USING HERBARIUM DATA TO INCREASE THE LIKELIHOOD OF FINDING FERTILE PLANTS IN THE FIELD

J. S. Silva[1]*, E. Lenza[2], A. L. C. Moreira[1] & C. E. B. Proença[1]

The Phenological Predictability Index (PPI) is an algorithm incorporated into BRAHMS, one of the most widely used herbarium database management systems. PPI uses herbarium specimen data to calculate the probability of the occurrence of various phenological events in the field. Our hypothesis was that use of PPI to quantify the likelihood that a given species will be found in flower bud, flower or fruit in a particular area in a specific period makes field expeditions more successful in terms of finding fertile plants. PPI was applied to herbarium data for various angiosperm species locally abundant in Central Brazil to determine the month in which they were most likely to be found, in each of five areas of the Distrito Federal, with flower buds, flowers or fruits (i.e. the 'maximum probability month' for each of these phenophases). Plants of the selected species growing along randomised transects were tagged and their phenology was monitored over 12 months (method 1), and two one-day field excursions to each area were undertaken, by botanists with no prior knowledge of whether the species had previously been recorded at these sites, to record their phenological state (method 2). The results showed that field excursions in the PPI-determined maximum probability month for flower buds, flowers or fruits would be expected to result in a > 90% likelihood of finding individual plants of a given species in each of these phenophases. PPI may fail to predict phenophase for species with supra-annual reproductive events or with high event contingency. For bimodal species, the PPI-determined maximum probability month is that in which a specific phenophase is likely to be most intense. In planning an all-purpose collecting trip to an area with seasonal plant fertility, PPI scores are useful when selecting the best month for travel.

Introduction

A common objective of field studies is to find fertile plants. Fertile plant material is needed if a floristic voucher specimen is to be prepared for long-term preservation in a herbarium as a representative of the species. In phytosociological and ecological studies, the presence of flower buds, flowers and fruits facilitates taxon identification. Additionally, for certain studies fertility is a *sine qua non* condition. For example, flower buds are a prerequisite for determination of gametic chromosome number, *n* (Costa & Forni-Martins, 2007a), and fruits are needed for determination of somatic chromosome number, 2*n*, from tissue prepared

[1] Departamento de Botânica, Instituto de Ciências Biológicas, Universidade de Brasília, 70919-900 Brasília − DF, Brazil.
[2] Departamento de Ciências Biológicas, Universidade do Estado de Mato Grosso, 78690-000 Nova Xavantina − MT, Brazil.
* Corresponding author. E-mail: julay_silvestre@yahoo.com.br.

from the root tips of germinating seeds (Costa & Forni-Martins, 2007a, 2007b); these chromosome counts are essential information on which to base the choice of parents for plant-breeding experiments (Bretagnolle & Thompson, 1995). Flower buds and flowers are obviously necessary for floral ontology studies (Gomes *et al.*, 2008), and seeds have been collected for propagation since ancient times, and in recent decades, for conservation in seed banks (Wishnie *et al.*, 2007).

A vast literature is available on how to successfully collect, preserve and germinate seeds (Willan, 1985; Vazquez-Yanes & Orozco-Segovia, 1993; Broadhurst *et al.*, 2008; and references therein). However, few studies have focused on how to find a given plant species in the field when it is in fruit.

Herbarium data have been made more widely accessible by projects to computerise the contents of herbaria and make the data available via online databases (Smith *et al.*, 2003) – a development that has aided traditional floristic (Harris *et al.*, 2012), ecological (Gimaret-Carpentier *et al.*, 2002), phytogeographical (Silva *et al.*, 2013) and morphological (Malhado *et al.*, 2009) studies. Other readily available sources of data for botanical researchers are reports of phenological (e.g. Boulter *et al.*, 2006) and plant conservation (e.g. van Hengstum *et al.*, 2012) studies. This information can be used to develop our understanding of how biological systems interact with the environment (Borchert, 1996; Miller-Rushing *et al.*, 2006). Several studies have combined scattered floristic and phenological information from herbaria or field excursions and successfully organized it into standardized tables (in other words, performed data structuring) (Barros & Caldas, 1980; Antunes & Ribeiro, 1999; Tannus *et al.*, 2006; Vasconcelos *et al.*, 2012; Pinheiro, 2013).

Developed over the past 20 years, Brahms (Botanical Research and Herbarium Management System) is a database system used in herbarium, botanic garden and seed bank settings in about 60 countries. For researchers working in the field of systematics or floristics, or carrying out botanical surveys or biodiversity studies, its wide-ranging functionality includes the ability to carry out extensive analyses, calculations and text formatting (Filer, 2010).

Brahms now incorporates an algorithm, the Phenological Predictability Index (PPI), that uses information held in the database to determine the month in which a phenological event is most likely to occur (hereafter referred to as the 'maximum probability month'). As an extreme example of the potential benefit of using the PPI tool, the reader is asked to imagine a novice botanist with little experience in northern temperate forest phenology. If a field excursion were planned for the middle of winter, it would probably result in no fertile collections. However, if PPI were used it would show records of flower bud, flower and fruit phenological events to be most concentrated in summer and autumn, and use of this information to guide planning would result in a more successful and economical field excursion by maximising the likelihood of finding fertile plants and thereby minimising the collecting effort and associated costs. The aim of the present study was to test how PPI

performs this way as a practical planning tool, therefore the research question was, 'Do PPI's predictions of the maximum probability month translate to increased likelihood of finding fertile plants?'

To reflect the different potential needs of Brahms users, the performance of PPI was tested using two methods. Method 1, carried out over the long term, was used to test the utility of PPI for ecologists and field biologists carrying out controlled experiments over an extended time (e.g. months). Method 2, conducted over the short term, was used to test the utility of PPI for foresters and specimen or seed collectors, whose field excursions are of shorter duration (e.g. days).

Materials and methods
Locality selection

The Distrito Federal, in Central Brazil, was chosen as an ideal area to test the performance of PPI, because it has high biological diversity and is one of the most well-collected regions in Central South America (Simon & Proença, 2000). Located between 15°30' and 16°03'S and 47°25' and 48°12'W, the area is a rich mosaic of biomes: grasslands, savannas, seasonal forests and gallery forests (Coutinho, 2006; Batalha, 2011). It has a tropical seasonal climate and varies in altitude from 750 m to 1336 m; tropical flowering patterns are generally more diverse than temperate ones (Newstrom *et al.*, 1994).

