Vol.66: e23220682, 2023
https://doi.org/10.1590/1678-4324-2023220682
ISSN 1678-4324 Online Edition

*Article - Biological and Applied Sciences*

# An Anatomical Ontology for the Class Collembola (Arthropoda: Hexapoda)

**Luis Antonio González Montaña**[1]*
https://orcid.org/0000-0002-9136-9932

**Diana Rueda-Ramírez**[2]
https://orcid.org/0000-0002-7642-2256

**Francisco Javier Serna Cardona**[3]
https://orcid.org/0000-0002-6143-9821

**Andreas Gaigl**[3]
https://orcid.org/0000-0003-0410-8644

[1]Universidad de los Llanos, Facultad de Ciencias Básicas e Ingeniería, Villavicencio, Meta, Colombia; [2]Humboldt University of Berlin, Institute of Biology, Ecology Group,Berlin, Germany; [3]Universidad Nacional de Colombia, Facultad de Ciencias Agrarias, Bogotá, Colombia.

Editor-in-Chief: Paulo Vitor Farago
Associate Editor: Paulo Vitor Farago

Received: 04-Sep-2022; Accepted: 19-Dec-2022

*Correspondence: luis.gonzalez.montana@unillanos.edu.co; Tel.: +57-3052119037 (L.A.G.M.).

**HIGHLIGHTS**

- A first ontology is developed for the Class Collembola (CLAO)
- CLAO holds 1551 anatomical terms for anatomical systems, chaetae, sclerites, among others.
- CLAO is directed to morphological descriptions and production of semantic annotations for the Class Collembola.
- CLAO is directed to the recognition of homologues structures and homology.

**Abstract:** Communication in science requires standardized terminology with concepts unified that facilitate the processing and exploration of information in any knowledge domain. The morphology is not the exception; however, it has challenged problems, called "the linguistic problem of the morphology", which is related to the processing of morphological data result of taxonomic work. Ontologies, standardized vocabularies expressed through language parsable (Resource Description Framework, RDF) by machines, are proposed to resolve the linguistic problems in morphology. Springtails are dominants in the soil and other types of habitats with ca 9000 described species. The anatomical terms employed in Collembola are not exempt from problems such as the presence of homonyms related to the chaetotaxy, idiosyncratic language employed in morphological descriptions, and lack of consensual definitions of anatomical terms, which difficult the comparison of morphological data. Collembola Anatomy Ontology (CLAO) is built according to principles established by The OBO Foundry and includes 1554 anatomical terms for the region of the cuticle, appendages, anatomical systems, and cells. CLAO is directed toward morphological descriptions and the production of phenotypic data produced from taxonomic and anatomical works before the obtaining of homologies in a phylogenetic framework. Also, to increase the interoperability with other anatomical ontologies for Arthropoda and knowledge domains.

**Keywords:** Anatomical ontology; Entomobryiomorpha; Neelipleona; Symphypleona; Poduromorpha; Resource Framework Description (RDF).

## INTRODUCTION

Processing and exploring morphological data involve standardized terminology and unified concepts to improve the communication of information [1–4]. However, this communication has obstacles and challenges as result of the lack of morphological terminology free of evolutionary and functional assumptions, standardized, and taxon-independent [3,5–7].

Anatomical ontologies are a solution to overcome these problems and whose formalization in morphological descriptions has reached an important development [3,8–12]. Anatomical ontologies are still scarce and only available for certain taxa, although, supported by the development of new computational tools for database building to store phenotypic data (see Glossary, Table 1) and new techniques for anatomical studies [4,11,13].

An anatomy ontology is a computable representation of the body, organs, organ parts, and tissues of an organism, including the properties and relationships among those parts [2,14–15]. This entails the usage of a structural language with a formal syntax, Resource Description Framework (RDF), a triple statement (subject-predicate-object), related to a unique uniform resource identifier (URI) [16]. The resources will be classes and subclasses, whose “parent-child” relationships constitute the hierarchical structure exhibited by the ontology (Figure 1).

The specification of classes and relationships between them allows to infer some conclusions from it, for example, assumptions of phenotypic instances or gaps from semantic-based morphological descriptions through reasoning techniques [2,17]. In the following RDF, where the subject and object are in bold and predicate in italic, “antenna is *part_of* head and head is *part_of* tagma” can be inferred that “antenna is *part_of* tagma”, however, the antenna could have other parental classes specified by some relationship, e.g., “antenna *is* an appendage”.

