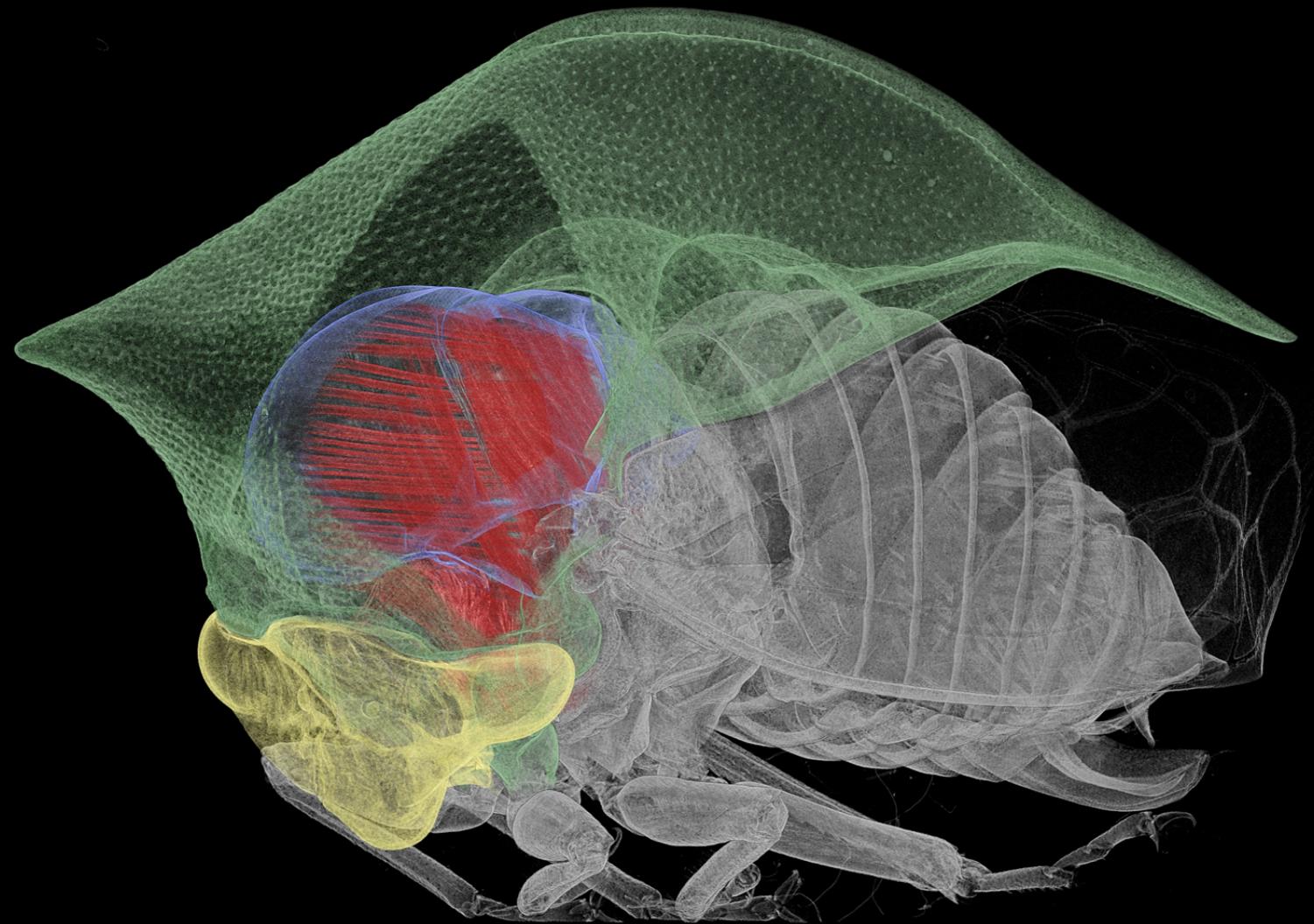


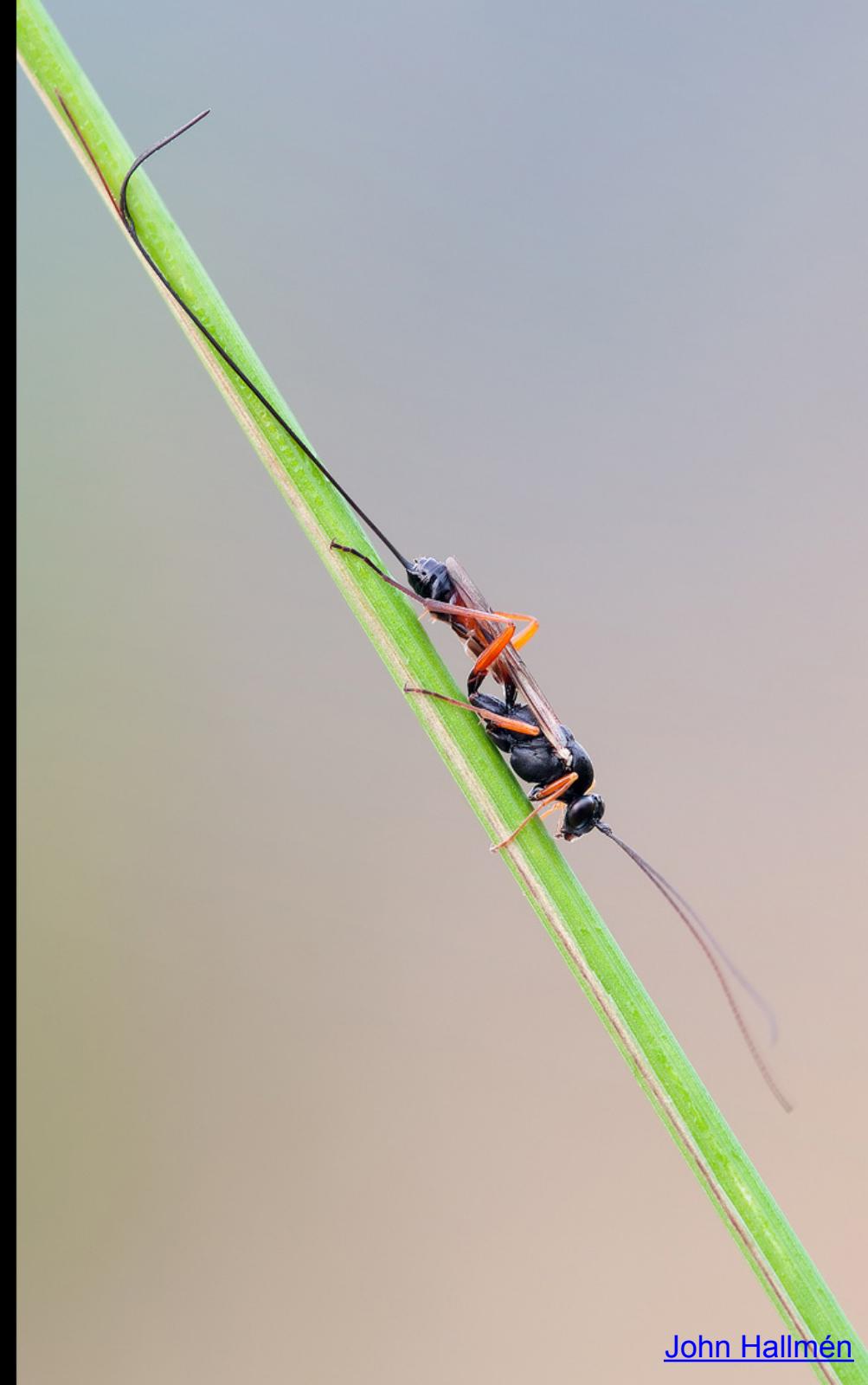
# Illuminating the Phenome: The Future of Morphology in Entomology



# Changing the way we describe biodiversity

Andy Deans  
Katja Seltmann  
Matt Bertone  
István Mikó  
Jim Balhoff  
Matt Yoder

*North Carolina State University*



# Acknowledgments

## funding:

- Improvements to Biological Research Collections (*NSF DBI-0847924*)
- Advances in Biological Informatics (*NSF DBI-0850223*)
- Research Coordination Networks (*NSF DEB-0956049*)
- Systematic Biology (*NSF DEB-0842289*)
- EOL BioSynC
- NESCent (*NSF EF-0423641*)



