Chapter 2

Lepidoptera Collection Curation and Data Management

Jurate De Prins

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.70925

Abstract

The collections of Lepidoptera often serve as foundational basis for a wide range of biological, ecological, and climate science disciplines. Species identification and higher taxa delimitation based on collection specimens and especially, on types test scientific hypotheses, provide multiple types of evidence for a broad range of users. Curation and data management approaches applied in Lepidoptera collections benefit greatly from many newly developed information techniques, which link and integrate data. Mostly attention is focused on clean verified collection and taxonomic literature mining data to obtain correct species-group and higher taxa names, as well as reliable data on the distribution of Lepidoptera and their trophic interactions. Collection creation and management became a subject of natural sciences itself. The chapter provides a historic overview on collection creation and curation together with a short discussion on collection goals and purposes. The creation of a virtual collection based on interlinked data is emphasized. Information science and data management tools became very important in Lepidoptera collection curation. The complexity of techniques and computing tools used in taxonomy and the increase in the amount of data that can be obtained by collection-based disciplines make it necessary to automate data gathering, manipulation, analysis, and visualization processes.

Keywords: integrated collection, virtual collection, collection management tools, taxonomic text mining, data mining, web-based platforms, online catalogs

1. Introduction

The diversity of Lepidoptera is one of the most fascinating subjects of biology. Evolution, natural selection, and many other biotic and abiotic factors have produced different species of butterflies and moths and the speciation process is going on continuously. Present studies on Lepidoptera embrace many aspects on their function within the communities of plants and animals and a lot of different inter-relational processes that affect Lepidoptera. There are

about 157,000 species of butterflies and moths currently described [1–3], in 135 families and 45 superfamilies [2]. Lepidoptera are a globally distributed, widely recognizable, and admired order of insects. Lepidoptera comprise ca. 10% of the total amount of described species of living organisms [3]. Lepidoptera are common in smaller or larger institutional collections and they are disproportionally abundant in private collections. Despite their popularity as one of the best known and most collected of all insect orders [4], there are no exact data available on how many species or specimens are deposited in natural history collections. At present, we know that about 17 million lepidopteran specimens are deposited in the collections of North America [5] and ca. 80% of all described Lepidoptera taxa are deposited in 60 European repositories [6], while the representatives of more than 38,000 species of moths described from the Afrotropical region are deposited in 158 natural history collections all over the world [7]. A rough estimation of the total Lepidoptera specimens in the depositories worldwide could be about 10% from the estimated 2.5 billion of natural history collection specimens [8–10].

Lepidoptera specimens in the collections of natural history document the present and historic delineation of species and higher taxa concepts which represent natural entities resulting from the differentiation of lineages through speciation with constantly changing boundaries [11]. In a collection, we deal with lepidopteran specimens sampled across vast geographical areas and through time [12, 13]. Another important aspect of the nomenclature of species is that it is based on type specimens, which means that any species name in Lepidoptera is eternally linked with the name-bearing type specimen, deposited in a public institutional collection which from the moment of publication serves as an unambiguous reference to the species name. Finally, collection is an endless source for large scale data:

i. taxonomic/nomenclatorial information related to the names of taxa;

ii. geographical data related to the distribution areas and biotopes;

iii. morphological data related to the delineation of species and higher taxa;

iv. biological/ecological data include valuable information on feeding and behavior habits of species within the complexity of interrelations; and

v. historical data related to the personalities of collectors and their activities.

Thus, the specimen is a natural history collection and its associated data serve as one of the most direct and reliable sources to answer numerous biodiversity research questions.

Novel technology adopted in Lepidoptera collections allows us to explore new horizons in the productivity of handling specimens, the curation of the collection and it significantly increases the quality of generated taxonomic data. The boom of digitization of natural history collections in recent years and the fast development of curatorial software allows easy access to the multiple collections of Lepidoptera spread over the world, and increases the use and reuse of valuable biodiversity data stored in those collections by providing access to species/specimen data through the Internet [8]. These data, ready to be incorporated into different models and virtual simulations, become crucial evidence for decision-making facing global problems such as climate change, species decline, habitat loss, pest monitoring, biological disaster predictions,

http://dx.doi.org/10.5772/intechopen.70925

and threats to agriculture and public health [14]. The major force driving the acceleration of interest and the use of data from the Lepidoptera collections is the unlimited digital access and interlinked visualized information. The usual practice of physical visits to a museum, negotiate with a curator for the access to the specimens, obtain permission following numerous internal regulations, and the financial and administrative restrictions related to them cause a serious bottleneck for collection-based research. Only a very limited number of people were privileged to have access to the valuable specimens and their associated data. This situation caused a huge taxonomic impediment, which means that despite the fact that collections contain a lot of novel data that need to be studied and incorporated into a broader pattern of the natural history data pool, these data were frozen in the collections for decennia [15–17]. In addition to this, collection curation became a dead-end professional career and a small group of people professionally engaged could not handle the broad scale of activities related to Lepidoptera biodiversity and collection data management in particular.

In this chapter, I intend to show that accessibility to a collection through the means of what new technology offers is the key to resolve a long-standing taxonomic impediment. A responsible and safe management of digitally interchangeable data provides new ways of handling different aspects and suggests new solutions for the complexity of problems related to biodiversity. Working with digital data, mined from the literature and Lepidoptera collections, is not the replacement of traditional methods by new ones, but rather it is the processing of the extracted data which have to pass the quality control by vetting and scrutinizing these data. It is the straight forward way to achieve what society needs at the moment: stable, long-lasting taxonomic decisions based on repeatable evidences influencing many aspects of society life.

2. Curation strategy of Lepidoptera collection

2.1. Historic approach

For centuries, Lepidoptera collections were created as part of curiosity objects, as a certain art of nature showing the interest of the owner to the world. The specimens were grouped in a certain order according to their size, geographic area, taxonomic knowledge of the owner, or other criteria. In the beginning of the twentieth century, many individual Lepidoptera collections moved as donations to museums or were purchased by public museums forming a major part of the holdings which the museums possess today. The role of a museum curator also developed in the course of time from the concept of ownership of collection cabinets (note: even the titles like Keeper (Natural History Museum, London) or Beheerder (Naturalis, Leiden)) indicate the attitude of possession, keeping, and administrating to the concept of a curator who collected and added specimens to the collection supporting taxonomic publications. The collections became reflections of the personal research of a scholar. Many expeditions were conducted for the need of finding new taxa for their taxonomic/curatorial research, based upon personal interests [18–20]. The concept of developing a Lepidoptera collection as a whole structural institutional unit reflects the result of historical taxonomic work done serving an integral and important part of the ongoing educational and research programmes of today,

which was not developed at that time [6, 21]. There were no rules on priorities of collecting and processing the material. Curators were left to their own personal expertise. This collection keeping and managing style had a very huge negative effect on the next generation of taxonomists creating cross- and intra-institutional conflicts of interests. Further to this, many curators and researchers began to complain about the state of the current collections, the work involved to maintain and manage them [22–25].

