# Supplementary Information

#### **Supplementary Information S1.** Details on taxonomic harmonization

Because we combined data across different trait databases (AusTraits, BIEN, GIFT, and TRY) with GloNAF, we had to harmonize the taxonomy across all of them. These databases however do not use a shared taxonomic backbone, and as such we had to align them along the same taxonomic backbone. The size of the combined dataset only allowed us for an automatic harmonization, with efficient tools. We leveraged the Taxonomic Name Resolution Service (TNRS) through its R package named TNRS with the eponymous function TNRS(). For each database we proceeded in a similar fashion: we first went back to the original taxon name recorded in the database, we then used TNRS separately for each database. We configured TNRS to only return the best matched name for each submitted taxon name. We only kept binomial names returned as the Accepted\_species field by TNRS for each database before merging all datasets. Because all databases provided taxon names that were already standardized for common entry errors, like double spaces or encoding issues, we didn’t process them further before submission to TNRS.

For AusTraits, we extracted the column taxon\_name from the taxa table of AusTraits, we then matched this column, which contains the complete raw taxon name (including author names and infraspecific), through TNRS. We got 33 494 names which matched to 24 454 binomial names through TNRS.

For BIEN, we directly used the column scrubbed\_species\_binomial from the traits table, as the entire database is already matched against TNRS.

For GIFT, we extracted the columns genus, species\_epithet, subtaxon, and author from the raw traits table obtained from GIFT::GIFT\_traits\_raw() function. We processed these 676 678 raw unmatched names through TNRS and obtained 286 818 names

For TRY, we extracted the SpeciesName field from the trait table. We initially had 509 366 names which matched 268 312 names through TNRS.

For GloNAF, we extracted the taxon\_orig column from the flora\_orig\_2\_0 table, which gave original referenced names. We had 57 156 names, which matched 14 539 binomial names through TNRS.

In the end we had a set of 320 539 matched binomial names across trait databases, of which 14 073 were found in GloNAF.

#### 

#### **Supplementary Information S2.** Details on trait harmonization

Only AusTraits offered direct correspondence between its trait definitions and TRY, through direct mention of TRY Trait ID and, for certain traits, links to standard definitions provided by the TOP trait Thesaurus (Garnier *et al.*, 2017; Falster *et al.*, 2021; Wenk *et al.*, 2024). For all other databases, we manually matched the names of traits of each pair of databases (giving a total of six correspondence tables). For each trait of an input database, we tried to match trait(s) in the target database based on their associated metadata. We then noted if traits were either identical or related, the latter meaning that they are not exactly measuring the same entity but could be considered similar in a certain context. For example, BIEN reports “leaf area per leaf dry mass”, i.e. LMA, while AusTraits uses a trait named “specific\_leaf\_area”, which corresponds to SLA. Even if both traits are named differently, they correspond to the same biological measurement as LMA = 1/SLA, they can be considered fully identical. TRY reports six different types of SLA/LMA depending if it includes the petiole and/or is done at the leaflet level for compound leaves or if it’s undetermined. At the date of the analyses, no other database differentiated SLA values as much, thus we considered them identical to the “undetermined” value (TraitID in TRY: 3117) of TRY and considered the five other types of SLA as similar but not identical. Another example is that GIFT references “maximum plant height” while BIEN reports “measured plant height”; these two traits can be considered related while not identical. We built a trait name network where nodes are trait names and edges are links between traits, with three categories of links, as in the Austraits Plant Dictionary ontology: exact, if the traits were exactly matching ; close, if the traits were close but not exactly matching ; related, if the traits are related more lightly. The network allowed us to jointly consider all links across all databases. We considered traits the same across databases if they were part of the same connected component in the network. Each connected component then got assigned a new consolidated name representing the final trait name. We obtained three different trait name networks based on how stringent we were to consider two traits connected:

* The “full” network considered all links (exact, close, and related) to obtain connected components.
* The “close” network considered only exact and close links.
* The “exact” network considered only exact links.