The Distrito Federal is in the centre of a 2 million-ha savanna-dominated ecological region, the Cerrado. The Cerrado is the most diverse savanna in the world (Klink & Machado, 2005), borders both the Atlantic and the Amazon Forests, and runs from the River Plate basin to the semi-desertic Caatinga scrubs of northeastern Brazil. Contact with such diverse habitats over time has favoured a high level of floristic and genetic exchange. This has increased the taxonomic and functional diversity of the Cerrado (Unesco, 2002; Mendonça *et al.*, 2008) to such an extent that it is now one of the world's biodiversity hotspots (Myers *et al.*, 2000).

Data from five protected areas within the Distrito Federal were used in the analyses described in the present study: Estação Ecológica de Águas Emendadas (ESECAE; average altitude, 1075 m), Parque Nacional de Brasília (PNB; average altitude, 1100 m), Reserva Ecológica do Instituto Brasileiro de Geografia e Estatistica (IBGE; average altitude, 1100 m), Jardim Botânico de Brasília (JBB; average altitude, 1090 m) and Campus da Universidade de Brasília (COUNB; average altitude, 1025 m). Together, these encompass the three biomes of the cerrado *sensu lato* (Coutinho, 2006; Batalha, 2011), with field, savanna and forest all represented.

The two methods used to test the performance of PPI (see *Testing strategy* for details) were carried out at different areas. Method 1 (phenological monitoring) was used at ESECAE, PNB and IBGE, and method 2 (spot-check field excursions to record phenological state) at JBB and COUNB.

Species selection

The 28 target species whose data were used in the present study met the following criteria:
1) wide taxonomic sampling across the angiosperms (13 orders, 23 botanical families);
2) wide ecological variability, as reflected by habitat, pollination and seed dispersal
syndromes, Raunkiaer system (life form classification), and leaf drop and flush strategy;
3) a sufficient number of herbarium specimens to make it likely that more than 50 unique
combinations of phenological event, month and year (hereafter referred to as 'unique records')
were in the databases of the herbaria visited; and 4) clear species circumscription (Table 1).

All the species were well known to the authors, and their identifications had been
confirmed both in the field and in herbaria. Regarding data from the *speciesLink* (2014)
herbarium database, only specimens whose identity had been determined by taxonomic
specialists were included.

Phenological predictability based on herbarium data

In PPI, default phenological events are flower buds, flowers, fruits (any stage), mature fruits,
leaf senescence and vegetative state; other periodic events, such as leaf flushing, galls or
fungal infections, may be added by the user (Brahms documentation, 2012). PPI avoids
some of the pitfalls of phenological scoring in herbaria described by Yost *et al.* (2018), such
as those arising from the use of words in different languages, different terms and different
abbreviations for the same phenological state. For example, 'flower', 'flowers', 'fleur', 'flores',
'fl' and 'flws.' could be inserted into the same field to indicate a flowering specimen, thus
making automatic interpretation by a program algorithm very difficult. PPI works on the
basis of a different field for each phenological state (buds, flowers, fruit, etc.) and requires
the recorder to simply insert an asterisk (*) into the relevant field to indicate its presence.

The PPI algorithm calculates a score for each month of the year, using an *ad hoc* formula
that takes into account the number of database records of phenological events for that
month and for its neighbouring months. This process is repeated 12 times, targeting each
month of the year in succession. The higher the PPI score, the greater the concentration
of records of relevant phenological events in or around the maximum probability month. In
any cases of the same phenological state having been recorded for two or more collections
as occurring in the same month and year, these 'duplicate' records are removed from the
calculations so that only unique records are used (see Proença *et al.*, 2012, for details).

The PPI results are obtained by submitting a query for the taxon of interest. The results
for each species−phenophase combination are displayed as: 1) the maximum probability
month; 2) the PPI score, ranging from 0.02 to 1 (i.e. the minimum to maximum likelihood of
finding individual plants of the species in the specified phenophase); and 3) a graph showing
the number of unique records (*y*-axis) plotted against the months of the year (*x*-axis), with
the maximum probability month circled (Figure 1).

Table 1. Characters of species of the Distrito Federal, Central Brazil, whose data were included in the present study to test the utility of the Phenological Predictability Index (PPI) tool in Brahms

Class, order, family and species[a] (voucher no.[b])	Habitat	Habit	Pollination and seed dispersal syndrome (-phily, -chory)	Raunkiaer system (-phyte/-phyll)	Leaf drop and flush strategy
Commelinidae, Poales					
Poaceae					
Echinolaena inflexa (Poir.) Chase[c] (696)	G, S, F	H	Anemo, auto	Hemicrypto, micro	EC
Core eudicots, Dilleniales					
Dilleniaceae					
Davilla elliptica A.St.-Hil.[d] (691)	S, F	Sh	Zoo, zoo	Phanero, noto	B
Rosidae I, Fabales					
Fabaceae					
Periandra mediterranea (Vell.) Taub.[c] (*Soares* 137, HEPH)	S, F	SSh	Zoo, auto	Chamae, micro	ES
Chamaecrista conferta (Benth.) H.S.Irwin & Barneby[c] (*Nóbrega* 2152, HEPH)	S, F	SSh	Zoo, auto	Chamae, micro	B
Stryphnodendron adstringens (Mart.) Coville[d] (490)	S, F	T	Zoo, auto	Phanero, micro	B
Calliandra dysantha Benth.[c] (693)	G, S	SSh	Zoo, auto	Chamae, nano	D
Rosidae I, Rosales					
Moraceae					
Brosimum gaudichaudii Trécul[c] (708)	S, F	Sh	Anemo, zoo	Phanero/Micro	ES
Rosidae I, Malpighiales					
Euphorbiaceae					
Dalechampia caperonioides Baill.[c] (697)	G	H	Zoo, auto	Hemicrypto/Micro	B
Ochnaceae					
Ouratea hexasperma (A.St.-Hil.) Baill.[d] (699)	G, S, F	Sh	Zoo, zoo	Phanero/Meso	ES
Malpighiaceae					
Byrsonima verbascifolia (L.) DC.[d] (500)	G, S, F	Sh	Zoo, zoo	Phanero/Macro	B
Banisteriopsis campestris (A.Juss.) Little[c] (637)	G, S	Sh	Zoo, anemo	Chamae/Noto	D
Salicaceae					
Casearia sylvestris Sw.[d] (702)	S	Sh	Zoo, zoo	Phanero/Noto	ES
Caryocaraceae					
Caryocar brasiliense Cambess.[c] (701)	S, F	T	Zoo, zoo	Phanero/Meso	ES
Rosidae II, Myrtales					
Lythraceae					
Diplusodon villosus Pohl[c] (*Zanatta & J.S. Silva* 1232)	S	H	Zoo, auto	Chamae/Micro	B
Vochysiaceae					
Vochysia elliptica Mart.[c] (*Haidar* 634, HEPH)	S, F	T	Zoo, anemo	Phanero/Noto	EC