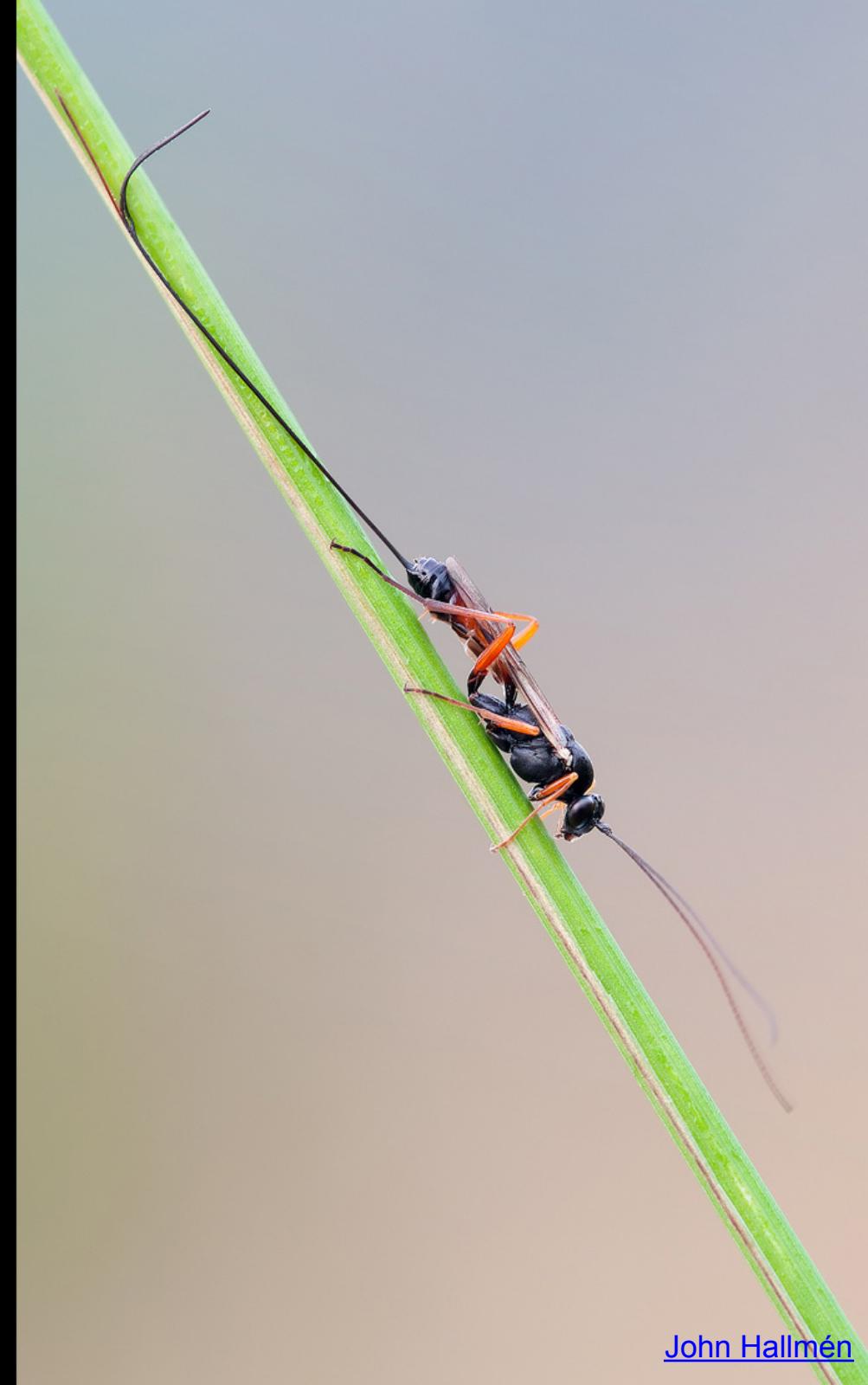
## intellect and/or enthusiasm:

- Amy Bader, Bob Blinn, Kelly Dew
- Trish Mullins, Andrew Ernst, Heather Campbell-Melvin
- Hilmar Lapp, Todd Vision, Wasila Dahdul (NESCent)
- Paula Mabee (USD)
- many, many others

Hymenoptera images:

<http://www.flickr.com/photos/johnhallmen/>  
<http://www.flickr.com/photos/eurythyrea/>

# Seeding a Revolution in descriptive taxonomy



[John Hallmén](#)



# How are taxa described?



ignita. 23. S. glabra nitida, thorace viridi, abdomine aureo : apice  
quadridentato.

*Fn. svec. 1004.*

*Frisch. inf. 9. t. 10. f. I.*

*Habitat in muris Europæ.*



= *Chrysis ignita* (Linnaeus, 1758)



ignita. 23. S. glabra nitida, thorace viridi, abdomine aureo : apice  
quadridentato.

*Fn. svec. 1004.*

*Frisch. inf. 9. t. 10. f. I.*

*Habitat in muris Europæ.*

216. SPHEX. Os maxillis absque proboscide.  
*Alæ* plano-incumbentes (nec plicatæ),  
in omni sexu.  
*Aculeus* punctorius reconditus.

\* *Abdomine petiolato*: petiolo elongato.



= *Chrysis ignita* (Linnaeus, 1758)

# The Quiet Revolution: Biodiversity Informatics and the Internet

Frank A. Bisby

The massive development of biodiversity-related information systems on the Internet has created much that appears exciting but chaotic, a diversity to match biodiversity itself. This richness and the arrays of new sources are counterbalanced by the maddening difficulty in knowing what is where, or of comparing like with like. But quietly, behind the first waves of exuberance, biologists and computer scientists have started to pull together in a rising tide of coherence and organization. The fledgling field of biodiversity informatics looks set to deliver major advances that could turn the Internet into a giant global biodiversity information system.

There is a resonance between the needs of biodiversity science and the opportunities for globalization and interoperability provided by the Internet. One is that biodiversity workers are distributed all over the globe, literally dotted about in every country and on every island. A second arises from our interdependence. Global events and global syntheses in biodiversity have an impact on all of us. People who set conservation priorities do not just access local information, they need to understand the whole; they need information, for instance, from neighboring regions and from climatically similar lands in distant continents. But third, and most important, the science of global biodiversity studies depends critically on high-level concepts—biomes, ecosystems, phyla, floras and faunas, hot-spots, genetic erosion, the impact of aliens—abstractions put together by synthesizing the myriad observations and studies by local observers, local teams, and local institutions. Hence, a central goal in biodiversity informatics is to develop systems that permit interoperability and knowledge synthesis across wide arrays of local systems, and to embed them in global knowledge architectures such as Species 2000 (1) and the Global Biodiversity Information Facility (GBIF) (2). Here, I give a brief picture of the research, techniques, and developments that are bringing these goals within reach.

