The COMPADRE plant matrix database: an open online repository for plant demography


Published in:
Journal of Ecology

DOI:
10.1111/1365-2745.12334

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE (Digital Academic Repository)
Schedules of survival, growth and reproduction are key life-history traits. Data on how these traits vary among species and populations are fundamental to our understanding of the ecological conditions that have shaped plant evolution. Because these demographic schedules determine population distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
growth or decline, such data help us understand how different biomes shape plant ecology, how plant populations and communities respond to global change and how to develop successful management tools for endangered or invasive species.

2. Matrix population models summarize the life cycle components of survival, growth and reproduction, while explicitly acknowledging heterogeneity among classes of individuals in the population. Matrix models have comparable structures, and their emergent measures of population dynamics, such as population growth rate or mean life expectancy, have direct biological interpretations, facilitating comparisons among populations and species.

3. Thousands of plant matrix population models have been parameterized from empirical data, but they are largely dispersed through peer-reviewed and grey literature, and thus remain inaccessible for synthetic analysis. Here, we introduce the COMPADRE Plant Matrix Database version 3.0, an open-source online repository containing 468 studies from 598 species world-wide (672 species hits, when accounting for species studied in more than one source), with a total of 5621 matrices. COMPADRE also contains relevant ancillary information (e.g. ecoregion, growth form, taxonomy, phylogeny) that facilitates interpretation of the numerous demographic metrics that can be derived from the matrices.

4. Synthesis. Large collections of data allow broad questions to be addressed at the global scale, for example, in genetics (GENBANK), functional plant ecology (TRY, BIEN, D3) and grassland community ecology (NUTNET). Here, we present COMPADRE, a similarly data-rich and ecologically relevant resource for plant demography. Open access to this information, its frequent updates and its integration with other online resources will allow researchers to address timely and important ecological and evolutionary questions.

Key-words: big data, comparative approach, elasticity, matrix population model, open access, plant population and community dynamics, population growth rate, sensitivity, transient dynamics

Introduction

Demography is central to the understanding of ecology and evolution (Metcalf & Pavard 2007). The environment affects populations through its impacts on the vital rates of individuals (e.g. survival, growth, development, reproduction, dispersal), and those effects commonly differ among life cycle stages. Indeed, genetic variance in vital rates represents the foundation of fitness differences and thus the building blocks on which natural selection operates (Lande 1982). Thus, projections of demographic performance such as population growth rate and structure, equilibrium density, viability, and risk of local extinction are appropriately calculated using stage-structured models. Structured population models (Tuljapurkar & Caswell 1996) provide a convenient mathematical framework to connect variation in vital rates throughout an organism’s life cycle to its population dynamics. These models have been formulated as partial differential equations (McKendrik 1926; Kermack & McKendrick 1939), matrix population models (Leslie 1945; Lefkovitch 1965; Caswell 2001), delay differential equations (Nisbet 1997; Gurney, Blythe & Stokes 1999), integro-difference equations (the most relevant of which are, in the context of this manuscript, integral projection models (IPMs); Easterling, Ellner & Dixon 2000; Ellner & Rees 2006) and other mathematical frameworks, as reviewed elsewhere (Keyfitz 1967; Metz & Diekmann 1986; Nisbet, Gurney & Metz 1989; Tuljapurkar & Caswell 1996; Caswell 2001).

Matrix population models (MPMs hereafter) are the most widely used structured population models among plant population ecologists (Salguero-Gómez & de Kroon 2010; Crone et al. 2011). The popularity of MPMs arose from their (i) straightforward formulation, (ii) value at assembling complex data in an analytically tractable framework, (iii) solid mathematical foundations and (iv) clear biological interpretation of their outputs (e.g. the dominant eigenvalue corresponds to the population growth rate λ). MPMs are constructed by identifying multiple stages into which a species’ life cycle can be classified, either based on biological knowledge or various optimization algorithms (Vandermeer 1978; Moloney 1988; Salguero-Gómez & Plotkin 2010). Individuals within each stage are then characterized by their likelihood of surviving and either remaining in that stage or transitioning to another, and their contributions to sexual or clonal recruitment stages (for a detailed treatment, see Caswell 2001). This division of the life cycle into stages allows for the explicit incorporation of one of the most fundamental aspects of the study of demography: not all individuals in a population contribute equally to its dynamics; for example, seedlings have low survival and no reproductive output, whereas large plants typically have high survival and a large reproductive output.

Matrix population models describe the dynamics of populations over a discrete projection interval:

\[ \mathbf{n}(t+1) = \mathbf{A} \mathbf{n}(t) \] (1)

where \( \mathbf{n}(t) \) and \( \mathbf{n}(t+1) \) are population vectors containing the number of individuals in each life cycle stage at times \( t \) and \( t+1 \), respectively, and \( \mathbf{A} \) is the population projection matrix. The projection interval can vary from days (Hamda, Jevtic & Laskowski 2012) to weeks (J. Metcalf, pers. comm.), months...
facilitate these new endeavours, we introduce the COMPADRE revisiting conclusions drawn on the basis of smaller data sets. To provide all the classical demographic results of linear models. Time-varying models include periodic, density-dependent and stochastic models, depending on the nature of the trend in \( A(t) \). In nonlinear models, \( A(n) \) depends on the population vector \( n \). The projection matrix may also depend on the state of the environment. All these models are reviewed in Caswell (2001). New analytical techniques for MPMs are being developed rapidly, and many of these have already led to applications for plants. Some of these applications include the study of stochastic population dynamics under the impact of fires, floods or hurricanes (Caswell & Kaye 2001; Horvitz, Tuljapurkar & Pascarella 2005; Smith, Caswell & Mettler-Cherry 2005); extraction of age-specific information from stage-specific models (Cochran & Ellner 1992; Lebreton 2005; Caswell 2006, 2009; Tuljapurkar & Horvitz 2008; Horvitz & Tuljapurkar 2008) to explore the evolution of senescence in plants (Baudisch et al. 2013; Caswell & Salguero-Gómez 2013); spatial models for stage-structured invasions (Neubert & Caswell 2000; Buckley et al. 2005; Caplat, Nathan & Buckley 2012); density-dependent models in plants (Ramula & Buckley 2009), as well as bifurcation and sensitivity analyses in plants (Shyu et al. 2013); periodic models for seasonal population dynamics (Caswell & Trevisan 1994; Pico, de Kroon & Retana 2002; Bacaer 2009) (Caswell 2001); and short-term population dynamics (Verdy & Caswell 2008; Stott et al. 2010; Stott, Townley & Hodgson 2011) applied to a variety of species, including threatened and invasive species (Le Corff & Horvitz 2005; Hahn, Buckley & Muller-Scharer 2012).

Hundreds of studies in plant population biology have been published using MPMs. This rapid accumulation of demographic data, together with a commitment to open-access information by funding agencies, journals and researchers (Van Noorden 2012), now allows us to (i) address questions not yet answered due to the lack of global demographic information and (ii) revisit conclusions drawn on the basis of smaller data sets. To facilitate these new endeavours, we introduce the COMPADRE Plant Matrix Database (COMPADRE, for short), an open-access online repository of plant (MPM-based) population dynamics. COMPADRE (version 3.0) contains MPMs from 468 studies with 598 plant species, as well as ancillary information that allows for in-depth interpretation of the species’ demographic, such as geographic location of the study populations, ecoregion, study periods, treatments and plant growth form descriptors (see ‘What is in the COMPADRE portal?’ below).

