

Semantic Species Description: TaxonWorks & PhenoScript

Sergei Tarasov

Beetle Curator
Finnish Museum of Natural History, Helsinki, Finland



• @tarasov_sergio

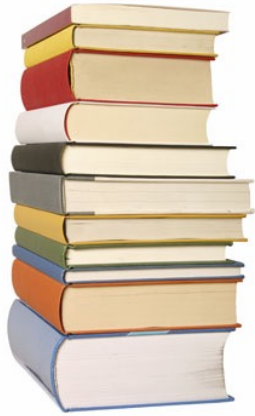


• sergei.tarasov@helsinki.fi



• <https://www.tarasovlab.com>

Natural Language Species Descriptions (~250 years of effort)



1.3 million of described species



NOT COMPUTER PARSABLE

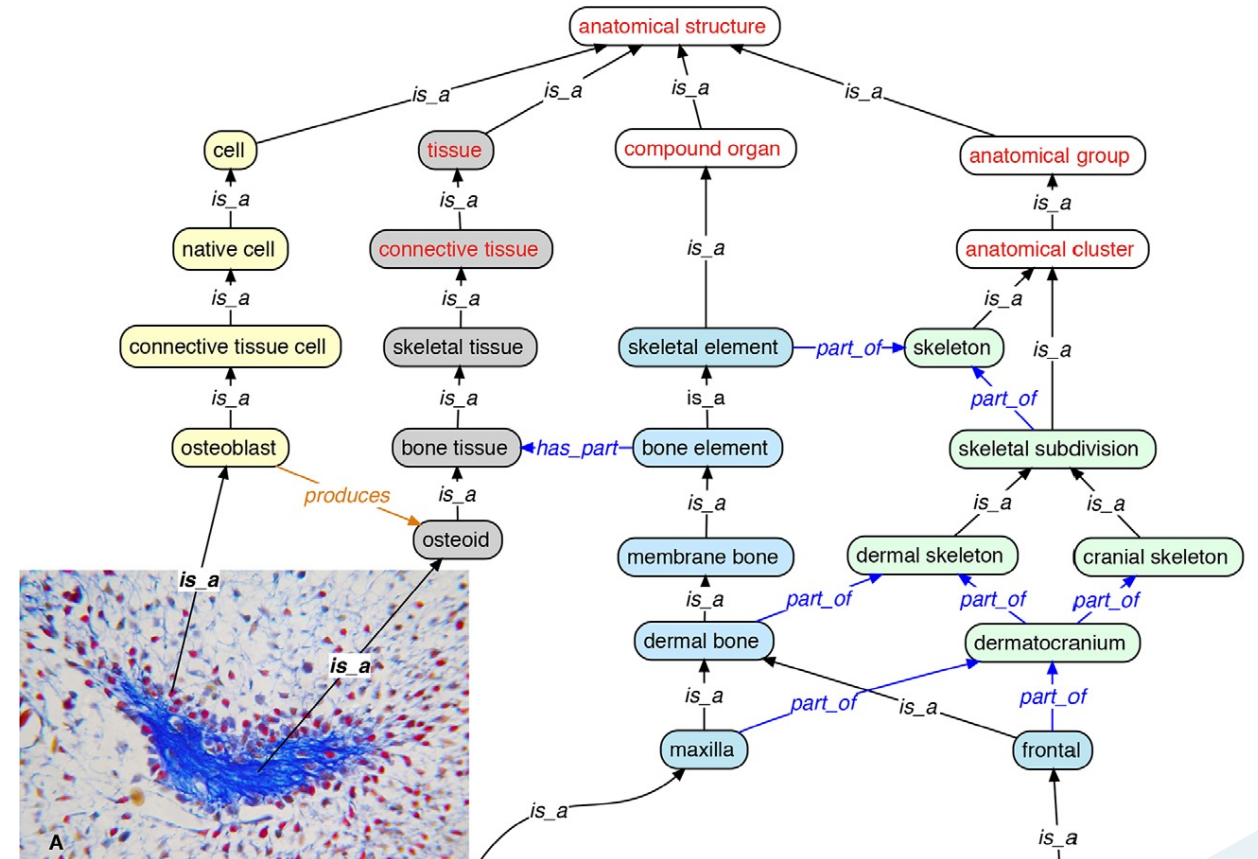


Carl von Linné, 1775

- Rarely reused
- Not computer-parsable
- This precludes comparative phenomics in broad sense

