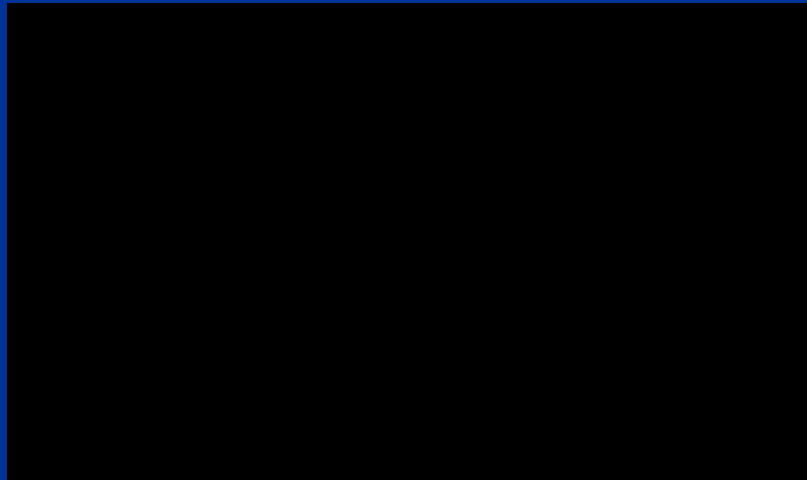


The 'cyber-centipede':

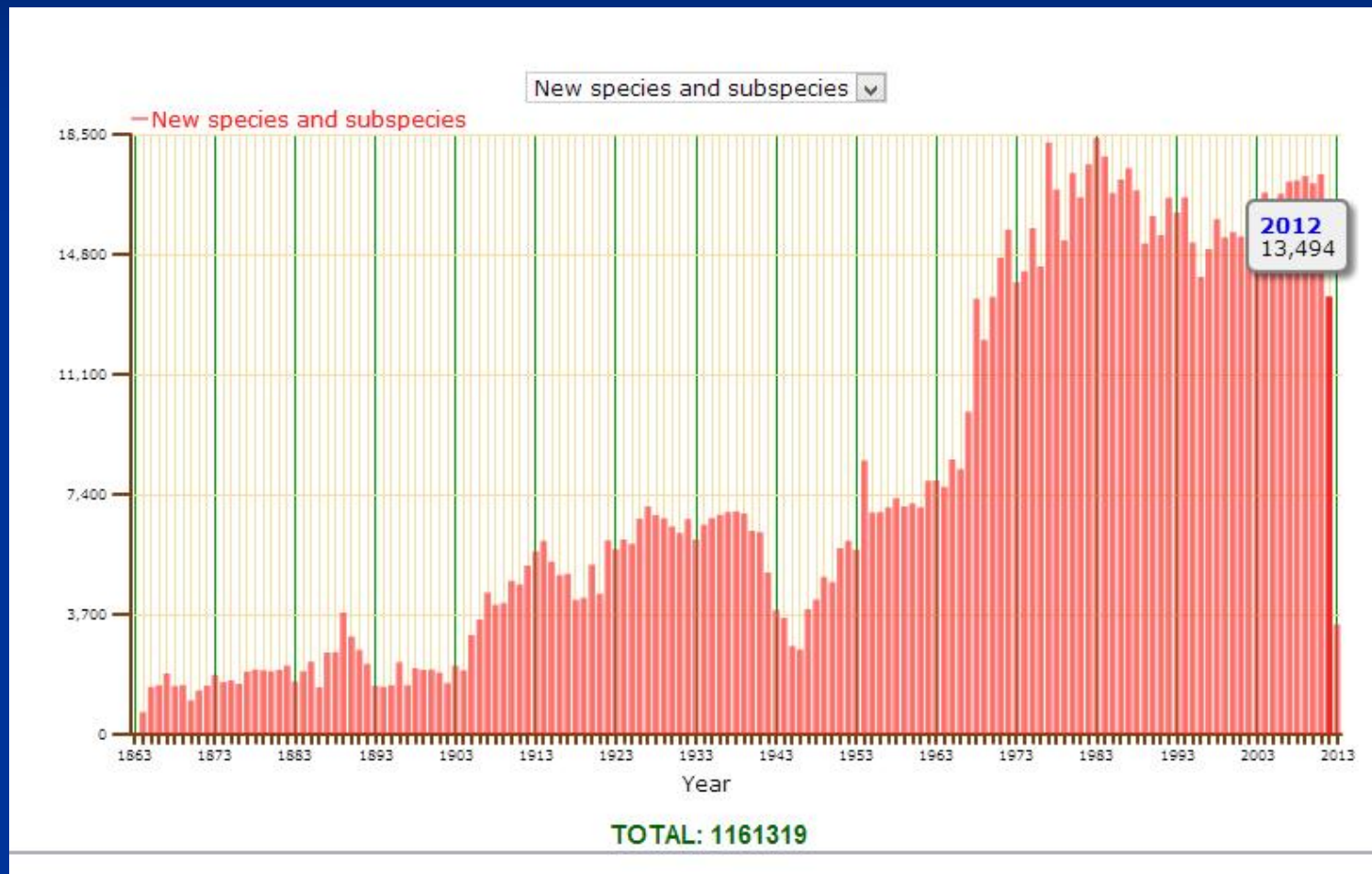
The first eukaryotic organism described with a fully sequenced transcriptome, DNA barcode and micro computed tomography