COMPADRE is the result of efforts initiated over 25 years ago (Franco & Silvertown 1990) and later continued in parallel by several research groups, before being recently integrated into a single repository. The information in the database has been standardized and error-checked to facilitate user analyses and made publicly available at www.compadre-db.org. The goal of this publication is to introduce and describe this resource. We first offer a historical description of its origins and development; next, we explain how the database is managed and organized, and detail its current content. Finally, we briefly highlight its research potential and suggest future directions towards improving our understanding of plant population dynamics world-wide.

A historical perspective: from Leslie to COMPADRE 3.0

Introduced by Bernardelli (1941) and Leslie (1945, 1948), MPMs were largely neglected by ecologists, evolutionary biologists and demographers for two decades. The pioneering work of Lefkovitch (1965) and Keyfitz (1964) (Fig. 1) indicated the potential of MPMs to examine how individuals contribute to a population’s dynamics as a function of attributes such as age, ontogeny, size, spatial location and causes of death. Plant ecologists, who by the 1960s had realized that plant demography often depends more on size or ontogeny than age (Harper 1967; Harper & White 1974; Werner 1975), started adopting stage-structured MPMs in their research (Usher 1966; Sarukhan & Gadgil 1974; Hartshorn 1975; Werner & Caswell 1977).

COMPADRE 1.0, which previously stood for comparative plant demographic research, was founded in 1989 (J. Silvertown & M. Franco). A decade after its creation, it archived 105 plant species with their corresponding MPMs averaged across periods and populations. This number contrasted sharply with the handful of publications with MPMs available by 1986 (Table 8 in Caswell 1986). COMPADRE 1.0 resulted in several seminal publications on comparative demography including the exploration of the fast–slow continuum in the plant kingdom (Franco & Silvertown 1997), the evolution of senescence in plants (Silvertown, Franco & Perez-Ishiwara 2001) and the broad state of the art of plant population ecology (Franco & Silvertown 1990). Silvertown and Franco’s approach of partitioning the elasticity of population growth rate into three main components – stasis, growth and reproduction – enabled inter-specific comparison of basic demographic properties across the plant kingdom (Silvertown, Franco & McConway 1992; Silvertown et al. 1993). Briefly, the authors calculated the relative contribution of the stasis, growth and reproduction matrix elements to the population growth rate (\( \lambda \)) (elasticities \( s_{\text{sen}} \) de Kroon et al. 1986) and used them to locate species onto a ternary plot space. Refinement of this method to estimate the elasticity of the basic vital rates survival, growth and fecundity, rather than that of matrix elements that only approximate them (Franco & Silvertown 2004) triggered interest in the comparative use of MPMs for hundreds of plant species (Crone et al. 2011). This approach also helped to establish links between demography and conservation biology (Silvertown, Franco & Menges 1996; Crone et al. 2013).

Following COMPADRE 1.0, various research teams continued MPM digitization, either building upon COMPADRE 1.0 (COMPADRE 2.0; R. Salguero-Gómez, D. Hodgson), or starting anew (e.g. Iriondo et al. 2009; Ellis et al. 2012). The development of various MPM repositories resulted in publications examining a wide range of topics. These include comparative
demography of native and invasive populations (Ramula et al. 2008), effects of environmental stochasticity on plant population dynamics (Buckley et al. 2010), plant life-history evolution (Burns et al. 2010), the importance of plant shrinkage for population dynamics (Salguero-Gómez & Casper 2007), the first comparative exploration of short-term dynamics (Stott et al. 2010) and an assessment of utility for recovery of endangered species (Zeigler, Che-Castaldo & Neel 2013).

The current database, COMPADRE 3.0, has been hosted by the Max Planck Institute for Demographic Research since 2011. There, the priorities have been to (i) minimize redundant efforts in data digitization by integrating independent research groups around the world, (ii) continue archiving plant demographic data, adding important ancillary information, and (iii) build an online portal to share the information on an open-access basis. The efforts have entailed incorporating parallel MPM databases into a single repository and significantly scaling up the digitization (See Internal Organization below) of both published MPMs and MPMs obtained directly from the authors or calculated by us (Table 1). The third priority, building an online portal, is described in more detail in the section What is in the COMPADRE portal? (below).

The COMPADRE Plant Matrix Database

INTERNAL ORGANIZATION

The continuous increase in published MPMs in the recent decades (Fig. 1) requires a coordinated effort to search, digitize, error-check and release the information (Fig. 2). To facilitate this, COMPADRE is supported by two committees and a digitization team. The core committee (Appendix S1 in Supporting Information) is responsible for creating and updating protocols for data search, digitization, error-check and release, and for the overall infrastructure of the database and the internal organization of COMPADRE. The science committee, composed of a group of expert demographers located world-wide, provides external advice to the core committee on future directions, helps expand the geographic reach of COMPADRE and supplies the COMPADRE digitization team with grey literature. Both committees work together to secure funding for COMPADRE in the long term.

The COMPADRE digitization team primarily digitizes published information containing plant MPMs and ancillary information. The team is composed of students and postdoctoral fellows primarily based at the MPIDR (but see Appendix S1). The team has been trained in population ecology, MPMs and database archiving by the project leaders. In addition to entering information in a standardized format (Table 1; Fig. 3; Appendix S2), they contact authors to request information that is not included in the publication, or for help interpreting demographic information and carrying out error-checks.

FROM THE FIELD TO THE SCIENTIFIC COMMUNITY

Our work in COMPADRE starts when the work of the authors ends. After authors have collected plant demographic data, parameterized MPMs and – for the most part – published
Table 1. Variable names and meaning contain in the COMPADRE Plant Matrix Database, organized by seven general aspects: taxonomy, plant architecture, source, details of study, geolocation and matrix population model. A more detailed description can be found in the user protocol of COMPADRE in www.compadre-db.org

