Arch Biol Sci. 2023;75(1):89-102 https://doi.org/10.2298/ABS230116008T

Standard descriptive matrices in the identification of ex-ophytophagous caterpillars

Aleksandra D. Trajković*, Maja J. Lazarević, Saša S. Stanković, Miloš A. Popović, Marijana M. Ilić Milošević and Vladimir A. Žikić

Faculty of Sciences and Mathematics, Department of Biology and Ecology, University of Niš, Višegradska 33, 18000, Serbia

***Corresponding author**: aleksandra.trajkovic2@pmf.edu.rs

Received: January 16, 2023; **Revised**: February 13, 2023; **Accepted**: February 20, 2023; **Published online**: March 8, 2023

Abstract: Identification of exophytophagous lepidopteran larvae is a necessity for researchers in biological disciplines ranging from biodiversity inventorying to research in parasitoid evolution and species monitoring. The lack of expertise in the field jeopardizes the outcomes of further investigations and recording of the multilevel plasticity of juvenile Lepidoptera. This paper offers an improvement to the existing haphazard approach by developing 41 simplified characters that include 150 morphological, behavioral and autecological states and their delineation, visual validation, and a descriptive matrix for 83 heterogeneous species. By combining the states into all possible identification scenarios, the matrix revealed 582 morphological, habitat and resource polyphenisms for the mentioned species. The categorical nature of the data implied the use of categorical principal component analysis to visualize the discriminative capacity without character relationship assumptions. The object-point biplot was used to derive the K value for K-mode clustering, while the cluster membership was introduced as a labeling variable to further inspect the grouping pattern. The results of this descriptive analytic research indicate that descriptive matrices will allow continuous expansion and fine examination of many different species assemblages. From interactive identification keys to machine learning training, the presented framework can make data storage and interpretation significantly more attainable.

Keywords: Lepidoptera; polyphenism; morphology; autecology

INTRODUCTION

The lack of traditional or computer-aided identification tools has made caterpillar identification, whether during fieldwork or on online biodiversity platforms, a process dependent on empirical knowledge or Gestalt morphology comparison. Consulting an expert or surveying online catalogs is time-consuming and often inefficient [1] for a group containing several thousand species [2]. While current technology offers a large pool of data regarding the specific lineage or an area, caterpillar identification keys are either overburdened or outdated [3].

Dyar [4-7] was the first to introduce the position of the primary setae as a taxonomically important character. Frohawk [8] made further progress with captive rearing and characterization of the butterfly species with detailed descriptions, illustrations and scientifically sound morphological terminology. As the Lepidoptera phylogeny gained importance in scientific circles, the description of the developmental stages became common but was limited by taxonomic or geographic coverage. Details about the biology and ecology of Lepidoptera were mostly published in different PhD theses [9-11], handbooks [12] or monographs [13], presumably because of the volume and complexity within the order. Most of the applicable keys [14-19] dealt with economically important species, which for the most part belong to the informal Microlepidoptera group and exhibit concealed feeding habits. Work on setal maps and the morphological aspects of molting and metamorphosis was continued in field guides [20] and short contributions [21,22]. Further elaboration came with Stehr's two-volume book [23], which included a detailed glossary of terms, setal arrangement nomenclature and a dichotomous key to families with 225 couplets. This publication is considered a standard tool in the research field to this day [24]. Subsequent publications [25,26] proposed diagnostic protocols and helped develop the

specialized terminology. Kristensen's edition of the Handbuch der Zoologie [27] and works cited therein, compiled the most itemized overview of micromorphological details, an expansion to the existing key to families [23] based on larval ecology and morphology. The author also expressed clear criticism of the scientific neglect of the diverse and taxonomically important developmental stage. Probably the most specialized work concerning the morphology of larval Noctuidae was Beck's highly illustrated two-volume publication [28,29] in the German language. More recent publications [30-32] included natural assemblages of caterpillars that share a geographic area or a functional trait, and thus provided general descriptions, keys to higher taxa and detailed pictorials suitable for both practitioners and experts.

The well-established identification technique, chaetotaxy [33], discriminates between the taxa by comparing standard setal maps [34]. The specimens must be prepared in laboratory conditions, which is not always in the interest of researchers [35], and the accuracy is sometimes questionable due to numerous asymmetric aberrations [36]. Though widely used in pest identification, referring to setal maps is not suitable for many Macrolepidoptera species, especially if their final instars have prominent secondary setae or change drastically from the first to the last molt. Other than setae, the micromorphology, and the arrangement of setal bases (pinaculae and verrucae), tonofibrillary platelets and mouth parts were recognized as useful in taxonomy and identification and have been included in descriptive schemes as well [27].

Reliable caterpillar identification requires rearing to the adult stage or employing molecular methods. This, however, is often inconvenient or impossible [37] as the equipment or other requirements might not be readily available [38]. The existing identification tools, both ink-on-paper and digitally aided, usually share the choice of diagnostic characters. The solving pathway fuses morphological, micromorphological, behavioral and ecological traits [14,15,17,23,27,39-41] but is often restricted to higher rank identification and accompanied by species fact sheets [42] for guidance. This applies to available online tools as well, e.g., the Lucid Software keys [48]. However, there are several limitations to this approach, primarily regarding the number of included taxa, geographical differences, polymorphism and polyphenism[43].

The primary goal of this article was to facilitate the development of more effective identification tools for exophytophagous caterpillars, and to provide a protocol that incorporates polymorphisms and polyphenisms in species description. This was achieved through simple feature interpretation, which explored all aspects of the habitus and habitat that can be confirmed during field specimen inspection. Besides serving as satisfactory examples, the species used for state presentation were fully profiled through categorical data matrices.

MATERIALS AND METHODS

Ethics statement

The study did not involve vertebrates nor invertebrates included in the Animal Welfare policy in Europe, nor human participants, and therefore does not require special permission.

Data acquisition

To form descriptors that can be used for standard, scheme-like appearance and life form, but also for more specific cases, the descriptors were conceived based on visual inspection and extensive literature search (Supplementary Table S1) into all European exophytophagous Macrolepidoptera. Exceptions that were taken into consideration were some members of the Zygaenoidea superfamily, and noctuid caterpillars that create a shelter at a certain point of their metamorphosis (cutworms hiding under litter or species living within the flowering part of the host plant).

The autecological, behavioral, and morphological profiling included the inspection of fluid-preserved specimens of the representative species from the collection of the University of Niš, Faculty of Sciences and Mathematics, Department of Biology and Ecology. Brief descriptions of the lower-than-family taxa, diagnostically important notes and personal observations, field and laboratory notes and photographs obtained through fieldwork and caterpillar rearing, were used as secondary data. To obtain an insight into the species-level habitus and autecological heterogeneity, online platforms were surveyed, most notably

Lepiforum [44] and public data from the citizen science databases, iNaturalist [45], Biologer [46] and Alciphron [47].