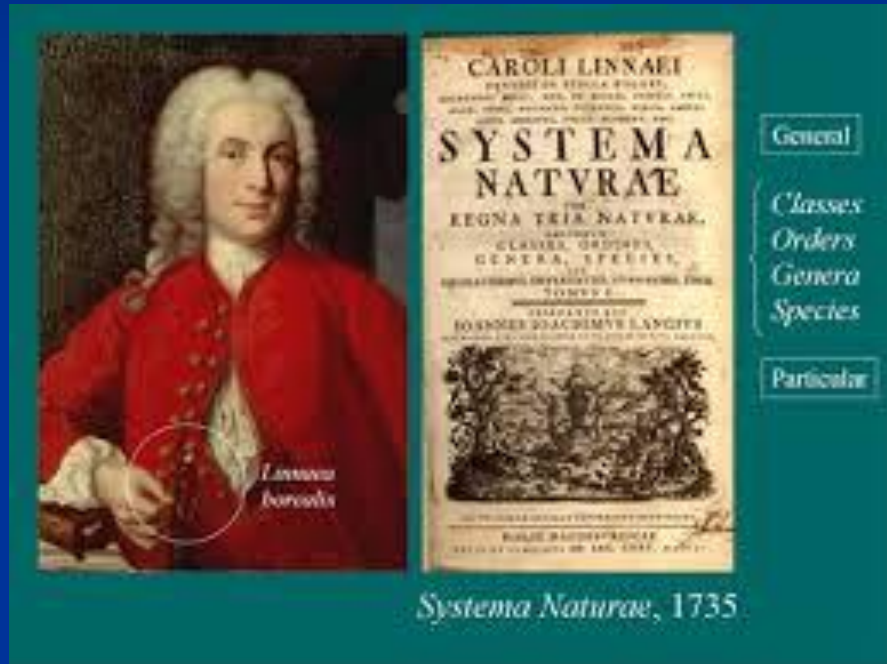
L. Penev, P. Stoev, A. Komerički, N. Akkari, S. Li, X. Zhou, S. Edmunds, C. Hunter, A.M. Weigand, D. Porco, M. Zapparoli, T. Georgiev, D. Mietchen, D. Roberts, V. Smith



“Only” 13,494 new animal species described in 2012 (Index of Organism Names)



How much have taxonomic descriptions changed since Linnaeus?



273. SCOLOPENDRA. *Antennae* setaceae.
Palpi duo filiformes, articulati, inter maxillas con-
 nati.
Labium dentatum fissum.
Corpus depressum.
Pedes numerosi, totidem
 utrinque, quot corpo-
 ris segmenta.

lagura. 1. Sc. pedibus utrinque duodecim, corpore ovali, cauda pe-
 nicillo albo. *Fn. Suec.* 2063.* *Geoffr. inf. par.* 2. p. 677.
n. 6. *t.* 22. *f.* 4. *Fabr. sp. inf.* 1. p. 531. *n.* 1. *mant. inf.*
1. p. 341. *n.* 1.
Scop. ent. carn. 1152. *Julus lagurus.*
Degeer inf. 7. p. 571. *n.* 1. *t.* 36. *f.* 1-3. *Julus penicillatus.*
Act. par. mill. 1. p. 532. *t.* 17. *f.* 4-10.
Habitat sub Europae muscis.