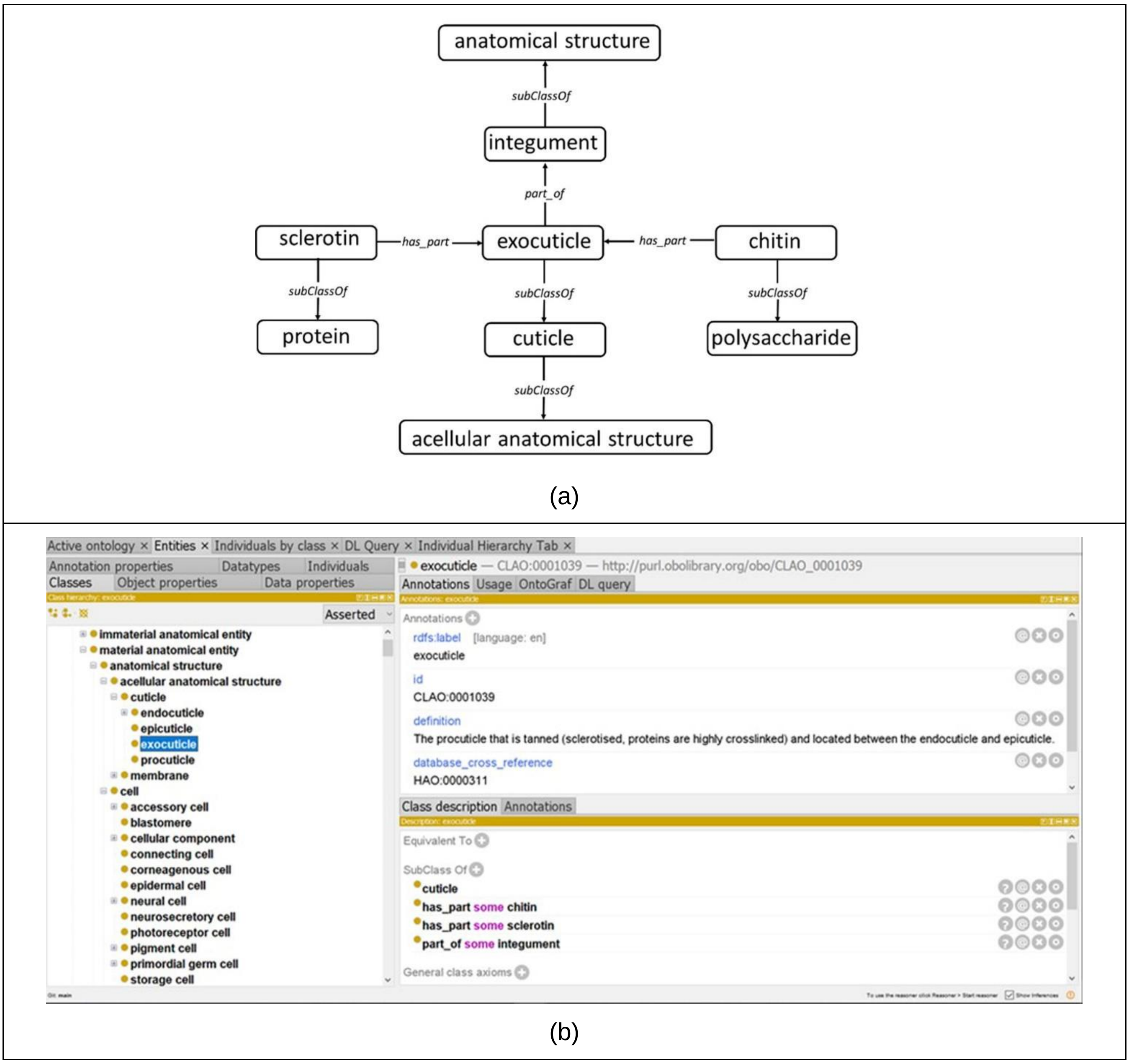
**Table 1.** Glossary of terms employed in this paper.

| Term | Definition |
|---|---|
| Ontology | Coding of knowledge domain in such a way that the knowledge can be understood by a computer (modified from [16]). |
| anatomical ontology | Coding of knowledge about anatomy. |
| RDF | RDF is a framework for representing information on the Web [16]. |
| class | A term defined in an ontology representing a concept [17]. |
| subclass | A term in an ontology that narrows down another (parent) term [17]. |
| instance | The individuals in the class extension. |
| resource | A resource is anything that is being described by RDF statements [16]. |
| annotation | Statements composed of ontology terms, linked to natural language descriptions such as characters and states [17]. |
| owl | A language standardized by the World Wide Web Consortium (WC3) for defining DL ontologies [17]. |
| database | A structured collection of data, usually organized as multiple data tables linked via identifiers into relational databases [28]. |
| term | A word that names or labels a particular concept as part of the specialized vocabulary of a field [28]. |
| granularity | Representation on different levels of detail in data, information, and knowledge that are located at their appropriate level [29]. |
| terminology | The body of terms and concepts used with a particular application in a subject of study, usually formalized in a thesaurus or ontology [28]. |
| URI | An unambiguous pointer to a unique resource on the Internet; used to refer to single terms of a thesaurus or ontology [28]. |

Anatomical ontologies for Arthropoda are scarce, i.e., Hymenoptera Ontology Anatomy (HAO) [5], Mosquito Gross Anatomy Ontology (TGMA) [18], *Drosophila* Gross Anatomy (FBBT) [19], Spider Ontology (SPD) [20], Tick Anatomy Ontology (TADS) [18], and anatomical systems as the Ontology of Arthropod Circulatory Systems (OARCS) [7]. Although each ontology has a scope or purpose, these have gained applicability in the development of semantic-based morphological descriptions [9–10,21–22], and homology reasoning in a phylogenetic framework [11,17,23], and where Collembola Anatomy Ontology (CLAO) follows the same principles [24].

The Class Collembola includes about 9000 species worldwide [25], occupying different habitats like soil, canopy, or seashores, where they are decomposers and regulators of biotic communities. The anatomical terminology employed in Collembola is not free of linguistic problems as the recurrent usage of synonyms (i.e., furcula, chaeta), homonyms (acronyms to name the chaetotaxy), and anatomical terms devoid of definitions. Also, anatomical terms employed in Collembola are highly variable because of different interpretations during the recognition of homologies.