The terminology used in this paper primarily follows the general key to families proposed by Stehr [23]. For an accurate but widely applicable description, the selected traits were evaluated for perceptibility, stability in the taxa and throughout development, and discriminative power. Descriptors were uniformly organized into tables that contain character codes, character states as well as notes and instructions. Each macromorphological and autecological section was assigned a coding capital letter, to avoid possible confusion with abbreviations used in larval micromorphology and imaginal morphological taxonomy. Characters were coded numerically, while lowercase letters were used for states. The morphological characters were explained through image plates in which a suitable species, regardless of taxonomic affiliation, represents each state. Autecological features were selected to cover general information such as the geographic region from which the sample originates, and the more specific details that can be recorded prior to collecting the specimen, e.g., the caterpillar's position on the host plant.

Descriptive matrix

The final matrix was created by assigning each species that was used as an example to a state for all other characters. As each state is a precise description of a certain feature, it was necessary to include all known variations at an interspecific, intra- and interpopulation level. The matrix was limited to full-grown caterpillars (L_x, with the $_x$ being the final instar), but it included drastic changes during hibernation, fresh molt and prepupation. Profiling polyphenisms was achieved through selective combination: if a species was recognized to have multiple states of the same morphological character, each morphological polyphenism was described in a separate row and named accordingly (the name of the habitus was derived from its most prominent attribute and additionally labeled with an "*m*" in the dataset). If a species had variation in autecological and behavioral characters, the outcomes were recorded with numbers, to prevent evolutional implications and the context of eco-type (e.g., *e1*, *e2*… etc., where the "*e*" stands for ecological/resource polyphenism). If there is variation in both classes of characters, all logically possible combinations were recorded with an additional "*me*" abbreviation prefix. The matrix consisted of 582 fully described cases of various polyphenisms, derived from 83 species and belonging to 14 families (Supplementary Material), each assigned to one of the states of 34 morphological and 7 autecological or behavioral characters imputed as categorical variables for data visualization.

Data analysis

The described states were imported into an SPSS statistical software platform dataset through a numeric encoding scheme. To investigate the structure of the data, categorical principal component analysis (CATPCA) was performed in SPSS desktop version 20.0 (SPSS Inc, Chicago, IL, USA). In the CATPCA, category variables represented by the model are in fact the transformed variables, which suggests that the fit of the model is predetermined by the optimal choice of (non)monotonic quantification. Accordingly, the 34 morphological characters were scaled through ordinal analysis level, as the described states can indeed be considered monotonic (e.g., see Table 1 with H3, 1 (a) being the score for a completely smooth head capsule, and 4 (d) being the score for a head capsule fully covered in setae). The scale displays the stretching of the character from the basic, schematic presentation of a caterpillar to the most modified (visually contrasting) state and under no circumstance implies an evolutionary context. The transformation of the 7 autecological and behavioral categorical variables was carried out in agreement with their actual nominal measurement as there is no reason to assume that there is a class membership to be preserved. A two-dimensional model was evaluated through the total VAF and Cronbach's α value. No categories were omitted, regardless of their contribution, because the goal of the methodology was not to perfect the model but to display the ordination of this particular species assemblage based on the provided descriptors. At this point, the CATPCA plotting was performed without any labels but case numbers.

Classification of the categorical data in the k-modes clustering algorithm available from the klaR library in the R Studio environment for statistical computing, partitions the data set into groups according to modes, instead of means, as it is impossible to calculate the mathematical distance between items when dealing

Table 1. General morphology, eidonomy with alignment and concise autecology explained by descriptors.

Ch. code and description	Ch. state description	Notes and instructions
H1 – head: T_1 ratio	**a** – equal	width ratio should be inspected by dorsal view, disproportion common among young instars
	b – head smaller	
	c – head bigger	
H2 – retractability	**a** – none	overlapping T_1 segment
	b – apparent	
H3 – hairiness	**a** – setae only visible around mouthparts	hairiness should be inspected in plenty of light and with contrasting background
	b – setae scarce	
	c – dense minute setae	
	d – full dense coverage	
H4 – general texture/ appearance	**a** – shiny	4b – can be a result of minute setae presence
	b – dull	
	c – rough	
	d – spinulose	
H5 – epicranial surface	**a** – mostly flat	5b – any observable level of incision
	b – protruding	
	c – bulbaceous	
	d – elongated, meet apically	
H6 – position	**a** – in line	refers to both anatomical properties and resting posture
	b – upward	
	c – downward	
H7 – shape	**a** – regular	7b – mandibles and lobes in the same plane
	b – flattened	
H8 – additional structures	**a** – none	additional structures lying on the epicranial lobes and should not be mistaken for any projections from T_1 segment
	b – fleshy projections	
	c – simple tubercles	
	d – branched tubercles	
H9 – pigment schemes	**a** – none/ monochromatic	9c – any defined, bordered, markings or pattern on each lobe (submedial arcs, genal edge)
	b – reticular	
	c – symmetrical	
	d – combined	
B1 – general body shape	**a** – cylindrical	assessment of T_1: A segments ratio 1a – both flattened and regular
	b – fusiform	
	c – oval, stocky	
B2 – general integument appearance	**a** – smooth	2a – both oily and velvety; 2b – spinulose, granulose; 2c – undetectable under setae
	b – textured	
	c – hairy	
B3 – prothoracic shield	**a** – shiny	3c – matches the rest of the integument, inconspicuous, absent
	b – dull	
	c – undetectable	
B4 – suranal segment shape	**a** – pointed	often positioned downwards
	b – rounded	
B5 – setae presence	**a** – absent or scarce	5b –setae most prominent around the frontal plane line 5d– includes scoli
	b – minute	
	c – lateral coverage	
	d – full coverage	
B6 – setae-bearing structure	**a** – indistinctive	6b – flat, bulbous, or pointed, carries single seta; 6c – multiple shapes, carries many setae; 6d – simple or branched, fleshy or sclerotized
	b – pinaculae	
	c – verruca base	
	d – scolus	

Table 1. continued

B7 – setal form	**a** – fine	applies to individual setae
	b – bristle	
	c – plumose	
B8 – setal grouping	**a** – individual	in case of multiple state presence, the most prominent one should be chosen
	b – tuft	
	c – pencil	
	d – verruca	
B9 – dorsoventral folding	**a** – none	common among arboreal species, usually subspiracular
	b – visible callosity	
B10 – segment folding	**a** – border clearly defined	character easily detectable by stimulating the caterpillar to move
	b – border undefined	
	c – intersegmental rings	
B11 – projection type	**a** – hump	11a – fleshy; 11b – sclerotized; 11c – osmeterium or mediodorsal papillae projections
	b – horn	
	c – glandular	
	d- none	
B12 – projection localization	**a** – T segments	T_1 segment projections often mistaken for head
	b – A segments	
	c – T and A segments	
	d – inapplicable	
M1 – abdominal prolegs presence	**a** – A_3-A_6	anal segment pair excluded
	b – A_4-A_6	
	c – A_5-A_6	
	d – A_6	
M2 – abdominal prolegs, degree of development	**a** – equal in size	2c – rudiments, small papillae, integument slightly changed in A_5 and A_6 proleg position
	b – one/two pairs smaller	
	c – one/two pairs vestigial	
M3 – anal prolegs modification	**a** – none	3b and 3c fleshy; 3c significantly longer than 3b
	b – fork	
	c – stemapoda	
M4 – forked anal plate	**a** – absent	not to be mistaken for anal prolegs, inspect for both; 4b – short projections above the anal prolegs
	b – present	
M5 – resting posture	**a** – elongated and flat	not to be confused with defensive curling in thanatosis
	b – hook	
	c – twig mimic	
	d – rear segments raised	
M6 – movement pattern	**a** – regular	causal character (number of prolegs)
I1 – integument color origin	**a** – pigment	1b – transparent, slimy integument, common for young instars; 1c– fully covered with setae
	b – subintegumental structures	
	c – not applicable	
I2 – dorsoventral contrast	**a** – strong	regardless of the color transition scheme
	b – slight	
I3 – mediodorsal line	**a** – solid	3b – irregular/linearly grouped elements; 3e– complete pigment deficiency or a line originating from subintegumental structures
	b – interrupted	
	c – double	
	d – bordered	
	e – unmarked	
I4 – mediodorsum	**a** – pattern	refers to the region between the mediodorsal and subdorsal lines
	b – sprinkled	
	c – heterogeneously marked	
	d – none of the above	