At the same time, the curators became individuals collecting more and more specimens because biotopes and habitats were disappearing in an ever increasing speed. The collected specimens of numerous expeditions were stored without any processing and associated records. The primary core of a scholarly master of the collection disappeared. The taxonomic community at the end of the twentieth and the beginning of the twenty-first century is known as entering into the "crisis period" [26–28] and the collections seriously needed an effective "crisis management" strategy. Museum professionals started to regard the collections as a burden which consume finances and place, and not as strength of the museum. Because of this, the museums started hiring administrators and collection managers to better control the physical care over the collections. Individuals in these roles were preoccupied to create administrative rules and regulations, keeping track as where particular specimens were going. As a result, the clash between the collection curation and its administration increased.

2.2. Collection aims, objectives, and concepts

It is widely understood that habitat loss, land transformation, and habitat destruction are the major factors leading to the biodiversity loss. To understand how ecological systems change and interchange through space and time, reliable indicators are needed recording the state-of-the art of diversity, community composition within the framework of biotic and abiotic environmental conditions that facilitate these communities. Butterflies and moths possess multiple qualities that make them ideal as indicators. They are hyper-diverse, colorful, liked very much by many collectors of different ages, fill a wide range of functional roles like pollination, pest control, serving as prey in the complexity of food chains, nutrient cycles, have different population sizes, and life cycles, and respond rapidly to environmental changes [29, 30]. Lepidoptera communities in healthy habitats are often characterized by higher diversity, comprising a wide variety of taxa. Certain evolutional history or ecological aspects can be clarified by the presence or absence of specific taxa. For example, the presence of the micro-moth genus *Triberta*, De Prins et al. [31] might indicate the islands of a pre-glacial distribution pattern as well as the more recent colonization facilitated by human activities, since the genus is associated with the plant family Cistaceae [31]. Because of the overwhelming amount of information present already on Lepidoptera, the data must be sifted in order to identify those key species assemblages that can reveal the condition of the whole system and novel interactions.

The manual collection curation uses combined sources which are disparate and not always linked. At the heart of the collection curation, there are few but major ambitions and aims:

- to facilitate the access to the biodiversity resources and to present the physical illustration of the Lepidoptera biodiversity knowledge;

http://dx.doi.org/10.5772/intechopen.70925

- to offer solutions to group and specify the biodiversity;
- to facilitate the direct and specific communication of collection users by providing a gate to the solid biodiversity platform related to all aspects of Lepidoptera;
- to provide inter-operability to data and resources at the highest level at the same time recognizing the fact that a well-curated and managed collection facilitates to further discoveries and application of novel methods in many areas of life sciences and disciplines of biological education.

Two domains, climate science and Big Data—which are directly connected to the Lepidoptera collection, are experiencing unprecedented attention, financing, and exponential growth. The curators of the collections address the Big Data challenges associated with climate science while incorporating the collection data into bigger data packages [14]. The main focus conceptually defining a collection is moved towards data analysis, because the taxonomic knowledge stored in a collection and gained from its interaction with other life and earth sciences produces biological Big Data that ultimately influence societal benefits [32]. The main concepts of the Lepidoptera collection can be defined as follows:

- national collection—reflecting the lepidopteran species of a certain country;
- continental or regional collection—reflecting the lepidopteran species of a certain continent or bio-geographical region;
- biological lepidopteran collection—reflecting species assemblages for agricultural or pest control purposes;
- taxonomic lepidopteran collection—reflecting the taxonomic accomplishment of a smaller or bigger taxonomic group, for example, family or superfamily;
- historic collection, which is usually obtained from the famous individual lepidopterist as his/her life achievement and reflects his/her personal views in defining the species concept within a particular group of Lepidoptera and at a certain epoch of time, for example, the Linnaeus collection.

It is important to acquire one of these abovementioned concepts of collection specialization and execute it creating a collection which becomes a tool-as-a-service within the framework of ongoing projects. Within a certain but defined framework of natural history problematics, a Lepidoptera collection plays an important role. Nevertheless, it is seen as a part of the constellation of sources which are a prerequisite to delivering the analytics to climate science as a service to society. Also a Lepidoptera collection serves as an educational tool, certain live textbook for a broad scale of people across many layers of society. Both elements of the Lepidoptera collection, (1) a tool-as-a-service and (2) a-tool-as-an-educational mean, are essential in handling and managing the Lepidoptera collection because in the aggregation with other domains of natural history, the data extracted from the collection lead to the generativity and assembly of interlinks which is the key of solving many of the Big Data challenges in the domains related to natural history. The creation of a Lepidoptera collection is an example of a building up verified data for a very long-term usage and enables a multi-sided retrospective analysis for research and applied lepidopterology.

Here below, I present a step-by-step methodology for an efficient, fast, reliable, easy searchable Lepidoptera collection curation. This approach can be easily applied and repeated by any mobile curatorial team in any museum of natural history housing Lepidoptera collections. The author has long-term experience curating large Lepidoptera collections of different ages and of different state in the biggest museums. The concept of a well-curated collection is that not only a curator but any authorized person easily finds any specimen he/she is looking for. The curated and completed collection expresses the concept adopted by a museum.

2.3. Delineating, identifying and describing the taxa

There are hot debates over the species concept and its biological reality [33]. While now there is a consensus to view species as natural entities delineated by multiple evidences resulting from differentiation of lineages through speciation [11], species boundaries are often much harder to discern especially in a Lepidoptera collection because specimens are sampled across vast geographical areas and through time [4, 11, 13, 18, 21]. Furthermore, the order Lepidoptera challenges species boundaries for regular potential or often occurring interbreeding [34]. Despite this, a species is a central concept in biology, conservation, legislation, and trade and therefore it has a particular social relevance to the museum collection and to the society.