Inclusion or not of homologies and functional assumptions in ontologies is constantly discussed [3,5,7 11–12,17], however, Collembola Anatomy Ontology (CLAO) is “neutral” in the inclusion of homology assumptions, where terms are defined only by their structural properties [26]. The homology assumptions make it difficult to identify what instances belong to a particular anatomical class, minimizing the applicability and interoperability between ontologies [5,7]. This neutrality is not strict in structural definitions, where homology criteria could be implicit when an anatomical entity is named [27].



- immaterial anatomical entity
- material anatomical entity
  - anatomical structure
    - acellular anatomical structure
      - cuticle
        - endocuticle
        - epicuticle
        - exocuticle
        - procuticle
      - membrane
    - cell
      - accessory cell
      - blastomere
      - cellular component
      - connecting cell
      - corneagenous cell
      - epidermal cell
      - neural cell
      - neurosecretory cell
      - photoreceptor cell
      - pigment cell
      - primordial germ cell
      - storage cell

exocuticle — CLAO:0001039 — http://purl.obolibrary.org/obo/CLAO_0001039

Annotations

rdfs:label [language: en]
exocuticle

id
CLAO:0001039

definition
The procuticle that is tanned (sclerotised, proteins are highly crosslinked) and located between the endocuticle and epicuticle.

database_cross_reference
HAO:0000311

Equivalent To

SubClass Of
cuticle
has_part some chitin
has_part some sclerotin
part_of some integument

General class axioms

**Figure 1.** Diagram of the internal structure of CLAO. **a)** each anatomical entity is related to another within or between levels of granularity. In the example, the exocuticle is composed of the molecules chitin and sclerotin and is part of the integument, **b**) screenshot from Protégé of the classification of the class **exocuticle** into the hierarchical structure within CLAO.

The goal of this paper is to describe the first anatomical ontology for the Class Collembola called Collembola Anatomy Ontology (CLAO), made available to the community of Collembola specialists. CLAO is an open anatomical glossary, where the anatomical terms can be imported or discussed, and the definitions reached through consensus by specialists. CLAO has the potential to integrate other knowledge domains or databases, e.g., molecular, and ecological data, and constituting a methodological tool that complements morphological descriptions from taxonomic work.

## MATERIAL AND METHODS

CLAO is designed according to the following steps [30–31]: 1. Scope and Concept Selection, 2. Literature Review and reuse of ontologies, 3. Ontology Creation, 4. Ontology Testing, and 5. Ontology Evaluation.

Step 1. CLAO is a multispecies ontology allowing queries, analyses, and description of the phenotypic variation in Collembola. Also, CLAO integrates other ontologies e.g. HAO, UBERON, AISM increasing the interoperability between anatomical ontologies.

Step 2. The anatomical terminology is compiled and curated manually from diverse sources: reused terms from anatomical ontologies available via http://www.ontobee.org and http://www.bioportal.biontology.org (Table 2), specialized anatomical and morphological literature as journals, books, or databases, and the anatomical glossary compiled by [25]. An anatomical term is a name for parts of organisms so that the terms related to phenotypes are omitted, likewise, these are defined according to their structural properties [26]. The preference of an anatomical term by an author, for instance, the use of the term “chaeta” and “setae” is common in the literature. To resolve the above, the anatomical term imported to CLAO was selected following the criteria of the term most frequent used [7], while secondary terms are included as synonyms. Anatomical terms devoid of structural definitions are defined for the first time or redefined to avoid evolutive or functional assumptions. These terms include generalized definitions to represent the morphological diversity exhibit by the species of Collembola. CLAO includes terms used in other taxa e.g., Hymenoptera, where its definitions are not applicable to Collembola, for example, **labial palp** (HAO:0000450) used in HAO is not fitted to the labial palp observed in Collembola, being a homonymy. However, the import of terms that are not exclusive of Collembola increase the interoperability with other ontologies, taxonomic scope in Arthropoda, and documentation of phylogenetic data related with assumptions of homology (see below).

Step 3. The relations between terms (classes and subclasses) were structured according to top-level foundational categories developed by the Multi-species Anatomy Ontology (UBERON) [15] and The Basic Formal Ontology (BFO) (https://basic-formal-ontology.org/). The classification of anatomical terms in this hierarchy follows structural definitions, granularity level, spatial boundaries, “bona fide”, and “fiat” [32]. CLAO employs OWL2 Language (Ontology Web Language) and Manchester Syntax (http://www.w3.org/TR/owl2-manchester-syntax/) to build axioms in Protégé 5.4.0 [33]. CLAO is available at http://www.ontobee.org and original files at https://github.com/luis-gonzalez-m. In this paper, the classes or subclasses are written in bold and relations in italics. Each class contains the following information, expressed in natural language through annotations, which are mandatory fields by The OBO Foundry (http://obofoundry.org):

Definition: Structural definitions of anatomical terms. Definitions follow the *genus-differentia* format [7].

Database_cross_reference: This annotation links other ontologies or databases where a term is found, e.g., http://www.wikipedia.com, http://collembola.org, HAO:0000450.

ID: Each class is represented by a unique identifier, e.g., dens (CLAO:0000877).

Has_exact_synonym: synonyms with the same meaning as the class name, e.g., chaeta and seta.

Has_relative_synonym: synonyms related to the class name, e.g., bristle.

Sensu: This annotation property defines a *sensu*, which is the combination of a bibliographic reference, a label, and a class (represented by verbatim and logical definitions) [5].

Contributor: An entity responsible for making contributions to the resource.