Ontology-based phenotype descriptions

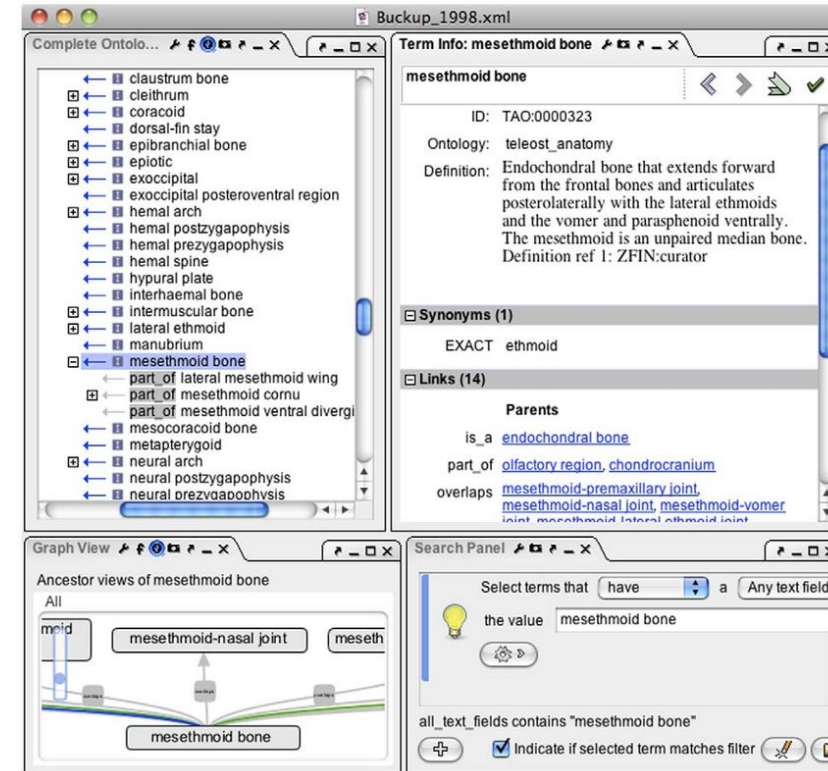
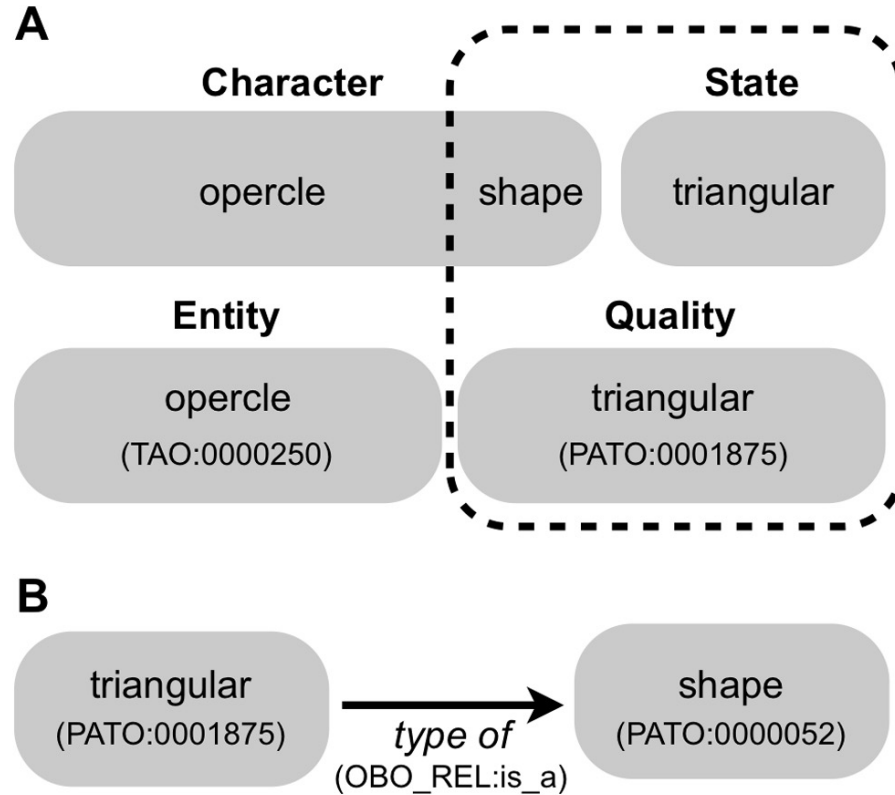
- Ontologies are networks representing our knowledge
- Part of Artificial Intelligence toolbox
- Ontology-driven approach allows computer parsable species descriptions
- Ontology-based descriptions is an active area of research



