Lewis, Christopher J Mungall, et al. 2005. "The Sequence Ontology: A Tool for the Unification of Genome Annotations." *Genome Biology* 6 (5): R44. https://doi.org/10.1186/gb-2005-6-5-r44.
2. Ashburner, Michael, Catherine A. Ball, Judith A. Blake, et al. 2000. "Gene Ontology: Tool for the Unification of Biology." *Nature Genetics* 25 (1): 25–29. https://doi.org/10.1038/75556.

3. Aleksander, Suzi A, James Balhoff, Seth Carbon, et al. 2023. “The Gene Ontology Knowledgebase in 2023.” *GENETICS* 224 (1): iyad031. https://doi.org/10.1093/genetics/iyad031.
4. Cooper, Laurel, Ramona L. Walls, Justin Elser, et al. 2013. “The Plant Ontology as a Tool for Comparative Plant Anatomy and Genomic Analyses.” *Plant and Cell Physiology* 54 (2): e1–e1. https://doi.org/10.1093/pcp/pcs163.
5. Cooper, Laurel, Justin Elser, Marie-Angelique Laporte, Elizabeth Arnaud, and Pankaj Jaiswal. 2023b. “Planteome 2024 Update: Reference Ontologies and Knowledgebase for Plant Biology.” *Nucleic Acids Research* 52 (D1): D1548–55. https://doi.org/10.1093/nar/gkad1028.
6. Shrestha, Rosemary, Luca Matteis, Milko Skofic, et al. 2012. “Bridging the Phenotypic and Genetic Data Useful for Integrated Breeding through a Data Annotation Using the Crop Ontology Developed by the Crop Communities of Practice.” *Frontiers in Physiology* 3: 326. https://doi.org/10.3389/fphys.2012.00326.
7. Hastings, Janna, Gareth Owen, Adriano Dekker, et al. 2016. “ChEBI in 2016: Improved Services and an Expanding Collection of Metabolites.” *Nucleic Acids Research* 44 (D1): D1214–19. https://doi.org/10.1093/nar/gkv1031.
8. Gkoutos, Georgios V, Paul N Schofield, and Robert Hoehndorf. 2017. “The Anatomy of Phenotype Ontologies: Principles, Properties and Applications.” *Briefings in Bioinformatics* 19 (5): 1008–21. https://doi.org/10.1093/bib/bbx035.
9. Buttigieg, Pier Luigi, Evangelos Pafilis, Suzanna E. Lewis, Mark P. Schildhauer, Ramona L. Walls, and Christopher J. Mungall. 2016. “The Environment Ontology in 2016: Bridging Domains with Increased Scope, Semantic Density, and Interoperation.” *Journal of Biomedical Semantics* 7 (1): 57. https://doi.org/10.1186/s13326-016-0097-6.
10. Buttigieg, Pier Luigi, Evangelos Pafilis, Suzanna E. Lewis, Mark P. Schildhauer, Ramona L. Walls, and Christopher J. Mungall. 2016. “The Environment Ontology in 2016: Bridging Domains with Increased Scope, Semantic Density, and Interoperation.” *Journal of Biomedical Semantics* 7 (1): 57. https://doi.org/10.1186/s13326-016-0097-6.

Table 2. Workflows available within the CartograPlant Analysis Panel, the general methodology they employ, and the software they utilize.

Workflow	Method	Software
Trait correlation analysis	PCA-based visualization of inferred correlation between selected traits	R *v3.6.0*[1], ggplot2 *v*3.3.6[2], ggbiplot *v*0.55[3]
Marker remapping	Flank sequence-based remapping	Python *v*3.6.8[4], samtools *v*1.22.1[5], bwa *v*0.7.19 (*mem*)[6], bedtools *v*2.31.1[7], pandas *v*2.3.2[8]
Marker overlap analysis	Identification of non-unique marker positions across datasets after remapping to a common reference assembly. Results visualized in an UpSet plot	bcftools *v*1.20[5], UpSet.js *v*1.11.0[9]
Marker and genotype filtering	Filtering of markers, genotypes, or individuals from a dataset based on user-defined statistical cutoffs. Distributions visualized in histograms	bcftools *v*1.20[5], D3.js *v*5.16.0[10]
Population structure inference	Inference of population structure from genotype data. Results visualized in an admixture bar plot	PLINK *v*1.90[11-12], fastStructure *v*1.0[13], R *v*3.6.0[1], ggplot2 *v*3.3.6[2]
Environmental data distribution and correlation analysis	Visualization of selected environmental variables, including their distributions, pairwise Pearson correlations, and pairwise scatterplots	R *v*3.6.0[1], psych *v*2.2.5[14]
Marker association analysis	GWAS or GEA for the inference of associations between genotypes and phenotypic or environmental variables, respectively. *P*-values are visualized within a Manhattan plot	GEMMA *v*0.98.3 (linear model, univariate linear model)[15], R *v*3.6.0[1], LEA *v*3.6.0 (*LFMM2*)[16], ggplot2 *v*3.3.6[2]

1. R Core Team. 2025. *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL Https://Www.R-Project.Org/.* URL https://www.R-project.org/.
2. Wickham, Hadley. 2011. "ggplot2." *Wiley Interdisciplinary Reviews: Computational Statistics* 3 (2): 180–85. https://doi.org/10.1002/wics.147.

3. Vu, Vincent Q., Michael Friendly, and Aghasi Tavadyan. 2024. *Ggbiplot: A Grammar of Graphics Implementation of Biplots*. https://CRAN.R-project.org/package=ggbiplot.
4. Python Software Foundation. Python. Version 3.6.8. 2023. https://www.python.org.
5. Danecek, Petr, James K Bonfield, Jennifer Liddle, et al. 2021. "Twelve Years of SAMtools and BCFtools." *GigaScience*10 (2): giab008. https://doi.org/10.1093/gigascience/giab008.
6. Li, Heng. n.d. "Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM." https://arxiv.org/abs/1303.3997.
7. Quinlan, Aaron R., and Ira M. Hall. 2010. "BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features." Bioinformatics 26 (6): 841–42. https://doi.org/10.1093/bioinformatics/btq033.
8. The pandas development team. 2019. pandas-dev/pandas: Pandas. Zenodo. https://doi.org/10.5281/zenodo.3509134
9. Lex, Alexander, Nils Gehlenborg, Hendrik Strobelt, Romain Vuillemot, and Hanspeter Pfister. 2014. "UpSet: Visualization of Intersecting Sets." IEEE Transactions on Visualization and Computer Graphics 20 (12): 1983–92. https://doi.org/10.1109/tvcg.2014.2346248.
10. Bostock, M., V. Ogievetsky, and J. Heer. 2011. "D³ Data-Driven Documents." IEEE Transactions on Visualization and Computer Graphics 17 (12): 2301–9. https://doi.org/10.1109/tvcg.2011.185.
11. Purcell, Shaun, Benjamin Neale, Kathe Todd-Brown, et al. 2007. "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses." *American Journal of Human Genetics* 81: 559–75. https://doi.org/10.1086/519795.
12. Slifer, Susan H. 2018. "PLINK: Key Functions for Data Analysis." *Current Protocols in Human Genetics* 97 (1): e59. https://doi.org/10.1002/cphg.59.
13. Raj, Anil, Matthew Stephens, and Jonathan K Pritchard. 2014. "fastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets." *Genetics* 197 (2): 573–89. https://doi.org/10.1534/genetics.114.164350.
14. Revelle, William. 2025. *Psych: Procedures for Psychological, Psychometric, and Personality Research*. https://CRAN.R-project.org/package=psych.
15. Zhou, Xiang, and Matthew Stephens. 2012. "Genome-Wide Efficient Mixed-Model Analysis for Association Studies." *Nature Genetics* 44 (7): 821–24. https://doi.org/10.1038/ng.2310.
16. Caye, Kevin, Basile Jumentier, Johanna Lepeule, and Olivier François. 2019. "LFMM 2: Fast and Accurate Inference of Gene-Environment Associations in Genome-Wide Studies." *Molecular Biology and Evolution* 36 (4): 852–60. https://doi.org/10.1093/molbev/msz008.