CLAO employs mainly the categorical relation *has_subclass* and the lexical relations *attached_to*, *continuous_with, has_part,* and *part_of*, which specify how the terms or parts are related [34]. The relations employed in CLAO are foundational and ruled by The *Basic Formal Ontology* (BFO) (https://basic-formal-ontology.org/). Finally, CLAO implements the Ontology Development Kit (ODK) [35] (https://github.com/INCATools/ontology-development-kit), which allows for automation in the building of ontologies.

Steps 4 and 5. The evaluation of ontologies is divided into verification and validation [36]. The verification of CLAO was evaluated through Hermit 1.4.8 [37] available at Protégé, which finds consistency, subsumption relations, and classification problems. Operations such as reasoner (logical validation and automatic classification), quality check, querying, and merge were run with ROBOT [38]. This tool generates a YAML output with violations as multiple definitions for an anatomical term, ID absent, among others, and employed during each ontology lifecycle. ROBOT has a command to verify violations of rules expressed by SPARQL SELECT. If the query produces results mean that violations are present [38]. Quality control check is run with the options “report” and “verify” in ROBOT [38]. Validation includes two approaches: evaluation by anatomy and arthropods experts, and application-based evaluation, which consists of the creation of description templates for organisms belonging to Collembola available at http://github/luis-gonzalez-m/Lepidocyrtus_RDF_store [39]. This evaluation allowed to check what requirements are essential in the building of CLAO. The quality assurance was conducted by The OBO Foundry (https://obofoundry.org), the organization that defined the principles in the building of ontologies in biological sciences.

## RESULTS

CLAO (v2021-09-27) holds 1516 classes, 26 object properties, and 1554 labels (terms). The anatomical terms refer to the region of the cuticle, appendages, cellular components, anatomical systems, parts of the embryo, cells, among others. 115 terms were imported from Hymenoptera Anatomy Ontology (HAO), 92 from Uberon Multi-species Anatomy Ontology (UBERON) [38], 76 from The Gene Ontology (GO) [40], and 26 from Ontology for the Anatomy of the Insect SkeletoMuscular system (AISM) [41] (Table 2). Quality control check of CLAO was run with “report” (Table 3) find violations related to the definitions missing, which were fixed during the lifecycle of CLAO. An example of verification is shown in the Table 4 with the subclasses for the term tagma **(**CLAO_0000952).

**Table 2.** Imported terms from other ontologies and used in the development of CLAO.

| Ontology | No. imported terms | URL |
|---|---|---|
| Uberon Multi-Species Anatomy Ontology (UBERON) | 92 | http://purl.obolibrary.org/obo/uberon.owl |
| Hymenoptera Anatomy Ontology (HAO) | 115 | http://purl.obolibrary.org/obo/hao.owl |
| Mosquito gross anatomy ontology (TGMA) | 27 | http://purl.obolibrary.org/obo/tgma.owl |
| Drosophila gross anatomy (FBbt) | 35 | http://purl.obolibrary.org/obo/fbbt.owl |
| Tick Anatomy Ontology (TADS) | 2 | http://purl.obolibrary.org/obo/tads.owl |
| Foundational Model of Anatomy Ontology (FMA) | 9 | http://purl.obolibrary.org/obo/fma.owl |

The anatomical terms included in CLAO to describe organisms follow the principle of “single anatomical structure” adopted by AISM, where each anatomical entity relates to another [41] and through various levels of granularity, for example, the class exocuticle (CLAO:0001039) is composed of chitin (CLAO_0001184) and sclerotin (CLAO_0001461), which is also part of the integument (CLAO_0001049) (Figure 1).

**Table 3.** Output of warnings report generated by ROBOT.

```
WARN  missing_definition    CLAO:0001215                    IAO:0000115
WARN  missing_definition    IAO:0000115                     IAO:0000115
WARN  missing_definition    dc11:title                      IAO:0000115
WARN  missing_definition    oboInOwl:default-namespace      IAO:0000115
WARN  missing_definition    oboInOwl:hasDbXref              IAO:0000115
WARN  missing_definition    oboInOwl:hasExactSynonym        IAO:0000115
WARN  missing_definition    oboInOwl:hasOBONamespace        IAO:0000115
WARN  missing_definition    oboInOwl:hasRelatedSynonym      IAO:0000115
```

**Table 4.** Output of verification generated by ROBOT.

```
PREFIX obo-term: <http://purl.obolibrary.org/obo/>
SELECT DISTINCT ?x ?label
from <http://purl.obolibrary.org/obo/merged/CLAO>
WHERE
{
?x rdfs:subClassOf obo-term:CLAO_0000952.
?x rdfs:label  ?label.
}
class,label
http://purl.obolibrary.org/obo/CLAO_0000956, "thorax"
http://purl.obolibrary.org/obo/CLAO_0001010, "abdomen"
http://purl.obolibrary.org/obo/CLAO_0001011, "head”
```

This principle is maintained for the terms that describe connections between anatomical systems and for the development stages. The connection of the muscles to the cuticle is described through *attached_to*, e.g., posterior craniomandibular muscle (CLAO_0001282) *attached_to* (RO:0002371) some mandible (CLAO:0000969). The innervation of anatomical structures by nerves as postantennal nerve (CLAO:0000119) *innervates* (RO:0002134) some postantennal organ (CLAO:0000029). For the development stages as antenna of embryo (CLAO:0001487) *develops_into* (CLAO_0001531) some antenna (CLAO:0001017). Other anatomical systems are described using different object properties.