Table 1. continued

I5 – subdorsum	**a** – solid line	5a – multiple lines possible; 5b– irregular/ linearly grouped elements
	b – interrupted	
	c – unmarked	
I6 – lateral markings	**a** – spiracular	6a,b,c – includes solid, interrupted, linearly grouped dots and stripes
	b – subspiracular	
	c – supraspiracular	
	d – oblique patterns	
I7 – spiracles	**a** – simple	7a – if not black, bordered
	b – ornamented	
P1 – host plant type	**a** – deciduous	algae, lichens, and mosses are treated as P1b., semi-deciduous as P1a
	b – evergreen	algae, lichens, and mosses are treated as P2a
P2 – host plant form	**a** – herbaceous	
	b – shrub	
	c – small tree	
	d – tree	
P3 – caterpillar position	**a** – leaf	trunk is treated as P3b; P3c explicitly refers to a position on a known host plant; ground and urban environments are treated as P3e
	b – stem/twig/branch	
	c – near root	
	d – flowering parts	
	e – inapplicable	
P4 – congregation	**a** – single	P4b - several specimens on one hostplant; P4d excludes silk threads
	b – multiple	
	c – palisade	
	d – cluster	
	e – web	
	f – procession	
C1 – **Lx** season	**a** – early to mid-spring	conjectural without experience or rearing, states represent approximations and highest encounter frequency
	b – mid to late spring	
	c – early to mid-summer	
	d – mid to late summer	
	e – fall	
	f – winter	
C2 – activity	**a** – diurnal	conjectural, refers to observed feeding. C2c covers ambivalence
	b – nocturnal	
	c – cathemeral	
C3 – subregion*	**a** – Palearctic	
	b – Nearctic	
	c – Neotropical	
	d – Afrotropical	
	e – Australian tropical	
	f – Oriental	
	g – Andean	
	h – Afrotemperate	
	i – Neoguinean	
	j – Australian temperate	
	k – Neozelandic	

*__H__ – head capsule descriptor, **B** – body descriptor (T – thoracic segments, A – abdominal segments); **M** – movement-related descriptor (A_x – abdominal segment with the corresponding ordinal number); **I** – integument descriptor (diagnostically important pigment distribution traits); **P** – host plant descriptor; **C** – conditions descriptor (diagnostically important host plant and sample-related conditions traits); **Ch. –** character.

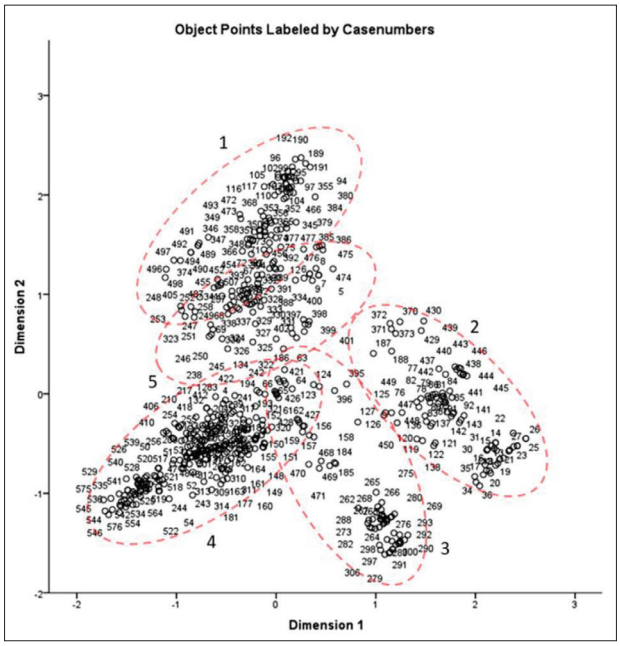


Fig. 1. CATPCA ordination of cases (582 morphological and resource polyphenisms derived from 83 exophytophagous caterpillar species) in PC1 (Dimension 1) and PC2 (Dimension 2); dashed line (with the associated clockwise numbering) indicates the observed grouping pattern.

with categorical data (see [49] for details and background). Exploration of the data set was performed by setting the k-value at 5 (the number of a random set of rows that serve as the initial modes), based on the grouping pattern of the plotted object scores obtained from the CATPCA analysis, as demonstrated in Fig. 1.

RESULTS

Character extraction

To standardize the descriptors, all descriptive accounts were reduced to appropriate coding. Table 1 offers an extensive overview of the macromorphological, behavioral and autecological traits observable in field conditions and applicable to the entire taxonomic assemblage addressed in this study. To accompany the generalized overview, all palpable descriptors were visually validated through representative species, which were further used for cross-character profiling in the data matrix. Head capsule, integument, and pigment distribution, as well as movement and resting posture-related traits, together with conventional marking topography and basic autecology, are represented through Figs. 2, 3, 4 and 5.

Ordination and classification of the derived polyphenism cases

The CATPCA model accounts for 33.756% of the variance through two principal components, 18.618% and 15.138%. According to the general rule of thumb for Cronbach's α, internal consistency with a 0.951 score indicates high reliability.

The PC1 ordination axis presented the gradient pattern between the highly condensed groups 1, 4 and 5 that represent the largest number of the analyzed items versus the more scattered and isolated groups 2 and 3 distributed on the lower-right portion of the plot. The refinement of the PC2 ordination axis primarily affects the first group and its overlap with groups 4 and 5.

The component loadings suggest the most influential variance contributors (Supplementary Table S2) belong to the morphological subclass (Figs. 2-5), namely HC hairiness (Fig. 2), segment and dorsoventral folding (Fig. 3), the presence of prolegs (Fig. 4) and resting posture for PC1, and general appearance of both HC and integument, HC pigment schemes and the presence of integument setae for PC2.

The plotted component loadings display the expected positive correlation between causal body characters such as setae presence, setal form and setae-bearing structure, their correlation to the hairiness of the head capsule, and the clearly inversed correlation with the existing HC pigment schemes. Apart from the many logical connections, such as the one between the number of prolegs and the movement patterns that are confirmed through the visualization, some of the relationships are more peculiar. This is especially noticeable with the subgroup of variables that describe group 5 cases: indices of positive correlation between C2 (activity), I6 (lateral marking; Fig. 5) and P4 (congregation).