**Interoperability.** One priority is to draw together basic biodiversity accession records from dispersed sites. How could we access the vast number of plant and animal records dispersed in the museums and herbaria of the world? The utility of doing this was first demonstrated by the Australian government's original Environmental Resources Information Network (ERIN) system, now part of Environment Australia Online (3), albeit by centralizing plant and animal distribution

Centre for Plant Diversity & Systematics, School of Plant Sciences, University of Reading, Reading RG6 6AS, UK. E-mail: f.a.bisby@reading.ac.uk

records. ERIN led the way by making the combined data available for Australia-wide Geographic Information System (GIS) analysis and modeling.

A number of interoperative systems are approaching the tasks originally offered by ERIN for its centralized data, but with the powerful possibility of extending to data from a vast range of autonomous institutions around the world. The Biological Collection Information Service for Europe (BioCISE) program (4) has established an extensive metadata system holding information centrally on the contents and locations of various collections. The idea is that intelligent software will lead users to this information, which will be retrieved using common interfaces. The University of Kansas team is developing its Species Analyst system (5), which interacts directly with an array of herbarium and museum accession databases. The Z39.50 protocol is used to locate and return records, and these are transformed into Extensible Markup Language (XML) for use by World Wide Web browsers and analytical software. The Z39.50 search profile used corresponds to the Darwin Core metadata standard (6) being developed informally among U.S. institutions. A request (such as for specimen records for a particular species) goes to all museums and herbaria selected by the user, and the dispersed databases return data, for instance, giving latitude, longitude, and date for every matching specimen. The assembled data set from mixed sources is then available for analysis using GIS mapping and modeling routines at the San Diego Supercomputer Center. Similar goals are being pursued by the TaxaServer group in

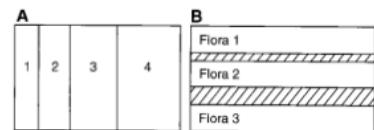
Australia (7) and by the European Natural History Specimen Information Network (ENHSIN) team in Europe (8).

A second area for networking and interoperability is the taxonomic framework itself. Again, there are centralized models from the 1990s where organizations bring together taxonomic treatments from authors and institutions to provide a centrally collated system. It now seems agreed that these taxonomic frameworks should be constructed “taxon-by-taxon” as in Species 2000 (1), the Integrated Taxonomic Information System (ITIS) (9), and the UNESCO-IOC Register of Marine Organisms (URMO) (10), thus avoiding the “flora-by-flora” work of integrating systems in which the taxonomies overlap, a contrast illustrated in Fig. 1. Only the International Organization for Plant Information (IOP) Global Plant Checklist (11), perhaps because of well-developed flora databases, is attempting the flora-by-flora route (12).

**Species 2000.** Species 2000 (1) is a global program to compile a “catalog of life” using distributed networking on the Internet. It has the ambitious aim of creating a uniform and validated index to the world’s known species for use as a practical tool in inventorying and monitoring biodiversity worldwide. The index will be used to provide (i) electronic baseline species lists for use in inventorying projects worldwide, (ii) the index for an Internet digital library of species databases worldwide, (iii) a reference system for comparison between inventories, and (iv) a comprehensive worldwide catalog for checking the status, classification, and naming of species.

The comprehensive index of all known plants, animals, fungi, and microorganisms is being constructed by accessing a distributed array of taxonomic indexes, one for each group of organisms. These are known as global species databases (GSDs), represented by boxes in the primary array of the Species 2000 architecture (Fig. 2). The taxonomic database organizations starting the program already provide such indexes for viruses, bacteria, archaea, corals, algae (red, green, and brown), cephalopods, crustaceans, scarabaeid

**Fig. 1.** Comparison between global taxonomies assembled from taxonomic treatments for complete taxa (taxon-by-taxon, no overlap) (A) and from floras or faunas (in this example, flora-by-flora, with overlaps) (B). [Adapted from (12)].



Downloaded from www.sciencemag.org on February 22, 2011





# Extending the diagnostic process ...



THE ROYAL  
SOCIETY

Received 29 July 2002  
Accepted 30 September 2002  
Published online 8 January 2003

## Biological identifications through DNA barcodes

Paul D. N. Hebert<sup>1</sup>, Alina Cywinska, Shelley L. Ball  
and Jeremy R. deWaard

Department of Zoology, University of Guelph, Guelph, Ontario N1G 2W1, Canada

Although much biological research depends upon species diagnoses, taxonomic expertise is collapsing. We are convinced that the sole prospect for a sustainable identification capability lies in the construction of systems that employ DNA sequences as taxon 'barcodes'. We establish that the mitochondrial gene cytochrome c oxidase I (COI) can serve as the core of a global biodiversity system for animals. First, we demonstrate that COI profiles, derived from the low-density sampling of higher taxonomic categories, ordinarily assign newly analysed taxa to the appropriate phylum or order. Second, we demonstrate that species-level assignments can be obtained by creating comprehensive COI profiles. A model COI profile, based upon the analysis of a single individual from each of 200 closely allied species of lepidopterans, was 100% successful in correctly identifying subsequent specimens. When fully developed, a COI identification system will provide a reliable, cost-effective and accessible solution to the current problem of species identification. Its assembly will also generate important new insights into the diversification of life and the rules of molecular evolution.