The handling of specimens with great care and the associated meticulous documentation assist to the accurate identification of species. However, the contrasting phenomena of "over-splitting" or "over-lumping" of lepidopteran species take their turn due to the application of different methods and approaches in species delineation as well as biological and non-biological reasons [11], such as:

- similar populations which have been considered as distinct species;
- species complexes with little genetic information;
- ecological forms of polymorphic species (altitudinal, latitudinal, habitat, host-plant associations);
- variable species;
- incomplete lineage sorting, introgression, hybridization;
- inaccurate taxonomy, misidentifications, labeling errors, etc.

Though historic collections were created mainly based on the biological species concept, we, contemporary curators, are dealing with more and more consensual species concepts. Some genetic factors might play a major role in the species delineation of insects, such as intraspecific variation and the extent of divergence between species [11]. While curating a collection and identifying the species curators deal with complexes of monophyletic and/or non-monophyletic (polyphyletic and paraphyletic) species complexes which are very often difficult to distinguish [11]. Summarizing, it needs to be stressed that preference for accurate curation and attentive identification is one of the most important approaches in the delimitation of species or higher taxa. At present more and more research is focused on studies which extend beyond the taxonomic species descriptions, which crosses the taxa among bio-kingdoms and

http://dx.doi.org/10.5772/intechopen.70925

bio-classes emphasizing the trophic chains of relationships between plants, Lepidoptera, and other orders of insects [21]. Though taxonomy represents one of the most classical fields of life sciences, the new technology provides unlimited possibilities within this discipline to embrace novelties and to combine multiple evidences into a holistic pattern for the delineation of taxa within the order Lepidoptera.

2.4. Transfer from physical specimen to digital data collection

Some present academic educational studies and projects on tropical Lepidoptera often involve the creation of a physical collection [35] which is designed for the primary purpose:

i. to correctly identify species;

ii. to preserve specimens;

iii. to document the information related to specimens and species; and

iv. to make specimens available for scientific studies.

The focus is on having a vouchered collection which facilitates the precise identification of species (**Figure 1**). Only when parts of the collection are properly curated the specimens can be quickly and efficiently digitized. This approach makes the physical specimen collection a ready-to-use tool for any project requesting digitized data within a short period of time and leaves time and space for research related to other collection items. The collection transfer to twenty-first century systematics happens in two phases:

Stage 1. Having the matrix of major families curated and temporarily leaving families of insects in their old place in the collection.

Stage 2. Within 2–4 days transfer the families according to the modern systematics, which is now stabilized after the publications of high standard molecular papers.

The further steps are related to the computing of the collection data since information science plays a more and more important role in the collection curation and data management. This computed collection-based biodiversity data presentation aims to study, design, and develop solutions to automate the steps in the data gathering and data curation in

Figure 1. Historic collection and structured identified specimens collection.

order to reduce the drawbacks and difficulties in handling the huge data sets that have been provided for a study. Any curator/taxonomist/researcher, even without deep knowledge of computing, using the proper taxonomic tools is able to design and built his/her data matrices, gather the needed data and see the results in a comparable and convenient way (**Figure 2**).

A lot of surveys consist of cross-sectional studies and use a great amount of information gathered from the collection by means of different taxonomy-related queries. While the information obtained from these queries is very useful for both researchers and collection curation professionals, the management and analysis of data in many occasions is cumbersome and leads to a long process of search carried out by humans which also implies a possibility of errors. Data gathering for the collection is usually preceded by data mining of the literature sources. After that follows the process and visualization of data related to the collection using the appropriate photographing techniques and software designed specifically for the collection items. So, the professionals are responsible for obtaining the information in a format that is useful for their work. This process in the collection is the most time-consuming and error prone and might be subjectively biased. For some Lepidoptera collections, we have designed the tool (s) which have already passed the time and application test [36, 37] in three institutions: Royal Museum for Central Africa (Belgium), Natural History Museum (United Kingdom), and Royal Belgian Institute of Natural Sciences (Belgium).

Figure 2. Intelligent computerized curation of the microlepidoptera collection at the Natural History Museum, London.

http://dx.doi.org/10.5772/intechopen.70925

The idea to use data mining tools in the collection-based research is getting the needed speed for many-sided approaches and studies: taxonomy, ecology, species interactions, biology, invasive patterns, host specifics or phylogenetic relationships. The obtained collection-based data assemblages demonstrate the complexity of biodiversity patterns. The further proper structuring of data by purposed collection data management tools enables to find the mechanisms which influence those particular biodiversity patterns as well as to obtain a clearer picture of the topic of ongoing research.

3. Digital data management of Lepidoptera collection

3.1. Creation of a virtual collection based on interlinked data

Collection digitization and putting the authorized data on the internet is a high priority at the moment since society not only needs but even demands to find authorized trustworthy data on the internet consultable at any moment and everywhere. The creation of a virtual collection has two main purposes:

- as a service to the community
- for data analysis for ongoing research.

There is no doubt that the future of collection management and collection consultation is digital. Much of society has already moved to the digital communication. Further development of a virtual collection of Lepidoptera has a strong emphasis on improving the existing digital collections and it is full heartily welcomed by the community of lepidopterists all over the world. Present achievement in informatics allows to operate the different aspects and relations concerning Big Data, and this is exactly what the Lepidoptera collection can provide for society and for any user beyond taxonomists. However, in order to succeed in creating a virtual collection, the steps taken should consequently follow a strict order and be completed:

Step 1. Data are **structured**, so they can be exported as Excel sheets and incorporated in any database.

Step 2. Data are presented in the same way **consequently**.

Step 3. Every data unit (species/specimen) should obtain a **unique code** and/or identifier, so the data are machine readable and operational.

Step 4. The biodiversity information is **intelligently text and data mined** from taxonomic literature and collection, intelligently verified and filtered.

Step 5. **Visualization** of a collection is a very important aspect, since it touches all layers and all groups of society. Furthermore, the visualized data becomes understandable for any user worldwide.

Step 6. Data are combined in a network of intelligent **relationships**. It not only enables to combine data in different groups and find correlations but also it saves a lot of time because any data are entered only once and the predefined reports are obtainable within seconds.

