

Manual de uso de la herramienta de extracción de información en literatura

Esta herramienta está siendo diseñada por el equipo informático, en colaboración con los expertos en abejas nativas, plantas hospederas, e interacciones abejas nativas - plantas, también del proyecto.

Su objetivo principal es simplificar la extracción de la información provenientes de distintas fuentes bibliográficas (artículos científicos, documentos, tesis, etc.).

A continuación se describen los principales aspectos.

Identificación de taxones

Son paquetes de archivos generados automáticamente que corresponden a los nombres de taxones obtenidos de una referencia bibliográfica (e.g., artículo científico, tesis). Muestran ocurrencias individuales, globales y la señalización visual (*highlight*) dentro del documento PDF.

Se generan utilizando 2 fuentes de información:

1. [El catálogo de especies de CONABIO](#). Para generar anotaciones acotadas a un subconjunto de especies.
2. [El catálogo global de nombres de especies GNRD](#). Que entrega un conjunto complementario de nombres de especies no tan estricto y que <<puede>> ser de utilidad.

La descripción técnica de los procesos se encuentra en [el siguiente enlace \(aún en construcción\)](#).

¿Para que sirve?

Su propósito es reducir el tiempo que una persona invierte en analizar el contexto de un taxón mencionado en una referencia bibliográfica.

Sus objetivos son:

1. Identificar y señalar los taxones en el documento.
2. Mostrar si es posible hacer un mapeo hacia un catálogo.
3. Proveer documentos intermedios que reporten los nombres de taxones.
4. Proveer una interfaz sencilla para dar retroalimentación a la metodología diseñada.

¿Como se usa?

Hay 3 archivos resultantes que presentan los taxones obtenidos. A describir:

1. **ocurrences**

Contiene las columnas que indican la posición dentro del documento (**page**, **ordinal**) de los matches **annotation-verbatim** y la fuente de la anotación (**ann_type**).

Este archivo está destinado ayudar en la interpretación del documento anotado resultante (**ID_DOCUMENTO.pdf**)

Columnas e información que representan

annotation: Anotación resultante

verbatim: Texto original en el documento

page: Página en donde está la anotación (*en índice 0*)

ordinal: Párrafo de la página `page` en donde está la anotación

most_probable_parents: *Si la anotación esta abreviada.* Taxón superior más probable

all_parents: *Si la anotación esta abreviada.* Taxones superiores relacionados **es_util.** Columna para indicar si la anotación obtenida es de utilidad.

parent_correcto: *Si la anotación esta abreviada y la columna most_probable_parents es incorrecta.* Columna para indicar el taxón superior relacionado.

Posición de las anotaciones.

(**page** y **ordinal**) indican la posición dentro del documento de una ocurrencia individual de una anotación (**annotation**) y el texto original (**verbatim**), así como la fuente de la anotación (**ann_type**).

Para ocurrencias abreviadas, del tipo *I. simulans*, además se muestran las columnas **most_probable_parents**, que indican cual es el taxón superior más probable de acuerdo al análisis del texto y, **all_parents**, que indican todos los taxones superiores que “podrían ser” los relacionados en el texto.

Ejemplo

Anotación: *I. simulans*

Taxón más probable: *leucophyllum simulans*

También podrían ser <<pero menos probables>>: *lycieae simulans, Ionicera simulans, litomegachile simulans, lantana simulans, ...*

Tipo de las anotaciones

Pueden identificarse por la columna **ann_type**, y corresponden a un color dentro de ``ID_DOCUMENTO.pdf``.

0. Son anotaciones acotadas al catálogo. Son las más estrictas y, potencialmente, las más confiables. Señaladas en color **AZUL** en ID_DOCUMENTO.
2. Proviene de [GNRD](#), se pueden considerar como un conjunto adicional de anotaciones de taxones que no están dentro del catálogo estricto. Sin embargo, son más fuzzies y la posición no es siempre la correcta (varían a lo máximo 1 o 2 ngramas de la posición actual). Señaladas en color **ROJO** en ID_DOCUMENTO.
3. Son las anotaciones menos confiables. Siempre son de 1-gram y también contienen palabras de vocabulario común. Véase <<sección>>. Señaladas en color **AMARILLO** en ID_DOCUMENTO

2. Documento

Es el documento anotado resultante, muestra las anotaciones y su tipo. Corresponde a los resultados en **taxa**.