**Keywords:** molecular taxonomy; mitochondrial DNA; animals; insects; sequence diversity; evolution

### 1. INTRODUCTION

The diversity of life underpins all biological studies, but it is also a harsh burden. Whereas physicists deal with a cosmos assembled from 12 fundamental particles, biologists confront a planet populated by millions of species. Their discrimination is no easy task. In fact, since few taxonomists can critically identify more than 0.01% of the estimated 10–15 million species (Hammond 1992; Hawksworth & Kalin-Arroyo 1995), a community of 15 000 taxonomists will be required, in perpetuity, to identify life if our reliance on morphological diagnosis is to be sustained. Moreover, this approach to the task of routine species identification has four significant limitations. First, both phenotypic plasticity and genetic variability in the characters employed for species recognition can lead to incorrect identifications. Second, this approach overlooks morphologically cryptic taxa, which are common in many groups (Knowlton 1993; Jarman & Elliott 2000). Third, since morphological keys are often effective only for a particular life stage or gender, many individuals cannot be identified. Finally, although modern interactive versions represent a major advance, the use of keys often demands such a high level of expertise that misdiagnoses are common.

The limitations inherent in morphology-based identification systems and the dwindling pool of taxonomists signal the need for a new approach to taxon recognition. Microgenomic identification systems, which permit life's discrimination through the analysis of a small segment of the genome, represent one extremely promising approach to the diagnosis of biological diversity. This concept has already gained broad acceptance among those working with the least morphologically tractable groups, such as viruses, bacteria and protists (Nanney 1982; Pace 1997;

Allander *et al.* 2001; Hamels *et al.* 2001). However, the problems inherent in morphological taxonomy are general enough to merit the extension of this approach to all life. In fact, there are a growing number of cases in which DNA-based identification systems have been applied to higher organisms (Brown *et al.* 1999; Bucklin *et al.* 1999; Trewick 2000; Vincent *et al.* 2000).

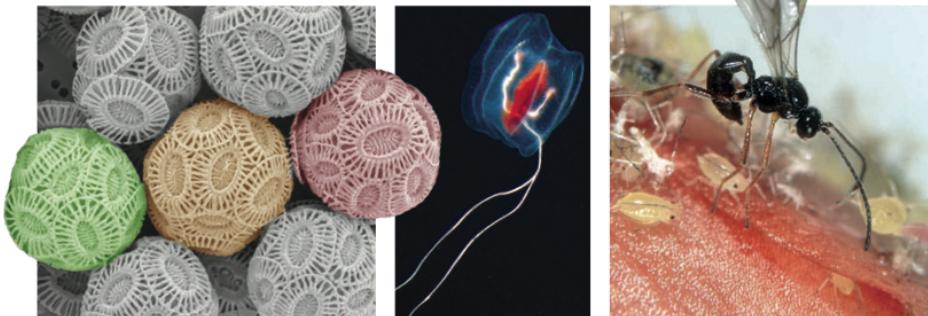
Genomic approaches to taxon diagnosis exploit diversity among DNA sequences to identify organisms (Kurtzman 1994; Wilson 1995). In a very real sense, these sequences can be viewed as genetic 'barcodes' that are embedded in every cell. When one considers the discrimination of life's diversity from a combinatorial perspective, it is a modest problem. The Universal Product Codes, used to identify retail products, employ 10 alternate numerals at 11 positions to generate 100 billion unique identifiers. Genomic barcodes have only four alternate nucleotides at each position, but the string of sites available for inspection is huge. The survey of just 15 of these nucleotide positions creates the possibility of  $4^{15}$  (1 billion) codes, 100 times the number that would be required to discriminate life if each taxon was uniquely barcoded. However, the survey of nucleotide diversity needs to be more comprehensive because functional constraints hold some nucleotide positions constant and intraspecific diversity exists at others. The impact of functional constraints can be reduced by focusing on a protein-coding gene, given that most shifts at the third nucleotide position of codons are weakly constrained by selection because of their four-fold degeneracy. Hence, by examining any stretch of 45 nucleotides, one gains access to 15 sites weakly affected by selection and, therefore, 1 billion possible identification labels. In practice, there is no need to constrain analysis to such short stretches of DNA because sequence information is easily obtained for DNA fragments hundreds of base pairs (bp) long. This ability to inspect longer sequences is significant, given two other biological considerations. First, nucleotide composition at third-position sites is often

\*Author for correspondence (phebert@uoguelph.ca).

nature

Vol 467 | 9 September 2010

## OPINION



## Time to automate identification

Taxonomists should work with specialists in pattern recognition, machine learning and artificial intelligence, say Norman MacLeod, Mark Benfield and Phil Culverhouse — more accuracy and less drudgery will result.