Table 3. CartograPlant hosts 998 environmental layers which can be classified into four broad categories. Layer groups, which are sets of similar layers, are given. The extent, description, resolution, and source of each layer group is described.

Layer group	Extent	Description	Resolution	Source	Reference
Biological, environmental, and climate data					
Biotic damage	United States, Canada	Presence and damage caused by Emerald Ash Borer, Forest Tent Caterpillar, Gypsy Moth, Hemlock Woolly Adelgid, Winter Moth; total defoliation and mortality	variable	USFS UVM EABIN	1 2 3
Climatic variables	Global, North America	Temperature, precipitation, aridity, seasonality (temperature, precipitation, aridity), and soil types, temperature and permafrost probability	1 km	Worldclim ClimateNA FAO PANGAEA	4 5 6 7
NDVI	Global (FAO), conterminous United States (MODIS)	Vegetation health index (Normalized Difference Vegetation Index) derived from satellite imagry, and phenology metrics	varies (FAO), 250m (MODIS)	NOAA-AVHRR USGS	8 9
Terrestrial Ecoregions	Global	Global classification of terrestrial ecoregions	coarse (ecoregion)	WWF	10
NEON field sites	United States (discrete sites)	Locations and monitoring data from the National Ecological Observatory Network	Site-specific	NSF	11
Low impact areas	Global	Landscapes with low current human density and impacts not primarily managed for human needs	1 km	Jacobson et al. 2019	12
Forest and vegetation data					
Forest fragmentation	North America	Forest fragmentation assessment based on an area's forest cover and adjacent pixel forest cover	1 km	USGS	13
Land cover	Global	Global land cover classification for vegetation and land use	30m	GLAD	14
National forests	United States	Boundaries of U. S. national forests	varies	USDA	15
Canopy height	Global	Canopy height data	30m (GLAD) - 1km (NASA)	NASA GLAD	16 14
Seed Zones	Eastern United States	Delineation of seed zones of the Eastern United States	regional	USFS	17
Intact forest landscape	Global	Extent of the intact forest landscapes for years 2000, 2013, 2016, and 2020	scale 1: 1,000,000	IFL	18
FIA land cover	United States	County-level statistics on forest land cover, timber volume, biomass, carbon stock, and forest growth and removals, along with sampling errors	coarse (county level)	USDA	19
Biodiversity and species data					
Biodiversity hotspots	Global	Areas of high biodiversity and conservation priority	30m	WRI	20

Species range maps	Global	Geographic distribution of plant species (GIS polygons)	NA	EUFORGEN USDA	21 22	
Density population	United States	Population density of various plant species within the United States	30m	BIEN	23, 24	
PET and aridity	Global	Global aridity index and evapotranspiration	30 arc-second	USDA Zomer et al. 2022	25 26, 27	
Agriculturally cultivated areas, protected areas, and soil data						
Protected and cultivated areas	Global	National parks and equivalents which hold economic and scientific importance or which will be preserved in their natural state	variable	IUCN, UNEP-WCMC	28	
General soil types	United States	Geographic distribution of soil types (GIS polygons)	1:12,000 - 1: 63,360	USDA	29	

Abbreviations: USGS = United States Geological Survey; USFS = United States Forest Service; EABIN = Emerald Ash Borer Information Network; FEMC = Forest Ecosystem Monitoring Cooperative; USDA = United States Department of Agriculture; FAO = Food and Agriculture Organization of the United Nations; EUFORGEN = European Forest Genetic Resources Programme; BIEN = Botanical Information and Ecology Network; GLAD = Global Analysis and Discovery; IFL = Intact Forest Landscapes; NSF = National Science Foundation; WWF = World Wildlife Foundation; NASA = National Aeronautics and Space Administration; WRI = World Resources Institute; IUCN = International Union for Conservation of Nature; UNEP-WCMC = United Nations Environment Programme World Conservation Monitoring Centre, UVM = University of Vermont

1. U.S. Department of Agriculture, and Forest Service. n.d.-a. "EAB Detected Counties through 2017: U.S. County-Level Data Containing the Year of Emerald Ash Borer (EAB) Detection and Live Ash Tree Volume."
https://usfs.maps.arcgis.com/home/item.html?id=d705c21c5afd4beebf93c6ea9e375865.

2. Duncan, James A., Emma R. Tait, Garrett W. Meigs, Alexandra M. Kosiba, Jennifer A. Pontius, and William Keeton. 2018. The Northeastern Forest Health Atlas – Technical Report. https://doi.org/10.18125/D2873C. https://www.uvm.edu/femc/forest-health-atlas/methods.
3. Emerald Ash Borer Information Network. Emerald Ash Borer Information Network. Michigan State University and USDA Forest Service. https://www.emeraldashborer.info
4. Fick, Stephen E., and Robert J. Hijmans. 2017. "WorldClim 2: New 1 Km Spatial Resolution Climate Surfaces for Global Land Areas." International Journal of Climatology 37 (12): 4302–15. https://doi.org/10.1002/joc.5086.
5. Wang, Tongli, Andreas Hamann, David Spittlehouse, and Carlos Carroll. 2016. "Locally Downscaled and Spatially Customizable Climate Data for Historical and Future Periods for North America." PLOS ONE 11 (6): e0156720. https://doi.org/10.1371/journal.pone.0156720.
6. Fischer, G., F. Nachtergaele, S. Prieler, H.T. van Velthuizen, L. Verelst, and D. Wiberg. n.d. "Global Agro-Ecological Zones Assessment for Agriculture (GAEZ 2008) - Harmonized World Soil Database (Version 1.2)." IIASA, Laxenburg, Austria and FAO, Rome, Italy. https://www.fao.org/soils-portal/data-hub/soil-maps-and-databases/harmonized-world-soil-database-v12/en/.
7. Obu, Jaroslav, Sebastian Westermann, Andreas Kääb, and Annett Bartsch. 2018. Ground Temperature Map, 2000--2016, Northern Hemisphere Permafrost. https://doi.org/10.1594/pangaea.888600.
8. Anyamba, Assaf, and Compton J. Tucker. "Analysis of Sahelian Vegetation Dynamics Using NOAA-AVHRR NDVI Data from 1981–2003." Journal of Arid Environments 63, no. 3 (2005): 596–614. https://doi.org/10.1016/j.jaridenv.2005.03.007.
9. Benedict, T. D., and M. B. Rigge. n.d. S-NPP 375-m eVIIRS Remote Sensing Phenology Metrics - across the Conterminous U.S. (Ver. 3.0, August 2025): U.S. Geological Survey Data Release. https://doi.org/10.5066/p9pztnbi. https://phenology.cr.usgs.gov/get_data_Aqua_C6_250e.php.
10. Olson, David M., Eric Dinerstein, Eric D. Wikramanayake, et al. 2001. "Terrestrial Ecoregions of the World: A New Map of Life on Earth." BioScience 51 (11): 933–38. https://www.worldwildlife.org/publications/terrestrial-ecoregions-of-the-world.
11. National Ecological Observatory Network. n.d. "NEON Data Portal: Explore Data Products." https://data.neonscience.org/data-products/explore.
12. Jacobson, Andrew P., Jason Riggio, Alexander M. Tait, and Jonathan E. M. Baillie. 2019. "Global Areas of Low Human Impact (Low Impact Areas) and Fragmentation of the Natural World." Scientific Reports 9: 14179. https://doi.org/10.1038/s41598-019-50558-6.
13. Riitters, Kurt, James Wickham, Robert O'Neill, Bruce Jones, and Elizabeth Smith. 2000. "Global-Scale Patterns of Forest Fragmentation." Conservation Ecology 4 (2): 3. http://www.consecol.org/vol4/iss2/art3/.
14. Potapov, Peter, Matthew C. Hansen, Adam Pickens, et al. 2022. "The Global 2000–2020 Land Cover and Land Use Change Dataset Derived from the Landsat Archive: First Results." Frontiers in Remote Sensing 3: 856903. https://doi.org/10.3389/frsen.2022.856903. https://glad.geog.umd.edu/dataset/GLCLUC2020.
15. U.S. Department of Agriculture, and Forest Service. n.d.-b. "Forest System Boundaries Map Service." https://apps.fs.usda.gov/arcx/rest/services/EDW/EDW_ForestSystemBoundaries_01/MapServer.