A source of conflict is the use of acronyms (notation systems for the chaetotaxy) because a machine cannot parser what type of anatomical entity is referred to. To solve this problem, each name in a nomenclatural system is associated with a term that responds to what the entity is, together with an acronym that relates the anatomical position when it is necessary, e.g., chaeta p4.ab3 (CLAO:0000367) refers to a chaeta p4 located on the abdominal tergite 3 (Table 5).

**Table 5.** Examples of homonyms founded in the anatomical terminology used in Collembola. The name “m2” refers to chaetae on body segments and is translated to standardized and parsable language.

| Term (natural language) | Term (in CLAO) | Related Anatomical entity | ID |
|---|---|---|---|
| m2 | chaeta m2.t2 | mesothorax | CLAO:0000533 |
| m2 | chaeta m2.t3 | metathorax | CLAO:0000534 |
| m2 | chaeta m2.ab1 | abdominal tergite 1 | CLAO:0001242 |
| m2 | chaeta m2.ab2 | abdominal tergite 2 | CLAO:0001243 |
| m2 | chaeta m2.ab3 | abdominal tergite 3 | CLAO:0001244 |
| m2 | chaeta m2.ab5 | abdominal tergite 5 | CLAO:0001246 |
| m2 | chaeta m2.lb | labium | CLAO:0000336 |
| A1 | chaeta A1 | abdominal tergite 4 | CLAO:0001245 |
| A1 | chaeta A1.h | head | CLAO:0000369 |
| A1 | area A1 | abdominal tergite 2 | CLAO:0001243 |

## DISCUSSION

CLAO is intended as a tool in morphology, however, there are conflicts between the terms usually employed in descriptions and the terms imported into CLAO. Traditionally, the chaetae in Collembola are named according to a nomenclatural system, which increases the level of homonyms, i.e., terms with multiple meanings. For instance, the name “p4” refers to chaetae located on different body segments, which makes querying ontology contents difficult. The above procedure results in a new terminology for the chaetotaxy that is parsable by a machine. It is contrary to the usage of natural language, which is most intuitive and comprehensible for human readers, nevertheless, a standardized common language contributes to sharing, reusing, and analyzing data stored in databases [13, 42].

The number of shared terms, including top-level and generalized classes, between CLAO and other anatomical ontologies is low (Figure 2). Shared terminology is a measure of interoperability between ontologies [43] and a prerequisite to query all morphological phenotypes applied to an anatomical entity and building multi-species ontologies [15]. For instance, the term chaeta has extensive use in the literature, with different definitions according to each taxon (e.g., Crustacea, Acari, Protura, Insecta) and with the systematic value given to these entities. Increasing interoperability might contribute to more efficiency in developing anatomical ontologies for Arthropoda, however, this is a challenging task and future work is necessary.

Anatomical ontologies, in addition to the role in phenotypic annotations, are focused on the documentation of homologies from literature and recognition of homologous structures through topological correspondence [2,10,11,44] employing object properties *homologous_to* and *not_homologous_to* [44–45]. The arm of Folsom (named also as Folsom’s anterior arm or anterior apodeme), an anatomical entity that is part of the cephalic endoskeleton of Collembola is recognized as homologous of the anterior tentorial plates in Archaeognatha (Insecta) [46]. The above could be annotated as arm of Folsom (CLAO:0000040) *homologous_to* some anterior tentorial plate, however, the last term must be declared in an anatomy ontology, e.g., AISM. Contrary, the hypopharyngeal fulturae, also part of the cephalic endoskeleton in Collembola, which is not homologous with the tentorium of Insecta [47], expressed as hypopharyngeal fulturae (CLAO:0000055) *not_homologous_to* some tentorium (AISM:0000191). Of course, the interpretation of homologous structures might change with new evidence or morphological data.

To document information about homologues anatomical structures, species-specific ontologies must be related to other ontologies through *homologous_to* and *not_homologous_to* (properties incorporated in

CLAO). The neotropical species *Lepidocyrtus caprilesi* Wray, 1953 and *L. decui* Gruia, 1984 are species with chaeta c3 on abdominal segment 4, which is homologous according to its topological position in both species. The above is expressed as: "chaeta c3 (CLAO:0000456) *part_of* some Lepidocyrtus caprilesi and (*homologous_to* some chaeta c3 (CLAO:0000456) *part_of* some Lepidocyrtus decui". Contrary, the chaeta c3 is not present in *Lepidocyrtus biphasis* Mari Mutt, 1986, which is expressed as: "chaeta c3 (CLAO:0000456) *part_of* some Lepidocyrtus caprilesi and (*not_homologous_to* some chaeta c3 (CLAO:0000456) *part_of* some Lepidocyrtus biphasis**"**.

The recognition and extraction of homologous requires phenotypic annotations, however, this step requires automation or methods where CLAO works as a database to capture morphological data in a computable way. A proposal is the use of lexical matching where identical morphological data (expressed as RDF triplets) are identified and extracted, in a similar way to the recognition of orthologues genes by similarity. Nevertheless, the above requires species-specific ontologies which are problematic due to high species diversity present in Arthropoda, but multispecies ontologies could solve this problem.

Recently, [17] delineate a semantic model to incorporate homologies, that instead of taking binary relations between two entities, considers ternary relations to encompass the concept of monophyletic group. In the future, CLAO could incorporate these ideas about the evaluation of methods for the expression of homologies to become a multi-species ontology. Notwithstanding, anatomical ontologies give place to the recognition of structural "similarity" with the first proposal of comparative homologous [11], while the obtention of homologies is only possible through building phylogenies [17].