An imaging system designed to identify marine zooplankton was recently adopted by scientists working for the US government to monitor the Deepwater Horizon oil spill. By measuring the size of oil droplets produced after chemical dispersants had broken up the oil, modellers could predict the depths at which the plume was accumulating. Only two instruments exist that can measure oil droplets while distinguishing them from other matter suspended in the water column, such as zooplankton, marine snow and gas bubbles at depths of down to 1,500 metres. The deployment, by the US National Oceanic and Atmospheric Administration, of one — the digital holographic imaging (DHI) system, developed jointly by the Massachusetts Institute of Technology in Cambridge and the Woods Hole Oceanographic Institution in Massachusetts — is a working example of something that should be happening on a grand scale: the shared use across diverse disciplines of generalized automated identification technologies.

Taxonomists who identify, describe and name species (who practise alpha taxonomy, as it is known in the trade) are central to many research programmes in applied biology, ecology and conservation. University cuts are shrinking this already small community. What's more, there is no tradition of — much less a requirement for — independent testing and verification of the accuracy of the identifications that taxonomists produce, unlike in

virtually every other scientific discipline.

Automating species identification using technologies developed by researchers in pattern recognition, artificial intelligence and machine learning would transform alpha taxonomy from a cottage industry dependent on the expertise of a few individuals to a testable and verifiable science accessible to anyone needing to recognize objects. Indeed, a concerted interdisciplinary research and development effort, within the next decade, could lead to automated systems capable of high-throughput identifications for hundreds or thousands of categories of living as well as non-living specimens.

### Human error

Many taxonomists use sophisticated technologies to capture images, sounds, even the smells and tastes of biological specimens. But most routine identifications involve a small group of experts scattered around the world assessing diagnostic data qualitatively — commonly the size, shape or texture of specimens, or the presence or absence of certain features. Surprisingly few blind-test studies have been published to assess the accuracy of taxonomists' findings objectively<sup>1,2</sup>. Those that have been carried out are worrying. For instance, a blind test to resolve a controversy about the pattern of extinction in one marine

"Humans miss items,  
count some objects  
more than once and  
misclassify others."

animal group about 65.5 million years ago<sup>1</sup> resulted in species lists that were so different as to make consensus impossible. Such inconsistencies shouldn't be a surprise given that, in controlled visual-cognition studies, humans frequently miss items presented in a scene, count some objects more than once and misclassify others.

Hopes are high among researchers and funding bodies that DNA bar-coding, by which a species is recognized according to a marker in its mitochondrial genome, will increase the accuracy of identifications — and ease bottlenecks resulting from a shortage of trained and experienced taxonomists. But bar codes are generally used to assign organisms to taxonomic categories that have already been defined on the basis of morphological traits. In other words, a bar code isn't useful until the reference species has already been identified by multiple experts.

The technique is still relatively expensive, slow and difficult to implement in the field except in certain situations — for example in laboratories on oceanographic research vessels. Moreover, researchers frequently need to identify non-living objects as well as living ones. Ecologists studying plankton, for example, commonly count 'fibres', 'detritus' or 'egg-like particles' that may or may not be alive.

In focusing on bar-coding, stakeholders have



What about taxonomic  
descriptions\*?

\* including diagnoses

**Female.** Body length 10.2 mm.

**Head.** Height 2.1 (1.8–2.1) mm, width 2.7 (2.3–2.7) mm, length 1.3 (1.1–1.3) mm, LID 1.5 (1.2–1.5) mm, MOD 0.2 mm. Frons greenish-blue. Scapal basin closely punctured, shining intervals between punctures present only around mediosagittal line. Upper half of frons is with a small but deep oval depression in the mediosagittal line. Transverse frontal carina well developed and slightly arced (Fig. 1). Vertex is coarsely punctured and black, without metallic luster. Genae bluish-green, malar space 0.3 mm long. Mandibles are black except basal and inner areas, which are metallic bluish-green and brown, respectively. Interior ridge of mandibles with one subapical tooth (Fig. 1). In lateral view, mandible fairly thick: basally 0.4 (0.3–0.4) mm and centrally 0.2 mm. Scapus, pedicellus and basal part of F-I dorsally blue, remainder of the antennae black. Lengths of antennal segments as follows: scapus 1.1 (0.9–1.1) mm, pedicel 0.3 (0.3–0.4) mm, F-I 0.6 (0.5–0.6) mm, F-II 0.4 (0.3–0.4) mm and F-III 0.3 (0.3–0.4) mm. Width of the scapus is 0.3 mm and first three antennal segments is 0.2 mm. F-IV and the following flagellomeres are distinctly longer than they are broad.

**Thorax.** Length 3.8 (3.2–4.1) mm. Pronotal collar 0.7 (0.5–0.7) mm long in the centre, black except metallic green on borders and area around pronotal furrow. Pronotal furrow well developed. Scutum is lusterless black, uniformly punctured. Tegulae are metallic green. Punctures on scutellum are slightly larger than those on scutum. Scutellum black, with metallic blue and green laterally. Metanotum unmodified, basal area centrally with black spot, rest of metanotum with violet and green reflections. Outer edge of lateral propodeal teeth is straight. Wing venation is as in other species of the *Chrysis ignita* group. Mid and hind tarsi are robust and shortened, hind tarsus is shorter than hind tibia. Ventral side of femora and tibiae are laterally bluish-green, tarsi dark brown. Apical spurs of the mid tibiae approximately equal to one another.