Data related to a Lepidoptera collection are arranged into different information packages: datasets (**Figure 3**) which are interlinked and integrated. However, at the same time these different datasets are easily independently consulted, independently displayed and new information can be independently added. These interlinked and integrated, but nevertheless independently operational datasets, serve as separate work packages for the simplified extraction of complex data. The protection of sensitive information within the interlinked work packages is foreseen also, since certain data and data packages can be made seen by authorized users only. I suggest to continue a well-tested formula of five interlinked and related data packages (**Figure 4**).

In this way it is easy to obtain clean data, to find trends in the present information packages, to keep track on specimens (types and vouchers) and to link taxonomic, morphological and DNA-related information to a concrete taxon (**Figure 5**): (species → subgenus → genus → tribus → subfamily → family → order).

The visualized dataset on Lepidoptera include computer-assisted automatic distribution atlases and unlimited possibilities in the presentation of image galleries, both types and verified voucher specimens.

Let me briefly mention the aspects of a database of Lepidoptera as a tool. There is a certain reluctance towards the databases in the community of curators, since the databasing is seen as an administratively imposed activity which takes a lot of time and gives little in return.

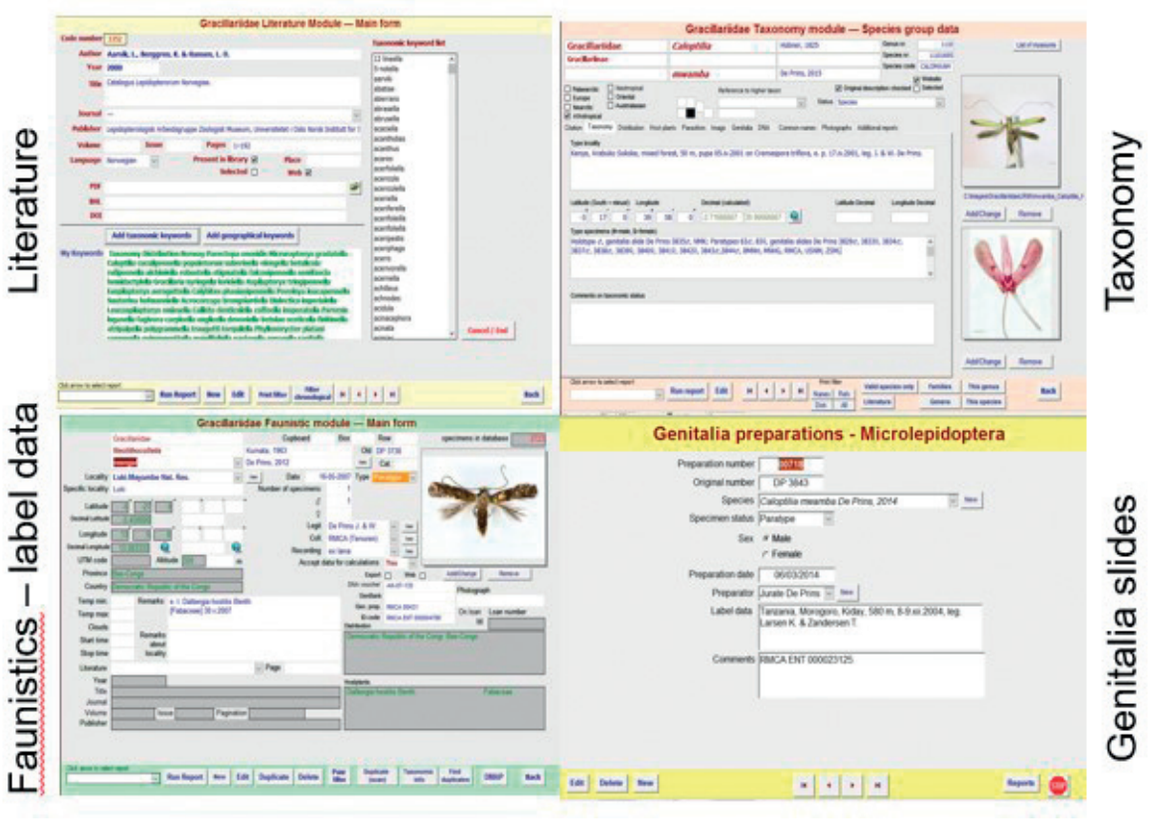

Figure 3. Simplified data entry behind the complex architecture.

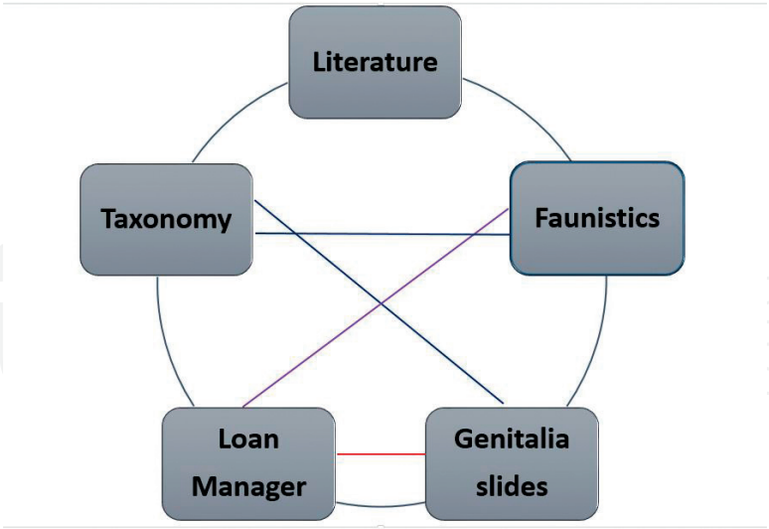

Figure 4. The structure of data: five interrelated and integrated datasets (following De Prins [36]).