We present only the result of the full network in the article as the results for the two other networks were qualitatively and quantitatively similar. We also focused on the “full” network as it allowed us to cast the widest net possible and show the most optimistic picture for trait data of non-native species.

#### 

#### **Supplementary Information S3.** Ecological justification of used trait combinations.

Westoby suggested the Leaf Mass per Area-Height-Seed Mass trio to quickly classify the ecology of many species with the fewest traits. We considered three ecologically meaningful trait combinations and assessed how often they were measured across non-native species, that are detailed in the table below:

#### **Table S3-1**. Description of the three considered trait sets

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name of trait combination | Source of trait combination | Number of traits | Traits in the set | Justification |
| Leaf-Height-Seed Mass (LHS) | Westoby, 1998 | 3 | * Leaf mass par area, * Plant height, * Seed mass | Compares the general life history strategy of species at global scale without being context-dependent |
| Global Spectrum of Plant Form and Function (GSPFF) | Díaz et al., 2016 | 6 | * Leaf mass par area, * Plant height, * Seed mass * Leaf area, * Stem Specific Density, * Leaf Nitrogen Content | Captures the main axes of variation in aboveground plant strategies worldwide |
| Root traits | Bergmann et al., 2020 | 4 | * Specific root length, * Root diameter, * Root tissue density, * Root nitrogen content | Positions species along the root economics spectrum and contrasts species along two axes: a “conservation” gradient corresponding to the fast-slow continuum and a “collaboration” gradient corresponding to reliance on mycorrhizal fungi to extract nutrients. |

### Supplementary Tables and Figures

#### **Figure S1**. Venn diagram showing the number of shared species across databases. AusTraits, BIEN, GIFT, and TRY are functional trait databases. GloNAF database references non-native species worldwide, while all other four are trait databases.

#### **Figure S2**. Complementarity of trait databases for the twenty-five most measured traits. Each circle has an area proportional to the number of species covered by a database for this trait, the intersections between circles is also proportional to the number of shared species. The numbers give the number of species in each section. Each database is always represented using the same color across traits: red for AusTraits, blue for BIEN, green for GIFT, and purple for TRY.

#### 

#### **Figure S3**. Partial residual plots of the trait knowledge model. These follow the negative-binomial model of the number of traits measured predicted with the 9 variables indicated as the x-axis of each panel (see Methods section for details). It shows the predicted variation of the number of traits in function of the variation of one predictor keeping all the other predictors equal to their mean value (here 0 as they were scaled).

#### 

#### **Figure S4**. Coefficients of alternative models of number of traits measured by species where variables were available for 70%, 80%, or 90% of the total range of species. The different models are thus fitted on a different number of species.

#### **Figure S5**. Comparison of estimates between phylogenetic and non phylogenetic models of trait knowledge. All variables showed p < 0.001 and are displayed with their confidence intervals.

#### **Figure S6**. Cumulative number and proportion of alien vascular plant species with at least a given number of traits indicated by the x-axis. The x-axis represents the number of traits. The way to read the graph is the following: “Close to 100% of species have at least 1 trait. More than 10,000 species have at least 10 measured traits”.

#### **Figure S7**. Treemap with number of traits measured per family. (A) All species ordered by number of traits with the respective numbers and proportions of species per amount of measured traits. (B) Ordered by family. The numbers correspond to the number of species within a family with the given trait knowledge. Way to read the graph: “There are 384 Poaeceae species which have 1 to 10 measured traits”.

#### **Figure S8**. Map showing the median number of measured traits across species per region (A) and the standard deviation of the number of traits across species per region (B). Note that the scale has been discretized for ease of reading. Empty circles represent islands and are not to scale for readability reasons. Light gray areas show areas with no information on alien vascular plant species, dark gray regions are where there were no variations in number of measured traits per species. The map is projected using the Equal-Earth projection.