Phenoscape Project



<https://phenoscape.org>



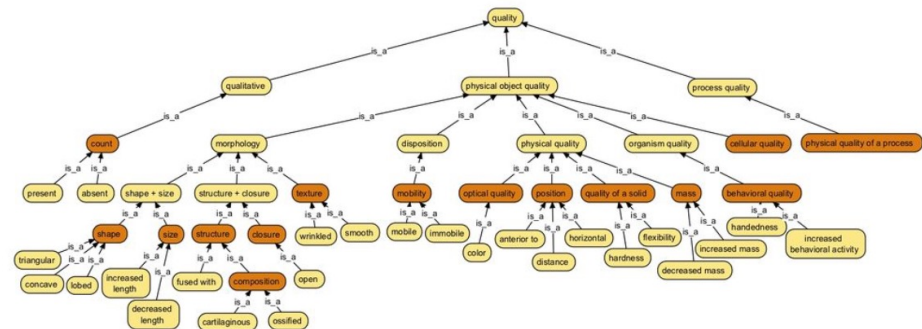
Balhoff, James P., et al. "Phenex: ontological annotation of phenotypic diversity." *PloS one* 5.5 (2010): e10500.

Balhoff, James P., et al. "A semantic model for species description applied to the ensign wasps (Hymenoptera: Evaniidae) of New Caledonia." *Systematic Biology* 62.5 (2013): 639-659.

Phenoscape Project: character annotation

PATO terms used to annotate systematic characters

Below is a simplified graph showing a subset of quality terms from the [Phenotype and Trait Ontology \(PATO\)](#)  used to curate systematic characters. Terms in orange represent attribute-level qualities.



PATO attribute-level term	Synonyms	Examples of child terms
shape		triangular, lobed, concave, protruding into
position	placement, location	horizontal, anterior to
size		thin, large, decreased height
structure		porous, fused with
composition	composed of, content	ligamentous, cartilaginous
texture		smooth, wrinkled
optical quality		color, brightness
mass		increased mass
quality of a solid		flexible, hard
mobility		mobile, immobile
closure		open, closed
behavioral quality		
count	number, amount	present, absent

Annotation Specificity

Annotations are made using the most specific term or post-composition possible for Entity and Quality. This may not be possible for some characters involving very complex phenotypes, particularly those describing shape variation. In these cases, annotations are made to 'attribute' level PATO terms (e.g., PATO: shape).