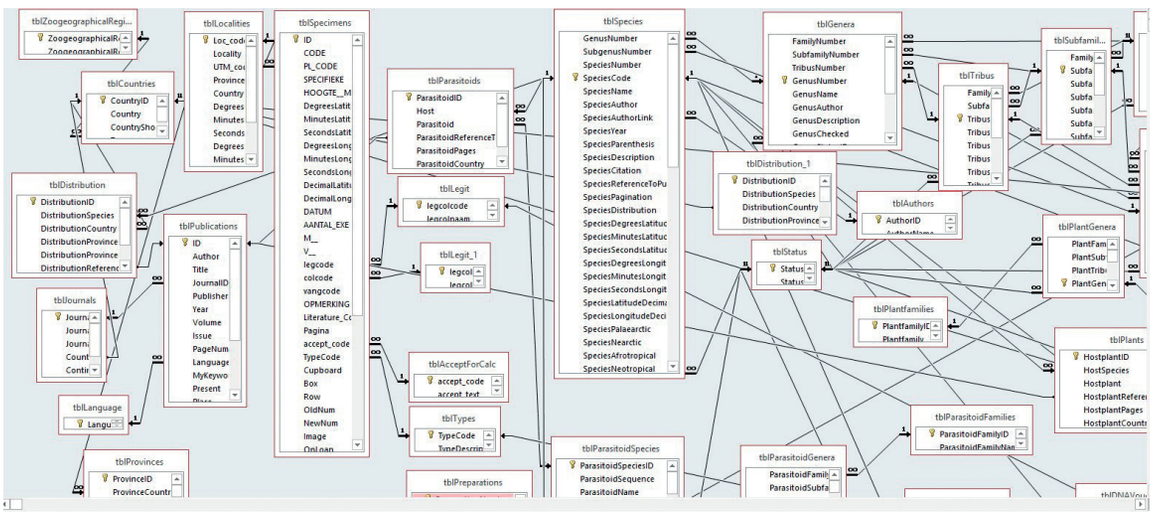

Figure 5. Part of the relationships of the dataset system (following De Prins [36]).

However, almost all curators-taxonomists working on a taxonomic group work with databases, because the information they have should be stored, accumulated, and consulted. What a good relational database can do for a curator:

- provide literature on taxa with reference to the exact pages of the original description and subsequent re-descriptions with indicated illustrations organized according to subject, in chronological or alphabetical order;
- provide correct taxonomic names, authors, dates;
- provide full and complete lists of synonyms with indicated sources;
- provide lists of taxa according to taxonomic classification or alphabetical order to arrange them in the collection;

- provide labels of taxa without typing errors or taxonomic mistakes;
- automatically make robust taxonomic catalogs and checklists and relates different biological data with references even with indicated pages;
- immediately assist locating the searched specimen in the collection, or if it is on loan to provide the details of the loan, so the curator always knows where any specimens under his/her care are located;
- show the type locality, even on Google Maps;
- show images of species, all their stages, host plants, habitats;
- provide data on genitalia or other micro-morphological structures, also images;
- provide lists of specimens belonging to the same species no matter where they are deposited;
- make a loan in a few minutes;
- show related species and help to make diagnoses and comparisons;
- quickly indicate the best time for organizing an expedition;
- show which types are deposited in which museum.

There can be a very strong motivation to compile the database because the outcomes of complete, related and verified data are rewarding for any taxonomist.

There are a number of specialized pre-defined queries and reports which are used very often in cyber cataloging which are immediately displayed without the need of creating them from scratch. All these reports in the suggested taxonomic database have a defined and fixed structure based on quantitative and qualitative extraction of data, mapping the data and interpreting these data based on the visual display of numerical charts. So both processes the curation of physical items (specimens) of the Lepidoptera collection and the integration of data into the predefined data matrices go together and can be largely automatized using unique digital ID scan-readable identifiers, predefined labels, loan forms, automatic monitoring of loans, automatic recognition of novelty in biological and distributional data, personalized data for collectors and donors, etc. This changes the way how Lepidoptera curators work in the collection and facilitates the process in order to eliminate errors due to human factors. When including societal and social media (e.g. www.waarnemingen.be or Facebook, Twitter, Instagram, ResearchGate) into the analysis of taxonomic data all the considerations presented above require greater relevance. Obtaining data on relationships in Lepidoptera from multiple digital communication records require to use the complex matrix and arrange data into smaller data packages (**Figure 4**) because the interrelated data is difficult to handle manually. The proper representation of basic faunistic data and associated species ID is crucial because this kind of information forms the basis for the later phase of analysis and visualization.

http://dx.doi.org/10.5772/intechopen.70925

3.2. Online searchable catalogs

Many studies on taxonomy, as well as phylogenies of higher Lepidoptera groups are hampered by the fact that the world fauna of Lepidoptera is still not inventoried. Moreover, the current research tries to study the wider range of factors that may be involved into the evolutionary processes of lepidopteran species. In particular many different environmental factors in which species are immersed are of special interest in present approaches. The online cataloging aims to fill at least partly the gaps and provides the following primary data:

- the diversity of species;
- the spatial distribution pattern;
- the taxonomic and phylogenetic distinctiveness;
- nomenclature and study of primary types;
- synonymy;
- concise records of natural history;
- concise records of taxonomic history; and
- DNA accession numbers.

We presented two online catalogs: the catalog of one family of moths on a global scale available from www.gracillariidae.net and the catalog of all species of moths from a defined bio-geographical region, in our case the Afrotropics, available from www.afromoths.net. For the creation of searchable taxonomic online catalogs we used two modules of the dataset: (1) taxonomic and (2) literature of the interlinked dataset (**Figure 3**). The online searchable catalogs provide the referenced taxonomic and life history information in the following fields:

- family;
- subfamily;
- checked and correct species name;
- status (species, subspecies, synonym, unavailable name etc.);
- author;
- description year;
- original combination;
- Google mapped type locality;
- type specimens, associated genitalia slides and depository;
- publication of original description and pagination;
- distribution per country;
- biological data.

An approach based on the analysis of taxonomic data along with modern imaging techniques may give more insight into the taxonomic situation and group relationships than narrower, more specialized traditional studies. Nevertheless, the widely adopted approaches include a number of well-known, standardized data packages. They have been included into the proposed digital tools for online cataloging:

i. **taxonomy** (taxonomic position, current species name, synonymy, original combination and the reference to the original description with indicated pagination);

ii. **types** (name bearing type specimen(s): holotype, syntypes, lectotypes, neotype), other verified type specimens (paratypes, paralectotypes) and other not defined by Code (ICZN 1999) verified type specimens with associated institutional numbers, associated microscopic preparation slides, their depository place;

iii. **distribution** and **habitats** (mapped and referenced distributional data);

iv. **biology** (referenced data on food plants, and concise life history data);

v. **DNA** (accession numbers of the GenBank and linked to them the DNA information).