Class, order, family and species[a] (voucher no.[b])	Habitat	Habit	Pollination and seed dispersal syndrome (-phily, -chory)	Raunkiaer system (-phyte/-phyll)	Leaf drop and flush strategy
Myrtaceae					
Psidium firmum O.Berg[c] (684)	S, F	Sh	Zoo, zoo	Chamae, noto	ES
Melastomataceae					
Miconia albicans (Sw.) Steud.[d] (686)	S, F	Sh	Zoo, zoo	Phanero, noto	EC
Rosidae I, Sapindales					
Burseraceae					
Protium ovatum Engl.[c] (707)	S, F	Sh	Zoo, zoo	Phanero, noto	EC
Anacardiaceae					
Anacardium humile A.St.-Hil.[c] (689)	G, S, F	Sh	Zoo, zoo	Chamae, meso	EC
Asteridae I, Ericales					
Styracaceae					
Styrax ferrugineus Nees. & Mart.[d] (690)	S	T	Zoo, zoo	Phanero, noto	ES
Asteridae I, Metteniusales					
Metteniusaceae					
Emmotum nitens (Benth.) Miers[c] (566)	S, F	T	Zoo, zoo	Phanero, meso	ES
Asteridae I, Gentianales					
Rubiaceae					
Palicourea rigida Kunth[c] (687)	S	Sh	Zoo, zoo	Phanero, macro	EC
Asteridae I, Solanales					
Solanaceae					
Solanum subumbellatum Vell.[c] (709)	G, S, F	SSh	Zoo, zoo	Chamae, micro	D
Asteridae I, Lamiales					
Lamiaceae					
Hyptis villosa Pohl ex Benth.[c] (698)	G, S, F	H	Zoo, auto	Chamae, micro	D
Acanthaceae					
Ruellia incompta (Nees) Lindau[c] (694)	G, S	SSh	Zoo, zoo	Chamae, micro	EC
Bignoniaceae					
Jacaranda ulei Bureau & K.Schum.[c] (692)	G, S, F	Sh	Zoo, anemo	Hemicrypto, micro	D
Asteridae II, Asterales					
Asteraceae					
Piptocarpha rotundifolia (Less.) Baker[c] (703)	S	T	Zoo, anemo	Phanero, meso	ES
Chresta sphaerocephala DC.[c] (*Martins* 1353, UB)	G, S	H	Zoo, anemo	Geo, meso	ES

B, Brevideciduous; D, deciduous; EC, evergreen continuous; ES, evergreen seasonal; F, forest; G, grassland; H, herb; S, savanna; Sh, shrub; SSh, subshrub; T, tree.

[a] Presented in linear APG IV order (Angiosperm Phylogeny Group, 2016).

[b] *J.S. Silva* collections at UB, unless otherwise indicated.

[c] Data from phenological monitoring carried out in 2012 (see text for details).

[d] Data from phenological monitoring carried out in 2001 (Lenza, 2005).
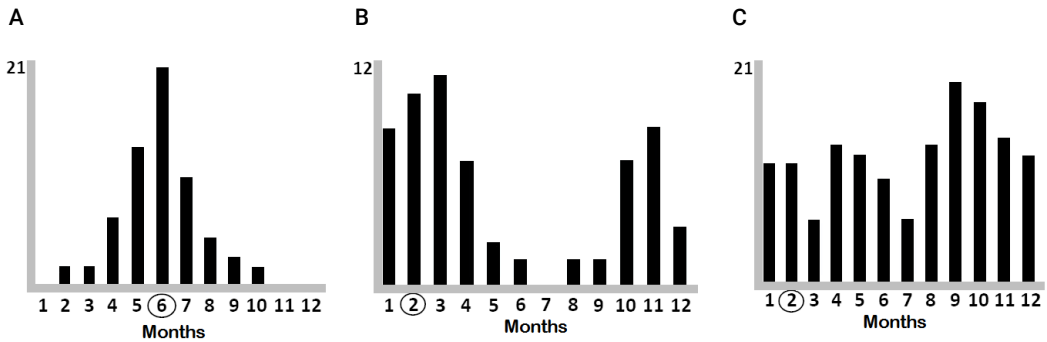
Figure 1. Graphs generated by the Phenological Predictability Index (PPI) tool in Brahms, which calculates the probability of a specific phenophase (in this case, flower buds) occurring in each month of the year (1, January, to 12, December): A, *Ruellia incompta* (unimodal phenology); B, *Emmotum nitens* (bimodal phenology); and C, *Dalechampia caperonioides* (multimodal phenology). The *y*-axis shows the number of unique records (each a unique combination of phenological event, month and year). The month with the highest peak in the graph may differ from the PPI-determined maximum probability month (circled); see text for explanation.

    For species with a unimodal strategy, the month with the highest peak in the graph usually coincides with the month with the highest PPI score (see **Figure 1**A). However, for species with bimodal or multimodal distributions (see **Figure 1**B and **Figure 1**C, respectively), the month with the highest peak in the graph does not always coincide with the month with the highest PPI score. In species with bimodal or multimodal phenological patterns, PPI scores are more strongly influenced by the neighbouring months, so the maximum probability month is not always the month with the most unique records.

    Mathematically, predictability can be broken down into constancy × contingency (Colwell, 1974). A perfectly *constant* event is invariable throughout the year (e.g. day length, 24 h). A perfectly *contingent* event has a fixed pattern (e.g. the once-yearly occurrence of Christmas, always on 25 December). Biologically, as applied to plant phenology, PPI score is influenced by three taxon-specific parameters: 1) phenophase length for an individual plant; 2) synchrony between individual plants; and 3) year-to-year variability in period, length and synchrony. The minimum PPI score (i.e. 0.02) may be interpreted as the lowest level at which a phenophase can be observed; when PPI = 0 (when the number of unique records [*f*] = 0), the phenological event cannot be predicted because it cannot be observed. The maximum PPI score (i.e. 1) indicates that all the phenological events in the database are in the same month in every year for which a unique record exists. Random modelling has shown that PPI scores are reliable provided the database contains more than 50 unique records for the relevant phenophase (Proença *et al.*, 2012).