Character Annotation Examples

The following are examples of character types commonly encountered in the systematic literature and how we annotate them using the **Entity-Quality (EQ)** model in Phenex. A character state can often require multiple EQs to capture all of the anatomical details described by an author. For example, a state may describe "premaxillary teeth round and multicuspidate." This is represented with the following two phenotypes:



https://wiki.phenoscape.org/wiki/Guide_to_Character_Annotation

The First Programming Language For Describing Species

Main aim

- provide a simple way to write semantic descriptions for species
- explores instance-based approach that is computationally cheaper



<https://github.com/sergeitarasov/PhenoScript/wiki>

From textual description to Species Programming

Traditional Description

Phenoscript

Species description

Parachorius semsanganus Tarasov & Keith sp. n.

urn:lsid:zoobank.org:act:E79FA444-725E-40E6-986A-9FB48B7D232A

http://species-id.net/wiki/Parachorius_semsanganus

Figs 1–5

Type locality. Laos, Xieng Khouang prov., Phou Sane Mt.

Material examined. *Holotype* (NHMB), male bearing the following labels:

LAOS-NE, Xieng Khouang prov., 19°38.20'N 103°20.20'E, Phonsavan (30 km NE): PHOU SANE Mt., 1420 m, 10.-30.v.2009, D. Hauck leg.

NHMB Basel, NMPC Prague Laos 2009 Expedition: M. Brancucci, M. Geiser, Z. Kraus, D. Hauck, V. Kubán

HOLOTYPE *Parachorius semsanganus* S. Tarasov & D. Keith det. 2011

Paratypes. 9♀, same data as holotype; 2♂, same data as holotype but Z. Kraus leg.; 4♂, 3♀, LAOS-NE, Xieng Khouang prov., 19°37-8.'N 103°20'E, 30 km NE Phonsavan: Ban Na, Lam Phou Sane Mt., 1300–1500 m, 10.-30.v.2009, M. Brancucci leg.

Description. Oval, convex, black, entirely shiny; mouthparts, antennae and legs red-brown. Dorsal body side covered with two types of punctures: larger (normal) punctures and very tiny punctures (which can be observed only under higher magnification of 40x or more) Dorsal and ventral body sides glabrous. Length 8.2–10.6

Male (Fig. 4). Head flat, punctuation fine; anterior margin notched medially; notch delimited by 2 prominent triangular obtuse teeth; clypeus laterad of each tooth very slightly notched; eyes completely divided by canthus into lower and upper lobes; lower lobes significantly larger than upper ones; genae and clypeus not distinctly separated from frons; genae rounded and protruding; antennae with 9 segments, antennal club with 3 segments.

```
1 OTU=[
2 OTU_model=
3 [
4   TU@CNCHYMEN_132936["to_Ophu_List"="has_phenotype", "rdfs:label"="Onthophagus taurus"]
5   denotes. female_organism@0_taurus ["rdfs:label"="Onthophagus taurus"];
6   TU@CNCHYMEN_132936 denotes. Onthophagus;
7 ]
8
9
10 Ophu_List=[
11   female_organism@0_taurus > head@h1 >> black;
12   female_organism@0_taurus > head@h1 !> cephalic_horn;
13   female_organism@0_taurus > clypeus > anterior_region >> elongate;
14   female_organism@0_taurus > tibial_tooth |=| 3;
15   female_organism@0_taurus > body >> length is_quality_measured_as. measurement_datum@1 has_measurement_unit_label. mm;
16   female_organism@0_taurus > measurement_datum@1 has_measurement_value. 0.5;
17   female_organism@0_taurus part
18 ]
19 ]
20
```

→I	P	part_of	BFO_0000050
→I	P	partner_in	RO_0002461
→I	C	partial_trisomy	PATO_0001390
→I	P	participates_in	RO_0000056
→I	C	partially_broken	PATO_0002082
→I	P	part_of_developmental_precursor_of	RO_0002287
→I	P	part_of_structure_that_is_capable_of	RO_0002329
→I	P	partially_overlaps	RO_0002151
→I	C	partially_paralysed	PATO_0001858
→I	C	partially_dislocated	PATO_0002157

a core relation that holds between a part and its whole

From textual description to Species Programming

- PhS parser and compiler are written in R & Python
- PhS offers a simple syntax to write semantic statements and export them into OWL (Web Ontology Language)
- It's like constructing a semantic graph
- Inspired by DOT language
- PhS does not run semantic inferences
- OWL statements can be subjected to semantic reasoning for automatic comparison and construction of character matrices (still underexplored field)