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taxonomy*</td>
<td>SpeciesAuthor</td>
<td>Taxonomic species name as used by the author(s) in the publication. When more than one study exists for the same species, these are given sequential numeric suffixes (e.g. <em>Cirsium_pitcheri, Cirsium_pitcheri_2, Cirsium_pitcheri_3</em>)</td>
</tr>
<tr>
<td></td>
<td>SpeciesAccepted</td>
<td>Currently accepted taxonomic name as used by The Plant List (<a href="http://www.theplantlist.org">www.theplantlist.org</a>). See the Appendix S3 for an R script to check accepted and synonym names from SpeciesAuthor above</td>
</tr>
<tr>
<td></td>
<td>Authority</td>
<td>Taxonomic authority of SpeciesAccepted</td>
</tr>
<tr>
<td></td>
<td>TaxonomicStatus</td>
<td>Whether SpeciesAuthor is currently accepted or synonym of SpeciesAccepted, as per The Plant List</td>
</tr>
<tr>
<td></td>
<td>TPLVersion</td>
<td>Version of The Plant List used for taxonomic validation</td>
</tr>
<tr>
<td></td>
<td>InfraspecificAccepted</td>
<td>Taxonomic intraspecific name of study species, as per The Plant List</td>
</tr>
<tr>
<td></td>
<td>SpeciesEpithetAccepted</td>
<td>Taxonomic species epithet of study species, as per The Plant List</td>
</tr>
<tr>
<td></td>
<td>GenusAccepted</td>
<td>Taxonomic genus of study species, as per The Plant List</td>
</tr>
<tr>
<td></td>
<td>Genus</td>
<td>Taxonomic genus of study species, as in SpeciesAuthor</td>
</tr>
<tr>
<td></td>
<td>Family</td>
<td>Taxonomic family of study species</td>
</tr>
<tr>
<td></td>
<td>Order</td>
<td>Taxonomic order of study species</td>
</tr>
<tr>
<td></td>
<td>Class</td>
<td>Taxonomic class of study species</td>
</tr>
<tr>
<td></td>
<td>DicotMonocot</td>
<td>Whether study species is a dicot, a monocot or neither</td>
</tr>
<tr>
<td></td>
<td>AngioGymno</td>
<td>Whether study species is an angiosperm, a gymnosperm or neither</td>
</tr>
<tr>
<td></td>
<td>Phylum</td>
<td>Taxonomic phylum of study species</td>
</tr>
<tr>
<td></td>
<td>Kingdom</td>
<td>Taxonomic kingdom of species. Note that while COMPADRE’s main focus is in the Plantae Kingdom, it also contains a few MPMs from species that do not belong to this kingdom. Nonetheless, these are included in COMPADRE due to taxonomic inertia and their demographic similarity with true plants (e.g. 68 MPMs from red algae, kingdom Chromalveolata)</td>
</tr>
<tr>
<td>Architecture*</td>
<td>AngioGymno</td>
<td>Whether species is an angiosperm, gymnosperm or neither</td>
</tr>
<tr>
<td></td>
<td>GrowthType</td>
<td>General functional type of the species (e.g. annual, fern, liana, herbaceous perennial; Table 2)</td>
</tr>
<tr>
<td></td>
<td>Authors</td>
<td>Last names of full authorship in study</td>
</tr>
<tr>
<td></td>
<td>YearPublication</td>
<td>Year of publication of study</td>
</tr>
<tr>
<td></td>
<td>DOI/ISBN</td>
<td>Digital object identifier (for manuscripts) or international standard book number (for books), when available; old publications do not have an assigned DOI. An R script is also provided to obtain full citation from manuscripts based on DOI (Appendix S3)</td>
</tr>
<tr>
<td></td>
<td>AdditionalSource</td>
<td>If additional information was obtained from a secondary source, the abbreviated citation is included here (first author’s first last name, abbreviated journal name and publication year; e.g. Godinez-Alvarez Bot Rev 2003 for <em>Escontria chiotilla</em>)</td>
</tr>
<tr>
<td>Source of information*</td>
<td>StudyDuration</td>
<td>Years of observation of the population dynamics of the species, calculated as StudyEnd – StudyStart + 1 (e.g. 2005–2000 + 1 = 6)</td>
</tr>
<tr>
<td></td>
<td>StudyStart</td>
<td>Year the study started</td>
</tr>
<tr>
<td></td>
<td>StudyEnd</td>
<td>Year the study ended</td>
</tr>
<tr>
<td></td>
<td>AnnualPeriodicity</td>
<td>Frequency with which seasonal or annual MPMs were constructed (e.g. 1: once per year; 2: twice per year; 0.2: once every five years)</td>
</tr>
<tr>
<td></td>
<td>NumberPopulations</td>
<td>Number of populations examined in the study – These may not match the number of populations with MPMs in COMPADRE 3.0 if the author has not made available all of the MPMs</td>
</tr>
<tr>
<td></td>
<td>MatrixCriteriaSize</td>
<td>Whether and on which biometric aspects of the species was the MPM constructed</td>
</tr>
<tr>
<td></td>
<td>MatrixCriteriaOntogeny</td>
<td>Whether some aspect of the developmental stage of the species was used to construct the MPM</td>
</tr>
<tr>
<td></td>
<td>MatrixCriteriaAge</td>
<td>Whether some aspect of the age of the species was used to construct the MPM</td>
</tr>
<tr>
<td></td>
<td>MatrixPopulation</td>
<td>Name(s) of populations from which the MPM was constructed. When no population name is provided in the source, the name of closest geographic landmark or letters in alphabetical (e.g. ‘A’, ‘B’, ‘C’...) or numerical order (e.g. ‘1’, ‘2’, ‘3’...) are used</td>
</tr>
<tr>
<td>Details of the study*</td>
<td>GPS location</td>
<td>Latitudinal degrees of study population</td>
</tr>
<tr>
<td></td>
<td>LatDeg</td>
<td>Latitudinal degrees of study population</td>
</tr>
<tr>
<td></td>
<td>LatMin</td>
<td>Latitudinal minutes of study population</td>
</tr>
<tr>
<td></td>
<td>LatSec</td>
<td>Latitudinal seconds of study population</td>
</tr>
<tr>
<td></td>
<td>LatNS</td>
<td>Latitudinal cardinal direction: North or South</td>
</tr>
<tr>
<td></td>
<td>LonDeg</td>
<td>Longitudinal degrees of study population</td>
</tr>
<tr>
<td></td>
<td>LonMin</td>
<td>Longitudinal minutes of study population</td>
</tr>
</tbody>
</table>

(continued)
Table 1. (continued)