    For each of the species whose data were used in the present study, the PPI tool incorporated into Brahms version 7.1 was used to determine the maximum probability

month for flower buds, flowers and fruits. Additionally, for each of these phenophases, graphs were generated showing PPI scores for all the target species plotted against month. Most of the data were from the four Distrito Federal herbaria: CEN (Embrapa Recursos Genéticos e Biotecnologia), HEPH (Jardim Botânico de Brasília), IBGE (Reserva Ecológica do IBGE) and UB (Universidade de Brasília). Fewer than 10% of the data were from the published studies (Barros & Caldas, 1980; Gribel, 1986; Oliveira, 1991; Barros, 1992; Oliveira & Gibbs, 1994; Silva, 1995; Barros, 1996; Felfili *et al.*, 1999; Ribeiro, 2003; Munhoz & Felfili, 2005; Kutschenko, 2009; Ramos, 2010; Alves & Silva, 2013), dated photographs, or other herbarium information available from *speciesLink* (2014).

Testing strategy

The performance of PPI was tested using two different methods: phenological monitoring (method 1) and spot-check field excursions to record phenological state (method 2). These were carried out over the long- and short-term, respectively.

Method 1. ArcGIS version 9.3 (ArcGIS, 2011) was used to divide vegetation maps of ESECAE and PNB into 10" × 10" (approximately 25 × 25 cm) grids, to each of which a unique number was assigned. Randomizer Research version 3.0 (Urbaniak & Plous, 2011) was then used to draw grid numbers at random until all habitat types had been drawn once; if the same habitat was drawn more than once, a new number was drawn. In each grid whose number had been drawn, a transect of c.500 m was marked. Individual plants of 21 of the 28 target species (see Table 1) growing up to 10 m from each transect were tagged.

   The phenology of these tagged individual plants was then monitored by the first or second author over the course of 12 months (January to December 2012). The presence of flower buds, flowers and fruits was recorded every 2 weeks, generating approximately 1440 records of phenological events per species (three phenophases × 20 tagged plants [mean] × two observations per month × 12 months). Added to these data from 2012 were those for an additional seven species that also satisfied the species selection criteria. The same method as that used in 2012 had been used at IBGE to monitor their phenology between January and December 2001 (Lenza, 2005), generating approximately 864 records of phenological per species (three phenophases × 12 tagged plants [mean] × two observations per month × 12 months). For each of the 28 species in total, success was recorded if any of the tagged plants were found in a specific phenological state in the maximum probability month determined by PPI for that state.

Method 2. On 6 July 2014, two botanists (including one of the authors), both familiar with the target species, carried out two one-day spot-check field excursions to record the phenological states of tagged plants growing along trails in JBB and COUNB. Neither botanist had prior knowledge of whether the target species had previously been recorded from these areas; they knew only the general geographical distribution of each species.

The JBB trail runs through dense and typical cerrado and campo sujo and was followed for c.2.5 km. The COUNB trail runs through dense and sparse cerrado and was followed for c.1 km.

Results and discussion

For each of the target species, we aimed to find in the herbarium databases more than 50 unique records for each of the three phenophases. This was achieved for all 28 species for the flower bud and flower phenophases, and for all but three species for the fruit phenophase (Table 2).

The mature fruit phenophase was not analysed. This was due not to lack of specimens but rather to difficulties in determining the maturity of fruit preserved as herbarium material. Fully grown, immature dry fruits tend to open precociously during the drying process, and fully grown yet immature fleshy fruits are hard to distinguish from mature ones because they may differ only in subtle differences in colour and texture that are not apparent in dehydrated material.

The mean maximum monthly predictability scores (PPI $\bar{x}$) were similar for the three phenophases analysed: $\bar{x}$ [*flower bud*] = 0.12 ± σ = 0.10, $\bar{x}$ [*flower*] = 0.12 ± σ = 0.07, and $\bar{x}$ [*fruit*] = 0.10 ± σ = 0.10. This result suggests that there are no significant differences between these three phenophases in terms of predictability of their occurrence in the target species found in the study areas.

Method 1

With phenological monitoring, the likelihood of finding plants in a specific phenophase in the relevant PPI-determined maximum probability month was > 90%: 100% for flower buds, 92.6% for flowers and 95.8% for fruits. For *Davilla elliptica* A.St.-Hil. and *Diplusodon villosus* Pohl, no flowering individuals were found in May and April, their respective PPI-determined maximum probability months for this phenophase (see Table 2).

The failure to observe *Davilla elliptica* flowering in May, its maximum probability month (see Table 2), is attributable to it having one of the lowest PPI scores for the maximum probability month for this phenophase, ranking 23rd among the 28 species (Figure 2). Field phenological studies of *Davilla elliptica*, carried out in five different years, have shown that flowering in this species is prolonged but varies between years, with several interruptions and beginnings (Oliveira, 1991; Lenza, 2005; Kutschenko, 2009).

The failure to observe *Diplusodon villosus* flowering in April, its maximum probability month (see Table 2), is attributable to this species having apparently supra-annual flowering, that is, intervals of over 1 year between flowering episodes. When transects were set up in November 2011, several individuals of this species that had been tagged were observed to be in fruit. However, only 2 of the 11 tagged plants flowered during the study period

Table 2. Data generated by the Phenological Predictability Index (PPI) tool in Brahms for species of the Distrito Federal, Central Brazil

Class, order, family and species[a] (voucher no.[b])	Flower buds			Flowers			Fruits		
	f	M	PPI	*f*	M	PPI	*f*	M	PPI
COMMELINIDAE, Poales									
Poaceae									
Echinolaena inflexa (Poir.) Chase[c] (696)	68	2, 12	0.065	58	2	0.120	53	2	0.028
CORE EUDICOTS, Dilleniales									
Dilleniaceae									
Davilla elliptica A.St.-Hil.[d] (691)	81	5	0.036	73	5	0.073	76	6, 5	0.06
ROSIDAE I, Fabales									
Fabaceae									
Periandra mediterranea (Vell.) Taub.[c] (*Soares* 137, HEPH)	108	4, 2–5	0.03	121	4, 2–5	0.029	76	5	0.03
Chamaecrista conferta (Benth.) H.S.Irwin & Barneby[c] (*Nóbrega* 2152, HEPH)	60	6	0.167	67	6	0.164	57	6	0.142
Stryphnodendron adstringens (Mart.) Coville[d] (490)	(41)	NA	NA	70	9	0.152	84	6	0.028
Calliandra dysantha Benth.[c] (693)	85	6	0.109	124	6	0.100	50	9	0.172
ROSIDAE I, Rosales									
Moraceae									
Brosimum gaudichaudii Trécul[c] (708)	59	8, 9	0.086	53	9	0.140	63	9, 10	0.139
ROSIDAE I, Malpighiales									
Euphorbiaceae									
Dalechampia caperonioides Baill.[c] (697)	155	9	0.026	176	10	0.026	62	4, 11	0.023
Ochnaceae									
Ouratea hexasperma (A.St.-Hil.) Baill.[d] (699)	87	8, 9	0.118	95	8, 9	0.161	51	10	0.215
Malpighiaceae									
Byrsonima verbascifolia (L.) DC.[d] (500)	63	10	0.087	78	10	0.084	55	1	0.032
Banisteriopsis campestris (A.Juss.) Little[c] (637)	100	1, 12	0.033	102	1, 12	0.035	57	2	0.077
Salicaceae									
Casearia sylvestris Sw.[d] (702)	61	8	0.198	60	8	0.197	(31)	NA	NA
Caryocaraceae									
Caryocar brasiliense Cambess.[c] (701)	63	9	0.092	63	9	0.097	52	11	0.133
ROSIDAE II, Myrtales									
Lythraceae									
Diplusodon villosus Pohl[c] (*Zanatta & J.S. Silva* 1232)	84	3, 2–4	0.156	97	4	0.159	69	4	0.129
Vochysiaceae									
Vochysia elliptica Mart.[c] (*Haidar* 634, HEPH)	102	5	0.039	114	5	0.039	52	8	0.144
Myrtaceae									
Psidium firmum O.Berg[c] (684)	(43)	NA	NA	(47)	NA	NA	67	9	0.143