Para identificar una anotación dentro del documento nos referimos a las columnas del documento de *ocurrences* **annotation-verbatim, page, ordinal, ann_type**,

Ejemplo 1: Anotación proveniente del catálogo de CONABIO

annotation	asclepias angustifolia
verbatim	asclepias angustifolia
page	1
ordinal	9
ann_type	0
most_probable_parents	—
all_parents	—

shape floral traits, acting as a barrier to interspecies gene flow in sympatric milkweeds (Kephart & Theiss 2003). However, certain milkweeds, e.g., *A. incarnata* appear to be less reliant on specialization and are successfully pollinated by a broad diversity of insects (Ivey et al. 2003). For most North American *Asclepias*, these traits and the insects they attract remain unknown.

This study addressed the following questions:

1. What insects are nectarivorous, phytophagous, or predatory in a native population of *A. angustifolia* in Madrean Pine-Oak Woodland, and under cultivation in nearby gardens?
2. What is its pollinator spectrum in Madrean Pine-Oak Woodland and nearby gardens and how do they differ?
3. Would planting *A. angustifolia* in butterfly gardens and habitat restoration sites provide a food resource for a diversity of insects, augmenting the plant's other values?

MATERIALS AND METHODS

Study species **Anotación**

Asclepias angustifolia Schweigger (Arizona milkweed, Spanish: *algodoncillo californico*) is a member of the clade **Proveniente del catálogo (Color azul)** (c. 8% of the c. 130 New sp.). Widespread in the Sierra ... ranges northward to three counties in south-eastern Arizona, where it has isolated populations in seven mountain ranges and is considered rare (SEINet 2018, Fishbein in prep., Woodson 1954). Occupying elevations of c. 1,066–2,133 m, it inhabits riparian areas in montane canyons, floodplain meadows, and wetland edges, growing as a low, bushy, perennial forb reaching c. 70 cm in height (Nabhan et al. 2015, SEINet

leaves, and stem (Aphididae), so (Formicidae), in species were eaten spiders (Thomisidae), butterflies, bees, enumerated. Insects without feeding, not considered, plants after hum could not be identified

TABLE I. Visits/site, total hours, percentage of total Garden Canyon, Ft. Canyon (RAM), B Casa de San Pedro

Year	Totals
2018	Visits hrs
2019	Visits hrs
	Σ Visits
	Σ Hours
	Σ Species
	Σ Uniq
	% Σ Hor
	% Σ Speu

Ejemplo 2: Anotación abreviada del catálogo de CONABIO

annotation	a. fascicularis
verbatim	a. fascicularis
page	4
ordinal	24
ann_type	0
most_probable_parents	asclepias
all_parents	['augusta', 'apidae', 'anothea', 'augochlorini', 'apocynaceae', . . .]

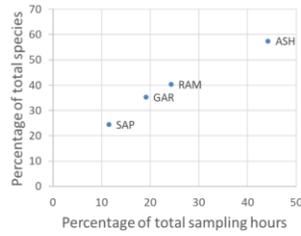
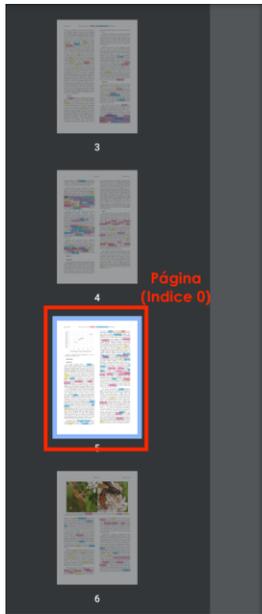


FIGURE 1. Percentage of total sampling hours (x axis) vs percentage of total species (y axis)

DISCUSSION

Pollination

The highly specialized flowers of *Asclepias* are characterized by a column-like gynoecium composed of fused pollen- and carpel-bearing structures, a short to tall whorl of cup-shaped, petal-like hoods containing rich nectar, and pollen grains aggregated into paired sacs (pollinia) that are joined by translator arms and a grooved corpusculum that attaches to an insect visitor. Unique to the family Apocynaceae, such specializations might suggest visitation by a guild of evolutionarily-dedicated insects; however, the pollinia of *Asclepias* are distributed by a characteristically diverse association of nectarivorous insects, primarily