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LonSec</td>
<td>Longitude seconds of study population</td>
<td></td>
</tr>
<tr>
<td>LonWE</td>
<td>Longitude cardinal direction: West or East</td>
<td></td>
</tr>
<tr>
<td>Altitude</td>
<td>Altitude of study population (in metres) obtained from Google Earth</td>
<td></td>
</tr>
<tr>
<td>Country</td>
<td>Country or countries where the study population was studied. Here, only countries currently accepted by the United Nations according to the ISO 3 list were employed (<a href="http://unstats.un.org/unsd/tradekb/Knowledgebase/Country-Code">http://unstats.un.org/unsd/tradekb/Knowledgebase/Country-Code</a>)</td>
<td></td>
</tr>
<tr>
<td>Continent</td>
<td>Continent of the study population</td>
<td></td>
</tr>
<tr>
<td>Ecoregion</td>
<td>Description of the terrestrial or aquatic ecoregion, corresponding to Olson et al.'s classification (2001), where the study took place. When the study is undertaken in its majority under controlled, indoor conditions (e.g. laboratory, glasshouse), this is noted as 'LAB'</td>
<td></td>
</tr>
<tr>
<td>Details of matrix population model†</td>
<td>StudiedSex</td>
<td>Sex(es) considered to construct the MPM</td>
</tr>
<tr>
<td>Details of matrix population model†</td>
<td>MatrixComposite</td>
<td>MPMs were differentiated among matrices that correspond to a given single population, single treatment and single annual period ('individual'; Fig. 4), to a single population, treatment and intra-annual period ('seasonal'), to a MPM that is the result of element-by-element arithmetic mean ('mean'), or where the individual-level data were pooled to construct a MPM over various periods, populations and/or treatments ('pooled'). We must note that by default we calculated the mean MPM when all individual MPMs in the study were made available. The pooled and mean matrices for all the individual, unmanipulated (see MatrixTreatment) MPMs coincide when the sample sizes and stage distributions at time $t$ are the same across all the individual MPMs. Mean MPMs were only calculated by us for unmanipulated individual matrices (see below)</td>
</tr>
<tr>
<td>Details of matrix population model†</td>
<td>MatrixTreatment</td>
<td>Treatment to which the demographic data used to parameterize the specific MPM was subjected. We specified 'unmanipulated' as those matrices where no human-led experimentation was carried out. Users are encouraged to carefully examine variable MatrixObservation (below) for additional pertinent information</td>
</tr>
<tr>
<td>Captivity</td>
<td>Whether the study species was in its wild setting, or under controlled conditions (e.g. greenhouse, botanical garden) for most of the demographic data that were collected</td>
<td></td>
</tr>
<tr>
<td>Start and end of study period</td>
<td>MatrixStartYear</td>
<td>Beginning year $t$ for MPM $A$ describing the population dynamics between time $t$ and year $t+1$</td>
</tr>
<tr>
<td></td>
<td>MatrixStartSeason</td>
<td>Beginning season $s$ for seasonal MPM $B$ describing the population dynamics between season $s$ and season $s+1$</td>
</tr>
<tr>
<td></td>
<td>MatrixStartMonth</td>
<td>Beginning month $m$ for seasonal MPM $B$ describing the population dynamics between month $m$ and month $m+1$</td>
</tr>
<tr>
<td></td>
<td>MatrixEndYear</td>
<td>End year $t+1$ for MPM $A$ describing the population dynamics between time $t$ and time $t+1$</td>
</tr>
<tr>
<td></td>
<td>MatrixEndSeason</td>
<td>End season $s+1$ for seasonal MPM $B$ describing the population dynamics between seasons $s$ and season $s+1$</td>
</tr>
<tr>
<td></td>
<td>MatrixEndMonth</td>
<td>End month $m+1$ for seasonal MPM $B$ describing the population dynamics between month $m$ and month $m+1$</td>
</tr>
<tr>
<td></td>
<td>MatrixSplit</td>
<td>To facilitate the calculation of various demographic properties (e.g. life expectancy $\eta_{P}$, mean age at first reproduction $L_{F}$, vital rate sensitivities), the MPM $A$ (matA, below) has been split into survival (matU), sexual (matF) and clonal reproduction (matUC) submatrices (Fig. 3) when sufficient information was provided in the source. In 4% of cases, insufficient information led to us not being able to split $A$ into $U$, $F$ and $C$. This matrix is referred to as Indivisible</td>
</tr>
<tr>
<td></td>
<td>Observation</td>
<td>Relevant observation that the user should bear in mind when analysing and interpreting the MPMs. In the present version, 43% of the matrices have observations. Observations include, for instance, warnings about the description by the author of an 'unmanipulated' population that some researchers may wish to treat as a treatment (e.g. natural fires), among others</td>
</tr>
<tr>
<td></td>
<td>MatrixClassAuthor</td>
<td>Classification of the stages in the life cycle of the study species as described by the author</td>
</tr>
<tr>
<td></td>
<td>MatrixClassOrganized</td>
<td>Standardization of MatrixClassAuthor into three stages: prop for seed banks, dorm for vegetatively dormant individuals and active for individuals photosynthetically active. We standardized MatrixClassAuthor in this way to facilitate cross-comparisons of various general life cycle stages. Note that other general classifications are possible, for instance, distinguishing reproductive individuals from non-reproductive individuals by evaluating the $F$ and $C$ submatrices</td>
</tr>
<tr>
<td></td>
<td>MatrixClassNumber</td>
<td>Sequence of numbered classes from 1 to MatrixDimension</td>
</tr>
<tr>
<td></td>
<td>MatrixDimension</td>
<td>Dimension of the MPM</td>
</tr>
<tr>
<td></td>
<td>SurvivalIssue</td>
<td>Reports maximum stage-specific survival in the submatrix $U$ (below). If this value $&gt; 1$, users are encouraged to carefully evaluate the matrix</td>
</tr>
</tbody>
</table>

(continued)
Table 1. (continued)

<table>
<thead>
<tr>
<th>Aspect</th>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population matrix model</td>
<td>matA</td>
<td>MPM including demographic processes that depend on survival (SubMatrixU below), sexual reproduction (if pertinent and available; SubMatrixF below), and clonal reproduction (if pertinent and available; SubMatrixC below; Fig. 3)</td>
</tr>
<tr>
<td></td>
<td>matU</td>
<td>Submatrix population model describing only survival-dependent demographic processes (e.g. seed bank, stasis, progression, retrogression, vegetative dormancy). Matrix elements corresponding to sexual and clonal reproduction are filled with zeros (Fig. 3)</td>
</tr>
<tr>
<td></td>
<td>matF</td>
<td>Submatrix population model describing only sexual reproduction. All other matrix elements are filled with zeros (Fig. 3)</td>
</tr>
<tr>
<td></td>
<td>matC</td>
<td>Submatrix population model describing only clonal reproduction. All other matrix elements are filled with zeros (Fig. 3)</td>
</tr>
</tbody>
</table>

*Information that is study specific.
†Information that is matrix population model specific.

their study (Fig. 2), the first step in database digitization is for COMPADRE project leaders and the digitization team to find published MPMs. Information reaches us through two main channels: (i) personal communications (currently < 2%) and (ii) periodic searches implemented by the project leaders in the core committee. Every week (ISI Web of Knowledge and Scopus) and every day (Google Scholar) automated literature searches are conducted with a combination of keywords (‘demography’, ‘elasticity’, ‘life cycle’, ‘matrix model’, ‘plant’, ‘population’ and/or ‘population growth rate’) that target manuscripts potentially containing MPMs. These searches are complemented with automatic alerts for publications that cite important MPM methodological advancements (e.g. de Kroon et al. 1986; de Kroon, van Groenendael & Ehrln 2000; Caswell 2001; Morris & Doak 2002). Finally, the project leaders and digitization team carry out searches with lower frequency (approx. 2 months) on ProQuest for PhD, MSc and BSc theses.

Once the publications containing MPMs have been acquired, the matrices and metadata are digitized into COMPADRE under a strict data entry protocol (Appendix S2) before being published online via the COMPADRE portal. Data provided via personal communications may alternatively be placed under an embargo period. The embargo date is chosen by the contributing author(s), and the data are released in the online portal only after the date has passed, and the authors have given explicit written permission.

WHAT IS IN THE COMPADRE PORTAL?

The fundamental piece of information in COMPADRE is the MPM. Each MPM describes the population dynamics of a given study × species × population × period × treatment combination. However, MPMs alone are of limited value. COMPADRE contains ancillary information that is study- or matrix-specific (Table 1) that allows users to interpret the MPM. A description of each variable can be found in Table 1 and in the user’s guide. The information can be broadly categorized into taxonomy of the species (and its phylogenetic position in the tree of life; see Appendices S2 and S5), species traits such as growth form or architectural organization, primary data source, study details and geographic location of the study populations. Further ancillary information specific to each MPM is also given (Table 1).

As described in the introduction, MPMs can be parameterized in various ways. The MPMs in COMPADRE contain only the matrix elements, and no underlying relationships that would describe, when available, the relationship between the vital rates and, for instance, density dependence, or environmental impacts. When possible, MPMs measured over a series of years are included, but COMPADRE does not include the stochastic models that might have been developed from that series of matrices (See The vision: beyond COMPADRE below).