Class, order, family and species[a] (voucher no.[b])	Flower buds			Flowers			Fruits		
	f	M	PPI	*f*	M	PPI	*f*	M	PPI
Melastomataceae									
Miconia albicans (Sw.) Steud.[d] (686)	67	9	0.087	55	9	0.096	62	10, 12	0.078
ROSIDAE I, Sapindales									
Burseraceae									
Protium ovatum Engl.[c] (707)	52	5	0.121	51	6	0.147	103	7	0.031
Anacardiaceae									
Anacardium humile A.St.-Hil.[c] (689)	58	8, 9	0.276	70	9	0.172	52	9, 10	0.252
ASTERIDAE I, Ericales									
Styracaceae									
Styrax ferrugineus Nees. & Mart.[d] (690)	91	5, 6	0.073	101	5	0.077	54	8, 9	0.091
ASTERIDAE I, Metteniusales									
Metteniusaceae									
Emmotum nitens (Benth.) Miers[c] (566)	63	2, 3	0.069	62	3	0.147	101	11	0.025
ASTERIDAE I, Gentianales									
Rubiaceae									
Palicourea rigida Kunth[c] (687)	97	11, 10	0.039	93	11	0.082	56	2	0.035
ASTERIDAE I, Solanales									
Solanaceae									
Solanum subumbellatum Vell.[c] (709)	79	10	0.041	84	10	0.040	59	11	0.031
ASTERIDAE I, Lamiales									
Lamiaceae									
Hyptis villosa Pohl ex Benth.[c] (698)	50	3, 2	0.037	58	3, 2	0.074	(42)	NA	NA
Acanthaceae									
Ruellia incompta (Nees) Lindau[c] (694)	59	6	0.201	72	6	0.188	(31)	NA	NA
Bignoniaceae									
Jacaranda ulei Bureau & K.Schum.[c] (692)	58	10	0.338	67	10	0.100	90	10	0.028
ASTERIDAE II, Asterales									
Asteraceae									
Piptocarpha rotundifolia (Less.) Baker[c] (703)	57	12	0.038	54	12	0.040	52	5	0.027
Chresta sphaerocephala DC.[c] (*Martins* 1353, UB)	55	6	0.207	64	6	0.192	(19)	NA	NA

f, Number of unique records (in parentheses if < 50); M, PPI-determined maximum probability month(s), that is, the month(s) in which plants of each species were most likely to be found in a specific phenophase (1, January; 2, February; 3, March; etc.) (in cases in which the month with the highest PPI score differed from the month with the highest peak in the graph of PPI scores, both months are shown); NA, not analysed; PPI, PPI score.
[a] Presented in linear APG IV order (Angiosperm Phylogeny Group, 2016).
[b] *J.S. Silva* collections at UB, unless otherwise indicated.
[c] Data from phenological monitoring carried out in 2012 (see text for details).
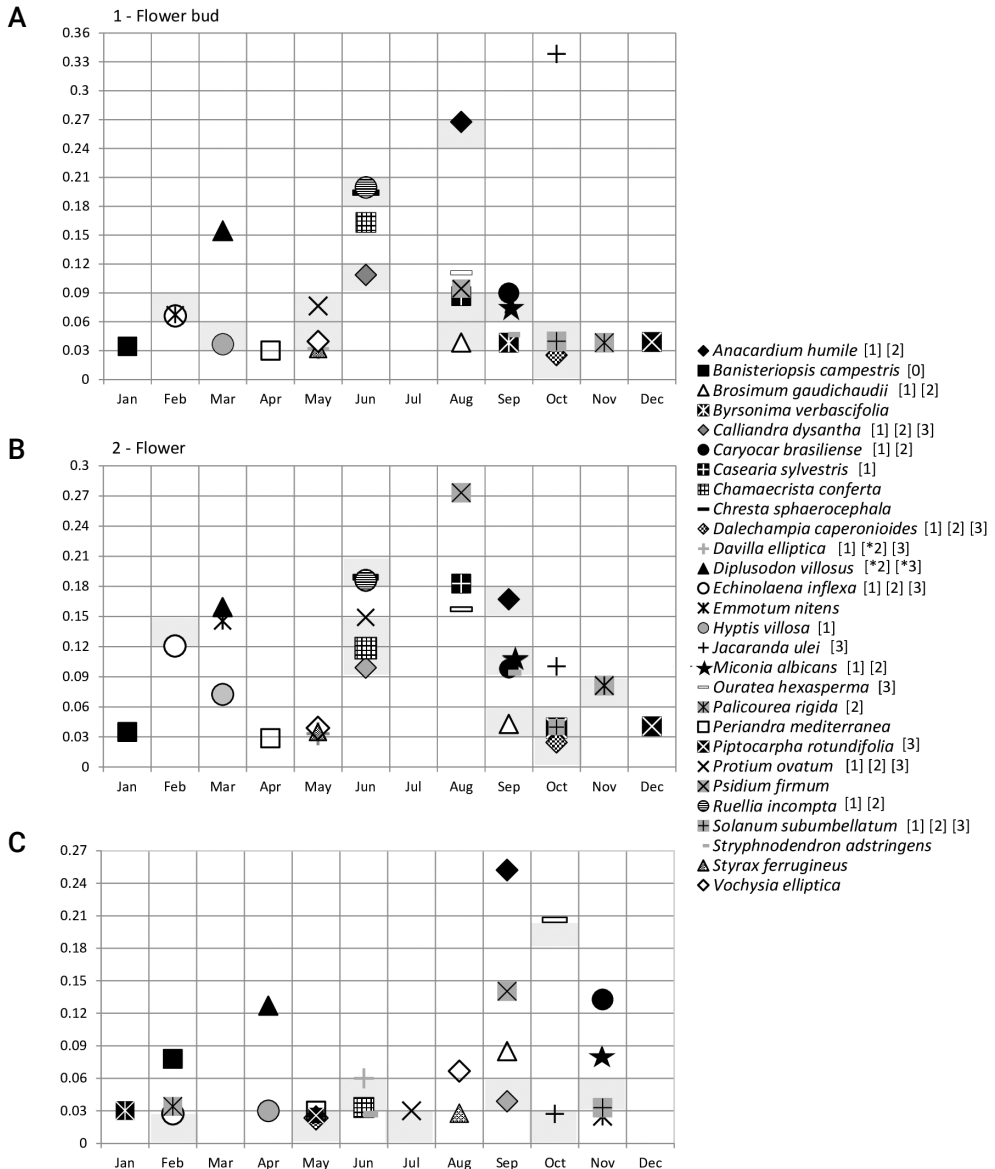[d] Data from phenological monitoring carried out in 2001 (Lenza, 2005).