Using the distinguishable clumps of cases in the CATPCA biplot as an argument for k-modes provided a new labeling variable for further composition inspection. The five clusters obtained through the added k-modes value are coherent with previous

Fig. 2. Head capsule characters (H), following the order of appearance: (1a) *Cucullia verbasci*, , (1b) *Lithophane socia*, (1c) *Carcharodus alceae*, (2a) *Colotois pennaria*, (2b) *Noctua comes*, (3a) *Cerura vinula*, (3b) *Diloba caeruleocephala*, (3c) *Antocharis cardamines*, (3d) *Euthrix potatoria*, (4a) *Cucullia lactucae*, (4b) *Saturnia pavonia*, (4c) *Agriopis bajaria*, (4d) *Vanessa atalanta*, (5a) *Cosmia affinis*, (5b) *Lycia hirtaria*, (5c) *Drymonia rufficornis*, (5d) *Mimas tiliae*, (6a) *Orthosia gracilis*, (6b) *Erannis defoliaria*, (6c) *Papilio machaon*, (7a) *Opheropera brumata*, (7b) *Chrysodeixis chalcites*, (8a) *Dasycorsa modesta*, (8b) *Apatura ilia*, (8c) *Catocala fulminea*, (8d) *Polygonia c-album*, (9a) *Asphalia ruficollis*, (9b) *Phlogophora meticulosa*, (9c) *Calyptra thalictri*, (9d) *Mythimna conigera*.

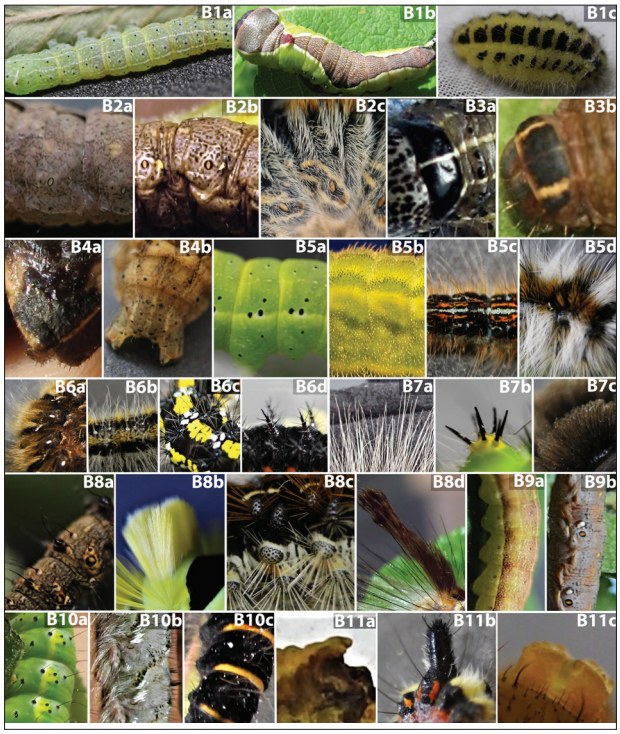


Fig. 3. Characters describing thoracic and abdominal segments (B) in accordance with the order of appearance: (1a) *Orthosia cerasi*, (1b) *C. vinula*, (1c) *Zygaena filipendulae*, (2a) *Polyphenis sericata*, (2b) *C. pennaria*, (2c) *Lasiocampa trifolii*, (3a) *C. affinis*, (3b) *Orthosia cruda*, (4a) *Agrius convolvuli*, (4b) *Mythimna unipuncta*, (5a) *Polymixis rufocincta*, (5b) *Polyommatus daphnis*, (5c) *Malacosoma castrensis*, (5d) *Eriogaster catax*, (6a) *Lasiocampa quercus*, (6b) *Aporia crataegi*, (6c) *Callimorpha dominula*, (6d) *Argynnis aglaja*, (7a) *Arctia caja*, (7b) *S. pavonia*, (7c) *Amata phegea*, (8a) *Phigalia pilosaria*, (8b, 8d) *Orgyia antiqua*, (8c) *Rhyparia purpurata*, (9a) *Agrochola lychnidis*, (9b) *C. pennaria*, (10a) *D. caeruleocephala*, (10b) *Dendrolimus pini*, (10c) *Macrothylacia rubi*, (11a) *Thyatira batis*, (11b) *Acronicta psi*, (11c) *Zerynthia polyxena*.

assumptions, but also informative in terms of the discriminatory ability of the categorical variables and the reasoning behind the overlap of the clusters and the distance between individual cases.

According to the k-mode analysis, group 4 matches cluster 1 (Fig. 6A), and partially overlaps with cluster 3. It includes smooth noctuid species, usually with a sketched mediodorsum and without a specialized diet preference. The large number of cases in the cluster is the result of many autecological varieties within the group. The second cluster (comparable to group 3) encompasses morpho-eco-types with a prominent projection as the main feature of the thoracic and abdominal segments, pattern pigment schemes and often a canopy habitat. Cluster 3 partially matches group 5

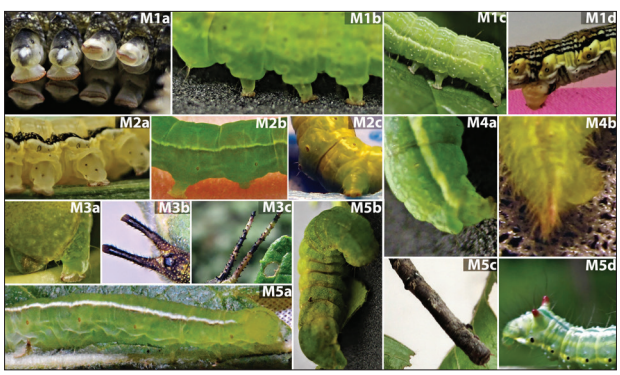


Fig. 4. Movement and prolegs related characters in accordance with the order of appearance (M): (1a) *Xestia c-nigrum*, (1b) *Hypena proboscidalis*, (1c) *C. chalcites*, (1d) *E. defoliaria*, (2a) *Orthosia gothica*, (2b, 6a) *Scoliopteryx libatrix*, (2c) *Alsophila aescularia*, (3a, 6b) *P. meticulosa*, (3b) *Dicranura ulmi*, (3c) *C. vinula*, (4a) *Orthosia incerta*, (4b) *Melanargia galathea*, (6c) *Biston betularia*, (6d) *Ptilodon capucina*.