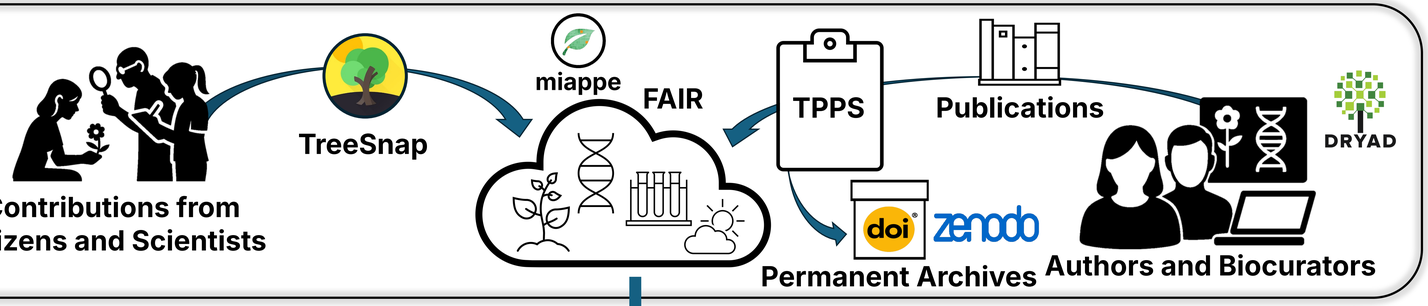
16. National Aeronautics and Space Administration. n.d. “Global Ecosystem Dynamics Investigation (GEDI) Lidar Data Access and Tools.” https://www.earthdata.nasa.gov/data/instruments/gedi-lidar/data-access-tools#toc-global-ecosystem%20dynamics%20investigation%20lidar-datasets.
17. Pike, Carolyn, Kevin M. Potter, Paul Berrang, et al. 2020. “New Seed-Collection Zones for the Eastern United States: The Eastern Seed Zone Forum.” Journal of Forestry 118 (4): 444–51. https://doi.org/10.1093/jofore/fvaa013. https://usfs.hub.arcgis.com/maps/usfs::eastern-seed-zones-development-map/about.
18. Potapov, Peter, Matthew C. Hansen, Lars Laestadius, et al. 2017. “The Last Frontiers of Wilderness: Tracking Loss of Intact Forest Landscapes from 2000 to 2013.” Science Advances3 (1): e1600821. https://doi.org/10.1126/sciadv.1600821. https://intactforests.org/data.ifl.html.
19. U.S. Department of Agriculture, and Forest Service. n.d.-c. “FIA Landcover County Estimates -- 2016.” https://catalog.data.gov/dataset/fia-landcover-county-estimates-2016-feature-layer-5a8c5.
20. Global Forest Watch, and Conservation International. n.d. “Biodiversity Hotspots (Global).” https://data.globalforestwatch.org/documents/gfw::biodiversity-hotspots/about.
20. EUFORGEN. n.d. “European Forest Genetic Resources Programme Species Distribution Maps.” https://www.euforgen.org/species.
22. Thompson, Robert S., Katherine H. Anderson, and Patrick J. Bartlein. n.d. “Digital Representations of Tree Species Range Maps from ‘Atlas of United States Trees’ by Elbert L. Little, Jr. (and Other Publications).” https://research.fs.usda.gov/feis/treedistributionmaps.
23. Maitner, Brian S., Brad Boyle, Nathan Casler, et al. 2018. “The Bien r Package: A Tool to Access the Botanical Information and Ecology Network (BIEN) Database.” Methods in Ecology and Evolution 9 (2): 373–79. https://doi.org/10.1111/2041-210x.12861.
24. Enquist, B. J., B. Sandel, B. Boyle, et al. n.d. “Botanical Big Data Shows That Plant Diversity in the New World Is Driven by Climatic-Linked Differences in Evolutionary Rates and Biotic Exclusion.” Unpublished Manuscript.
25. U.S. Department of, and Forest Service Agriculture. n.d.-c. “Individual Tree Species Parameter Maps.” https://www.fs.usda.gov/science-technology/data-tools-products/fhp-mapping-reporting/individual-tree-species-parameter-maps.
26. Zomer, Robert J., and Antonio Trabucco. 2019. Global Aridity Index and Potential Evapotranspiration (ET0) Climate Database V2. https://doi.org/https://doi.org/10.6084/m9.figshare.7504448.v7. https://figshare.com/articles/dataset/Global_Aridity_Index_and_Potential_Evapotranspiration_ET0_Climate_Database_v2/7504448.
27. Zomer, Robert J., Antonio Trabucco, Deborah A. Bossio, Lalisa A. Duguma, Meine van Noordwijk, and Jianchu Xu. 2022b. “Version 2 of the Global Aridity Index and Potential Evapotranspiration Database.” Scientific Data 9: 409. https://doi.org/10.1038/s41597-022-01493-1.
28. UNEP-WCMC, and International Union for Conservation of Nature. n.d. “Protected Planet: The World Database on Protected Areas (WDPA).” https://www.protectedplanet.net/en/thematic-areas/wdpa?tab=WDPA.
29. U.S. Department of Agriculture, and Natural Resources Conservation Service. n.d. “Web Soil Survey.” https://websoilsurvey.sc.egov.usda.gov/app/WebSoilSurvey.aspx.