Figure 2. Graphs generated by the Phenological Predictability Index (PPI) tool in Brahms, showing the PPI score for each of the 28 species whose data were used in the present study plotted against each species' PPI-determined maximum probability month for: A, flower buds; B, flowers; and C, fruits. PPI scores were calculated only for species with more than 50 unique records. Method 1 (phenological monitoring studies carried out in 2001 and 2012): [*], species for which only sterile plants were found in the maximum probability month. Method 2 (spot-check field excursions in July 2014 to record phenological state): species found [0], sterile; [1], with flower buds; [2], with open flowers; or [3], with fruits. Grey-shaded cells, at least one species found in its maximum probability month (July 2014 field excursions).

of January to December 2012. The hypothesis that *Diplusodon villosus* is a supra-annual flowerer is further supported by the findings of a 2-year phenological field study of the species in the Distrito Federal, Brazil (Barros, 1996). In that study, flowering occurred only in the second year; selected individual plants may have been reproductively immature in the first year. However, there is evidence that the plants whose data were used in the present study were mature, because fruits from previous flowering episodes were still attached to the tagged individuals. Furthermore, *Diplusodon villosus* was the only species without fruits in the maximum probability month for this phenophase, a consequence of most individuals not having flowered.

Method 2

July was the maximum probability month for flower buds or flowers for none of the target species, and the maximum probability month for fruits for only one species (see Table 2). Therefore, our field excursions provided us with the opportunity to learn how PPI would perform under challenging conditions. We expected to find fruits on individual plants of the species for which PPI determined July to be the maximum probability month for fruiting (i.e. *Protium ovatum* Engl.; see Table 2), as well as two classes of species: 1) those with high PPI scores for specific phenophases in the neighbouring months of June or August (particularly June, because the field excursions of method 2 were carried out in early July); and 2) those with low PPI scores for specific phenophases due to the occurrence of year-round reproductive episodes or multiple reproductive episodes throughout the year.

A total of 18 (64%) of the 28 target species were found on the field excursions undertaken in early July 2014. Results for flower buds, flowers and fruits were interpreted separately. As predicted, most species found had either high PPI scores for neighbouring months (i.e. May and June, or August and September) or low PPI scores for more distant months.

- Thirteen species were found in the flower bud phenophase. For three of these, the peak PPI scores for flower buds were for the months closest to July (i.e. June or August); the remaining 10 had low PPI scores for this phenophase (i.e. in the lower quarter of the range; see Figure 2A).
- Eleven species were found in the flower phenophase. Their peak PPI scores for flowering were for June for three of these species and for September for another; all the rest (except *Echinolaena inflexa* Poir. Chase) had low PPI scores for this phenophase (i.e. in the lower third of the range; see Figure 2B).
- Nine species were found in the fruit phenophase. For one, the peak PPI score for fruiting was for July; all the rest (except *Ouratea hexasperma* (A.St.-Hil.) Baill.) had low PPI scores for this phenophase (i.e. in the lower quarter of the range; see Figure 2C).

Palicourea rigida Kunth was found with flowers, despite the PPI score for July being low for this species (the maximum PPI score for flowers, 0.082, being for November; see

Table 2). The unexpected flowering of *Palicourea rigida* in July is attributable to its being a heterostylous and bimodal species in which intense flowering, dominated by the pin morph, occurs during the rainy season, and a second, less intense flowering event, dominated by the thrum morph, occurs in mid-July, in the dry season; additionally, the pin morph was found at greater frequency in ESECAE (Silva, 1995; Machado *et al.*, 2010). It was the second flowering event that was recorded in our field excursion. PPI had predicted that flowering of *Palicourea rigida* would peak in November because most herbarium specimens of this species had been collected during the first flowering event. The less showy flowers of the second flowering event are presumably less likely to be collected.

A single *Echinolaena inflexa* individual was also found in flower, despite the PPI score for July being zero for this species (the maximum PPI score for flowers, 0.120, was for February; see Table 2). The presence of this flowering individual is attributable to the unexpected rains that had occurred shortly before. In two phenological field studies, this species had been found to flower during the rainy season (Almeida, 1995; Ramos, 2010); this finding is consistent with our finding of February being the PPI-determined maximum probability month for this phenophase. June 2014 was unusual in that it rained for 3 days (INMET, 2014); this occurrence, shortly before the July field excursions, may have triggered the unseasonal flowering of *Echinolaena inflexa* that we observed.

Conclusions

We conclude that by basing the timing of field excursions on PPI-determined maximum probability months for specific phenophases (based on robust PPI scores, i.e. scores calculated from more than 50 unique records), a > 90% likelihood of finding a given species in the desired phenophase can be expected. However, this PPI-based approach to maximising the success of field excursions may be inappropriate for species with supra-annual flowering or with low PPI scores (< 0.04) for the maximum probability month, which indicate high event contingency. For bimodal species, that is, those with two phenophase peaks that differ in intensity, PPI scores will indicate the month of highest phenophase intensity as the maximum probability month, because most herbarium specimens will have been collected during this phenophase peak, when individual plants of the species are most visible to collectors (Miller-Rushing *et al.*, 2004). Therefore, graphs generated by PPI should be examined for subsidiary peaks possibly indicating less intense phenophases.

It is worth noting that the herbarium records used in the present study to calculate PPI scores were less than 1° latitude × 1° longitude from the places where the plants were searched for, because it is well known that phenology varies geographically (Borchert, 1996; Menzel *et al.*, 2006; Giuliani *et al.*, 2014). As global phenological patterns such as climatic and photoperiod-induced fluctuations become better understood, correction for geographical variation may be possible in future versions of BRAHMS. If habitat destruction

continues at its present rate, finding rare and endangered species with fruits for propagation and *ex situ* conservation may increasingly be considered a priority.