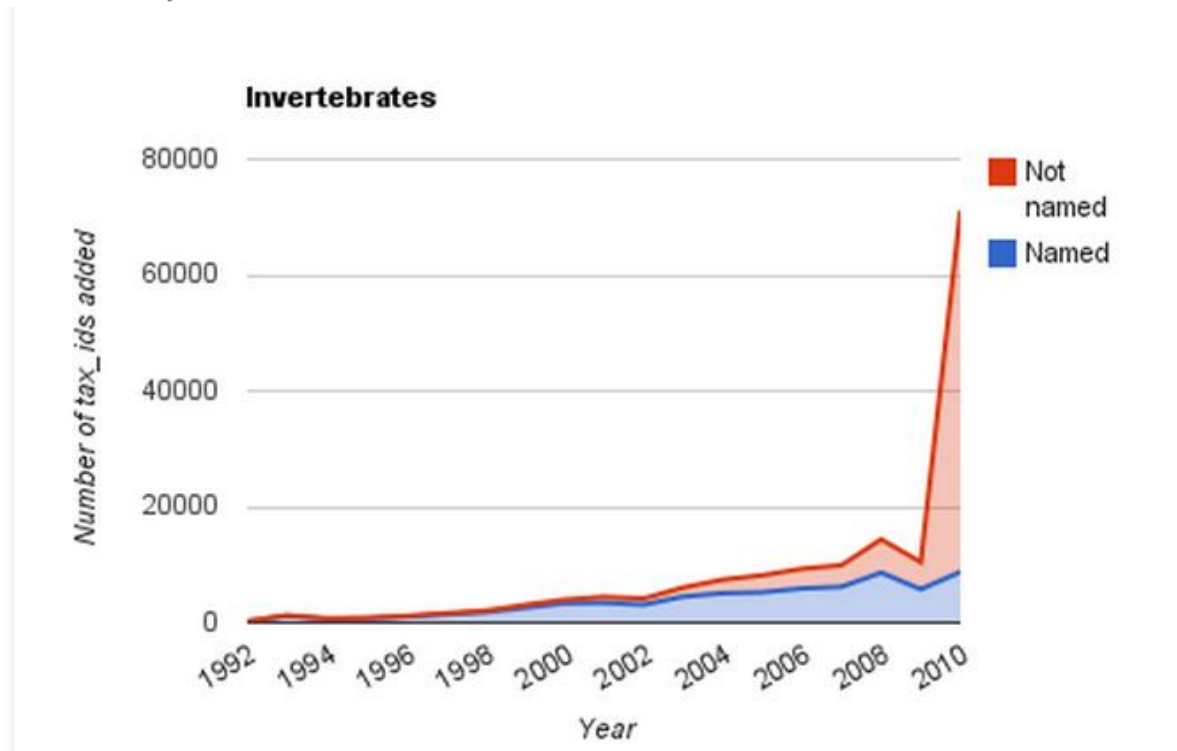
The data deluge

Massive sampling/data gathering



Proportion of dark taxa increase dramatically!

For "invertebrates" 2010 saw an explosive growth in the number of new taxa sequenced, with nearly 71,000 new taxa added to GenBank.



This coincides with a spectacular drop in the number of properly-named taxa, but even before 2010 the proportion of named invertebrate species in GenBank was in decline: in 2009 just over a half of the species added had binomials.

Low uptake of open data publishing/sharing



Solutions?

Turbo-taxonomy: publish fast and many!

Pipeline descriptions of new species
(Butcher et al. 2012, Riedel et al. 2013b):

- molecular data
- concise morphological descriptions
- digital imaging

66. *Trigonopterus ptolycoides* Riedel, sp. n.

[urn:lsid:zoobank.org:act:37E45438-FBB4-4FD0-BC0E-B72E4D9979F8](http://zoobank.org/act:37E45438-FBB4-4FD0-BC0E-B72E4D9979F8)

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

Diagnostic description. Holotype, male (Fig. 66a). Length 4.28 mm. Color ferruginous, elytral disk with some irregular darker spots; integument partly covered with yellowish-brownish tomentum. Body dorsally flattened, with irregular lateral ridge; with constrictions at middle of pronotum, between pronotum and elytron, and in basal third of elytron. Rostrum in basal half with distinct median ridge and pair of submedian ridges; basally above eyes with pair of protrusions; apical $\frac{1}{4}$ subglabrous, punctulate. Pronotum with subapical constriction; with marked lateral flanges; coarsely rugose-punctate; punctures containing each one upcurved scale and much finer tomentum. Elytra with distinct striae; intervals with irregular tomentose tubercles; with marked lateral flanges; base bisinuate; apex extended ventrad, slightly beak-shaped. Femora narrow, parallel-sided, edentate. Meso- and metafemur dorsally with fringe of erect scales; metafemur subapically with stridulatory patch. Meso- and metatibia tapering from base to apex. Tarsi asymmetrical; tarsomere 3 with anterior lobe much larger than posterior lobe. Abdominal ventrite 5 apically deeply emarginate. Aedeagus (Fig. 66b) with body flattened, sides subparallel, apex pointed; transfer-apparatus short, spiniform; ductus ejaculatorius without bulbus. **Intraspecific variation.** Length 3.92–4.28 mm. Female abdominal ventrite 5 apically rounded.

Material examined. Holotype (SMNK): ARC1415 (EMBL # HE615964), PAPUA NEW GUINEA, Eastern Highlands Prov., 37 km S Goroka, Hogave vill., Mt. Michael, $S06^{\circ}22.798'$, $E145^{\circ}15.427'$ to $S06^{\circ}22.925'$, $E145^{\circ}16.645'$, 2179–2800 m, 09-15-VII-2009, sifted. Paratypes (NAIC, SMNK, ZSM): 6 exx, ARC1416 (EMBL # HE615965), ARC1417 (EMBL # HE615966), same data as holotype.

Distribution. Eastern Highlands Prov. (Mt. Michael). Elevation: ca. 2179–2800 m.

Biology. Sifted from leaf litter in primary forest.

Etymology. This epithet is a combination of the genus name *Ptolycus* and the Latin suffix *-oides* (having the form of) and refers to the species' resemblance in habitus.

Notes. *Trigonopterus ptolycoides* Riedel, sp. n. was coded as "*Trigonopterus* sp. 68" by Tänzler et al. (2012).