These interlinked data packages serve to obtain qualitative online catalogs that present data on Lepidoptera biodiversity displayed into a number of categories. Online cataloging standardizes taxonomy all over the world and makes the collection curation a fast and finalized endeavor in any museum at any place in the world. It becomes also possible to combine data obtained from many different museums into one huge data network and to fill the gaps in taxonomy and other related disciplines.

3.3. Data management and analysis

The proposed approach of data management [36, 37] is based on possibilities that any professional researcher or citizen scientist retrieves personalized taxonomic, trophic, and distribution data needed for research or study. The proposed architecture of taxonomic data management has a web application function and a robust image gallery. On server side it has been developed with the assistance of GBIF, BeBIF, Catalog of Life as the up-to-date database management system. The literature records have been used for displaying the referenced data which in many cases are also linked with the Biodiversity Heritage Library.

Four different user roles can be defined in the taxonomic data management system:

i. Supervisor Administrator who has full permissions to manage and manipulate data, to create queries and reports, answer any taxonomic question or produce structured data sets in an exportable format (excluding access to data described in data protection laws).

ii. Taxonomist-Administrator who validates the taxonomic information, nomenclatural issues, homonymy, synonymy, availability of names and links with the checked reference; produces robust global or regional interlinked taxonomic catalogs.

iii. Taxonomist-Identifier who adds associated images to species pages, host plant(s), and distribution information.

iv. Collection Administrator who manages collection specimen data and administrates loans.

The display of results through the data management system eases the work of curators and presents the data in the way the user needs:

Standardized data management. The user obtains all the data in a strictly standardized format and can deeply inspect them as well as the general information meaningful to the search question.

Customized data management. The user can design and create the combinations of data of interest. Also data can be grouped into sets based on predefined formula in the similar way to the validated queries. The data management system will automatically perform the display of data from the requested search fields.

Visualized survey of data. This part of data display shows the imaged specimens once they are identified and curated. As well as the plain qualitative and quantitative taxonomic results obtained from data curation and data management there is part of data presentation which is visualized. This type of information showing shapes, patterns and colors of Lepidoptera and representing different characteristics of taxa/specimens has been carried out by the application of micro/macro photography techniques. Digital imaging analysis tools for Lepidoptera have been extensively used for a number of studies and websites during the first decade of this century. New imaging tools exist today that have been designed for collection specimens with capacities to overcome the shortcomings of traditional micro/macro photography and are applicable for mobile devices. In addition to microphotography, the microtomography has been proven as a promising technique to allow the visualization of internal morphological structures in non-destructive way and present them three-dimensionally.

3.4. Intelligently assembled and curated collection

The internet-linked data analysis of expert assembled and curated Lepidoptera collection integrates observational data into many possible models. An intelligently assembled and curated Lepidoptera collection represents a data product that is of growing importance to researchers working in the domain of climate science and preoccupied with a wide range of applications which need fast decision procedure. The Lepidoptera collection, as a more or less completed institutional unit, brings together the following set of elements:

1. high reliability and performance of data analysis;

2. data management on smaller or bigger scale;

3. appliance of virtualization and user attractive visualization;

4. possibilities to adaptive usage of data;

5. harmonization of data within the domain of natural sciences.

The effectiveness of the internet-linked data extracted from the Lepidoptera collections has been demonstrated in several successful case studies [7, 8, 12, 38]. Structuring and digitization

of collection data and presenting them in an internet-linked environment lowers the barriers for obtaining the taxonomic scholarship, democratizes the taxonomic community of lepidopterists, fosters innovation and experimentation with the collection data, facilitates the usage of technology with the collection items, and provides the agility required to meet the multi-purpose needs of users. The structured, internet-linked taxonomic and collection-based data are providing new data service within natural history that helps to connect academic and computational resources. Moreover, the structured, internet-linked Lepidoptera collection data engage the multinational communities of naturalists and climate science specialists in the construction of new capabilities. The provision of such interchangeable multi-purpose data probably is one of the most important changes in the way we, museum-based curators and taxonomists work within the modern society.

New technological and societal developments shifted significantly the paradigm what collection curation and what curators are. For 200 years lepidopteran taxonomists followed individual strategies. Later the taxonomy of Lepidoptera saw the inclusion of molecular approaches and techniques. Despite these new technologies, the integrative delimitation of lower taxa at the genus and species level proceeds to be a continuously ongoing process parallel with the descriptions of novel taxa and requires a taxonomic consensus which actually needs to be included into the data matrices on a daily basis. Also, an intelligently curated collection of Lepidoptera is a live tool changing daily due to changes in taxonomy or additions in biogeography. Taxonomists/curators may feel uneasy with increasingly time consuming and laborious work while delineating taxa, but also scientists find it difficult for obtaining comparative data if the reference collection is not curated or is curated poorly.

A curated and managed high quality collection will overcome the problems that hamper data search and interoperability between taxonomic and research labs. The use of incorrect and invalid names will result in heterogeneous, incomplete, fragmented datasets which need verification [39]. The application of a unique numbering system for taxa facilitates machine-readable linkages to data sources, easily detects any human-caused error [36, 38] and speeds up the connectivity in ever expanding mass of taxa and data affiliated to them. Digital collection curation is fully in line with ideas to link and share collection-based data and to explore all resources available in public institutions and private holdings. Collection curation based on a long tradition as a non-profit occupation allows and facilitates public access to lepidopteran biodiversity worldwide and adds a needed value to the already known and still unknown lepidopteran biodiversity in such a way that a Lepidoptera collection becomes a ready-to-use tool for science, scientists, and society. The collection allows us to exploit the sources and knowledge from different aspects, to discover and disclose new findings within the framework of global science of natural history.

4. Conclusions

Information science and data management tools have become very important in the curation of Lepidoptera collections. The complexity of techniques and computing tools used in taxonomy

and the increase in the amount of data that can be obtained from collection-based disciplines make it necessary to automate processes in data gathering, manipulation, analysis and visualization. Much data used in taxonomy and Lepidoptera collection management comes from unverified offline taxonomic datasets and specimen labels. This can lead to time-consuming and error-prone processes that can be easily automated. In this sense, the collaboration between researchers, taxonomists, citizen scientists, collection curators, and computing/information science is crucial to build and to use the proper approaches in taxonomy needed to avoid error-prone situations and to obtain qualitative results without the need of being experts in a certain taxonomic group or in the techniques underlying the automated processes. Modern approaches towards Lepidoptera collections and data management help to focus on the goals and studies that can be finalized.