Leaving aside the role of anatomical ontologies in phylogenetics, CLAO has other potential applications. Connecting phenotypic data with genes that regulate development might address evolutionary and developmental questions [2,23]. An initiative was developed by [48], who recognized functional gene families in Collembola using functional annotations from the Gene Ontology (GO). Although phenotypic annotations were not included, CLAO could be integrated to recognize "functional units" (phenotypic and genomic) not only in Collembola but in other hexapods.

## CONCLUSION

The use of anatomical terms is highly variable within the anatomy knowledge domain, explained by historic legacy, interpretation of anatomical structures, and how the anatomical information is employed, which could affect the internal validity of CLAO. But this is open and collaborative, where the consensus reduces the subjectivity in the concept choice or terms preference. CLAO is directed toward morphological descriptions of organisms and obtaining of phylogenetic characters, but the implementation by non-expert users is problematic. Initiatives such as proto.morphdbase.de to create phenotypic annotations are available, where CLAO could be integrated in the future. Other options include the adoption of a secondary format in the standardized language during the process of publishing in journals, such as XML language, but it requires a semi- or automatic process and conciliating the author's necessities (specialist or taxonomist) and the technical requirements requested by ontology engineers. CLAO will continue its development over the years, constituting a collaborative effort between specialists in the morphology and taxonomy of Collembola as a methodological tool in morphological descriptions.

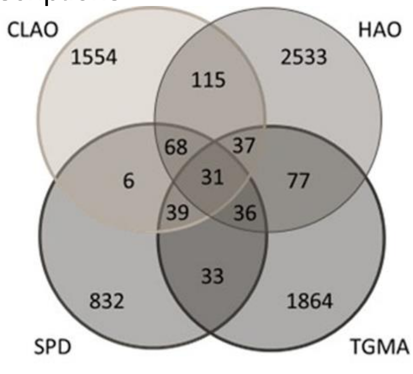


**Figure 2.** Venn diagram showing the approximate number of shared terminologies between CLAO and selected anatomical ontologies for Arthropoda, Hymenoptera Anatomy Ontology (HAO), Spider Anatomy Ontology (SPD), and Mosquito Gross Anatomy Ontology (TGMA). Shared terminology is a measurement of interoperability. In this example, 31 refers to top-level classes and generalized classes that extend to all ontologies, for instance, the term muscle tissue. Names such as "a1", "a2", "a3", "p4", used in natural language could be imported into CLAO through the annotation *has_exact_synonym*, while the user must specify the relationship between instances and ontology class.

Various aspects or requirements could increase the efficiency of CLAO. 1. a reference database to compute synonyms and homonyms under the *sensu* model proposed by [5], 2. inclusion of logical definitions to incorporate CLAO to the anatomical ontologies family in developing currently, 3. image database (i.e., Morphobank) could be linked to CLAO for building character matrices [13], 4. During the implementation of CLAO, the development of RDF stores is essential for the management of the morphological data, and 5. The acceptance and use of anatomy ontology by taxonomists.

**Funding:** This research was funded by MINCIENCIAS, grant number 727”.
**Acknowledgments:** To Nico Matentzoglu, Jennifer Girón, István Mikó, and Lars Vogt for comments about improving CLAO.
**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## REFERENCES