16 b



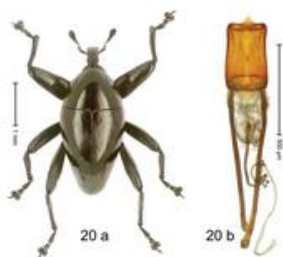
17 b



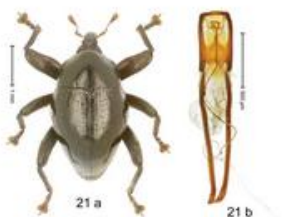
18 b



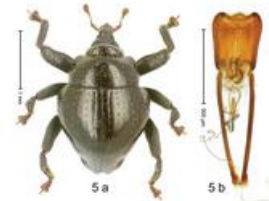
19 b



20 b



21 b



5 a

5 b



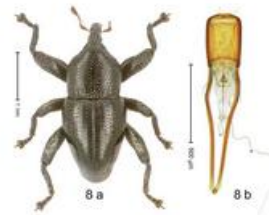
6 a

6 b



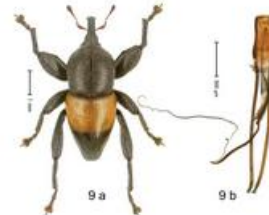
7 a

7 b



8 a

8 b



9 a

9 b



10 a

10 b



Biodiversity Data Journal 1: e1013 (28 Oct 2013)
doi: 10.3897/BDJ.1.e1013



Taxonomic paper

Eupolybothrus cavernicolus Komerički & Stoev sp. n. (Chilopoda: Lithobiomorpha: Lithobiidae): the first eukaryotic species description combining transcriptomic, DNA barcoding and micro-CT imaging data

▼ Pavel Stoev, Ana Komerički, Nesrine Akkari, Shanlin Liu, Xin Zhou, Alexander M. Weigand, Jeroen Hostens, Christopher I. Hunter, Scott C. Edmunds, David Porco, Marzio Zapparoli, Teodor Georgiev, Daniel Mietchen, David Roberts, Sarah Faulwetter, Vincent Smith, Lyubomir Penev

Abstract

We demonstrate how a classical taxonomic description of a new species can be enhanced by applying new generation molecular methods, and novel computing and imaging technologies. A cave-dwelling centipede, *Eupolybothrus cavernicolus* Komerički & Stoev sp. n. (Chilopoda: Lithobiomorpha: Lithobiidae), found in a remote karst region in Knin, Croatia, is the first eukaryotic species for which, in addition to the traditional morphological description, we provide a fully sequenced transcriptome, a DNA barcode, detailed anatomical X-ray microtomography (micro-CT) scans, and a movie of the living specimen to document important traits of its ex-situ behaviour. By employing micro-CT scanning in a new species for the first time, we create a high-resolution morphological and anatomical dataset that allows virtual reconstructions of the specimen and subsequent interactive manipulation to test the recently introduced 'cybertype' notion. In addition, the transcriptome was recorded with a total of 67,785 scaffolds, having an average length of 812 bp and N50 of 1,448 bp (see GigaDB). Subsequent annotation of 22,866 scaffolds was conducted by tracing homologs against current available databases, including Nr, SwissProt and COG. This pilot project illustrates a workflow of producing, storing, publishing and disseminating large data sets associated with a description of a new taxon. All data have been deposited in publicly accessible repositories, such as GigaScience GigaDB, NCBI, BOLD, Morphbank

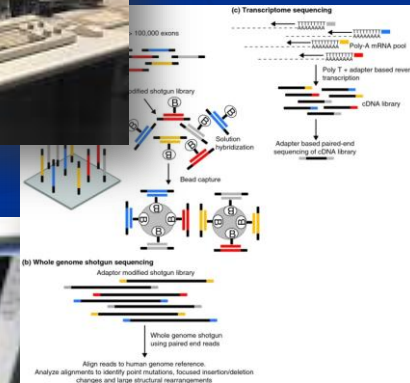
[Contents](#)
[Article info](#)
[Citation](#)
[Metrics](#)
[Share](#)
[Review it](#)
[Figures](#)
[Tables](#)
[Map](#)
[Taxa](#)
[Data](#)
[References](#)
[Article title](#)
[Abstract](#)
[Introduction](#)
[The study group](#)
[Materials and methods](#)
[Collected material and morphological study](#)
[Molecular experiments and sequencing](#)
[Micro-CT scanning](#)
[Abbreviations](#)
[Taxon treatments](#)
[Eupolybothrus cavernicolus](#)
[Eupolybothrus leostygis](#)
[Identification key to the species of Eupolybothrus \(Schizopolybothrus\) based on adult males](#)
[Analysis](#)
[Molecular delimitations](#)
[Transcriptome analysis and annotation](#)
[Discussion](#)
[Taxonomic affinities](#)
[Micro-computed tomography and 'cybertype' notion](#)
[Data management and release](#)
[Conclusions](#)
[Acknowledgements](#)
[References](#)
[Supplementary files](#)