Phenoscript

```
1
2 OTU=[
3 OTU_model=
4 [
5   TU@CNCHYMEN_132936["to_Ophu_List"="has_phenotype", "rdfs:label"="Onthophagus taurus"]
6   denotes. female_organism@0_taurus ["rdfs:label"="Onthophagus taurus"];
7   TU@CNCHYMEN_132936 denotes. Onthophagus;
8 ]
9
10 Ophu_List=[
11   female_organism@0_taurus > head@h1 >> black;
12   female_organism@0_taurus > head@h1 !> cephalic_horn;
13   female_organism@0_taurus > clypeus > anterior_region >> elongate;
14   female_organism@0_taurus > tibial_tooth | = | 3;
15   female_organism@0_taurus > body >> length is_quality_measured_as. measurement_datum@1 has_measurement_unit_label. mm;
16   female_organism@0_taurus > measurement_datum@1 has_measurement_value. 0.5;
17   female_organism@0_taurus part
18 ]
19 ]
20
```

→I	P	part_of	BFO_0000050
→I	P	partner_in	RO_0002461
→I	C	partial_trisomy	PATO_0001390
→I	P	participates_in	RO_0000056
→I	C	partially_broken	PATO_0002082
→I	P	part_of_developmental_precursor_of	RO_0002287
→I	P	part_of_structure_that_is_capable_of	RO_0002329
→I	P	partially_overlaps	RO_0002151
→I	C	partially_paralysed	PATO_0001858
→I	C	partially_dislocated	PATO_0002157

a core relation that holds between a part and its whole

PhenoScript Syntax: writing phenotype semantic graphs

Atom Editor

```
1 OTU=[
2 OTU_model=
3 [
4   TU@CNCHYMEN_132936["to_Ophu_List"="has_phenotype", "rdfs:label"="Onthophagus taurus"]
5   denotes. female_organism@0_taurus ["rdfs:label"="Onthophagus taurus"];
6   TU@CNCHYMEN_132936 denotes. Onthophagus;
7 ]
8
9
10 Ophu_List=[
11   female_organism@0_taurus > head@h1 >> black;
12   female_organism@0_taurus > head@h1 !> cephalic_horn;
13   female_organism@0_taurus > clypeus > anterior_region >> elongate;
14   female_organism@0_taurus > tibial_tooth |=| 3;
15   female_organism@0_taurus > body >> length is_quality_measured_as. measurement_datum@1 has_measurement_unit_label. mm;
16   female_organism@0_taurus > measurement_datum@1 has_measurement_value. 0.5;
17   female_organism@0_taurus part
18 ]
19 ]
20
```

→I	P	part_of	BF0_0000050
→I	P	partner_in	RO_0002461
→I	C	partial_trisomy	PATO_0001390
→I	P	participates_in	RO_0000056
→I	C	partially_broken	PATO_0002082
→I	P	part_of_developmental_precursor_of	RO_0002287
→I	P	part_of_structure_that_is_capable_of	RO_0002329
→I	P	partially_overlaps	RO_0002151
→I	C	partially_paralysed	PATO_0001858
→I	C	partially_dislocated	PATO_0002157

a core relation that holds between a part and its whole

- Atom Editor
- Syntax Highlight
- Aliases (>,>>...) for commonly used terms
- Snippets to guide term selection