1. Cabré MT. Terminology: Theory, methods, and applications (Vol. 1) [Internet]. Philadelphia: John Benjamins Publishing; 1999 [cited 2020 Nov 7]. 252 p. Available from: doi.org/10.7202/004006ar.
2. Mabee PM, Ashburner M, Cronk Q, Gkoutos GV, Haendel M, Segerdell E, et al. Phenotype ontologies: the bridge between genomics and evolution. Trends Ecol Evol [Internet]. 2007;22:345–50. Available from: doi.org/10.1016/j.tree.2007.03.013.
3. Vogt L, Bartolomaeus T, Giribet G. The linguistic problem of morphology: structure versus homology and the standardization of morphological data. Cladistics [Internet]. 2010;26:301–25. Available from: doi.org/10.1111/j.1096-0031.2009.00286.x.
4. Vogt L. Organizing phenotypic data a semantic data model for anatomy. J Biomed Semantics [Internet]. 2019;10:1–14. Available from: doi.org/10.1186/s13326-019-0204-6.
5. Yoder MJ, Mikó I, Seltmann KC, Bertone MA, Deans AR. A gross anatomy ontology for Hymenoptera. PloS ONE. 2010;5:e15991. Available from: doi.org/10.1371/journal.pone.0015991.
6. Richter S, Wirkner CA. Research program for Evolutionary Morphology. J Zoolog Syst Evol [Internet]. 2014;52:338–50. Available from: doi.org/10.1111/jzs.12061.
7. Wirkner C, Göpel T, Runge J, Keiler J, Klussmann-Fricke B, Huckstorf K, et al. The first organ-based ontology for arthropods (Ontology of Arthropod Circulatory Systems - OArCS) and its Integration into a novel formalization scheme for morphological descriptions. Syst Biol [Internet]. 2016;66:754–68. Available from: doi.org/10.1093/sysbio/syw108.
8. Edgecombe GD. Anatomical nomenclature: homology, standardization and datasets. Zootaxa [Internet]. 2008;1950: 87–95. Available from: doi.org/10.11646/zootaxa.1950.1.8.
9. Mullins P, Kawada R, Balhoff J. Deans A. A revision of Evaniscus (Hymenoptera, Evaniidae) using ontology-based semantic phenotype annotation. ZooKeys [Internet]. 2012;223:1. Available from: doi.org/10.3897/zookeys.223.3572.
10. Balhoff J, Mikó I, Yoder M, Mullins P, Deans A. A semantic model for species description applied to the ensign wasps (Hymenoptera: Evaniidae) of New Caledonia. Syst Biol [Internet]. 2013; 62:639–59. Available from: doi.org/10.1093/sysbio/syt028.
11. Vogt L. Assessing similarity: on homology, characters and the need for a semantic approach to non-evolutionary comparative homology. Cladistics [Internet]. 2017;33:513–39. Available from: doi.org/10.1111/cla.12179.
12. Göpel T, Wirkner C. Morphological description, character conceptualization and the reconstruction of ancestral states exemplified by the evolution of arthropod hearts. PloS ONE [Internet]. 2018;13:e0201702. Available from: doi.org/10.1371/journal.pone.0201702.
13. O’Leary M, Kaufman S. MorphoBank: phylophenomics in the “cloud”. Cladistics [Internet]. 2011;27:529–537. Available from: doi.org/10.1111/j.1096-0031.2011.00355.x.
14. Kalet I. Modeling Biological Structure, pp. 303-323. In: Kalet I, editor. Principles of Biomedical Informatics [Internet]. London: Academic Press; 2009 [cited 2021 Nov 3]. Available from: doi.org/10.1016/B978-0-12-369438-6.00011-8.
15. Mungall C, Torniai C, Gkoutos G, Lewis S, Haendel M. Uberon, an integrative multi-species anatomy ontology. Genome Biol [Internet]. 2012;13:R5. Available from: doi.org/10.1186/gb-2012-13-1-r5.
16. Yu L. A developer’s guide to the semantic Web. Heidelberg: Springer Science & Business Media; 2014 [cited 2020 Jun 10]. 829 p. Available from: doi.org/10.1007/978-3-662-43796-4.
17. Mabee P, Balhoff J, Dahdul W, Lapp H, Mungall C, Vision T. A logical model of homology for comparative biology. Syst Biol [Internet]. 2020;69:345–62. Available from: doi.org/10.1093/sysbio/syz067.
18. Topalis P, Tzavlaki C, Vestaki K, Dialynas E, Sonenshine DE, Butler R, et al. Anatomical ontologies of mosquitoes and ticks, and their web browsers in VectorBase. Insect Mol Biol [Internet]. 2008;17:87–9. Available from: doi.org/10.1111/j.1365-2583.2008.00781.x.
19. Costa M, Reeve S, Grumbling G, Osumi-Sutherland D. The Drosophila anatomy ontology. J Biomed Semant [Internet]. 2013;4:1–11. Available from: https://doi.org/10.1186/2041-1480-4-32.