Eupolybothrus cavernicolus



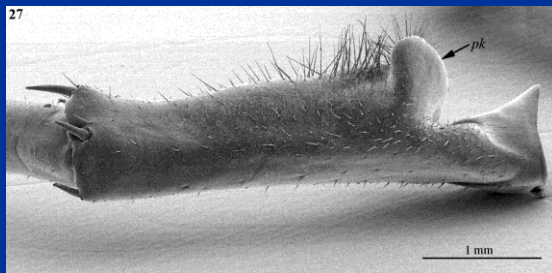
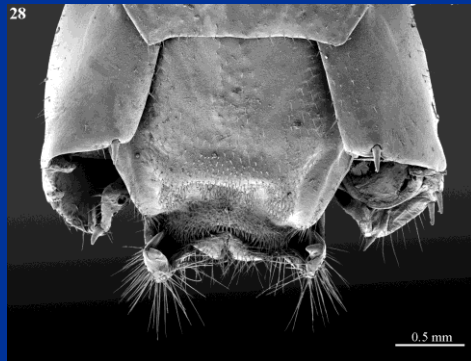
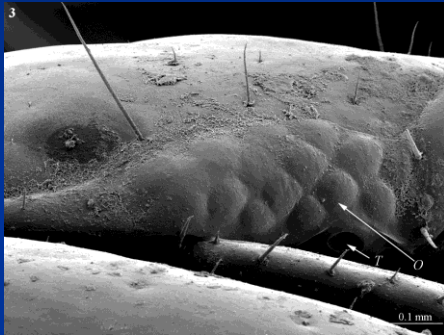
Methods

- Morphological study with a Zeiss microscope
- SEM images
- Mitochondrial Cytochrome C Oxidase Subunit I gene sequencing
- Full transcriptome sequencing
- Micro-CT scanning
- Video of living specimen

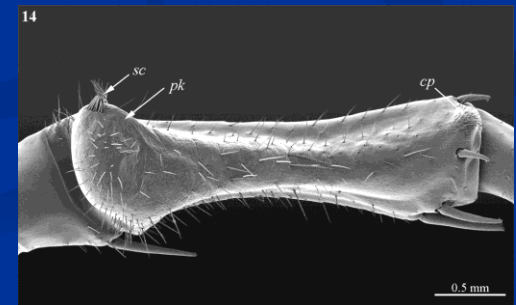
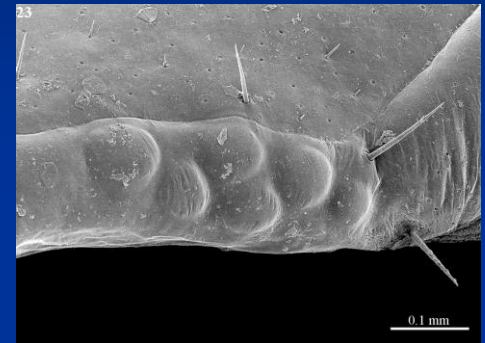