In planning a field excursion in which general collecting for herbarium enrichment or floristic inventory are the aims, PPI can be used to identify the best time of the year to travel. Furthermore, in ecological studies PPI may also help identify times of peak fertility in a plant community, thereby increasing the likelihood of correct identification of specimens. Two factors must be considered in combination: 1) the months with the highest PPI scores, and 2) the concentration of unique records of relevant phenological events in those months. Obviously, this level of planning is worth while only if the reproductive activity of plants in the area to be visited varies significantly throughout the year.

Acknowledgements

References

Almeida SP. 1995. Grupos fenológicos da comunidade de gramíneas perenes de um campo cerrado no Distrito Federal, Brasil. Pesquisa Agropecuária Brasileira. 30(8):1067–1073.

Alves MVP, Silva JCS. 2013. Fenologia de *Emmotum nitens* (Benth.) Miers (Icacinaceae) na Reserva Ecológica Cerradão – Embrapa Cerrados, Planaltina – DF. Revista Verde. 8(1):125–131.

Angiosperm Phylogeny Group. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. Botanical Journal of the Linnean Society. 181(1):1–20.

Antunes NB, Ribeiro JF. 1999. Aspectos fenológicos de seis espécies vegetais em matas de galeria do Distrito Federal. Pesquisa Agropecuária Brasileira. 34(9):1517–1527.

ArcGIS. 2011. Software (version 9.3). http://www.esri.com/software/arcgis/index.html

Barros MAG. 1992. Fenologia da floração, estratégias reprodutivas e polinização de espécies simpátricas do gênero *Byrsonima* Rich (Malpighiaceae). Revista Brasileira de Biologia. 52(2):343–353.

Barros MAG. 1996. Biologia reprodutiva e polinização de espécies simpátricas de *Diplusodon* (Lythraceae). Acta Botanica Mexicana. 37:11–21.

Barros MAG, Caldas LS. 1980. Acompanhamento de eventos fenológicos apresentados por cinco gêneros nativos do cerrado (Brasília-DF). Brasil Florestal. 10(42):7–14.

Batalha MA. 2011. O cerrado não é um bioma. Biota Neotropica. 11(1):21–24.

Borchert R. 1996. Phenology and flowering periodicity of neotropical dry forest species: evidence from herbarium collections. Journal of Tropical Ecology. 12(1):65–80.

Boulter SL, Kitching RL, Howlett BG. 2006. Family, visitors and the weather: patterns of flowering in tropical rain forests of northern Australia. Journal of Ecology. 94(2):369–382.

BRAHMS documentation. 2012. Phenological Predictability Index in Brahms: A Tool for Herbarium-Based Phenological Studies. http://herbaria.plants.ox.ac.uk/bol/content/documentation/BRAHMS-PPI.pdf

Bretagnolle F, Thompson JD. 1995. Gametes with the somatic chromosome number: mechanisms of their formation and role in the evolution of autopolyploid plants. New Phytologist. 129(1):1–22.

Broadhurst LM, Lowe A, Coates DJ, Cunningham SA, McDonald M, Vesk PA, Yates C. 2008. Seed supply for broadscale restoration: maximizing evolutionary potential. Evolutionary Applications. 1:587–597.

Colwell RK. 1974. Predictability, constancy and contingency of periodic phenomena. Ecology. 55(5):1148–1153.

Costa IR, Forni-Martins E. 2007a. Karyotype analysis in South American species of Myrtaceae. Botanical Journal of the Linnean Society. 155(4):571–580.

Costa IR, Forni-Martins E. 2007b. Chromosome studies in *Gomidesia*, *Marlierea*, *Myrceugenia* and *Myrcia* (Myrtaceae, subtribe Myrciinae). Kew Bulletin. 62(1):113–118.

Coutinho LM. 2006. O conceito de bioma. Acta Botanica Brasilica. 20(1):1–11.

Felfili JM, Silva Júnior MC, Dias BJ, Rezende AV. 1999. Estudo fenológico de *Stryphnodendron adstringens* (Mart.) Coville no cerrado *sensu stricto* da Fazenda Água Limpa no Distrito Federal, Brasil. Revista Brasileira de Botânica. 22(1). https://doi.org/10.1590/s0100-84041999000100011

Filer DL. 2010. Botanical Research and Herbarium Management System – BRAHMS: Training Guide with Worked Examples. http://herbaria.plants.ox.ac.uk/bol/content/documentation/BRAHMStraining2010.pdf

Gimaret-Carpentier C, Dray S, Pascal JP. 2003. Broad-scale biodiversity pattern of the endemic tree flora of the Western Ghats (India) using canonical correlation analysis of herbarium records. Ecography. 26:429–444.

Giuliani AL, Kelly EF, Knapp AK. 2014. Geographic variation in growth and phenology of two dominant Central US grasses: consequences for climate change. Journal of Plant Ecology. 7(3):211–221.

Gomes SM, Kinoshita LS, Castro MM. 2008. Hemisincarpia e nectário apendicular enfocados através de ontogênese floral em *Mandevilla velame* (A.St.-Hil.) Pichon, Apocynoideae. Revista Brasileira de Botânica. 31(1):81–93.

Gribel R. 1986. Ecologia da polinização e da dispersão de Caryocar brasiliense (Caryocaraceae) na região do Distrito Federal. Masters dissertation, University of Brasília.

Harris KM, Foard MB, Marsico TD. 2012. Understanding floristic diversity through a database of Greene County specimens. Journal of the Arkansas Academy of Science. 66:94–105.

INMET [Instituto Nacional de Meteorologia]. 2014. Boletim Agroclimatológico Mensal de Junho, n. 6, v. 8. Coordenação-Geral de Desenvolvimento e Pesquisa, Brasília, Distrito Federal.

Klink CA, Machado RB. 2005. A conservação do Cerrado brasileiro. Megadiversidade. 1:147–155.

Kutschenko DC. 2009. Fenologia e caracterização de frutos e sementes de um cerrado sensu stricto, Jardim Botânico de Brasília, Distrito Federal, Brasil, com ênfase nas espécies com síndrome ornitocórica. Masters dissertation, University of Brasília.

Lenza E. 2005. Fenologia, demografia foliar e características foliares de espécies lenhosas em um cerrado sentido restrito no Distrito Federal e suas relações com as condições climáticas. Doctoral dissertation, University of Brasília.