The information available via the COMPADRE portal is completely open access after registration, and users are encouraged to employ it at their own discretion for research or teaching, but not for commercial uses. Registration and login required to access and download the data provide us with basic information (e.g. status, country, institution, email address) regarding users. This information allows the COMPADRE team to notify users if/when necessary (e.g. new data release, correcting errors, new/updated R scripts for matrix manipulations and analyses), as well as to obtain user statistics that will help justify grant support for the database in the long term. For these reasons, users should not share the downloaded data, but rather encourage other potential users to register and obtain the latest data set directly from COMPADRE portal.

Upon login, users are able to download four files: an R data object file (COMPADRE_Data-Nov_10_2014.RData) that contains the study-specific and matrix-specific information described in Table 1; the user’s guide, which details how the previous file is organized (Appendix S2); a zip file containing R scripts for data subsetting and manipulation (Appendix S3); and a nexus file containing the phylogeny of the species included in the first file (Appendix S3).

Once a significant volume of studies has been digitized, standardized and error-checked (see Data quality below), that section of the COMPADRE working version is pushed to the COMPADRE portal. Following the initial release of data from 468 studies, at the time of publication of this manuscript,
updates of the database will be made publicly available every 6 months – we currently have a total of 992 species in the process of being digitized and error-checked. For every new release, any errors found in previous releases will be corrected and accompanied by a note in the variable Observation (see COMPADRE User’s Guide and Table 1). In order to allow scientific reproducibility, outdated versions will be archived and accessible via the COMPADRE portal. Unpublished information provided to the digitization team by the author(s) under an embargo agreement (< 0.1% currently) is only released once the embargo period has expired and we have received consent from the author(s).

DATA QUALITY

Various aspects of the information are checked for quality assurance before the information is released in the COMPADRE portal. Below, we outline the most important error-checks and standardization procedures conducted on the content of COMPADRE (more details are provided in Appendix S2):

- Taxonomic names: For each study species, we report two values: the name used in the original source by the author(s) (SpeciesAuthor in Table 1) and the name currently accepted by The Plant List (www.theplantlist.org). In 15% of the species in COMPADRE 3.0, names provided by the authors are not the currently accepted names (i.e. no match, synonym or unresolved as per The Plant List). Citing the correct taxonomic name is vital to cross-database research (see The vision: beyond COMPADRE, below). Because taxonomic names may be updated frequently, we provide an R script to check on name spelling, synonyms and accepted taxonomic names based on the R package Taxonstand (Appendix S3).

- Phylogenetic tree: A phylogenetic tree for species in COMPADRE has been constructed to allow users to account for
phylogenetic ancestry if they wish to conduct comparative analyses (Appendices S3 and S5).

- Geolocation: The location where the demographic information was collected is important to the interpretation of MPMs. When GPS coordinates were made available, these were checked for consistency with the region and countries where the research took place (e.g. terrestrial plants cannot exist in the middle of an ocean). Information about the location of each population was then used to establish the ecoregion of the study as per the classification by Olson et al. (2001; Fig. 1 within).

- Architectural organization: The anatomic and physiological organization of plants can constrain their demography (Silvertown et al. 1993; Franco & Silvertown 1997; Silvertown, Franco & Harper 1997; Stott et al. 2010; Baudisch et al. 2013; Adler et al. 2014). In COMPADRE, plant habit is characterized according to the classification GrowthType (Table 1).

- Division of demographic processes: Whenever possible, each projection matrix $A$ in COMPADRE has been partitioned into a submatrix $U$ that contains only transitions and survival of existing individuals, a submatrix $F$ that contains sexual reproduction and a submatrix $C$ that contains clonal reproduction, as described in Fig. 3. This often requires communication with the authors to clarify what proportion of each element of $A$ corresponds to each process. In the current version of COMPADRE, 3% of the MPMs $A$ have not yet been divided into these components due to lack of information (see variable MatrixSplit in Table 1).

- Stage-specific survival: Stage-specific survival is given by the column sums of the submatrix $U$; it is constrained to lie between 0 and 1. Values greater than 1 render most analyses of survival and longevity impossible. While rounding errors may result in stage-specific survival probabilities slightly greater than 1, when any probabilities were much greater than 1, authors were contacted for clarification. In some cases, this resulted in a correction in the assignment of proportions of each matrix element in $A$ to the submatrices $U$, $F$ and $C$ described in Fig. 3, which was then noted in Observations (Table 1). MPMs pending of this correction have a value $>1$ in variable SurvivalIssue (Table 1).

- Classification of stages: To help automate comparative analyses, the life cycle stages in each MPM in COMPADRE have been classified into a standardized set of categories. Although we make available the exact description of the stages as described by the authors (see variable MatrixClassAuthor in Table 1), we have, based on this information, classified stages into ‘prop’ (propagules/seed bank stage), ‘dorm’ (vegetatively dormant individuals) and ‘active’ (individuals photosynthetically active, neither prop nor dorm). This allows the user to, for instance, derive life expectancy considering the beginning of life to be the point where seeds germinate (Caswell & Salguero-Gómez 2013;
Jones et al. 2014) (we provide some R scripts as basic examples in Appendix S3).

- Types of MPMs: When available, we have digitized the MPMs for each study × species × population × period × treatment combination. We refer to this unique combination of an MPM as an ‘individual’ matrix (variable MatrixComposite in Table 1), if they are based on an annual or lower frequency (e.g. every 5 years as in Shimatani et al. 2007; see AnnualPeriodicity in Table 1), and ‘seasonal’, if they explore the intra-annual population dynamics of a species. The seasons/months corresponding to intra-annual matrix models (Table 1) are identified to permit multiplying the seasonal matrices in the correct order to calculate statistics of interannual survival and population growth. Additionally, to facilitate comparative studies, we have also calculated, when not already available, element-by-element arithmetic ‘mean’ matrices across each of the aforementioned categories (Fig. 4). In addition to ‘individual’, ‘seasonal’ and ‘mean’ MPMs, COMPADRE also includes ‘pooled’ MPMs. The latter is only made available through the user, as they require individual-level data of the study, which COMPADRE does not have (see ‘What COMPADRE is not about’ below). A pooled MPM results from parameterizing the matrix directly from individual-level data, pooling a set of periods and populations, and thus by-passing the aforementioned element-by-element arithmetic mean approach. Depending on the estimation procedures used, the pooled and mean matrices may be identical if each matrix is based on the same number of individuals; they will, however, tend to differ when the sample sizes are different.

The COMPADRE team makes every effort to ensure data accuracy. However, we take no responsibility for consequences of improper use of the database by the user, nor for the possibility of potential typographical errors and omissions. Users who may detect incongruencies are encouraged to contact us at compadre-contact@demogr.mpg.de.

Fig. 4. COMPADRE contains the matrix population models per study corresponding to as many populations, study periods and treatments as the authors have made available, as well as those matrices that we have been able to derive from underlying demographic information. We refer to each matrix under a given study × species × population × period × treatment combination as the ‘individual’ matrix when the species’ population dynamics have been examined between years (a) or ‘seasonal’ matrix when the study explored intra-annual dynamics (b). Here, we showcase two examples of how matrices are classified in COMPADRE and, when necessary, calculated as summary matrices of individual or seasonal matrices.