Morphology

E. cavernicolus sp.n.

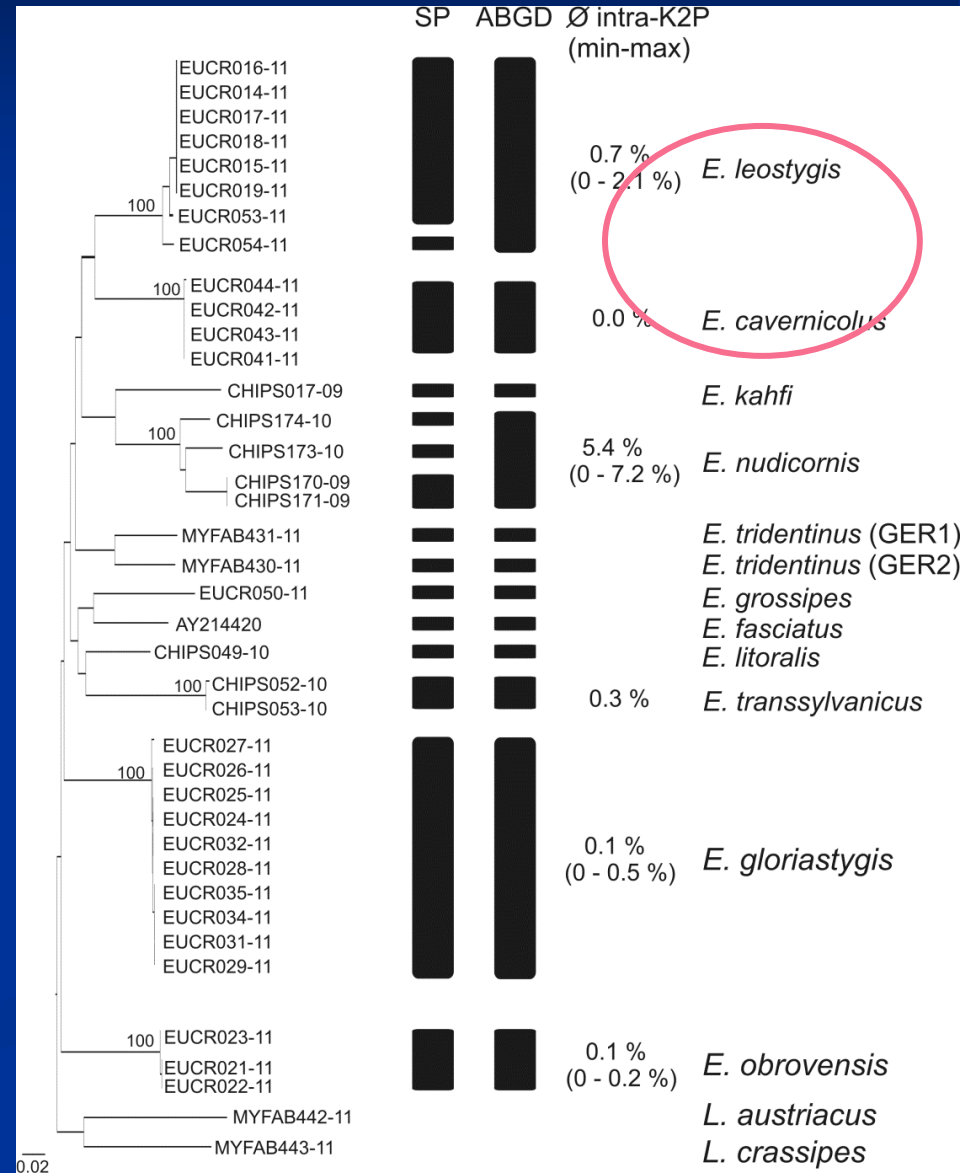


E. leostygis



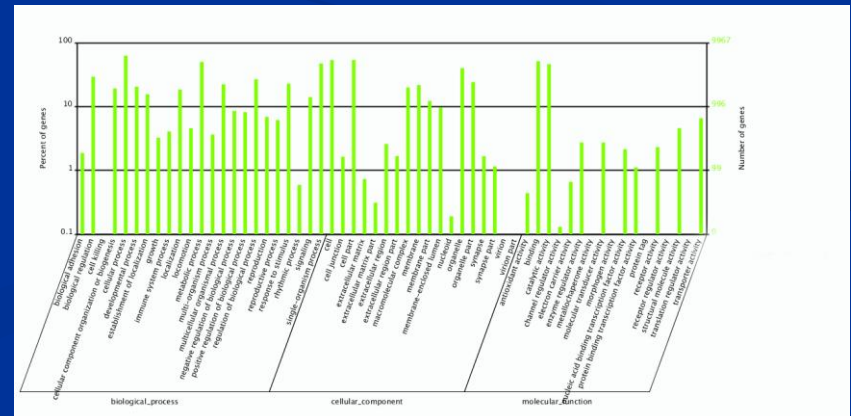
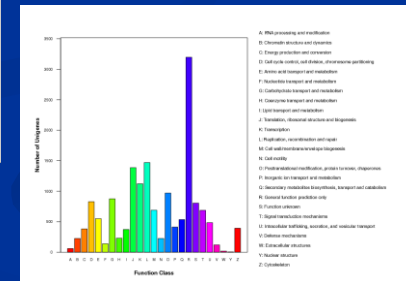
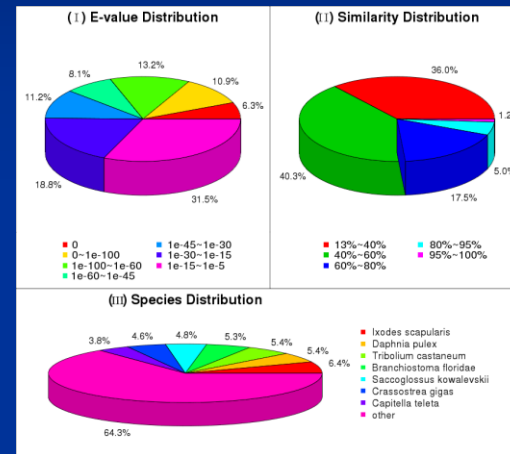
Molecular delimitation

1. The final dataset comprises 39 COI sequences from 10 species.
2. Molecular delimitation through ABGD algorithm - Puillandre et al. (2012)
3. The new species clusters with *E. leostygis*, its morphologically closest neighbour
4. Intraspecific K2P genetic distance was 0.0 % for *E. cavernicolus*



Transcriptome sequencing

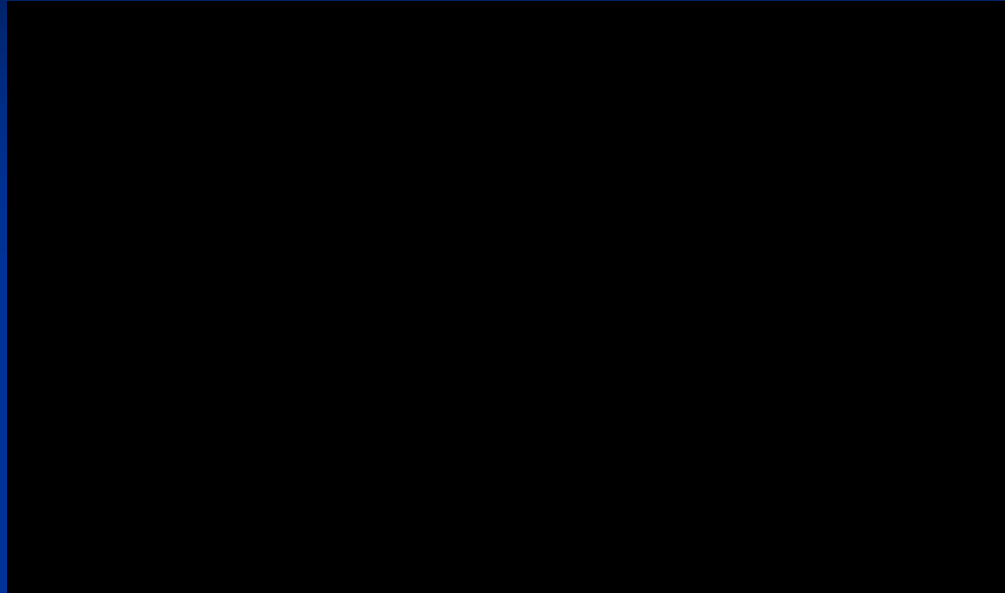
- i. A total of 67,785 scaffolds were produced with an average length of 812 bp.
- ii. Subsequent annotation was conducted by tracing homologs against current available databases
- iii. Using orthoDB, 2,188 one to one orthologs were filtered out from four selected arthropod genomes: *Drosophila melanogaster*, *Daphnia pulex*, *Ixodes scapularis* and *Strigamia maritima*