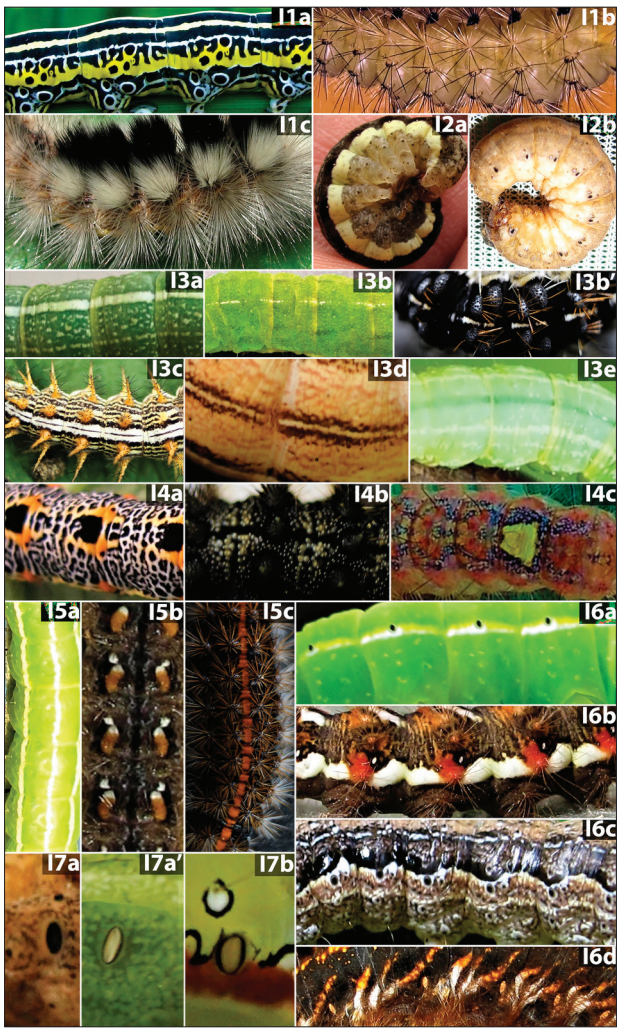


Fig. 5. Integument characters in two-dimensional space in accordance with the order of appearance (I) (character state codes with prime symbol stand for another variant of the same state): (1a) *Apopestes spectrum*, (1b) *Diaphora mendica*, (1c) *Dicallomera fascelina*, (2a) *Ammoconia caecimacula*, (2b) *Agrotis exclamationis*, (3a) *Cosmia pyralina*, (3b) *P. meticulosa*, (3b', 7a') *R. purpurata*, (3c) *Brenthis daphne*, (3d) *M. conigera*, (3e) *O. brumata*, (4a) *Artiora evonymaria*, (4b) *Melitaea arduinna*, (4c) *Subacronicta megacephala*, (5a) *Amphipyra tragopoginis*, (5b) *Eilema complana*, (5c) *Coscinia striata*, (6a) *Amphipyra pyramidea*, (6b) *Acronicta rumicis*, (6c) *Anorthoa munda*, (6d) *E. potatoria*, (7a) *Noctua pronuba*, (7b) *Xylena exsoleta*.

and is presented with the largest number of outliers, as well as with a significant overlap with other clusters (especially cluster 1) in the domain of integumental and autecological variation. Patterns and markings are rare among the cluster, but the constituents include both hairy and smooth appearances, making this cluster especially unsuitable for interpretation. Almost all cases with proleg modifications and tree or small tree feeding preferences are grouped within cluster 4 (coherent with group 2). The final cluster, which is partially compatible with group 1, almost exclusively consists of cases with full setae or scoli coverage and the same (diurnal) feeding habits.

The final relabeling introduced the family variable, and the plotted data (Fig. 6B) agrees with the general approach towards characterizing lepidopteran families, thus Erebidae, Noctuidae and Notodontidae are reported. Most of the characters handled in the descriptive matrix were previously used to create identification keys up to the family level but are not discussed as regards their discriminative abilities.

DISCUSSION

Most of the characters used in the descriptive matrix were previously used to create identification keys up to the family level but were not discussed with respect to their discriminative abilities. As an example, Stehr [23] coupled H1c with the family Hesperiidae but did not provide insight into the other variants of this character state or the morphological traits accompanying the particular appearance. Rather than representing a state, the narrow T1 segment was treated as an unusual feature. Uneven discrimination of taxa based on a single distinctive feature was applied to all characters and within all the existing macromorphological traits. The term hairiness is often entangled with setae categorization [16], which can overcomplicate the macromorphology-based identification (characters B1 and B2). Due to the morphoanatomical properties of the caterpillar's integument, each seta starts its growth from a raised, sclerotized, or modified surface [27]. Since this is not easily observable, the state B6a is introduced as a morphologically less precise, but less demanding choice. Caterpillars' small dimensions and invasive collecting methods could limit detailed inspection during fieldwork, but the rearing observations suggest that locomotion and grasping patterns could be used as discriminative characters if properly described. Some examples include specific anterior segment stretching common in some noctuids, and the tight grasp observed in lasiocampids.

Higher resolution in macromorphology-based identification often requires the use of double-natured

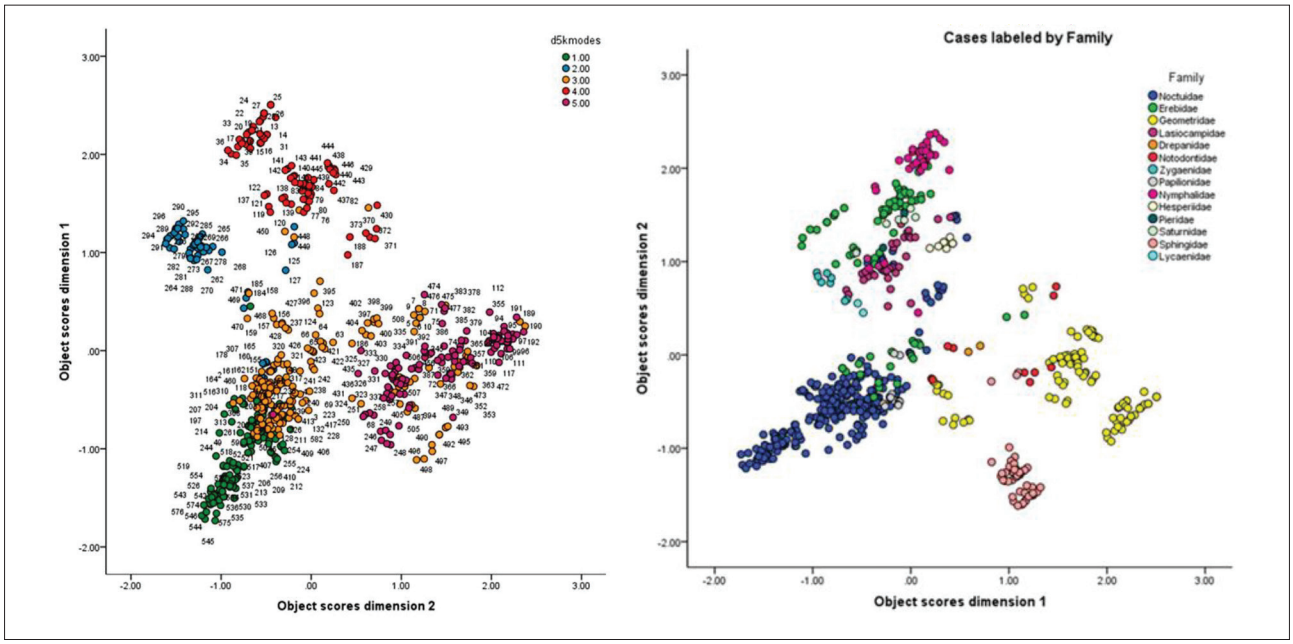


Fig. 6. A – CATPCA: two-dimensional ordination of the derived polyphenism cases by case numbers, color labeled by k-modes cluster membership value (d5kmodes). **B** – CATPCA: two-dimensional ordination of derived polyphenism cases in PC1 (Dimension 1) and PC2 (Dimension 2), color labeled by taxonomic affiliation value (family).