For future work, I see a much closer integration of different disciplines related to life and climate sciences and inclusion of new functionalities into the offline and online tools that could provide much deeper insights into the diversity of Lepidoptera as well as into the complexity of relationships, thus improving the usefulness of these tools for research and identification purposes. The structured, searchable global, and regional databases of Lepidoptera have already been of significant assistance in the evaluation of Lepidoptera diversity at national and international levels and in the curation of large institutional collections. The novel approaches in curation, data management, and collection-based science can also be incorporated into educational programs so that the lepidopterist community and society in general can test, use and explore all possible benefits from Lepidoptera collections.

Acknowledgements

I thank my lepidopterists colleagues and curators of Lepidoptera collections for valuable discussions during many years of collaboration. My special thanks are extended to Willy De Prins, Stefan Kerkhof, and Wouter Dekoninck (Royal Belgian Institute of Natural Sciences, Brussels, Belgium) for their inspiring talks and strong support for Lepidoptera collection-based issues. The IT team of the Belgian Biodiversity Platform and the Global Biodiversity Information Facilities is thanked for their technical assistance in launching the structured taxonomic data on Lepidoptera online. Donald and Mignon Davis (Smithsonian Institution, Washington DC, USA) are cordially acknowledged for their valuable comments.

Author details

Jurate De Prins

Address all correspondence to: jurate.deprins@gmail.com

Royal Belgian Institute of Natural Sciences, Brussels, Belgium

References

[1] Kristensen NP, editor. Lepidoptera, Moths and Butterflies. Volume 1: Evolution, Systematics, and Biogeography. Handbook of Zoology IV (35). Berlin, New York: Walter de Gruyter; 1998. 487 p

[2] van Nieukerken EJ, Kaila L, Kitching IJ, Kristensen NP, Lees DC, Minet J, Mitter C, Mutanen M, Regier JC, Simonsen TJ, Wahlberg N, Yen S-H, Zahiri R, Adamski D, Baixeras J, Bartsch D, Bengtsson BÅ, Brown JW, Rae Bucheli S, Davis DR, De Prins J, De Prins W, Epstein ME, Gentili-Poole P, Gielis C, Hättenschwiler P, Hausmann A, Holloway JD, Kallies A, Karsholt O, Kawahara AY, Koster S(JC), Kozlov MV, Lafontaine JD, Lamas G, Landry J-F, Lee S, Nuss M, Park K-T, Penz C, Rota J, Schintlmeister A, Schmidt BC, Sohn J-C, Solis MA, Tarmann GM, Warren AD, Weller S, Yakovlev RV, Zolotuhin VV, Zwick A. Order Lepidoptera Linnaeus, 1758. In: Zhang Z-Q, editor. Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. Zootaxa. 2011;**3148**:212-221

[3] Mallet J. The Lepidoptera Taxome Project [Internet]. 2014. Available from: http://www.ucl.ac.uk/taxome/ [Accessed: 2017-07-17]

[4] Kawahara AY, Pyle RM. An appreciation for the natural world through collecting, owning and observing insects. In: Lemelin RH, editor. The Management of Insects in Recreation and Tourism. Cambridge University Press; 2012. p. 138-152

[5] Seltmann KC, Cobb N, Gall LF, Barlett CR, Basham A, Betancourts I, Bills C, Brandt B, Brown RL, Bundy C, Caterino M, Chapman C, Cognato A, Colby J, Cook SP, Daly KM, Dyer LA, Franz NM, Gelhaus JK, Grinter CC, Harp CE, Hawkins RL, Heydon SL, Hill GM, Huber SH, Johnson N, Kawahara AY, Kimsey LS, Kondratieff BC, Krell F-T, Leblanc L, Lee S, Marshall CJ, McCabe LM, McHugh JV, Menard KL, Opler PA, Palffy-Muhoray N, Pardikes N, Peterson M, Pierce NE, Poremski A, Sikes DS, Weintraub JD, Wikle D, Zaspel J, Zolnerowich G. LepNet: The Lepidoptera of North America network. Zootaxa. 2017; **4247**:73-77. DOI: 10.11646/zootaxa.4247.1.10

[6] CETAF–Consortium of European Taxonomic Facilities. Available from: http://cetaf.org [Accessed: 2017-07-17]

[7] De Prins J, De Prins W. AfroMoths, an Online Database of Afrotropical Moth Species (Lepidoptera). 2017. Available from: http://www.afromoths.net/ [Accessed: 2017-07-17]

[8] Güntsch A, Hyam R, Hagedorn G, Chagnoux S, Röpert D, Casino A, Droege G, Glöckler F, Gödderz K, Groom Q, Hoffmann J, Holleman A, Kempa M, Koivula H, Marhold H, Nicolson N, Smith VS, Triebel D. Actionable, long-term stable and semantic web compatible identifiers for access to biological collection objects. Database – The Journal of Biological Databases and Curation. 2017;**2017**:1-9. DOI: 10.1093/database/bax003

[9] Duckworth WD, Genoways HH, Rose CL. Preserving Natural Science Collections: Chronicle of Our Environmental Heritage. Washington, DC: National Institute for the Conservation of Cultural Property; 1993. 140 p

[10] Nudds JR, Pettitt CW, editors. The Value and Valuation of Natural Science Collections. London: The Geological Society; 1997. 240 p

[11] Mutanen M, Kivelä SM, Vos RA, Doorenweerd C, Ratnasingham S, Hausmann A, Huemer P, Dincă V, van Nieukerken EJ, Lopez-Vaamonde C, Vila R, Aarvik L, Decaëns T, Efetov KA, Hebert PDN, Johnsen A, Karsholt O, Pentinsaari M, Rougerie R, Segerer A, Tarmann G, Zahiri R, Godfray HCJ. Species level para- and polyphyly in DNA barcode gene trees: Strong operational bias in European Lepidoptera. Systematic Biology. 2016;**65**:1024-1040. DOI: 10.1093/sysbio/syw044

[12] Chapman AD. Principles of Data Quality, Version 1.0. Report for the Global Biodiversity Information Facility: Copenhagen; 2005