Micro-CT in Taxonomy

By employing micro-CT scans in taxonomy, important morphological and anatomical characters can be examined in their natural position **without damage to the original specimen**

Micro-CT scan




Faulwetter et al. (2013) suggested the creation of high-resolution virtual morphological and anatomical data libraries allowing reconstruction and interactive manipulation of type specimens, the so-called 'cybertypes'.

Data repositories

- GigaDB (complete database)
- NCBI (COI and transcriptome sequences)
- BOLD (COI)
- Morphbank (morphological images)
- Morphosource (micro CT row data, 3D model)
- Plazi (taxon treatments of legacy literature related to genus *Eupolybothrus*)

Data management pilot



Revolutionizing data dissemination, organization, and use

[Home](#) | [About](#) | [Contact](#) | [Terms of use](#)

[Help](#) [Login](#) [Create account](#)

Data released on January 01, 2014

The first data and images of the centipede *Eupolybothrus cavernicolus* sp.n.

The species description of a cave-dwelling centipede, *Eupolybothrus cavernicolus* Komenčki & Stoev sp. n. (Chilopoda: Lithobiomorpha: Lithobiidae), found in a remote karst region in Knin, Croatia. This is the first eukaryotic species for which, in addition to traditional morphological description, a fully sequenced transcriptome, DNA barcode (i.e., mitochondrial Cytochrome C Oxidase Subunit I gene; COI), and detailed anatomical X-ray microtomography (micro-CT) scan has been provided. This project illustrates a workflow of producing, storing, publishing and disseminating large data sets associated with a description of a new taxon.


This dataset consist of the transcriptomic and annotation data as well as the Micro-CT scans, photographs, scanning electron micrographs (SEM) and video footage of the species.

Accessions (data included in GigaDB):
AE: [E-MTAB-1859](#)
ERA: [ERP003841](#)

Accessions (data not in GigaDB):
BOLD: [/10.5883/DS-EUPCAV](#)

Samples:

Sample ID	Taxonomic ID	Common name	Genbank name	Scientific name	Sample attributes
Centipede_Paratype2	1391732	Eupolybothrus sp. PS-2013		Eupolybothrus sp. PS-2013	description="One centipede collected from cave Milkacka II, near Hydroelectric power plant Miljacka, Kistanje, Krka, Knin, Croatia. Immediatly crushed"

Genomic, Transcriptomic, Imaging

GigaDB in this example provides direct links between the genomic and micro-CT data, through a Darwin Core CSV dataset describing the type specimens, as well as capturing all of the metadata in the interoperable ISA-TAB format

Next generation sequencing
and imaging require
next generation publishing!

Journal features

- Focus and Scope
- Globally unique innovations
- Criteria for publication
- Peer review
- For authors
- Data publication
- Publication fees
- Licenses and Copyright
- Frequently Asked Questions (FAQ)
- Contacts
- Editorial team

Follow us



This work is licensed under the Creative Commons Attribution 3.0 (CC-BY).



Resolving the publishing bottleneck for biodiversity

Science is a combination of gathering facts and making theories; neither can progress on its own. In the history of science, the laborious accumulation of facts is the dominant mode, not a novelty.

Peter Norvig



Making "small" data big

- No lower/upper limit of manuscript size
- Publish all kinds of biodiversity related data
- Reduced page charges affordable by all



More than just data journal

- Integrated text and data publishing
- Completely online revisions and
- Community ownership of data

Free of charge in launch phase



PENSOFT® Writing Tool

Authoring, reviewing and publishing in one place, for first time!

How it works

Online manuscript authoring

- ✓ Collaborate online with coauthors and peers
- ✓ No author guidelines, the tool guides you
- ✓ Easy to use, flexible article templates

Getting started

- ✓ Write a manuscript online
- ✓ Submit it at the click of a button
- ✓ Easy online revisions and editing

Free to use

Data and text publishing integrated

- ✓ Gateway to the Biodiversity Data Journal
- ✓ Data into text conversion and vice versa
- ✓ Novel workflow reduces publication costs

Dr. Lyubomir Penev

Start a manuscript

My recent manuscripts

Test checklist
 Title

See more

Why publish my data?