descriptors [50]. Pigment distribution can either confirm the identification or be extremely variable, unreliable and biased. Pigments do, however, play a significant role in empirical identification. Exophytophagous caterpillars offer many pigmentation details that could be incorporated into keys if concisely presented or vectorized for computer-aided identification. More recent expert literature [28, 29] replaces the color description with image plates. Future identification tools could find the presented pigmentation characters particularly useful, as they stand between traditional descriptions and large-scale pictorials. The correlation between activity interval, lateral markings and congregation observed in the disjointed cluster 3 can be inspected individually for each case until a potential pattern is exposed; however, such implications are beyond the scope of this paper.

Autecology and behavior characters, as interpreted herein, were not previously brought to a descriptor level for the Macrolepidoptera. Although the list of possible characters is not final, it can reveal the patterns among the taxonomically distant species that occupy the same niche, including elevation ranges or habitat specification, preferably in accordance with the standardized hierarchies, such as EUNIS, which could provide options for data modeling. True and partially exophytophagous caterpillars are versatile, most frequently encountered by non-experts and very convenient for this approach. The developed descriptive matrix brings together a heterogenous species association, with uneven distribution of taxonomic categories and different levels of variability. If the intention of the author is to create an identification key, the final taxa would be selected based on a high similarity, and thus more appropriate for discussions of the character's importance, plasticity, and correlation.

CONCLUSIONS

Unorganized morphological data and the omission of microhabitat and lifecycle details affect identification accuracy. Phenotypic diversity is often a norm in generalists, as their appearance reflects the life form rather than the taxonomic affiliation. The mechanisms through which macromorphology, autecology and behaviorism interact in survival strategies of Macrolepidoptera are not yet fully understood. The descriptive matrix that was developed employing a variety of traits and with respect to their dynamics is an exemplary model of data interpretation that can enable the analysis of these interactions.

Funding: This study was supported by the Ministry of Education, Science and Technological Development of the Republic of Serbia, Contract No.: 451-03-47/2023-01/200124.

Acknowledgments: The authors would like to thank Prof. Dr Đurađ Milošević for his immense help with the statistical procedures. Assistance with the photography of the database by Ivan Tot and Boženka Hric (HabiProt) is greatly appreciated. We extend our thanks to Aleksandra Petrović, Iva Stojanović, Marija Krstić and Valentina Krstić, students of the Faculty of Sciences and Mathematics, University of Niš, for their technical assistance.

Author contributions: Conceptualization, VŽ, and AT; methodology, SS and ML; software, AT, and MP; formal analysis, AT; investigation, MIM, SS, and ML; resources, ML, MIM, SS and MP; data curation, AT; writing, original draft preparation, AT; writing, review and editing, VŽ and MP; visualization, ML and MP; supervision, VŽ. All authors have read and agreed to the published version of the manuscript.

Conflict of interest disclosure: The authors have no financial or non-financial interests or conflicts to disclose.

Data availability statement: All data underlying the reported findings are available in a publicly accessible repository. The data presented in this study are openly available in Mendeley Data at: https://doi.org/10.17632/jg8wcxw4ws.2, and at: https://www.narcis.nl/dataset/RecordID/oai%3Aeasy.dans.knaw.nl%3Aeasy-dataset%3A271624.

REFERENCES

1. Capinera, JL. Handbook of Vegetable Pests, 1st ed. San Diego, California: Academic Press; 2001. Order Lepidoptera: Caterpillars, Moths, and Butterflies; pp. 353-510. https://doi.org/10.1016/B978-012158861-8/50011-9
2. Van Nieukerken EJ, Kaila L, Kitching IJ, Kristensen NP, Lees DC, Minet J, Mitter C, Mutanen M, Regier JC, Simonsen TJ, Wahlberg N. Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness. Zootaxa. 2011;3148(1):212-21. https://doi.org/10.11646/zootaxa.3148.1.41
3. Clark BR, Godfray HC, Kitching IJ, Mayo S, Scoble MJ. Taxonomy as an eScience. Philos Trans Royal Soc. 2009;367(1890):953-66. https://doi.org/10.1098/rsta.2008.0190
4. Dyar HG. The number of molts of lepidopterous larvae. Psyche. 1980;5(175-176):420-2. https://doi.org/10.1155/1890/23871
5. Dyar HG. A Classification of lepidopterous larvae. Ann N Y Acad Sci. 1984;8(1):194-232. https://doi.org/10.1111/j.1749-6632.1894.tb55421.x
6. Dyar HG. Descriptions of the larvae of some moths from Colorado. Proc U S Natl Mus. 1902;25 (1290):369-412. https://doi.org/10.5479/si.00963801.25-1290.369
7. Dyar HG. Life Histories of North American Geometridae - XLII. Psyche. 1903;10(325-326):190-200. https://doi.org/10.1155/1903/60850
8. Frohawk, FW. Natural History of British Butterflies, Vol. 2, 1st ed. London: Hutchinson & Co.; 1924.
9. Hensley SD. Comparative Study of the Immature Stages of Three Species of the Diatraea Complex. [disseration]. [Oklahoma]: Oklahoma State University; 1960. 90 p.
10. Barbier R. Recherches sur la morphologie tegumentaire d'un insecte holométabole: Galleria mellionella L. (Lepidoptera, Pyralidae). [disseration]. [Rennes]: Université de Rennes; 1971. 205 p.
11. Fink LS. Color polymorphism in Sphingid caterpillars (Lepidoptera: Sphingidae). [disseration]. [Gainesville, Florida]: University of Florida; 1989. 273 p. https://doi.org/10.5962/bhl.title.115051
12. Capps, HW. Keys for the identification of some lepidopterous larvae frequently intercepted at quarantine. United States Department of Agriculture, Agricultural Research Service; 1956. 37 p. https://doi.org/10.5962/bhl.title.67164
13. Gerasimov, AM. Caterpillars. Fauna of the USSR. Lepidoptera Insects. Moskva, Leningrad: Zoological Institute, Academy of Sciences USSR, 1952. 338 p.
14. Hinton HE. The larvae of the Lepidoptera associated with stored products. Bull Entomol Res. 1943;34(3):163-212. https://doi.org/10.1017/S0007485300023701
15. Aitken AD. A key to the larvae of some species of Phycitinae (Lepidoptera, Pyralidae) associated with stored products, and of some related species. Bull Entomol Res. 1963;54(2):175-88. https://doi.org/10.1017/S0007485300048720
16. Carter DJ, Kristensen NP. Classification and keys to higher taxa. Vol. 1, Lepidoptera, Moths and Butterflies. Berlin: Walter de Gruyter; 1998. 358 p. https://doi.org/10.1515/9783110804744.27
17. Gorham JR. Insect and mite pests in food: An illustrated key. US Department of Agriculture, Agricultural Research Service; 1991. 665 p.
18. Weisman DM. Keys for the identification of some frequently intercepted lepidopterous larvae. US Department of Agriculture, APHIS; 1986. 64 p. https://doi.org/10.5962/bhl.title.137947
19. Solis M. Key to selected Pyraloidea (Lepidoptera) larvae intercepted at US ports of entry: revision of Pyraloidea in "Keys to some frequently intercepted Lepidopterous larvae" by Weisman 1986. Washington, D.C.: USDA Systematic Entomology Laboratory; 2006. 58 p.
20. Peterson A. The Larvae of Insects, Part I, Lepidoptera and Hymenoptera. Ohio: Columbus; 1948. 315 p.
21. Hinton HE. On the homology and nomenclature of the setae of lepidopterous larvae, with some notes on the phylogeny of the Lepidoptera. Trans R Entomol Soc. 1946;97(1):1-37. https://doi.org/10.1111/j.1365-2311.1946.tb00372.x
22. Godfrey GL, Jahn OL, Fleming, E. A review and reclassification of larvae of the subfamily Hadeninae (Lepidoptera, Noctuidae) of America north of Mexico. Washington, D.C.: U.S. Dept. of Agriculture; 1972. 265 p.
23. Stehr FW. Immature Insects. Vol. 1. Iowa: Kendall/Hunt Publishing Company; 1987. 770 p.
24. Marquis RJ, Passoa CS, Lill TJ, Whitfield BJ, Le Corff J, Forkner ER, Passoa AV. Illustrated guide to the immature Lepidoptera on Oaks in Missouri. Morgantown, WV: U.S.