(a) An imaginary study species where the population dynamics were followed annually between 2000 and 2004, resulting in three annual periods for two populations, each with a baseline (‘unmanipulated’; light grey matrix) control for a treatment, in this case, herbivory (pink matrix). Each set of individual matrices within a given population and under a specific treatment level was used to calculate the mean MPM for that site and treatment across all periods (unmanipulated: grey; herbivory: red); each of the population × treatment mean matrices was then averaged to produce a grand mean matrix of all periods and populations for a given treatment level (dark grey for unmanipulated; dark red for treatment). Different treatment levels are not averaged due to the lack of foreseeable utility. (b) An imaginary study species where the population dynamics were followed every summer and fall of 2000–2002, resulting in four seasonal matrices (summer–fall: light green hatched pattern; fall–summer: light brown hatched pattern). When not reported by the authors, we calculated the season-specific mean MPM per population (summer–fall: green hatched pattern; fall–summer: brown hatched pattern) and the season-specific mean matrix across all populations (summer–fall: dark green; fall–summer: dark brown). An annual MPM for an annual period of interest, population(s) and treatment(s) (not shown here) can be calculated by back-multiplying seasonal matrices as described by Caswell (2001, p. 349).

GLOBAL REPRESENTATION IN COMPADRE

The current version in the COMPADRE portal contains an unprecedented sample size for information on plant population dynamics: 468 studies with 598 plant (and algae) species, resulting in 5621 MPMs. This represents a substantial improvement in sample sizes available from previous publications (17 in Jones et al. 2014; 290 in Baudisch et al. 2013; 207 in Mbeau-Ache & Franco 2013; 36 in Caswell & Salguero-Gómez 2013; 222 Adler et al. 2014; 417 in R. Salguero-Gómez, O.R. Jones, E. Jongejans, S.P. Blomberg, D. Hodgson, C. Mbeau-Ache, P.A. Zuidema, H. de Kroon & Y.M. Buckley, in review). The individual MPMs and seasonal MPMs together total 3614 matrices, representing unique combinations of studies × species × populations × treatments × periods (Fig. 5a). The remaining 1997 MPMs are the product of element-by-element arithmetic means of other individuals or seasonal matrices. Given the temporal and population replication in each study (Fig. 5f and g), the relatively high ratio of individual and seasonal MPMs to mean and pooled MPMs (approx. 3:2; Fig. 5a) highlights that not all authors release all the individual/seasonal MPMs they have produced; there seems to be a greater tendency towards making only summary matrices available.

The information in COMPADRE 3.0 is well represented across the phylogenetic tree of plants. The current version contains 5597 (99.5%) MPMs for plants and green and red algae (Kingdom Plantae), as well as 24 (0.4%) MPMs for brown

---

Fig. 5. The current online version of the COMPADRE Plant Matrix Database (3.0) contains 5621 matrix population models. Breakdown of number of MPMs by (a) type of MPM (see Fig. 4), (b) plant growth form, (c) continent, (d) geographic location, where dots represent the approximate locations of each study population (when available), (e) matrix dimension, (f) duration of study and (g) number of populations per study. In (e), (f) and (g), the grey and red vertical lines represent the median and mean values, respectively.

algae (Kingdom Chromalveolata). Within vascular plants, there are 5323 (96.3%) MPMs for angiosperms, and 206 (3.7%) MPMs for gymnosperms; this representation approximates the extant angiosperms/gymnosperms ratio in the world (www.theplantlist.org). Furthermore, the most abundant taxonomic divisions of the Plantae Kingdom are represented in COMPADRE 3.0: mosses (Bryophyta: 2 MPMs, 0.0004%), ferns (Pteridophyta: 65, 0.75%), cycads (Cycadophyta: 43, 0.8%), conifers (Coniferophyta: 163, 2.9%) and flowering plants (Anthophyta: Liliopsida 1163, 20.7%, and Magnoliopsida 4150, 73.8%). Yet, no demographic information exists for liverworts (Hepatophyta), hornworts (Anthocerophyta), club mosses (Lycophtyta), horsetails (Sphenophyta), whisk ferns (Psilotopsida), the charismatic Ginkgo biloba, the only representative of the Ginkgoopsida, nor gnetae (Gnetophyta). In total, 117 taxonomic families are represented. The number of MPMs within taxonomic families varies up to three orders of magnitude, with Adoxaceae, Moraceae or Vochysiaceae containing a single MPM each, but other groups such as the Poaceae (286 MPMs), Primulaceae (297), Leguminosae (465) or Compositae (720) contain many more.

COMPADRE also includes information on general plant growth forms (Table 2; Fig. 5b). The number of MPMs in vascular, non-woody species (herbaceous perennials, annuals, algae, lianas, epiphytes and succulents) sums to 4478 (80.6%), whereas the number of woody species (trees, palms and shrubs) results in 1078 (19.4%). These values reflect the relative diversity as indicated by a recent estimate of woody species world-wide (FitJohn et al. 2014).

Overall, COMPADRE 3.0 offers a robust geographic coverage of plant population dynamics. The MPMs in COMPADRE are well spread geographically, although countries with a higher gross domestic product clearly appear to have had more opportunities to implement plant demographic studies using MPMs (Salguero-Gómez, unpublished data). Information in COMPADRE 3.0 includes plant and algae population dynamics on all continents except Antarctica (Fig. 5d). Yet, clear gaps exist in our knowledge of plant demography in certain regions, including Africa and Asia, North America (Canada, the USA and Mexico), Europe, Australia and Brazil together provide 89.3% of the MPMs in COMPADRE 3.0 (Fig. 5c). Unfortunately, remarkably few studies report information from countries with high biodiversity such as Honduras, the Democratic Republic of Congo, Paraguay or Indonesia. Furthermore, even some developed countries, such as Saudi Arabia, Turkey, Greece or Ireland, are under-represented.

Replication of studies through time and space is highly variable, but overall is poor. The mode of duration of studies in COMPADRE 3.0 is 4 years (Fig. 5c), corresponding to the length of an average PhD project, as well as that of most funding agencies. Only a handful of studies have followed plant population dynamics for ≥10 years for the online portal. This short duration is particularly limiting for our knowledge of plant demography, as many plants rank among the longest-lived organisms (Penuelas & Munné-Bosch 2010), with many achieving life expectancies much longer than a decade (Bruns-stein & Yamaguchi 1992; Bowers, Webb & Rondeau 1995).

The mode of number of populations studied per publication

<table>
<thead>
<tr>
<th>Ecoregion</th>
<th>Alga</th>
<th>Bryophyte</th>
<th>Fern</th>
<th>Annual</th>
<th>Herb</th>
<th>Shrub</th>
<th>Succulent</th>
<th>Epiphyte</th>
<th>Liana</th>
<th>Palm</th>
<th>Tree</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>BOR</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>176</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>178</td>
</tr>
<tr>
<td>DES</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>42</td>
<td>277</td>
<td>112</td>
<td>74</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>520</td>
<td></td>
</tr>
<tr>
<td>FGS</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>42</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>LAB</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>28</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>MED</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>425</td>
<td>93</td>
<td>6</td>
<td>0</td>
<td>33</td>
<td>18</td>
<td>576</td>
<td></td>
</tr>
<tr>
<td>MON</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>POE</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>TBM</td>
<td>1</td>
<td>0</td>
<td>63</td>
<td>197</td>
<td>1351</td>
<td>70</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>107</td>
<td>1792</td>
<td></td>
</tr>
<tr>
<td>TCF</td>
<td>19</td>
<td>0</td>
<td>2</td>
<td>595</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>95</td>
<td>714</td>
<td></td>
</tr>
<tr>
<td>TDB</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>85</td>
<td>111</td>
</tr>
<tr>
<td>TEG</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>TGS</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>700</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>725</td>
<td></td>
</tr>
<tr>
<td>TGV</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>98</td>
<td>16</td>
<td>56</td>
<td>0</td>
<td>13</td>
<td>17</td>
<td>202</td>
<td></td>
</tr>
<tr>
<td>TMB</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>228</td>
<td>45</td>
<td>0</td>
<td>42</td>
<td>3</td>
<td>91</td>
<td>152</td>
<td>563</td>
</tr>
<tr>
<td>TSC</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>14</td>
<td>13</td>
<td>0</td>
<td>25</td>
<td>5</td>
<td>60</td>
</tr>
<tr>
<td>TSS</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>TUN</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>Sum</td>
<td>36</td>
<td>2</td>
<td>63</td>
<td>256</td>
<td>3972</td>
<td>350</td>
<td>157</td>
<td>60</td>
<td>3</td>
<td>179</td>
<td>498</td>
<td>5576</td>
</tr>
</tbody>
</table>