CartograPlant

Ecosystem Architecture

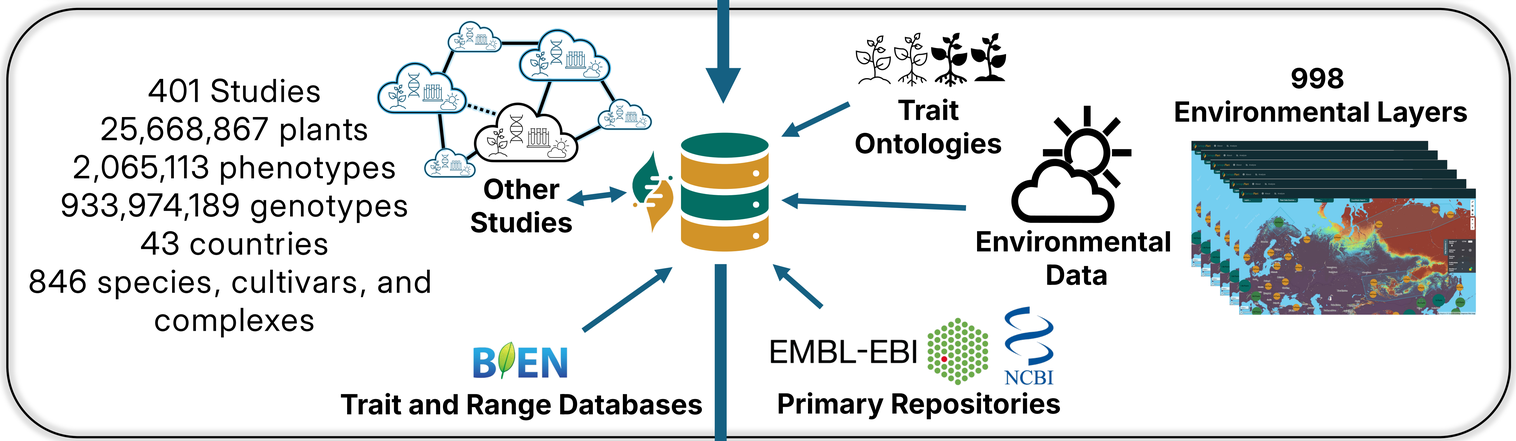
A

Data Ingestion & Curation



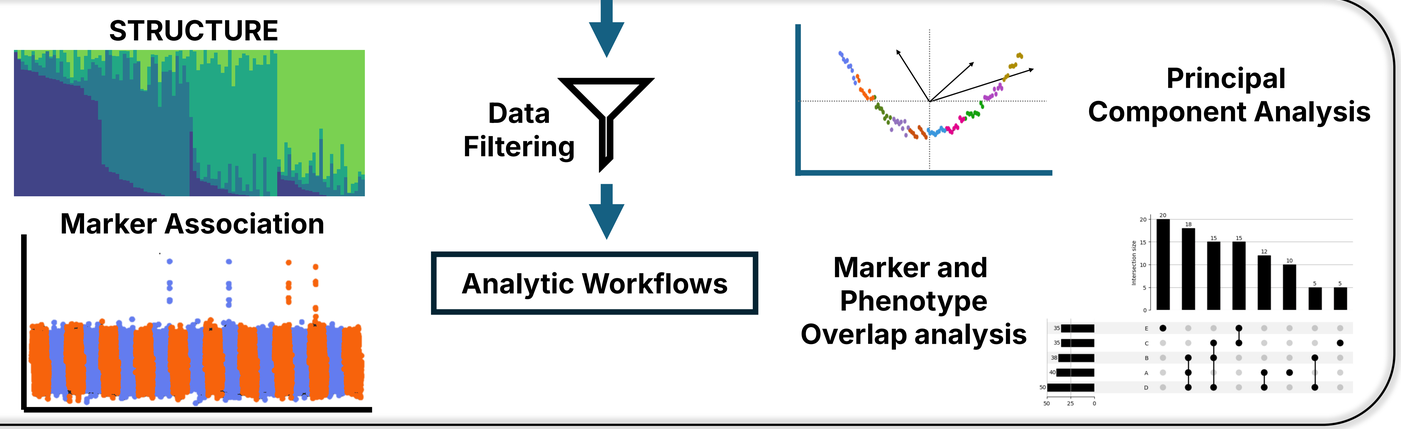
B

Data Integration & Storage

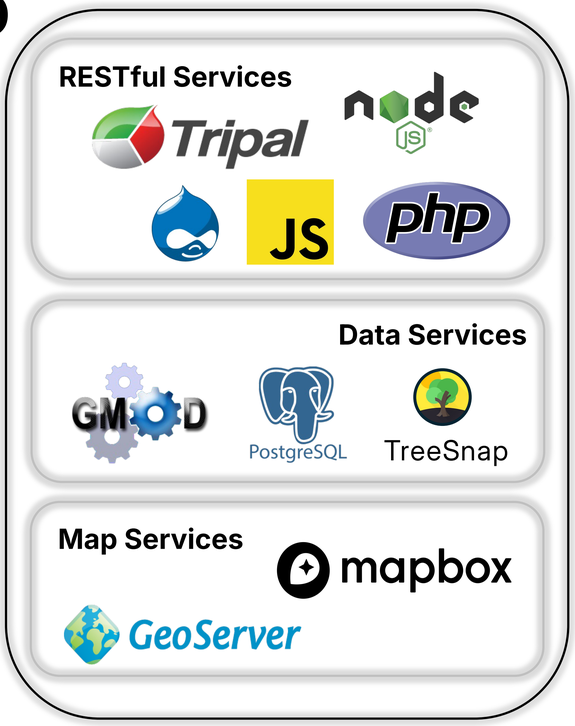


C

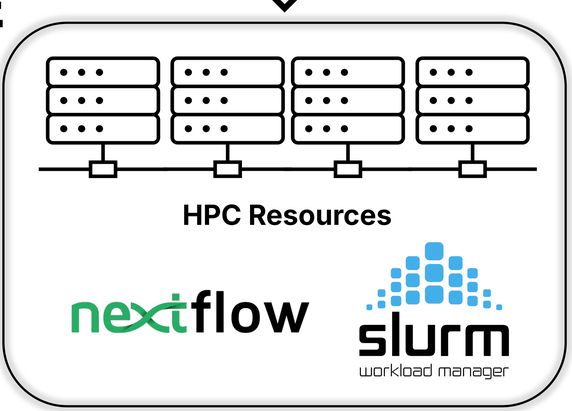
Data Collection & Visualization



D

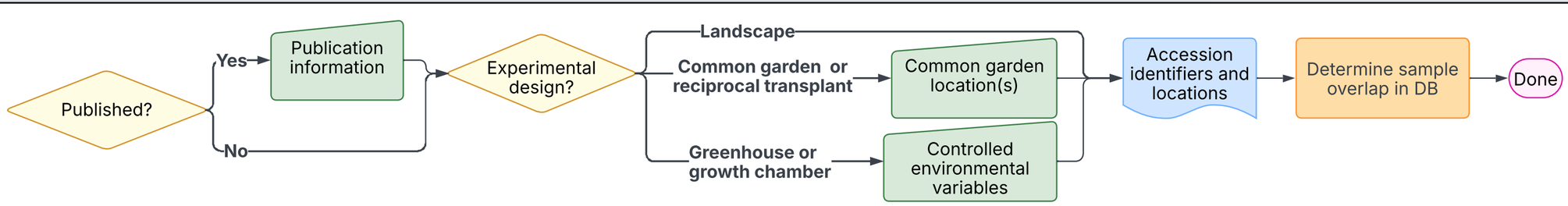


E

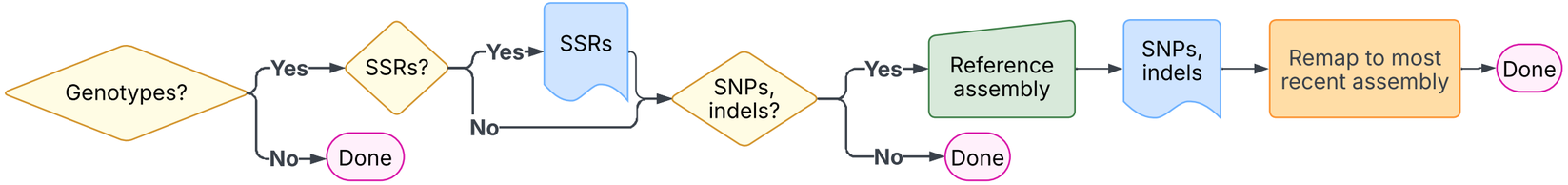


Data Ingestion (TPPS Workflow)