and wasps (i.e., *Bombus*, *Apis*, *Xylocopa*, *Sphex*) were the most frequent pollinators; all were present on *A. verticillata*, whose flowers resemble *A. angustifolia*. Butterflies were important visitors, but only half as common on the smaller, pale flowers of *A. verticillata*. All three species showed some degree of specialization favouring long-tongued Hymenoptera and Lepidoptera. Owing to differences (e.g., search image, response to floral chemistry) pollinators discriminated between milkweeds, occurring in varying percentages and potentially influencing gene flow. Other factors affecting pre- and postzygotic isolation included position of pollinia on insects' legs, size of flowers, pollinia and pollen tube incompatibility, flowering phenology, distance and duration of flights between blossoms, and presence of Anotación sympatric milkweeds. Hatfield & Kephart (2004) reported pollinators in sympatric Oregon populations of *A. fascicularis* and *A. speciosa*. Again, bees (*Apis*, *Bombus*, *Halictus*) and wasps (Vespidae) were the Posibles taxones superiores de *A. en most_prob_parents* isolation was maintained by the smaller stigmatic chambers of *A. fascicularis*, and differences in flowering phenology; however, there is a tantalizingly-brief mention of insects (selectively?) not flying between the two milkweeds in mixed populations. On *A. verticillata* growing in Illinois, the majority of pollinators (80–93%) were Hymenoptera and Lepidoptera, including a significant contribution (2–9%) from introduced *Apis mellifera* and cabbage butterfly, *Pieris rapae* (Willson et al. 1979). Other prominent pollinators were wasps (Sphecidae, Vespidae, and Tiphidae), and moths (Erebidae, Noctuidae). In south-eastern Arizona, long-tongued insects were prominent visitors at *A. tuberosa* (Fishbein & Venable 1996); certain taxa (e.g., *Apis*, *Bombus*, and medium Lepidoptera)

Ejemplo 3: Anotación del catálogo global GNRD

annotation	Melacoryphus lateralis
verbatim	Melacoryphus lateralis
page	7
ordinal	9
ann_type	2
most_probable_parents	—
all_parents	—



Thrips, likely in the family Phlaeothripidae, were encountered three times (GAR, RAM, ASH). Beating flower clusters to dislodge smaller cryptic insects would likely have generated additional records.

Hemiptera: True bugs

Fourteen families of Hemiptera representing c. 37 species (including two kinds of unidentified nymphs) were recorded. Numerically and visually prominent, were the seed bugs (Lygaeidae). Several genera (milkweed bugs), are usually adorned in bold red or orange and black aposematic (warning) patterns advertising chemical defences provided by toxic cardenolides ingested from their host plants-with which they may have coevolved (Agrawal 2017). Lygaeid nymphs were often present by the hundreds, their feeding activities damaging pods and seeds and compromising a plant's reproductive potential. Adults, often present by the dozens (Fig. 4A), frequently took nectar at flowers. Three species carried pollinia and two more were likely pollinators. Adults of *Lygaeus kalmii* and *O. fasciatus* sometimes bore pollinia. Found on 85 and 86 site visits respectively, the **Anotación** (Ollerton et al. 2019). Occasionally, *Melocoryphus lateralis* was numerous, at times bearing pollinia. *Nemobius f. fulvipes*, *O. fasciatus*, and oleander aphid, **Del catálogo global GNRD** *Aphis nerii*, were often present by **(Color rojo)** from the milkweed's phloem. This clustering aphid specializes on Apocynaceae; its yellow coloration is considered aposematic (Agrawal 2017). Other Hemiptera present in small numbers were predators (Geocoridae, Nabidae, Reduviidae), fed on nectar and plant tissues (Cicadellidae, Alydidae, Largidae, Coreidae, Rhopalidae, Thyreocoridae); or both (Berytidae, Miridae, Pentatomidae).

too small to remove pollinia from flowers (Chrysomelidae and Melyridae). Representative families transported pollinia. In Southeastern Arizona (Fig. 4B) and net-winged (Lycaidae), are diverse families whose members number by the hundreds, nearly obscuring the flowers upon cluster during feeding or mating aggregations. They are known pollinators of many plants. They are chemically with noxious tastes and odours characteristic aposematic colours and patterns such as Batesian and Müllerian mimicry complexes with other families, and occasionally other families or order mimicking moths (Eisner 2008, Linsley et al. 1979). The seven *Chauliognathus* soldier beetles and two *Lycus* net-winged beetles carried pollinia. During *Loricipes* was especially abundant, blanketing some (RAM especially) where many individuals carried several pair of pollinia, suggesting it and other members of Lycaidae are potentially important pollinators (Fig. 4C). Additional observations of less-numerous species members of both families may eventually be of pollinia. Three species of scarab beetles were not insects whose slow, lumbering movements suggest they be effective in gathering and transporting pollen. The Western rose chafer, *Macrodactylus* (Melolonthinae), was noted carrying pollinia (?). Longhorn beetles (Cerambycidae), were present. *Rhopalophora meesekei* is aposematically coloured and black. Two milkweed beetles, *Tetraope* (widespread) and *T. linsleyi* (once, ASH) were. Like the milkweed bugs mentioned previously *Tetraopes* belong to the guild of milkweed-feeding way of convergent evolution, they have characteristics of the guild: aposematic coloration