Represented ecoregion abbreviations: BOR, boreal forests/Taiga; DES, deserts and xeric shrublands; FGS, flooded grasslands and savannas; LAB, laboratory/glasshouse controlled conditions; MED, Mediterranean forests, woodlands and savannas; MON, montane grasslands and shrublands; POE, polar ecoregions; TBM, temperate broadleaf and mixed forests; TCF, temperate coniferous forests; TDB, tropical and subtropical dry broadleaf forests; TEG, temperate upwellings; TGS, temperate grasslands, savannas and shrublands; TGV, tropical and subtropical moist broadleaf forests; TSC, tropical and subtropical coniferous forests; TSS, temperate shelf and seas ecoregions; TUN, tundra.

For other ecoregions not represented in this version of COMPADRE, see the User’s guidelines (Appendix S2).
in COMPADRE is one. This is also a particular concern if we are to gain a basic understanding of the intraspecific demographic variability and to implement optimal conservation management measurements. Indeed, some species can exhibit a great deal of demographic variability, which in some cases can be even greater than among non-closely related species (Oostermeijer et al. 1996; Silvertown, Franco & Menges 2000; Menges et al. 1996; Jongejans et al. 2010).

INFORMATION NOT CONTAINED IN COMPADRE 3.0

The niche of COMPADRE is matrix-based plant (and algae) demography. COMPADRE does not contain the individual-level records of survival, changes in stage, reproduction and death from which MPMs are often parameterized. Likewise, COMPADRE does not contain integral projection models (Easterling, Ellner & Dixon 2000), as another IPM database is currently under construction (63 plant species, PADRINO; R. Salguero-Gómez, unpubl. data). In addition to the MPMs, COMPADRE mostly contains supporting demographic data. Regarding questions for which other information not contained in COMPADRE (e.g. reproductive system, microhabitat conditions, seed mass) would be critical, users are encouraged to explore existing databases [e.g. TRY (Katge et al. 2011), BIEN (http://bien.nceas.ucsb.edu/bien), D3 (Hintze et al. 2013), BioFlor (http://www2.ufz.de/biolflor/overview/merkmale.jsp)].

The potential of COMPADRE is not in the MPMs per se, but in the outputs that can be derived from them. We have made a number of simple R scripts available to manipulate and interact with matrices, derive demographic outputs and correct for phylogenetic relationships (Appendix S3). Users are welcome to explore these or other more developed open-source libraries (Stubben & Milligan 2007; Stott, Hodgson & Townley 2012; Metcalf et al. 2013) and to carry out their own calculations based on compendia of methods for analysis of MPMs (e.g., Caswell 2001, 2009; Morris & Doak 2002). Users are also encouraged to avail themselves of other existing tools for teaching and mentoring such as the matrix workflows of BioVel (https://portal.biовel.eu). Lastly, users must note that the COMPADRE team provides the information and some basic tools, but no technical support.

USES AND MISUSES

Users of COMPADRE, in accordance of scientific ethical standards, are encouraged to acknowledge those who collected the field data, parameterized the MPMs, and made them public. To facilitate study citation, in addition to the variables described in Table 1 under the aspect ‘Source’, we have also made available an R script that produces a citation list of the studies based on their DOI or ISBN (Appendix S3). Users are also encouraged to visit the Supporting Information Appendix S4 for a comprehensive list of studies made public in the present release. Significant efforts have been made to digitize, standardize, error-check and supplement demographic information with relevant ancillary information such as ecoregion, growth form, taxonomy and phylogeny. Consequently, the information available in the COMPADRE online portal (www.compadre-db.org) is released under the Creative Commons Attribution-NonCommercial-ShareAlike CC BY-NC-SA license (http://creativecommons.org/licenses/by-nc-sa/4.0/). This requires users to cite the present manuscript and the version of the online portal in their works.

To help in the interpretation of the MPM, users are also encouraged to consult the notes provided under the variable Observation associated with each MPM. Here, we provide three examples to illustrate the importance of understanding the context and construction of each MPM. First, although we use the term ‘unmanipulated’ for those MPMs used as baseline for treatments (note the intentional avoidance of the term ‘control’ here), the environment under which field data used to build these matrices were collected may not be representative of typical conditions for that population (e.g. fires, herbivory, droughts). These events may or may not be representative of long-term normal conditions for that population. Secondly, plant demographers are now aware of accidental addition of a year delay in the population dynamics of plants by specifying an unnecessary seed stage where no permanent seed bank exists (Caswell 2001; p. 60). This artefact has been corrected in most – but perhaps not all – MPMs in COMPADRE. A third aspect to consider here is that some submatrices F and/or C may have all zero values in their matrix elements; in some cases, quantifying sexual and/or clonal reproduction may not have been the goal of the authors or it may have been logistically impossible. Submatrices F or C containing only zeros (Fig. 3) may not necessarily represent a failure of reproduction in the population, but rather that reproduction might not have been measured (Kubota 1997). In such cases, calculations involving reproduction (e.g. calculation of population growth rates from \( A = U + F + C \) are of questionable relevance. Finally, the information contained in the variables NumberPopulations and Population (Table 1) assumes that each studied site is a separate population. The term ‘population’ here is used loosely, following the authors’ need to differentiate between sites where the same species was studied, regardless of whether or not those sites constitute biologically the same population or not. Users may want to carefully inspect the published source of each MPM for that purpose. Alternatively, it is possible to derive geographic distances among populations based on GPS coordinates provided here and integrate that information with dispersal kernels (Bullock et al. 2012; Hintze et al. 2013) for a better understanding of population or meta-population dynamics.