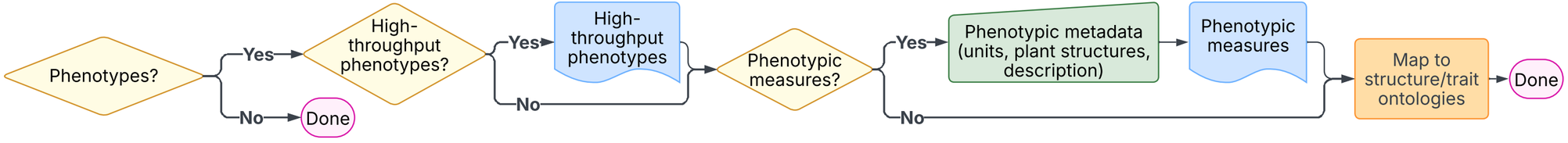
Experimental design



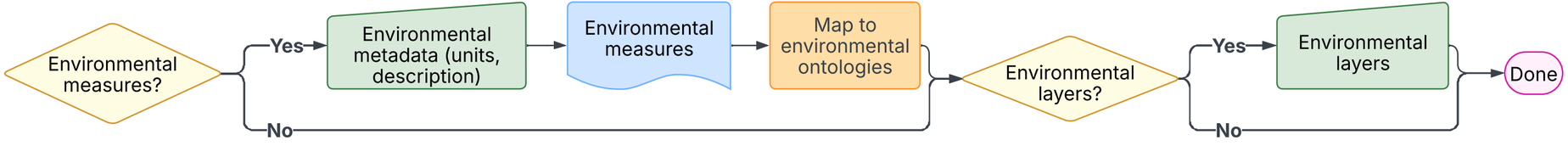
Genotype



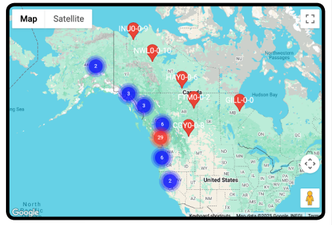
Phenotype

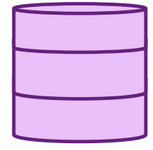


Environment



Study Record





STUDY IN RELATIONAL DB

STUDY FLAT FILES

TGDR

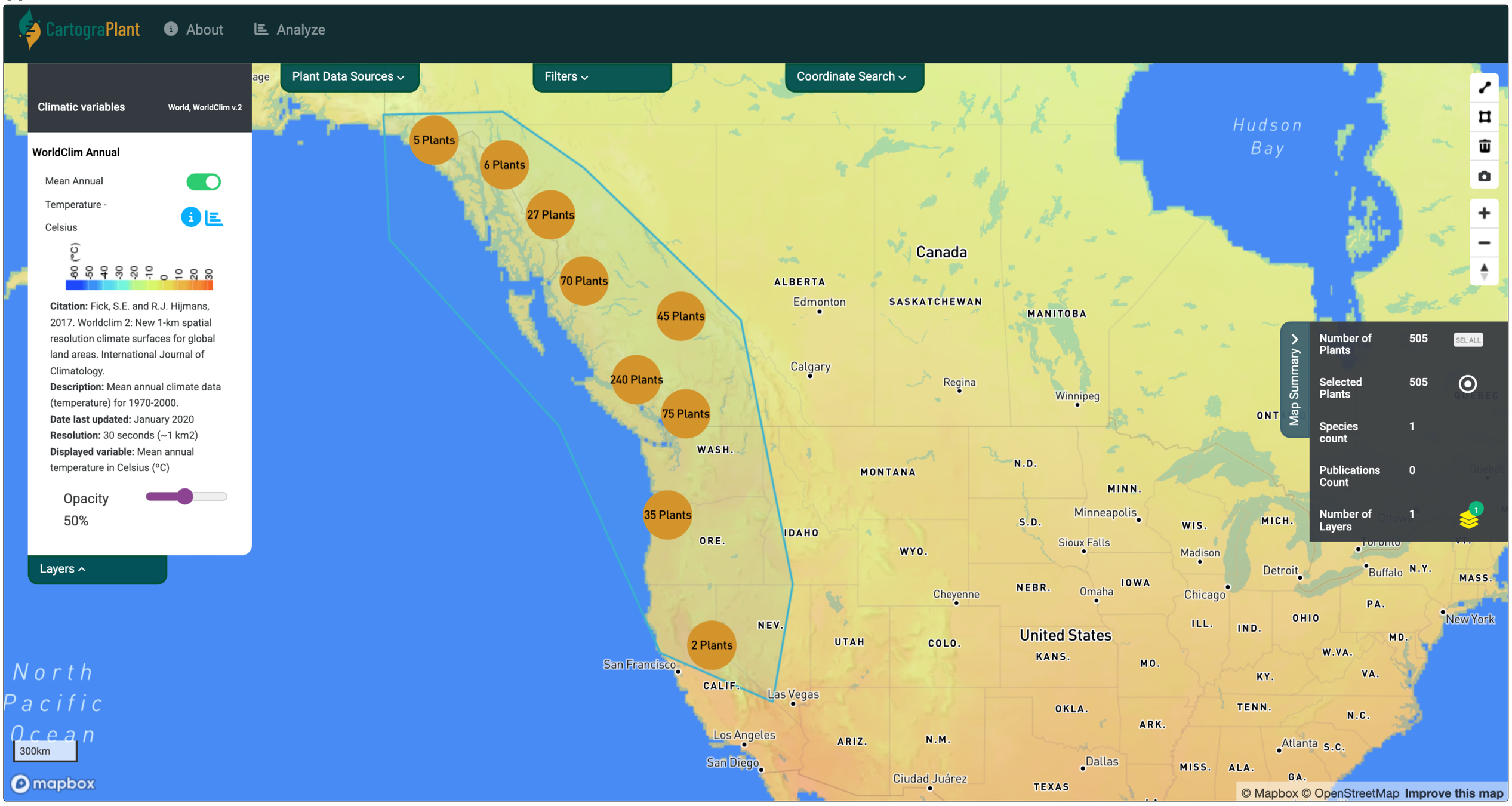
DOI

INTERNAL AND EXTERNAL STUDY ACCESSIONS

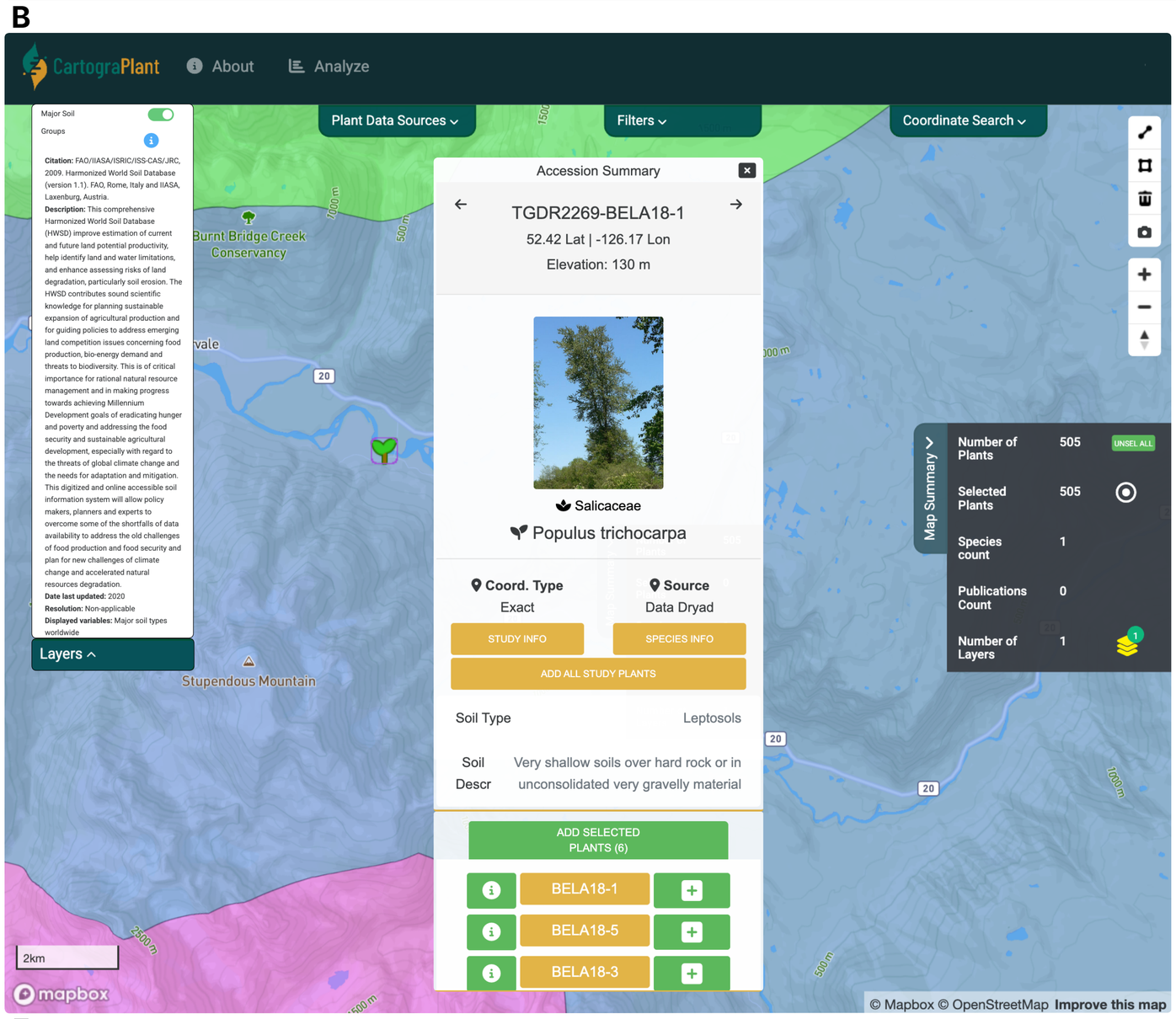
Key

- Decision
- Manual input
- File
- Automated process
- Subworkflow terminator
- Database

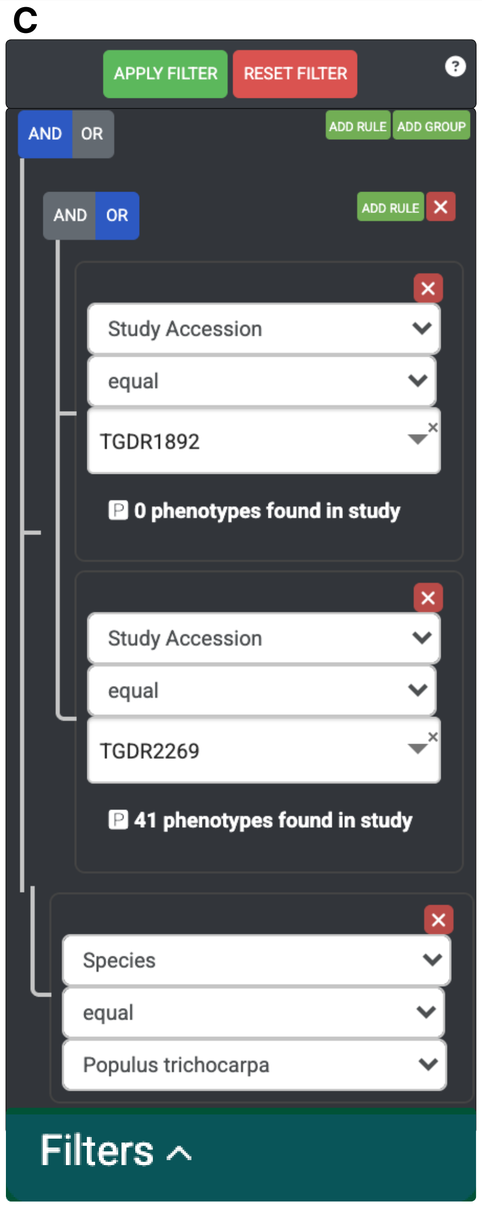
A


CartograPlant
About
Analyze
Plant Data Sources
Filters
Coordinate Search
Climatic variables
World, WorldClim v.2
WorldClim Annual
Mean Annual Temperature - Celsius
(°C)
-60 -50 -40 -30 -20 -10 0 10 20 30
Citation: Fick, S.E. and R.J. Hijmans, 2017. Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology.
Description: Mean annual climate data (temperature) for 1970-2000.