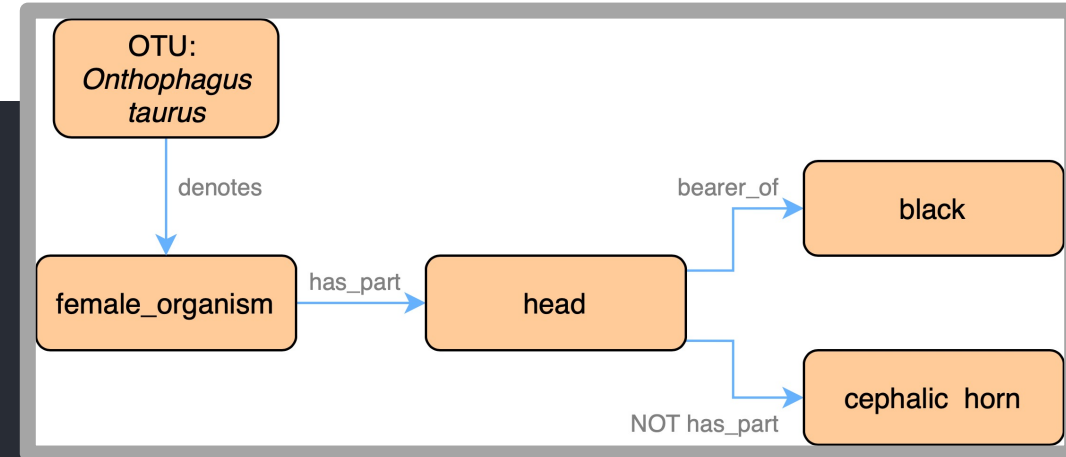


PhenoScript Syntax: writing phenotype semantic graphs

```
1
2 OTU=[
3 OTU_model=
4 [
5   TU@CNCHYMEN_132936["to_Ophu_List"="has_phenotype", "rdfs:label"="Onthophagus taurus"]
6   denotes. female_organism@0_taurus ["rdfs:label"="Onthophagus taurus"];
7   TU@CNCHYMEN_132936 denotes. Onthophagus;
8 ]
9
10 Ophu_List=[
11   female_organism@0_taurus > head@h1 >> black;
12   female_organism@0_taurus > head@h1 !> cephalic_horn;
13   female_organism@0_taurus > clypeus > anterior_region >> elongate;
14   female_organism@0_taurus > tibial_tooth |=| 3;
15   female_organism@0_taurus > body >> length is_quality_measured_as. measurement_datum@1 has_measurement_unit_label. mm;
16   female_organism@0_taurus > measurement_datum@1 has_measurement_value. 0.5;
17   female_organism@0_taurus part
18 ]
19 ]
20
```

→I	P	part_of	BF0_0000050
→I	P	partner_in	RO_0002461
→I	C	partial_trisomy	PATO_0001390
→I	P	participates_in	RO_0000056
→I	C	partially_broken	PATO_0002082
→I	P	part_of_developmental_precursor_of	RO_0002287
→I	P	part_of_structure_that_is_capable_of	RO_0002329
→I	P	partially_overlaps	RO_0002151
→I	C	partially_paralysed	PATO_0001858
→I	C	partially_dislocated	PATO_0002157

a core relation that holds between a part and its whole



- Atom Editor
- Syntax Highlight
- Aliases (>,>>...) for commonly used terms
- Snippets to guide term selection

Integrating TW & PhS

(currently under development)

Expectations:

- Simple interaction between ontologies and users
- Facilitate automated comparative phenomics
- Link OTUs with phenotypes
- Facilitate reuse of the description across various biological fields



Integrating



TW Content Editor tab

TaxonWorks v0.22.3 Issues

Dung_Beetles

New content

PhenoScript - *Parachorius thomsoni* Harold, 1873 ✓

B *I* <> **H** | “ ” ≡ ≡ | 🔗 📄 👁 Data links

OTU=[
OTU_model=
[
TU@CNCHYMEN_132936["to_Ophu_List"="has_phenotype", "rdfs:label"="Onthophagus taurus"]
denotes_female_organism@Parachorius_thomsoni ["rdfs:label"="Onthophagus taurus"];
TU@CNCHYMEN_132936 denotes_Onthophagus;
]

Ophu_List=[
female_organism@Parachorius_thomsoni > head@h1 >> black;
female_organism@Parachorius_thomsoni > head@h1 > cephalic_horn;
female_organism@Parachorius_thomsoni > clypeus > anterior_region >> elongate;
female_organism@Parachorius_thomsoni > tibial_tooth |=| 3;
female_organism@Parachorius_thomsoni > body >> length_is_quality_measured_as_measurement_datum@1 has_measurement_unit_label.mm;
female_organism@Parachorius_thomsoni > measurement_datum@1 has_measurement_value.0.5;
female_organism@Parachorius_thomsoni part
]
]
]