THE VISION: BEYOND COMPADRE

COMPADRE contributes to an ambitious programme envisioned decades ago. Much of that vision has been realized, and here, we detail various expansions that we envisage for the future. Some arise from new technologies or analytical methodologies, while others from the increased interest in gathering natural history records of species and the expansion of biological data repositories.
Several projects have recently enabled users to obtain the real-time biological information of species observed in natural settings, such as iSpot (http://www.ispotnature.org). Phenological events such as first flowering are recorded by members of the general public and students via online Websites and apps and verified by experts before ingesting into public databases; NaturesCalendar in the UK (http://www.naturescalender.org.uk) and ClimateWatch in Australia (http://www.climatewatch.org.au) are two rather successful examples. We believe that COMPADRE has potential for research, teaching and outreach in a similar way as the aforementioned projects. Our vision is that, after some initial training, any user will be able to obtain information on important demographic attributes of a species/population such as life expectancy, the abiotic and biotic agents that most affect its population dynamics, time to quasi-extinction and potential to invade all within the reach of a smartphone with GPS to determine locality and a built-in camera to determine the species of interest and microhabitat conditions. To that end, we are currently developing an R library, COMPARE-DB (Salguero-Gómez & Jones, unpublished data), that interacts with the online portal to manipulate MPMs, derive demographic output, and interact with other databases.

Naturally, this potential requires integration of COMPADRE with other existing data repositories and their interconnectivity through a central, interconnected platform. Central repositories are already available for most of the relevant disciplines: genetics (GenBank), taxonomy (Catalogue of Life; The Plant List), plant anatomic and physiological traits (TRY, BIEN, i3), occurrence data (GBIF), conservation status (IUCN) and climate information (Fetch). What is missing currently is the central platform and the user interface protocols to allow for real-time interactivity. New online repositories such as the Australian Ecological Knowledge and Observation System (http://www.ecoinformatics.org.au) have made progress at integrating rich ecological data at a national level via online tools for data description and publication and serve as a model for future efforts.

The potential of COMPADRE also resides in cross-taxonomic, comprehensive demographic studies. Other databases are available or under development that will help us progress towards this goal: birds [BIDDABA, n = 857 species (Lebretton et al. 2012)], mammals [Pantheria [life-history traits of all extinct and recently extinct mammals, (Jones et al. 2009)]; or all animals: DATLIFE (life tables for 293, age-specific fertility for 61, maximum life span for 2659, adult/juvenile mortality for 532 animal species; A. Scheuerlein & Vaupel, unpubl. data), COMADRE (MPMs for over 1300 animal species; R. Salguero-Gómez and COMADRE core committee unpubl. data), PADRE [Integral Projection Models (IPMs) for approx. 100 plant species; R. Salguero-Gómez, unpubl. data] and the Global Population Dynamics Database (NERC Centre for Population Biology 2010).

To date, COMPADRE 3.0 and its predecessors have led to over 30 publications. These have addressed key ecological (Buckley et al. 2010; Salguero-Gómez et al. 2012; Adler et al. 2014), evolutionary (Pfister 1998; Baudisch et al. 2013; Jones et al. 2014), methodological (Salguero-Gómez & Plotkin 2010; Caswell & Salguero-Gómez 2013) and conservation biology questions (Ramula et al. 2008). However, these studies have realized only a small proportion of the full potential of COMPADRE. Many pressing questions remain for which COMPADRE will be a useful tool. For example, the coupling of climatic models with MPMs can better inform species distribution models to gain insights into the demographic mechanisms behind range boundaries (Guisan & Zimmermann 2000; Huntley et al. 2010). The power of COMPADRE’s extensive, comprehensive information will also improve the estimation of extinction risk based on non-demographic characteristics, which can be used to inform the management of threatened and endangered species (Traill, Bradshaw & Brook 2007; Sodhi et al. 2008).

Analyses based on the COMPADRE database will also help to identify lacunae in our current understanding of plant demography. For example, comparative analyses would benefit from greater spatial and temporal replication in demographic studies. Variation in demographic parameters within species across years or sites representing its distribution across gradients (Buckley et al. 2010; Zeigler, Che-Castaldo & Neel 2013) may be as large as variation among species (Jongejans & de Kroon 2005), limiting ability to draw conclusions or to identify patterns across taxa. In the working version of COMPADRE (Fig. 2), 37% of studies (365/992) contain data on the same species at two or more sites, and only 23% of studies (229/992) contain data for more than 5 years. Yet, such data are critical for improving inferences and incorporating spatial and temporal stochasticity in analyses of MPMs.

As cross-disciplinary and cross-taxonomic research develops with the contributions of COMPADRE, our team will do all within our reach to remain updated on information and user needs. Scientific methods come and go, and we cannot predict how MPMs will be used in the coming decades, or what new ways of parameterizing or formulating MPMs will become common practice (e.g. vital rates are starting to be parameterized as density-dependent, climate-dependent variables). This uncertainty may pose some challenges but may also create opportunities for COMPADRE and for the users. The database and its associated protocol for digitizing, standardizing and error-checking information will likely change to respond to user needs; researchers are encouraged to provide us with feedback. At the same time, we hope that users will be inspired to maximize the usefulness and impact of their own data by collecting the ancillary information that is standardized in COMPADRE; for instance, we found that most studies do not report GPS coordinates of the study populations (Johnson & Barton 2004). To facilitate integration into future versions of COMPADRE, all MPMs and associated metadata should be made available in the publication’s supplementary information, or as separate data papers (Ellis et al. 2012).

The ultimate goal of COMPADRE is to encourage plant population ecologists to share information and think not only about their specific questions, but also about how addressing a given question with their specific study species fits into the complex collage of what we know and do not know about
plant demography. The 932 species currently in the working version of COMPADRE (Fig. 1) represent 0.374% of the ~250,000 extant species of plants (Govaerts 2001). Given how costly these data are to obtain, this is a fairly impressive representation, particularly when considering the substantial phylogenetic coverage in the data set (Appendix S5), and the youth of plant population ecology as a discipline (Harper 1977). The exponential increase in the number of these types of studies provides further cause for optimism (Fig. 1). Nonetheless, COMPADRE shows gaps and opportunities for the development of research in geographic regions where no plant studies are yet included, for example, in most African countries and the Middle East (Fig. 5d).

One of the great strengths of comparative studies is their use as a tool for identifying generalities that can be extrapolated to poorly known species (Shea et al. 1998; Ramula et al. 2008). With many plant species threatened with extinction both locally and globally, COMPADRE provides an easily accessible tool for obtaining data for these species, or for identifying useful generalizations for particular taxonomic groups, life-history strategies and regions which should help guide the management and forecasting of threatened populations. The COMPADRE Plant Matrix Database represents a step closer to achieving a global repository of biologically important data for the management of the planet’s biodiversity.

Acknowledgements
A list of authors whose fieldwork and analytical efforts constitute the very core of the repository Plant Matrix Database is provided in the Supporting Information Appendix S4. The authors’ willingness to share unpublished data and clarify questions on published materials is an inspiration for the ‘big data, big ecology’ era. COMPADRE is currently supported by the Evolutionary Demography Laboratory and the Max Planck Research Group on Modeling the Evolution of Phylogenetics in the Max Planck Institute for Demographic Research (MPIDR); for a list of funding support since its inception, in 1989, see Supporting Information. Logistical and computing support was provided by the IT team at the MPIDR.

Data accessibility
The data associated with this manuscript can be accessed at www.compadre-db.org.

References


R. Salguero-Gómez et al.


Received 21 July 2014; accepted 6 October 2014

Handling Editor: Mark Rees

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Constituency of COMPARE.

Appendix S2. COMPARE user’s guide.

Appendix S3. COMPARE R scripts.

Appendix S4. Extended literature used in COMPARE 3.0.

Appendix S5. COMPARE phylogeny.

Appendix S6. COMPARE funding support and extended acknowledgements.

Appendix S7. Author contributions.

Appendix S8. Supplementary information references.