**Abdomen.** length 4.6 (4–4.6) mm (T-I 1.5 (1.1–1.6) mm, T-II 2.1 (1.6–2.3) mm, T-III 1.5 (1.3–2.1) mm), maximum width 2.8 (2.3–3) mm. Basal and lateral edge of T-I greenish, gradually becoming metallic red towards distal edge, rest of abdomen metallic

red. Punctures on T-I and scutum are almost sub-equal in size, intervals with small dots. T-II, especially its basal part, is finely and densely punctured. T-II is with well developed longitudinal central carina. T-III is with very dense and fine punctures, and interspaces between the punctures with microsculpture. Apical margin of T-III is with four triangular teeth (Fig. 3). Central teeth are sharper and closer to each other than to lateral ones. Subapical pit row with 17 black pits, some pits are confluent. Subapical transverse swelling is weak. S-II is metallic green, with black spots as in Figure 5, S-III black. Ovipositor is short and thick, gastral terga IV–VII and sterna IV–VI as in Figures 9–15.

**Diagnosis.** Female of *Chrysis tripotini* sp. n. is separated from other members of *C. ignita* species group by a combination of following characters: inner edge of mandible with subapical tooth, shortened mid and hind tarsi (tarsus shorter than tibia) and lusterless black coloration to vertex and most of the dorsal thorax.

**Remarks.** The holotype is missing two distal tarsal segments of the right midleg.

**Male.** Unknown.

**Type material.** Holotype. ♀, labeled: South Korea, Chungnam, Keumsan, Nami-myeon, Seokdong, Pohyeonsa, N 36° 03.494' / E 127° 27.225' 6 malaise traps in forested area, 14.v–1.vi.2005, leg P. Tripotin. Depository: Natural History Museum, University of Tartu, Tartu, Estonia (TUZ000001), tissue sample of holotype is stored at the same depository (VS 123).

**Paratypes.** 1 ♀, same data as holotype (Pierre Tripotin's private collection, Rouen, France), 1 ♀, same data as holotype, except date 21.–31.vii.2005 (Same depository as holotype (TUZ000002), extracted DNA is stored at the same depository (BR02)). 2 ♀, labeled: South Korea, Kangwondo, Chuncheon, Nam-Myeon, Hodong-li 6.–31.vii.2003 Malaise trap, leg P. Tripotin, Malaise trap in pastured area on trail in half shade close to edge of forest (author's private collection, Tartu, Estonia). 1 ♀, labeled: South Korea, Chungbuk, Okcheon Dongi, Saesan-li, around old house, 23.v.2006 leg P. Tripotin (Pierre Tripotin's private collection, Rouen, France). 1 ♀, labeled: South Korea, Gyeonggi-do, Gapyeong-gun, Seo-myeon, Magok-li, on old pile of logs, N 37° 42.97' / E 127° 35.47' alt 50 m, 7.vi.2006 leg P. Tripotin (author's private collection, Tartu, Estonia).



*Lyneborgimyia magnifica* sp.n  
(Figs 1–20)

Description

**Body length.** Approximately 18 mm (measured before dissection, but accurate measure not attainable due to bent abdomen). Wing length: 13.35 mm (from proximal margin of costigum to apex).

**Head.** (Figs 1–4). In frontal view almost circular, in lateral view approximately twice as high as long, with a depression of the postocular orbit and the posterior part of the eye lateral to the dorsal eye corner (Figs 3, 4). Dorsal and ventral halves of eye of about the same size, its posterior margin almost straight, only faintly curved anteriorly above the vertex, holoptic, eye contiguity approximately half the length of frons, the eye rims actually separated by an extremely narrow stripe of the frontal cuticle, anterior angle of eyes ca. 60°, ommatidia near eye contiguity insignificantly enlarged. Eyes rather sparsely setose, the margins and the eye bridge broadly bare, the setae 0.2 mm long, glistening silvery. Cuticle of head capsule (including back) entirely covered by dense