20. Ramírez MJ, Michalik P. The spider anatomy ontology (SPD) a versatile tool to link anatomy with cross-disciplinary data. Diversity [Internet]. 2019;11:202. Available from: https://doi.org/10.3390/d11100202.
21. Mikó I, Masner L, Johannes E, Yoder M, Deans A. Male terminalia of Ceraphronoidea: morphological diversity in an otherwise monotonous taxon. Insect Syst Evol [Internet]. 2013;44:261–347. Available from: doi.org/10.1163/1876312X-04402002.
22. Trietsch C, Mikó I, Notton D, Deans A. Unique extrication structure in a new megaspilid, Dendrocerus scutellaris Trietsch & Mikó (Hymenoptera: Megaspilidae). Biodivers Data J [Internet]. 2018;6:e22676. Available from: doi.org/10.3897/BDJ.6.e22676.
23. Tarasov S. Integration of anatomy ontologies and evo-devo using structured Markov models suggests a new framework for modeling discrete phenotypic traits. Syst Biol [Internet]. 2019;68:698–716.
24. González-Montaña LA. Collembola Anatomy Ontology [Internet]. EE.UU:GitHub;2021 [updated 2021 June 15; cited 2012 Nov 5]. repository. Available from: doi.org/10.5281/zenodo.4660386.
25. Bellinger PF, Christiansen KA, Janssens F. Checklist of the Collembola of the World [Internet]. Belgium; 1996–2021 [updated 2022 July 31; cited 2021 Feb 10] Available from: http://www.collembola.org.
26. Mahner M, Bunge M. 1997. Foundations of Biophilosophy. Berlin. Heidelberg, New York: Springer-Verlag. 423 p.
27. Haendel MA, Balhoff JP, Bastian FB, Blackburn DC, Blake J, Bradford Y, et al. Unification of multi-species vertebrate anatomy ontologies for comparative biology in Uberon. J Biomed Semant [Internet]. 2014;5:21. Available from: doi.org/10.1186/2041-1480-5-21.
28. Schneider FD, Fichtmueller D, Gossner MM, Güntsch A, Jochum M, König-Ries B, et al. Towards an ecological trait-data standard. Methods Ecol Evol [Internet]. 2019;10:2006–2019. Available from: https://doi.org/10.1111/2041-210X.13288.
29. Keet C. Granularity. In: Dubitzky W., Wolkenhauer O., Cho KH., Yokota H, editors. Encyclopedia of Systems Biology [Internet]. New York: Springer; 2013 [cited 2021 Jul 7] Available from: doi.org/10.1007/978-1-4419-9863-7_65.
30. Ali NM, Khan HA, Amy Y, Then H, Ching CV, Gaur M, Dhillon SK. Fish Ontology framework for taxonomy-based fish recognition. PeerJ [Internet]. 2017;5:e3811. Available from: https://doi.org/10.7717/peerj.3811.
31. Bouzidi R, De Nicola A, Nader F, Chalal R. OntoGamif: A modular ontology for integrated gamification. Appl Ontol [Internet]. 2019;14:215–249. Available from: https://doi.org/10.3233/AO-190212.
32. Smith B, Varzi A. Bona Fide and Fiat Boundaries. Philos Phenomenol Res [Internet]. 2000; 60:401. Available from: doi.org/10.2307/2653492.
33. Musen MA. The Protégé project: A look back and a look forward. AI Matters [Internet]. 2015;1(4). Available from: doi.org/10.1145/2557001.25757003.
34. Smith B, Ceusters W, Klagges B, Kohler J, Kumar A, Lomax J, et al. Relations in biomedical ontologies. Genome Biol [Internet]. 2005;6:5, Available from: doi.org/10.1186/gb-2005-6-5-r46.
35. Matentzoglu N, Mungall C, Osumi-Sutherland D, Balhoff J, Douglass E, Vasilevsky N, et al. (2021). INCATools/ontology-development-kit. Version 1.2.27[software]. 2013 [cited 2021 April 04]2021-04-04. Available from: doi.org/10.5281/zenodo.4662067
36. Amith M, He Z, Bian J, Lossio-Ventura JA, Tao C. Assessing the practice of biomedical ontology evaluation: Gaps and opportunities. J Biomed Inform [Internet]. 2018;80:1–13. Available from: doi.org/10.1016/j.jbi.2018.02.010.
37. Glimm B, Horrocks I, Motik B, Stoilos G. Optimizing Ontology Classification. In: Patel-Schneider P, Pan Y, Hitzler P, Mika P, Zhang L, Pan J, Horrocks I, Glimm B, editors, Proceedings of the 9th international semantic web conference on the semantic web.2010 Nov 7–11; Shangai, China, Springer: 2010. p. 225–240.
38. Jackson R, Balhoff J, Douglass E, Harris N, Mungall C, Overton J. ROBOT: A tool for automating ontology workflows. BMC Bioinformatics [Internet]. 2019;20: 407. Available from: doi.org/10.1186/s12859-019-3002-3.
39. González-Montaña, LA. Semantic-based methods for morphological descriptions: An applied example for Neotropical species of genus Lepidocyrtus Bourlet, 1839 (Collembola: Entomobryidae). Biosyst Ecol [Internet]. 2021;1:e71620. Available from: doi:10.1553/biosystecol.1.e71620.
40. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. Nat Genet [Internet]. 2000;25:25–9. Available from: https://doi.org/10.1038/75556.
41. Girón JC, Tarasov S, Montaña LA, Matentzoglu N, Smith AD, Koch M, et al. Formalizing Insect Morphological Data: A Model-Based, Extensible Insect Anatomy Ontology and Its Potential Applications in Biodiversity Research and Informatics. Preprints [Internet]. 2022; 2022010254. Available from: https://doi.org/10.20944/preprints202201.0254.v1.
42. Cui H, Xu D, Chong S, Ramirez M, Rodenhausen T, Macklin J, et al. Introducing explorer of taxon concepts with a case study on spider measurement matrix building. BMC Bioinformatics [Internet].2016;17:471. Available from: https://doi.org/doi.org/10.1186/s12859-016-1352-7.
43. Bittner T, Donnelly M, Winter S. Ontology and semantic interoperability. In: Zlatanova S, Prosperi D, editors. Large-scale 3D data integration. Boca Raton: CRC Press; 2005. p. 139-160.
44. Haendel MA, Neuhaus F, Osumi-Sutherland D, Mabee PM, Mejino JL, Mungall CJ, et al. (2008). CARO–the common anatomy reference ontology. In: Burger A, Davidson D, Baldock R, editors. Anatomy ontologies for bioinformatics. London: Springer; 2008. p. 327-349.

45. Travillian R, Malone J, Pang C, Hancock J, Holland PW, et al. The vertebrate bridging ontology (VBO). In: Bodenreider O, Martone M, Ruttenberg A, editors. Proceedings of the Second International Conference on Biomedical Ontology (Bio-Ontologies 2011). 2011 Jul 26–30; Buffalo, New York. Aachen, Germany: CEUR Workshop Proceedings (CEUR-WS.org): 2011. p. 416–421.
46. Bitsch C, Bitsch J. The endoskeletal structures in arthropods: cytology, morphology and evolution. Arthropod Struct Dev [Internet]. 2002;30:159–77. https://doi.org/10.1016/s1467-8039(01)00032-9.
47. Blanke A, Machida R. The homology of cephalic muscles and endoskeletal elements between Diplura and Ectognatha (Insecta). Org. Divers. Evol. 2016;16:241–57. Available from: doi.org/10.1007/s13127-015-0251-5.
48. Faddeeva A, Studer R, Kraaijeveld K, Sie D, Ylstra B, Mariën J, et al. Collembolan transcriptomes highlight molecular evolution of hexapods and provide clues on the adaptation to terrestrial life. PLoS ONE [Internet]. 2015;10:e0130600. Available from: doi.org/ 10.1371/journal.pone.0130600.