Department of Agriculture, Forest Service, Forest Health Assessment and Applied Sciences Team; 2019. 369 p.
25. Kitching IJ, Rawlins JE. The Noctuoidea. In: Kristensen NP, editor. Handbook of Zoology, Lepidoptera, Moths and Butterflies, Vol. 1: Evolution, Systematics and Biogeography. Berlin: W. de Gruyter; 1999. p. 355-402. https://doi.org/10.1515/9783110804744.355
26. McGuffin WC. Guide to the Geometridae of Canada (Lepidoptera). III, IV and V. Subfamilies Archiearinae, Oenochrominae, and Geometrinae. Mem Ent Soc Can. 1998;120(S145):3-56. https://doi.org/10.4039/entm120145fv
27. Kristensen NP, editor. Lepidoptera, moths and butterflies. Vol. 1: Evolution, Systematics, and Biogeography. New York: Walter de Gruyter; 1999. 491 p. (Handbook of Zoology; Vol. 4; Part 35).
28. Beck H. Die Larven der europäischen Noctuidae - Revision der Systematik der Noctuidae (Lepid., Noctuidae) Vol. 1 and 2, Herbipoliana 5. Marktleuthen: Verlag Dr. Ulf Eitschberger; 1999. 859 p.
29. Beck H. Die Larven der europäischen Noctuidae - Revision der Systematik der Noctuidae (Lepid., Noctuidae). Vol. 3 and 4, Herbipoliana 5. Marktleuthen: Verlag Dr. Ulf Eitschberger; 2000. 525 p.
30. Ahola M, Silvonen K. Larvae of Northern European Noctuidae. Vol. 3. Finland: KuvaSeppala Yhtiot Oy; 2010. 672 p.
31. Wagner DL, Schweitzer DF, Sullivan JB, Reardon RC. Owlet Caterpillars of Eastern North America (Lepidoptera: Noctuidae). Princeton and Oxford: Princeton University Press;2011. 576 p. https://doi.org/10.1515/9781400838295
32. Marquis RJ, Lill TJ, Forkner ER, Le Corf J, Landosky MJ, Whitfield BJ. Declines and Resilience of Communities of Leaf Chewing Insects on Missouri Oaks Following Spring Frost and Summer Drought. Front Eco Evol. 2019;7:396. https://doi.org/10.3389/fevo.2019.00396
33. Ahola M, Silvonen K, Kronholm H, Lundsten KE, Sundell, PR. Description of the larva of Xestia atrata (Morrison, 1874) (Lepidoptera: Noctuidae) with notes on the biology of the species. Entomol Fenn. 2006;17(2):65-72. https://doi.org/10.33338/ef.84289
34. Singh GP, Goel SC. Body chaetotaxy of three noctuid caterpillars. (Lepidoptera). Deut Entomol Z. 1987;34(4-5):373-81. https://doi.org/10.1002/mmnd.4800340414
35. Duarte M, Robbins RK. Immature stages of Calycopis bellera (Hewitson) and C. janeirica (Felder) (Lepidoptera, Lycaenidae, Theclinae, Eumaeini): Taxonomic significance and new evidence for detritivory. Zootaxa. 2009;39-61. https://doi.org/10.11646/zootaxa.2325.1.4
36. Ahola M, Kuisma M, Leinonen R, Saarenmaa H, Silvonen K. Description of immature stages of Xestia brunneopicta (Matsumura, 1925), with a key to the mature larvae of the European species of Xestia (Pachnobia) (Lepidoptera, Noctuidae). Nota Lepidopterol. 2015;38(1):75-88. https://doi.org/10.3897/nl.38.4621
37. Pires C, Marinoni L. DNA barcoding and traditional taxonomy unified through integrative Taxonomy: A view that challenges the debate questioning both methodologies. Biota Neotrop. 2010;10(2):339-46. https://doi.org/10.1590/S1676-06032010000200035
38. Sharkey M, Brown B, Baker A, Mutanen M. Response to Zamani et al. (2020): The omission of critical data in the pursuit of "revolutionary" methods to accelerate the description of species. Zookeys. 2021;1033:191-201. https://doi.org/10.3897/zookeys.1033.66186
39. Carter DJ. Pest Lepidoptera of Europe: with special reference to the British Isles. Vol. 31, Springer Science & Business Media; 1984. 438 p.
40. Eichlin TD, Cunningham HB. The Plusiinae (Lepidoptera: Noctuidae) of America north of Mexico, emphasizing genitalic and larval morphology: U.S. Department of Agriculture, Agricultural Research Service. 1978; 122 p.
41. Turčáni M, Patočka J, Kulfan J. How to identify larvae of the protected species: Dioszeghyana schmidtii (Diószeghy 1935) and survey its presence and abundance (Lepidoptera: Noctuidae; Hadeninae). J For Sci. 2010;56(3):121-9. https://doi.org/10.17221/55/2009-JFS
42. Gilligan TM, Passoa SC. LepIntercept - An identification resource for intercepted Lepidoptera larvae. Identification Technology Program (ITP). 2014- [cited 2023 Jan 14]. Available from: https://idtools.org/id/leps/lepintercept/index.html
43. Anderson S, Majka, G. Biodiversity and biosystematic research in a brave new 21st century information-technology world. Zookeys. 2009;22:1-4. https://doi.org/10.3897/zookeys.22.222
44. Rennwald E, Rodeland E. Lepiforums-Europaliste Lepiforum's Checklist of European Lepidoptera [Internet]. Version 6, January 1st, 2019- [cited 2023 Jan 15]. Available from: https://lepiforum.org/
45. iNaturalist.org web application. [Internet]. [cited 2022 Jun 10]. Available from: http://www.inaturalist.org
46. Popović M, Vasić N, Koren T, Burić I, Živanović N, Kulijer D, Golubović A. Biologer: An open platform for collecting biodiversity data. Biodivers Data J. 2020;8:e53014. https://doi.org/10.3897/BDJ.8.e53014
47. HabiProt Alciphron - database about the insects of Serbia. [Internet]. 2014- [cited 2021 Nov 1]. Available from: https://alciphron.habiprot.org.rs
48. Moyle D, Byrne C. The Caterpillar Key. [Internet]. 2020- [cited 2022 Jun 25]. Available from: https://keys.lucidcentral.org/keys/v3/the-caterpillar-key/index.html
49. Huang Z. Clustering categorical data with k-Modes. In: Wang J, editor. Encyclopedia of Data Warehousing and Mining, 2nd ed. New York: IGI Global; 2009, p. 246-50. https://doi.org/10.4018/978-1-60566-010-3.ch040
50. Futahashi R, Banno Y, Fujiwara H. Caterpillar color patterns are determined by a two-phase melanin gene prepatterning process: new evidence from tan and laccase2. Evo Dev. 2010;12(2):157-67. https://doi.org/10.1111/j.1525-142X.2010.00401.x