[13] Hofmann AF, Tremewan WG. The Natural History of Burnet Moths. Museum Witt Munich & Nature Research Center Vilnius: Vilnius; 2017. 630 p

[14] Schnase JL, Duffy DQ, Tamkin GS, Nadeau D, Thompson JH, Grieg CM, McInermey MA, Webster W. MERRA analytic services: Meeting the big data challenges of climate science through cloud-enabled climate analytics-as-a-service. Computers, Environment and Urban Systems. 2017;**61**:198-211. DOI: 10.1016/j.compenvurbsys.2013.12.003

[15] Hebert PDN, deWaard JR, Zakharov EV, Prosser SWJ, Sones JE, McKeown JTA, Mantle B, La Salle J. A DNA 'barcode blitz': Rapid digitization and sequencing of a natural history collection. PLoS One. 2013;**8**:e68535. DOI: 10.1371/journal.pone.0068535

[16] Cavallin EKS, Munhoz CBR, Harris SA, Villarroel D, Proença CEB. Influence of biological and social-historical variables on the time taken to describe an angiosperm. American Journal of Botany. 2016;**103**:2000-2012. DOI: 10.3732/ajb.1600120

[17] Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin WD, Vogler AP, Hedin M. Sequence-based species delimitation for the DNA taxonomy of underscribed insects. Systematic Biology. 2006;**55**:595-609. DOI: 10.1080/10635150600852011

[18] Epstein ME. Moths, Myths & Mosquitoes. The Eccentric Life of Harrison G. Dyar Jr. New York: Oxford University Press; 2016. 325 p

[19] Clarke JFG. Catalogue of the Type Specimens of Microlepidoptera in the British Museum (Natural History) Described by Edward Meyrick. London: British Museum; 1955. 332 p

[20] Gilbert P. A Source Book for Biographical Literature on Entomologists. Leiden: Backhuys Publishers; 2007. 694 p

[21] Strategy to 2020. Advancing the Science of Nature. London: The Natural History Museum; 2015. 22 p

[22] Suarez AV, Tsutsui ND. The value of museum collections for research and society. Bioscience. 2004;**54**:66-74. DOI: 10.1641/0006-3568(2004)054[0066:TVOMCF]2.0.CO;2

[23] Ponder WF, Carter GA, Flemons P, Chapman RR. Evaluation of museum collection data for use in biodiversity assessment. Conservation Biology. 2001;**15**:648-657

[24] Holmes MW, Hammond TT, Wogan GOU, Walsh RE, LaBarbera K, Wommack EA, Martins FM, Crawford JC, Mack KL, Bloch LM, Nachman MW. Natural history collections as windows on evolutionary processes. Molecular Ecology. 2016;**25**:864-881. DOI: 10.1111/mec.13529

[25] Cho S, Epstein SW, Mitter K, Hamilton CA, Plotkin D, Mitter C, Kawahara AY. Preserving and vouchering butterflies and moths for large-scale museum-based molecular research. PeerJ. 2016;**4**:e2160. DOI: 10.7717/peerj.2160

[26] Evans Walter D, Winterton S. Keys and the crisis in taxonomy: Extinction or reinvention? Annual Review of Entomology. 2007;**52**:193-208. DOI: 10.1146/annurev.ento.51.110104.151054

[27] Mallet J, Willmott K. Taxonomy: Renaissance or tower of babel? Trends in Ecology and Evolution. 2003;**18**:57-59

[28] Agnarsson I, Kuntner M, Paterson A. Taxonomy in a changing world: Seeking solutions for a science in crisis. Systematic Biology. 2007;**56**:531-539. DOI: 10.1080/10635150701424546

[29] Walther G-R, Post E, Convey P, Menzel A, Parmesan C, Beebee TJC, Fromentin J-M, Hoegh-Guldberg O, Bairlein F. Ecological responses to recent climate change. Nature. 2002;**416**:389-395

[30] Parmesan C. Ecological and evolutionary responses to recent climate change. The Annual Review of Ecology, Evolution and Systematics. 2006;**37**:637-669. DOI: 10.1146/annurev.ecolsys.37.091305.110100

[31] De Prins J, Davis D, De Coninck E, Sohn J-C, Triberti P. Systematics, phylogeny and biology of a new genus of Lithocolletinae (Lepidoptera: Gracillariidae) associated with Cistaceae. Zootaxa. 2013;**3741**:201-227. DOI: 10.11646/zootaxa.3741.2.1

[32] Ford JD, Tilleard SE, Berrang-Ford L, Araos M, Biesbroek R, Lesnikowski AC, MacDonald GK, Hsu A, Chen C, Bizikova L. Opinion: Big data has big potential for applications to climate change adaptation. PNAS. 2016;**113**:10729-10732. DOI: 10.1073/pnas.1614023113

[33] De Queiroz K. Species concepts and species delimitation. Systematic Biology. 2007;**56**: 879-886. DOI: 10.1080/10635150701701083

[34] Zhang L, Reed RD. A practical guide to CRISPR/Cas9 genome editing in Lepidoptera. bioRxiv preprint first posted online 2017-04-24. DOI: 10.1101/130344

[35] Brito R, De Prins J, De Prins W, Mielke OHH, Gonçalves GL, Moreira GRP. Extant diversity and estimated number of Gracillariidae (Lepidoptera) species yet to be discovered in the Neotropical region. Revista Brasileira de Entomologia. 2016;**60**:275-283. DOI: 10.1016/j.rbe.2016.06.002

[36] De Prins J. An integrated taxonomic tool for online dissemination of concise, verified and visualized information on biodiversity, retrieved from data and text mining of natural history collections and libraries. JSM Bioinformatics, Genomics and Proteomics. 2016;**1**: 1006

[37] De Prins J. Filling the gaps on trophic interactions between Lepidoptera and hymenoptera using information retrieved from interlinked and integrated taxonomic datasets. JSM Anatomy and Physiology. 2017;**2**:1006

[38] GBIF–Global Biodiversity Information Facility. Available from: http://www.gbif.org/ [Accessed: 2017-07-18]

[39] de Jong R. Fossil butterflies, calibration points and the molecular clock (Lepidoptera: Papilionoidea). Zootaxa. 2017;**4270**:001-063. DOI: 10.11646/zootaxa.4270.1.1

IntechOpen