Date last updated: January 2020
Resolution: 30 seconds (~1 km2)
Displayed variable: Mean annual temperature in Celsius (°C)
Opacity
50%
Layers
5 Plants
6 Plants
27 Plants
70 Plants
45 Plants
240 Plants
75 Plants
35 Plants
2 Plants
Hudson Bay
Canada
ALBERTA
Edmonton
SASKATCHEWAN
MANITOBA
Calgary
Regina
Winnipeg
WASH.
MONTANA
N.D.
MINN.
Minneapolis
S.D.
WIS.
MICH.
IDAHO
ORE.
WYO.
Sioux Falls
Madison
Detroit
Buffalo
N.Y.
MASS.
Cheyenne
NEBR.
Omaha
IOWA
Chicago
PA.
New York
NEV.
UTAH
COLO.
United States
ILL.
IND.
OHIO
MD.
W.VA.
KANS.
MO.
KY.
VA.
San Francisco
CALIF.
Las Vegas
OKLA.
TENN.
N.C.
ARK.
Los Angeles
ARIZ.
N.M.
Atlanta
S.C.
San Diego
Dallas
MISS.
ALA.
GA.
Ciudad Juárez
TEXAS
Map Summary
Number of Plants 505 SEL ALL
Selected Plants 505
Species count 1
Publications Count 0
Number of Layers 1
North Pacific Ocean
300km
mapbox
© Mapbox © OpenStreetMap Improve this map