Save

Clone

Browse OTUs *Parachorius nudus* (Sharp, 1875) ✓

Root > Scarabaeinae > Parachorini > Parachorius thomsoni Harold, 1873

Parachorius thomsoni Harold, 1873 ✓ 🔍 ⚙️ 🔄 📄

Timeline | Descendants | Images | Common names | Type specimens | Specimen records | Description | Content | Asserted distribution | Biological associations | Distribution | Annotations

TIMELINE

All

Nomenclature

Protonym

OTU (biology)

Citations (6)

- Parachorius thomsoni* Harold, 1873: 103.
- Parachorius thomsoni* Harold, 1873 ✓ in Gillet, 1911: 54.
- Parachorius thomsoni* Harold, 1873 ✓ in Boucomont, 1929: 761.
- Parachorius lannathai* Hanboonsong & Masumoto, 2001: 138.
- Parachorius thomsoni* Harold, 1873 ✓ in Tarasov, 2017: 142. revision
- Parachorius lannathai* Hanboonsong & Masumoto, 2001 (subjective synonym of *Parachorius thomsoni* Harold, 1873) in Tarasov, 2017: 142.

References

- ☐ Harold, Evon (1873) Diagnosen neuer Coprophagen. *Coleopterologische Hefte* 11, 102–105. 🔍 🔄
- ☐ Gillet, J.J.E. (1911) 19 pars. 38 *Scarabaeidae: Coprinae I*. Junk and Schenkling. 100 pp. 🔍 🔄
- ☐ Boucomont, A. (1929) A list of the coprophagous Coleoptera of China. *Lingnan Scientific Journal*. Canton 7, 759–794. 🔍 🔄
- ☐ Hanboonsong, Y. & Masumoto, K. (2001) Dung beetles (Coleoptera, Scarabaeidae) of Thailand. Part 4. Genera Phacosoma, Cassolus and Parachorius (Canthonini and Dichotomini). *Elytra* 29, 129–140.
- ☐ Tarasov, S. (2017) A cybertaxonomic revision of the new dung beetle tribe Parachorini (Coleoptera: Scarabaeidae: Scarabaeinae) and its phylogenetic assessment using molecular and morphological data. *Zootaxa* 4329, 101–149. <https://doi.org/10.11646/zootaxa.4329.2.1> 🔍 🔄

DESCRIPTION

CONTENT

- PhenoScript**
OTU=[
OTU_model=
[
TU@CNCHYMEN_132936["to_Ophu_List"="has_phenotype", "rdfs:label"="Onthophagus taurus"]
denotes_female_organism@Parachorius_thomsoni ["rdfs:label"="Onthophagus taurus"];
TU@CNCHYMEN_132936 denotes_Onthophagus;
]

Ophu_List=[
female_organism@Parachorius_thomsoni > head@h1 >> black;
female_organism@Parachorius_thomsoni > head@h1 > cephalic_horn;
female_organism@Parachorius_thomsoni > clypeus > anterior_region >> elongate;
female_organism@Parachorius_thomsoni > tibial_tooth |=| 3;
female_organism@Parachorius_thomsoni > body >> length_is_quality_measured_as_measurement_datum@1 has_measurement_unit_label.mm;
female_organism@Parachorius_thomsoni > measurement_datum@1 has_measurement_value.0.5;
female_organism@Parachorius_thomsoni part
]
]
]

Acknowledgement

Thanks to:

- Mat Yoder
- Dmitry Mozzherin

for a fruitful discussion about integration
of TW & PhenoScript

- and Academy of Finland for funding

Relative Materials:

Watch Recent talk by *Jennifer Giron*

“Unlocking the power of morphological data with the Coleoptera Anatomy Ontology”

<https://www.youtube.com/watch?v=Nwpk3pBHn1U>