- Citable publication
- Increase collaboration
- Establish scientific priority
- Link data to a bigger network



Novel concepts

- Authoring, reviewing and publishing **in one place**, for first time!
- **Upfront** markup!
- “Small” **data** imported **in** and downloadable **from text**
- Large datasets published as data papers
- **Text** stored and treated **as data**
- **Community peer-reviewed** and **community owned** interoperable data
- **Text and data integrated publishing!**

Download data



Biodiversity Data Journal 1: e1013 (28 Oct 2013)
doi: 10.3897/BDJ.1.e1013



prefemur, F – femur, T – tibia, a, m, p stand for spines in respectively, anterior, medial and posterior position.

Taxon treatments

Eupolybothrus cavernicolus Komerički & Stoev, 2013, sp. n.

- ZooBank urn:lsid:zoobank.org:act:6F9A6F3C-687A-436A-9497-70596584678C
- SRA project accession <http://www.ebi.ac.uk/ena/data/view/ERP003841>
- ArrayExpress accession <http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-1859>
- GigaDB <http://dx.doi.org/10.5524/100063>
- GenBank [KF715043](https://www.ncbi.nlm.nih.gov/GenBank/KF715043)
- GenBank [KF715049](https://www.ncbi.nlm.nih.gov/GenBank/KF715049)
- GenBank [KF715050](https://www.ncbi.nlm.nih.gov/GenBank/KF715050)
- GenBank [KF715059](https://www.ncbi.nlm.nih.gov/GenBank/KF715059)
- BOLD dataset <http://dx.doi.org/10.5883/DS-EUPCAV>
- MorphBank [999021821&tsn=true](http://morphobank.org/999021821&tsn=true)
- MorphoSource http://morphosource.org/index.php/Detail/SpecimenDetail/Show/specimen_id/514

Materials

Holotype:

- a. country: Croatia; stateProvince: Knin; locality: NP Krka, village Kistanje, Hydroelectric power plant Miljacka, cave Miljacka II; verbatimElevation: 115 m; verbatimLatitude: [44°00'01.1"N](#); verbatimLongitude: [16°00'58.5"E](#); samplingProtocol: hand collected under clay sediment; eventDate: 9 February 2013; individualCount: 1; sex: male; lifeStage: adult; recordedBy: M. Lukić; institutionCode: CBSS; collectionCode: CHP536

Paratypes:

- a. country: Croatia; stateProvince: Knin; locality: NP Krka, village Kistanje, Hydroelectric power plant Miljacka, cave Miljacka II; verbatimElevation: 115 m; verbatimLatitude: [44°00'01.1"N](#); verbatimLongitude: [16°00'58.5"E](#); samplingProtocol: hand collected under lump of clay; eventDate: 18 October 2012; individualCount: 1; sex: female; lifeStage: adult; recordedBy: A. Komerički; institutionCode: CBSS; collectionCode: CHP517
- b. country: Croatia; stateProvince: Knin; verbatimLocality: NP Krka, village Kistanje, Hydroelectric

Download as CSV



Contents

Article info

Citation

Metrics

Share

Review it

Figures

Tables

Map

Taxa

Data

References



Tables and Figures, if present, can be downloaded from the article.

Download all occurrences as Darwin Core Archive

Supplementary material 1

Raw data used for COI delineation of the *Eupolybothrus* species

Authors: Stoev et al. 2013

Data type: genomic

Brief description: The archive contains the following data: 1) fasta-Alignment as the basis for all analyses (.FASTA), 2) mega-file for the calculation of the genetic distances and the NJ tree (.MDSX), 3) NJ-tree in Newick format (.NWK), 4) graph of the TCS Software for the Statistical Parsimony method (.GRAPH)

Filename: E_cavernicolus.rar

[Download file](#) (10.35 kb)

Interactive mapping of localities



Biodiversity Data Journal 1: e979 (01-09-2013)
doi: 10.3897/BDJ.1.e979



Feeds on: Adults and larvae feed on horse dung (Fig. 5).



Figure 5.

Aphodius bimaculatus. Dosang environs, Astrakhan Province, Russia.

Distribution: Central and Eastern Europe, Western Asia up to East Kazakhstan in the east. The species is mostly occur in forest-steppe and steppe zones.

Aphodius (Melaphodius) caspius Ménétriés, 1832

Material

[Download as CSV](#)

- a. country: Russia; stateProvince: Astrakhan'; locality: Dosang environs, left bank of Akhtuba River, floodplain; decimalLatitude: [46.91](#); decimalLongitude: [47.91](#); samplingProtocol: horse dung washing; eventDate: 2006-10-06; individualCount: 1; recordedBy: A. V. Frolov, L. A. Akhmetova; collectionID: urn:lsid:biocol.org:col:35005; institutionCode: ZIN; collectionCode: *Coleoptera*

Feeds on: Cattle dung.

Distribution: Steppe zone from Caucasus to West Siberia.

Contents

Article info

Citation

Related

Metrics

Share

Figures

Map

Taxa

Data

References



☐ All

☒ All taxa

☒ *Aphodius (Erytus) aequalis*

☒ *Aphodius (Chilothorax) badenkoi*

☒ *Aphodius (Acrossus) bimaculatus*

☒ *Aphodius (Melaphodius) caspius*

I have a dream...



Imagine how taxonomic data
(if published this way) can be re-
used...

... across different
domains....

I



Open Access!



ViBRANT