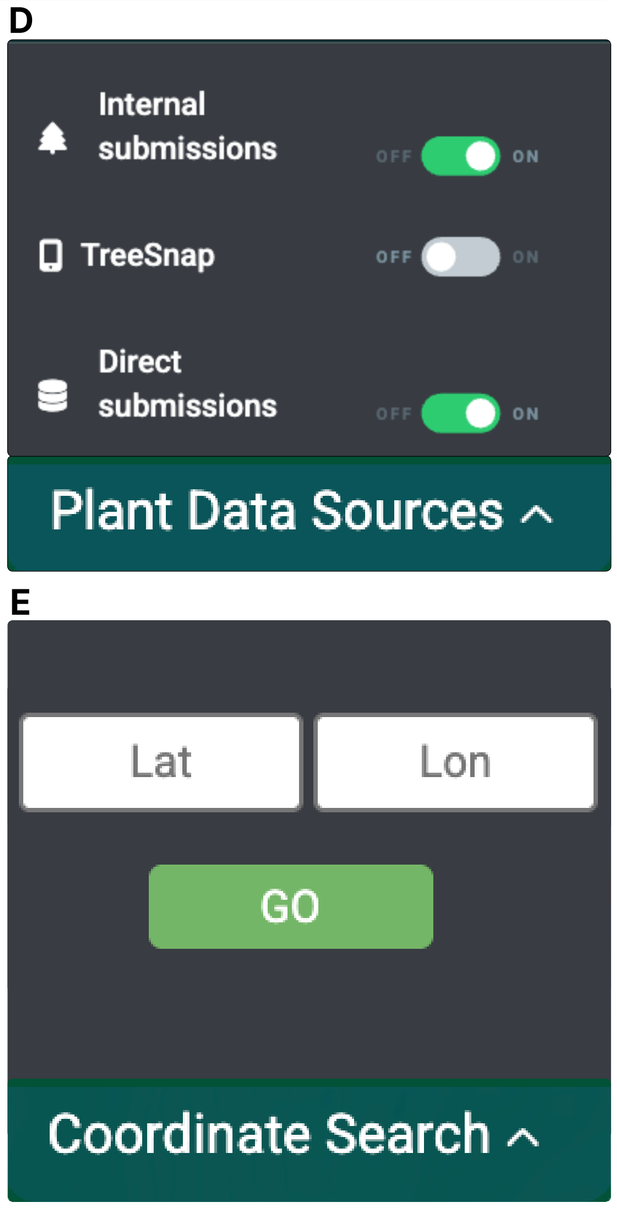
B


CartograPlant
About
Analyze
Plant Data Sources
Filters
Coordinate Search
Major Soil
Groups
Citation: FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009. Harmonized World Soil Database (version 1.1). FAO, Rome, Italy and IIASA, Laxenburg, Austria.
Description: This comprehensive Harmonized World Soil Database (HWSD) improve estimation of current and future land potential productivity, help identify land and water limitations, and enhance assessing risks of land degradation, particularly soil erosion. The HWSD contributes sound scientific knowledge for planning sustainable expansion of agricultural production and for guiding policies to address emerging land competition issues concerning food production, bio-energy demand and threats to biodiversity. This is of critical importance for rational natural resource management and in making progress towards achieving Millennium Development goals of eradicating hunger and poverty and addressing the food security and sustainable agricultural development, especially with regard to the threats of global climate change and the needs for adaptation and mitigation. This digitized and online accessible soil information system will allow policy makers, planners and experts to overcome some of the shortfalls of data availability to address the old challenges of food production and food security and plan for new challenges of climate change and accelerated natural resources degradation.
Date last updated: 2020
Resolution: Non-applicable
Displayed variables: Major soil types worldwide
Layers
Accession Summary
TGDR2269-BELA18-1
52.42 Lat | -126.17 Lon
Elevation: 130 m
Salicaceae
Populus trichocarpa
Coord. Type
Exact
Source
Data Dryad
STUDY INFO
SPECIES INFO
ADD ALL STUDY PLANTS
Soil Type
Leptosols
Soil Descr
Very shallow soils over hard rock or in unconsolidated very gravelly material
ADD SELECTED PLANTS (6)
BELA18-1
BELA18-5
BELA18-3
Burnt Bridge Creek Conservancy
Stupendous Mountain
Map Summary
Number of Plants 505 UNSEL ALL
Selected Plants 505
Species count 1
Publications Count 0
Number of Layers 1
2km
mapbox
© Mapbox © OpenStreetMap Improve this map

C


APPLY FILTER
RESET FILTER
AND OR
ADD RULE ADD GROUP
AND OR
ADD RULE
Study Accession
equal
TGDR1892
0 phenotypes found in study
Study Accession
equal
TGDR2269
41 phenotypes found in study
Species
equal
Populus trichocarpa
Filters

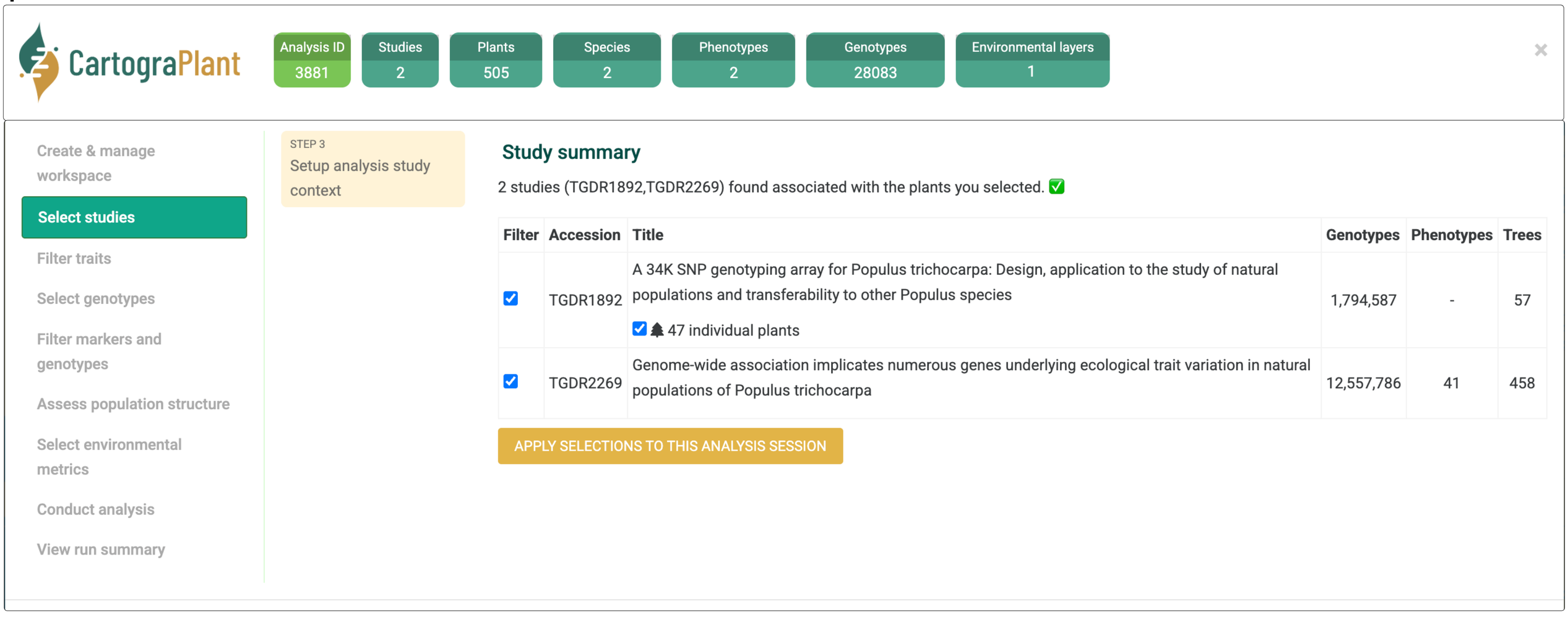
D


Internal submissions
OFF ON
TreeSnap
OFF ON
Direct submissions
OFF ON
Plant Data Sources

E

Lat
Lon
GO
Coordinate Search

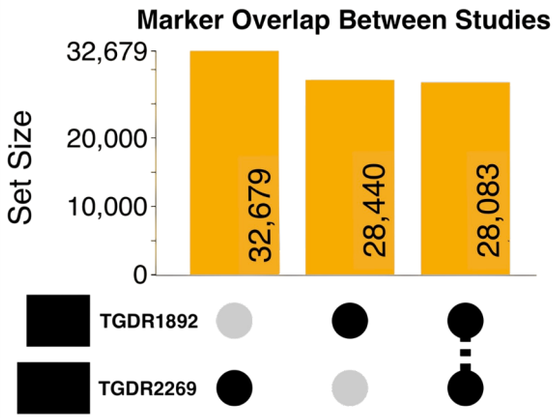
F


CartograPlant
Analysis ID 3881
Studies 2
Plants 505
Species 2
Phenotypes 2
Genotypes 28083
Environmental layers 1
Create & manage workspace
Select studies
Filter traits
Select genotypes
Filter markers and genotypes
Assess population structure
Select environmental metrics
Conduct analysis
View run summary
STEP 3
Setup analysis study context
Study summary
2 studies (TGDR1892,TGDR2269) found associated with the plants you selected.
Filter
Accession
Title
Genotypes
Phenotypes
Trees
TGDR1892
A 34K SNP genotyping array for Populus trichocarpa: Design, application to the study of natural populations and transferability to other Populus species
47 individual plants
1,794,587
-
57
TGDR2269
Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of Populus trichocarpa
12,557,786
41
458
APPLY SELECTIONS TO THIS ANALYSIS SESSION