Ejemplo 4: Anotación fuzzy de ambos catálogos.

annotation	pollinia
verbatim	pollinia.
page	9
ordinal	2
ann_type	3
most_probable_parents	—
all_parents	—



field or from photos, and usually too small to pollinate *Asclepias*. Table 1 lists seven identified only to subgenus; twice as many may have been present. Allowing for the quantity of unidentified species, the overall number of bees encountered, at least 13-17 species, was similar between sites.

The introduced Western honey bee, *Apis mellifera*, is a dominant pollinator of many *Asclepias*, e.g., *A. fascicularis*, *A. speciosa*, *tuberosa*, *syriaca*, and *A. verticillata*, and its importance as a pollinator milkweeds is significant (Robertson 1887a, Willson et al. 1979, Fishbein & Venable 1996, Ollerton & Liede 1997, Hatfield & Kephart 2003, Kephart & Theiss 2003). Entrapment in the anther wings of *Asclepias* flowers is often documented (e.g., Robertson 1887a, 1891, G. Thomas 2020, in litt.); Frost (1965) noted struggling bees generally escaped from flowers, often sacrificing leg segments, and Willson et al. (2021) found overburdened *Apis* may remove accumulated pollinia. In this study, where both wild nests and domesticated hives are present, honey bees, including several distinct individuals, were observed on *A. angustifolia* (Color amarillo) visits. Nearly every bee carried pollinia on its tarsomeres, tarsal claws, or (infrequently) tibiae. Some individuals were burdened with 25 or more pairs of pollinia or corpuscula with empty translator arms, underscoring their importance as pollinators of the small-flowered *A. angustifolia*.

Insects removing larger numbers of pollinia are also more likely to insert more pollinia per flower visited (Ivey et al. 2003). Here, larger bees (e.g., *Aphis*, *Anthophora*, *Melissodes*, *Tachusa*, *Megachile*) and wasps (e.g., *Prionyx*, *Isodontia*, *Colpa*, *Delis*, *Myzinium*, *Polistes*, and *Scolia*) carried the largest loads of pollinia. Individuals were frequently noted to feed on the same plant for long periods (c. 15–30 + min.), which may account for lost pollinia sitting atop flowers. Taken to an extreme, long feeding visits may increase the likelihood of self-pollination (i.e., autogamy or geitonogamy).

the lower elev grassland surr Jacobson 2000 *angustifolia* w Higher, in GA on the larger, y nearby *A. an Xylocopa pre tuberosa* (*Apis* introduced *C. petiolaris*, *J. Plectrocephalus* (Asteraceae) and introduced introduced *Sa.* and introduced *Leucophyllum* chemical comp present only ; ignored or it preferentially ; e.g., cuckoo ; *Anthophora*, a *Pseudopanurg* study area, but were some was

Lepidoptera

Ollerton & survey of pol pollination as (including *As* phenomenon. abundant nect certainly have their diverse as

3. template

Este es un documento intermedio que será complementado por medio de las anotaciones del usuario y contiene las ocurrencias globales del documento. Tiene como fin construir [un template de reporte como el mostrado](#).

¿Por que es un reporte intermedio? Si existiera la premisa de que todas las anotaciones entregadas son útiles y todos los parents son correctos, se pudiera implementar la construcción automática de este template. En la práctica este escenario todavía no es factible.

Premisas

Los dos procesos principales, marcar una especie y mapearla a un catálogo, parten de lo siguiente.

1. Para las anotaciones del catálogo de CONABIO. Las anotaciones pueden ser asociadas a un ID de categoría, es decir, para generarse previamente debieron existir y mapearse con el texto de la referencia.

¿Como se puede dar la retroalimentación?

Normalmente un proceso de aprendizaje automático necesita retroalimentación del usuario. Esto se obtiene con las columnas **es_util** y **parent_correcto** del documento **ocurrences**.

- Si la anotación es de utilidad la columna **es_util**, para ese renglón, se marca con 1.
- En caso de que la anotación provenga de un especie abreviada. Se sugerirá un taxon superior que sea lo más probable, de acuerdo a la metodología, en **most_probable_parents**. En caso de que esta sugerencia no sea correcta, se agregará el nombre del taxón que lo sea utilizando **all_parents**.