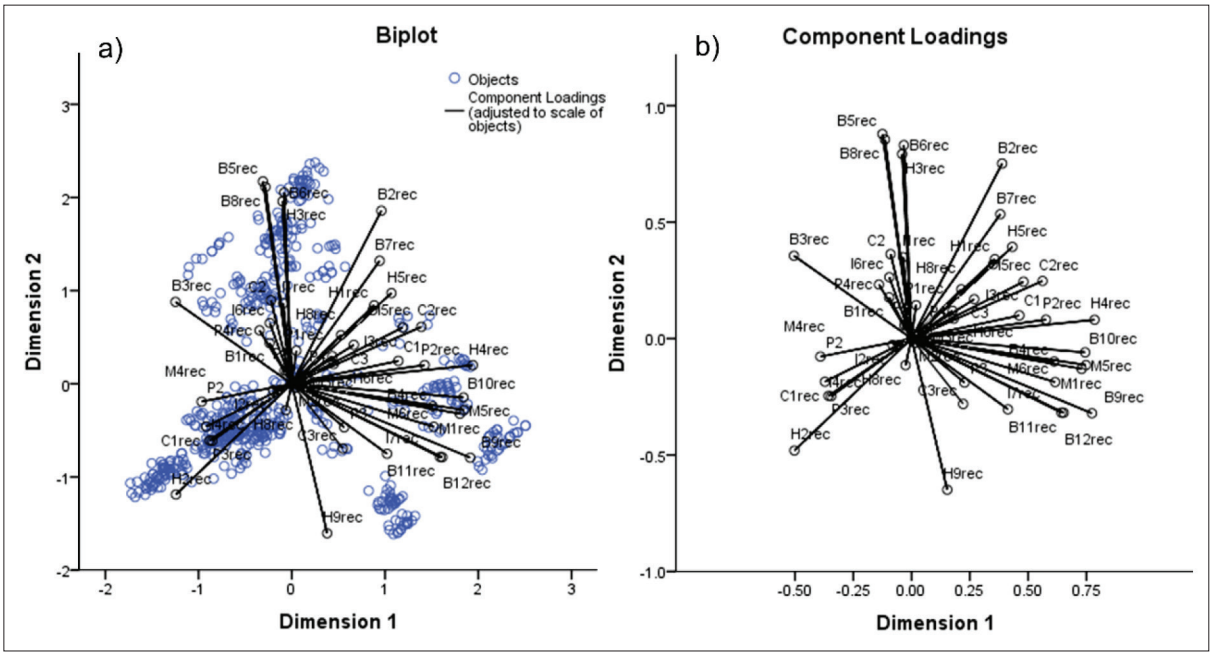
SUPPLEMENTARY MATERIAL

Supplementary Table S1. The number of species that entered the profiling in the descriptive matrix and the number of polyphenism cases described with regard to their taxonomic affiliation.

Family	No. of species	No. of polyphenisms
Noctuidae	29	251
Erebidae	14	71
Geometridae	10	92
Lasiocampidae	7	34
Notodontidae	4	10
Drepanidae	2	2
Saturniidae	1	5
Sphingidae	2	48
Nymphalidae	7	38
Pieridae	2	2
Papilionidae	2	11
Lycaenidae	1	2
Hesperiidae	1	8
Zygaenidae	1	8

Supplementary Table S2. CATPCA character-wise loadings for PC1 and PC2 and total variance accounted for; values in bold are highly loaded and responsible for higher resolution on the biplot.

Character	Component Loadings		Total VAF
	1	2	
H1rec	0.375	0.316	0.240
H2rec	-0.528	-0.434	0.467
H3rec	0.005	**0.806**	0.649
H4rec	**0.801**	0.074	0.647
H5rec	0.452	0.359	0.333
H6rec	0.203	0.089	0.049
H8rec	-0.011	-0.124	0.016
H8rec	0.207	0.19	0.079
H9rec	0.134	**-0.669**	0.465
B1rec	-0.068	0.209	0.049
B2rec	0.421	**0.704**	0.673
B3rec	-0.481	0.341	0.348
B5rec	-0.09	**0.888**	0.797
B6rec	0.006	0.833	0.694
B7rec	0.418	0.486	0.411
B8rec	-0.072	0.87	0.761
B9rec	**0.766**	-0.34	0.702
B10rec	**0.745**	-0.087	0.563
B4rec	0.607	-0.116	0.382
B11rec	0.625	-0.324	0.496
B12rec	0.633	-0.327	0.508
M1rec	**0.736**	-0.219	0.589
M2rec	0.01	-0.026	0.001
M3rec	0.063	0.007	0.004
M4rec	-0.022	0.059	0.004
M5rec	**0.748**	-0.198	0.599
M6rec	0.627	-0.265	0.464
I1rec	-0.038	0.347	0.122
I3rec	0.49	0.271	0.313
I4rec	-0.333	-0.277	0.187
I5rec	0.329	0.357	0.236
I6rec	-0.115	0.314	0.112
I7rec	0.4	-0.28	0.239
I2rec	-0.073	-0.023	0.006
P1rec	0.029	0.138	0.020
P2rec	0.586	-0.019	0.344
P3rec	-0.392	-0.241	0.212
P4rec	-0.118	0.246	0.075
C1rec	-0.59	-0.259	0.415
C2rec	0.573	0.371	0.466
C3rec	0.192	-0.256	0.103



Supplementary Fig. S1. CATPCA: **a** – relationships of the polyphenism cases and the categorical variables (characters); **b** – component loadings and ordination within PC1 and PC2; see Tables 2,3,4,5,6 for abbreviations, rec – recoded.