Collection creation

Collection of plants with associated study metadata

Marker preparation



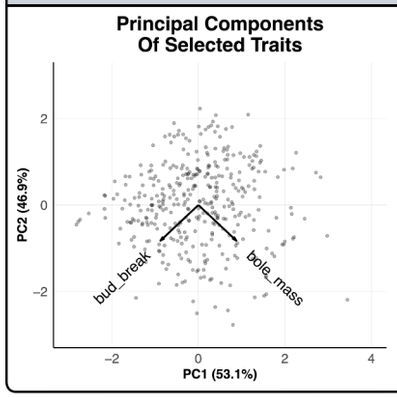
Access markers mapped to most recent reference assembly

Characterize marker overlap between studies

Define marker stastical cutoffs

Prune markers for LD (*PLINK*)

Trait preparation

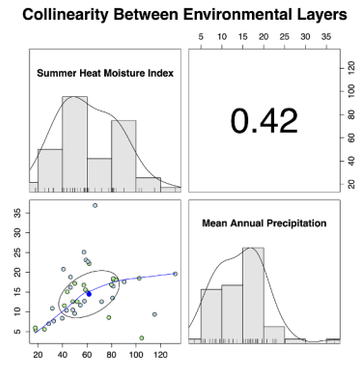


Standardize units across comparable traits

Trait correlation analysis

Select non-collinear trait measures

Environmental metric preparation



Environmental metric correlation analysis

Select non-correlatied environmental metrics from one or more stored layers

Finalized data selection

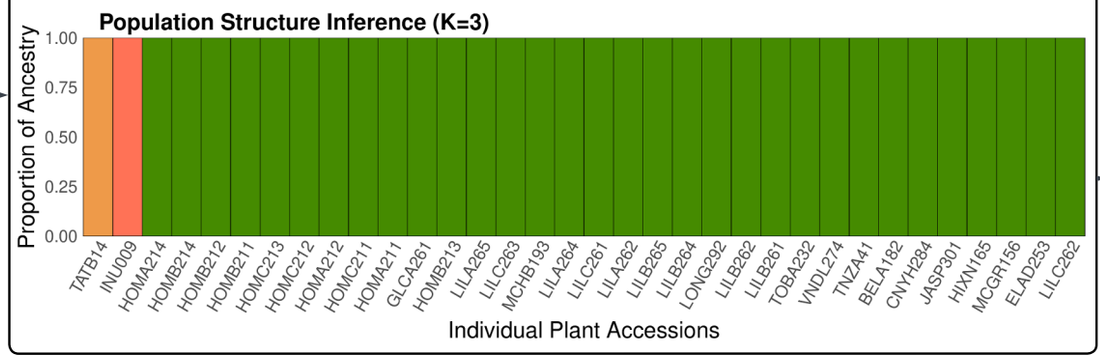
Markers

Traits

Environmental measures

Population structure analysis

fastSTRUCTURE



Marker association analysis

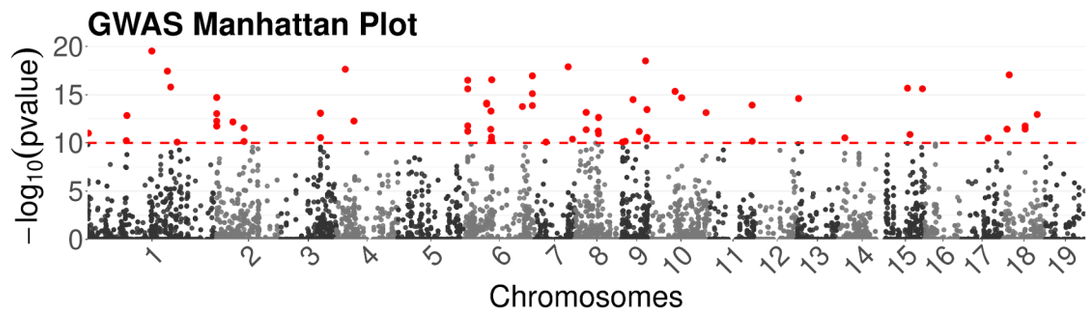
Genotype-trait association

GEMMA LM

GEMMA univariate LMM

Genotype-environment association

LEA



Supplemental File S1

CartograPlant: Bridging genomic, phenotypic, and environmental data to advance plant resilience and eco-evolutionary insight

Brandon M. Lind[1,2*†], Irene Cobo-Simón[1,3†], Meghan Myles[1†], Gabe Barrett[1], Emily Grau[1], Risharde Ramnath[1], Vlad Savitsky[1], Jill L. Wegrzyn[1,2]

[1]Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269, USA
[2]Institute for Systems Genomics, University of Connecticut, Storrs, CT 06269, USA
[3]Departamento de Ecología y Genética Forestal, Instituto de Ciencias Forestales (ICIFOR), Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria - Consejo Superior de Investigaciones Científicas (ICIFOR-INIA, CSIC), Madrid, Spain

26 September 2025

Running title: *Eco-evolutionary insight with CartograPlant*

Keywords: data integration, data interoperability, meta-analysis, biodiversity informatics, plant adaptation

[†]These authors contributed equally

***Corresponding Author**

Brandon M. Lind
ORCID: 0000-0002-8560-5417

Supplemental Text - Automated marker remapping

CartograPlant performs automated remapping of all marker data to all conspecific reference assemblies, including new assemblies as they are generated. It uses a custom Python (Rossum 1995) *v3.6.8* script which wraps a number of common bioinformatics tools to perform flank-based marker remapping. This script takes, as input, genotype data in either VCF or tabular format as well as a FASTA assembly to remap them to. It generates a VCF which contains genotype data as mapped to a different assembly. The following text will describe the software and methods used to complete this process.

The custom Python script initially computes indices for the target reference FASTA, using *samtools faidx* (Danecek et al., 2021) and *bwa index* (Li 2013). Then, the script checks whether it has been given marker flank sequence data. If not, for each SNP, it uses *samtools faidx* (Danecek et al., 2021) to obtain an 81-bp sequence from the reference genome to which the markers were initially mapped. This sequence represents the 40 nucleotides before the SNP, the SNP itself, and the 40 nucleotides after the SNP. If marker flank sequence data was given as an input, no new flanks are obtained.

The script then preprocesses SNP flank sequences by changing the nucleotide at the position of the variant to an *N*. Next, it creates a FASTA containing all of the SNP flank sequences. It uses a default *bwa mem* (Li 2013) command to align the flank sequence FASTA to the new assembly version FASTA. It then uses *samtools view* to parse the output binary alignment map (BAM) and *samtools sort* to sort it. It indexes the BAM with *samtools index* (Danecek et al., 2021) and converts it to a Browser Extensible Data (BED) file with *bedtools bamtobed* (Quinlan and Hall 2010). It obtains remapped marker positions from the BED file and stores them within a *pandas* (The pandas development team) dataframe. It then uses *samtools faidx* (Danecek et al., 2021) to determine the new reference allele for each marker from the assembly to which it is being remapped and adds it to the dataframe. For each SNP, it obtains a list of all called alleles in the input dataset, subtracts from this set the new reference allele, and adds this list to the dataframe as the possible alternate alleles. Finally, the genotype calls themselves are added to this dataframe, and numeric SNP calls are updated to reflect each SNP's new reference and alternate alleles.

The Python script then uses the generated dataframe to create a VCF. In this way, flank-based marker remapping can be completed for the integration of genotypic datasets which are mapped to different reference assemblies.

References

Danecek, P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, and H. Li. 2021. Twelve years of SAMtools and BCFtools. GigaScience 10:giab008.

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM (v2). https://doi.org/10.48550/arXiv.1303.3997

Quinlan, A. R., and I. M. Hall. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

Rossum, G. 1995. Python reference manual. CWI (Centre for Mathematics and Computer Science), NLD.

The pandas development team. 2019. pandas-dev/pandas: Pandas. Zenodo. https://doi.org/10.5281/zenodo.3509134