Machado ADO, Silva AP, Consolaro H, Barros MA, Oliveira PE. 2010. Breeding biology and distyly in *Palicourea rigida* HB & K. (Rubiaceae) in the Cerrados of Central Brazil. Acta Botanica Brasilica. 24(3):686–696.

Malhado ACM, Malhi Y, Whittaker RJ, Ladle RJ, ter Steege H, Phillips OL, Butt N, Aragão LEOC, Quesada CA, Araujo-Murakami A, Arroyo L, Peacock J, Lopez-Gonzalez G, Baker TR, Anderson LO, Almeida S, Higuchi N, Killeen TJ, Monteagudo A, Neill D, Pitman N, Prieto A, Salomão RP, Vásquez-Martínez R, Laurance WF. 2009. Spatial trends in leaf size of Amazonian rainforest trees. Biogeosciences. 6:1563–1576.

Mendonça RC, Felfili JM, Walter BMT, Silva-Junior MC, Rezende AV, Filgueiras TS, Nogueira PE, Fagg CW. 2008. Flora vascular do Cerrado: checklist com 12.356 espécies. In: Sano SM, Almeida SP, Ribeiro JF, editors. Cerrado: Ecologia e Flora, 2nd edition. Planaltina: Embrapa–CPAC. pp. 417–1279.

Menzel A, Sparks TH, Estrella N, Roy DB. 2006. Altered geographic and temporal variability in phenology in response to climate change. Global Ecology and Biogeography. 15(5):498–504.

Miller-Rushing AJ, Primack D, Primack RB, Imbres C, Del Tredici P. 2004. Herbarium specimens as a novel tool for climate change research. Arnoldia. 63(2):26–32.

Miller-Rushing AJ, Primack RB, Primack D, Mukunda S. 2006. Photographs and herbarium specimens as tools to document phenological changes in response to global warming. American Journal of Botany. 93:1667–1674.

Munhoz CBR, Felfili JM. 2005. Fenologia do estrato herbáceo-subarbustivo de uma comunidade de campo sujo na Fazenda Água Limpa no Distrito Federal, Brasil. Acta Botanica Brasilica. 19(4):979–988.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J. 2000. Biodiversity hotspots for conservation priorities. Nature. 403:853–858.

Newstrom LE, Frankie GW, Baker HG. 1994. A new classification for plant phenology based on flowering patterns in lowland tropical rain forest at La Selva, Costa Rica. Biotropica. 26(2):141–159.

Oliveira PEAM. 1991. The pollination and reproductive biology of a cerrado woody community in Brazil. Doctoral dissertation, University of St Andrews.

Oliveira P, Gibbs P. 1994. Pollination biology and breeding systems of six *Vochysia* species (Vochysiaceae) in Central Brazil. Journal of Tropical Ecology. 10:509–522.

Pinheiro MHO. 2013. Fenologia reprodutiva de comunidade savânica através de acervo de herbário. Brazilian Geographical Journal. 4:233–254.

Proença CEB, Filer D, Lenza E, Silva JS, Harris SA. 2012. Phenological Predictability Index in Brahms: a tool for herbarium-based phenological studies. Ecography. 35:289–293.

Ramos DM. 2010. Comportamento fenológico de gramíneas em um campo sujo de Cerrado: da indução de floração à emergência de plântulas. Masters dissertation, University of Brasília.

Ribeiro MF. 2003. Biologia reprodutiva e polinização de *Jacaranda ulei* Bureau and K. Schum. (Bignoniaceae). Masters dissertation, University of Brasília.

Silva AP. 1995. Biologia reprodutiva e polinização de *Palicourea rigida* H.B.K. (Rubiaceae). Masters dissertation, University of Brasília.

Silva WLS, Gurgel ESC, Santos JUM, Silva MF. 2013. Inventário e distribuição geográfica de Leguminosae no arquipélago de Marajó, PA, Brasil. Hoehnea. 40(4):627–647.

Simon MF, Proença C. 2000. Phytogeographic patterns of *Mimosa* (Mimosoideae, Leguminosae) in the *Cerrado* biome of Brazil: an indicator genus of high-altitude centers of endemism? Biological Conservation. 96(3):279–296.

Smith GF, Steenkamp Y, Klopper RR, Siebert SJ, Arnold TH. 2003. The price of collecting life: overcoming the challenges involved in computerizing herbarium specimens. Nature. 422:375–376.

speciesLink. 2014. Distributed Information System for Biological Collections. http://www.splink.cria.org.br [Accessed August 2014.]

Tannus JLS, Assis MA, Morellato LPC. 2006. Fenologia reprodutiva em campo sujo e campo úmido numa área de Cerrado no sudeste do Brasil, Itirapina-SP. Biota Neotropica. 6(3). https://doi.org/10.1590/s1676-06032006000300008

UNESCO. 2002. Vegetação do Distrito Federal – Tempo e Espaço. Brasília: UNESCO Brasil.

Urbaniak GC, Plous S. 2011. Research Randomizer (version 3.0). https://www.randomizer.org/

Van Hengstum T, Lachmuth S, Oostermeijer JGB, den Nijs H(J)CM, Miermans PG, van Tienderen PH. 2012. Human-induced hybridization among congeneric endemic plants on Tenerife, Canary Islands. Plant Systematics and Evolution. 298:1119–1131.

Vasconcelos TNC, Silva JS, Proença CEB. 2012. Testando a função PPI: análise de previsibilidade fenológica utilizando dados de herbário. Heringeriana. 6(1):66–69.

Vazquez-Yanes C, Orozco-Segovia A. 1993. Patterns of seed longevity and germination in the tropical rainforest. Annual Review of Ecology and Systematics. 24:69–87.

Willan RL. 1985. A Guide to Forest Seed Handling – With Special Reference to the Tropics. FAO Forestry Paper no. 20/2. Rome: Food and Agriculture Organization.

Wishnie MH, Dent DH, Mariscal E, Deago J, Cedeño N, Ibarra D, Condit R, Ashton PMS. 2007. Initial performance and reforestation potential of 24 tropical tree species planted across a precipitation gradient in the Republic of Panama. Forest Ecology and Management. 243:39–49.

Yost JM, Sweeney PW, Gilbert E, Nelson G, Guralnick R, Gallinak AS, Ellwood ER, Rossington N, Willis CG, Blum SD, Wallis RL, Haston EM, Denslow MW, Zohner CM, Morris AB, Stucky BJ, Carter JR, Baxter DG, Bolmgren K, Denny EG, Dean E, Pearson KD, Davis CC, Mishler BD, Soltis PS, Mazer SJ. 2018. Digitization protocol for scoring reproductive phenology from herbarium specimens of seed plants. Applied Plant Science. 6(2):e1022.