microtrichia, which are mostly white, but brown on some part of the back (especially near occipital foramen). Face concave, anterior mouth edge protruding, tentorial pit slit-like, facial ridge indistinct, border with antennal sockets straight, with a small weak tubercle below each antenna, ground colour blackish, with a reddish tinge laterally, difficult to observe under the dense layer of microtrichia, sparsely covered by short erect white setae laterally, leaving about the median third bare at about half height of face, and receding dorsally and ventrally, with a narrow gap between the facial and the frontal setosity, anterior corner of mouth edge bare. Subcranial cavity wide, anteriorly notched, lateral corner of mouth edge somewhat receding. Anteclypeus wide, about twice as long medially as wide proximally, widening distally, completely microtrichose, postclypeus sclerotized. Labela of proboscis externally without striking peculiarities (compared with *Merodon* spp.). Frons covered by a dense fringe of long, thick, subappressed, creamy white setae, anteriorly with erect, thin, brown setae, that are decreasing in length ventrally, ventrally reaching level of the ventral margin of antennal sockets, with a strong, point-like pit at half length, cuticle black, except for a triangular, non-setose area between lunule and frontal pit, which is red. Lunule dark brown, with slightly raised margins laterally, somewhat thickened, and a little produced in the middle, where it bears a small, microtrichose spot, otherwise without any vestiture, ventrally reaching level of ventral base of antenna. Antennal sockets approximately 1.5 times as wide as deep, without a median sclerotized stripe (see, e.g. Hippa & Ståhls, 2005: fig. 5c).

Postocular orbit dorsally only little wider than its narrowest part, without a postcranial carina, dorsally horizontal, below the depression sloping towards the posterior surface of the head, with which it has an angled border, parasagittal sulci well developed, long, white setae present at the posterior margin between the lateral ends of the depression, but leaving the vertex (i.e. the area between the parasagittal sulci) bare, and at about the ventral quarter of head, dark brown setae mixed with short, light setae are present anteriorly, leaving a narrow stripe (about as wide as an ocellus) along the eye margin bare, some of the short setae more or less spinose (without clear distinction to the slender setae). Vertical triangle long and narrow, with evenly diverging margins, its posterior part from the anterior ocellus to the posterior margin of the head distinctly raised above the level of the eyes, ocellar triangle equilateral, distance between posterior ocelli and eye corners equals the length of the ocellar triangle, posterior ocelli elongated (about twice as long as wide), nearly touching eye margin. Vertical triangle rather densely covered with long, dark brown setae, a few white setae present at the anterior end of the setosity, except for a number of bare areas: at the posterior end, where a stripe as long as the width of an ocellus is bare, narrow stripes along the eye margins from the posterior ocelli to the eye corners, the anterior half of the area between the anterior ocellus and the eye contiguity, and a narrow median stripe from the posterior end of the posterior ocelli towards the posterior margin of the vertical triangle. Postvertex narrow, with almost parallel margins, bare, with a deep pit at its ventral end. Hypostomal bridge with a sharp, transverse ariciform crest near the ventral margin

of the occipital foramen (Fig. 4). The dorsentral bulge wider than usual in *Merodon*, gradually narrowing ventrally (Fig. 4). Upper two-thirds of tempora setose, postgena and hypostomal bridge almost bare, with sparse, inconspicuous, short setae. The bare spot at the eye margin behind the ill-defined genal suture narrow (i.e. as usual within Eumerini, but in contrast to *Platynocheila* spp., where this spot is much enlarged).

Antenna blackish, postpedicel dorsally reddish, short, pedicel much shorter than deep medially, about as long as deep laterally, postpedicel about as long as deep. Scape about as long as deep, concave ventrally, with a strong bump on median surface. Pedicel (Fig. 18) laterally with a nearly rectangular excavation of its apical margin, its median surface nearly flat. Setae on scape and pedicel black to dark reddish brown, scape with two long apicoventral setae, the longer one nearly as long as width of scape, a group of five short setae somewhat irregularly arranged near the dorsal apex medially, and a larger group of short setae in line near the apex laterally, pedicel on medial surface almost entirely setose except for a small, bare, subapical area at the distal margin, setae on lateral surface restricted to dorsal and distal third, with a narrow gap between the dorsal and the apical patches, the dorsopalpal setae short (less than 1.5 times maximum width of arista), the apicoventral setae as long as about half depth of pedicel. Postpedicel (Fig. 18) with a well-demarcated dorsal fossette, with a narrow basal section differentiated (but difficult to see, especially laterally, with two to three sensory pits on median and lateral surface close to distal margin of basal section, one of which is as wide as maximum width of arista). Arista (Fig. 18) inserted dorsally near base, long, with nearly circular cross-section, gradually narrowing from base to apex, dark, reddish brown, microtrichose throughout, aristomere 1 not visible externally (but present), aristomere 2 shorter than wide.

**Thorax.** (Figs 1, 2, 7–10). Black to dark reddish brown, but scutellum and mediotergite light reddish brown, scutellar rim yellow, entirely densely covered by microtrichia, except for the posterior half of the scutellum, which is largely bare of microtrichia. Postpronotum, notopleuron, anepisternum on posterodorsal half of both parts, and anterior part of anepimeron (except for the posteroverentral corner), including the dorsomedian part, with a dense cover of long, thick (but not feathered), glistening golden setae, similar but less bright pale present on postalar callus and scutellum, remaining parts of scutum with rather dense, uneven, almost erect black setae composed of setae of two lengths, the shorter present on entire surface, the longer missing sublaterally, a few black setae present at the posterior end of the notopleuron, the anterior end of the postalar callus, and anteroventral end of the anepisternal setosity, median half of premetacoxal bridge rather densely black setose, sparse and inconspicuous short silvery or black setae are present on the proepimeron, pleurae otherwise bare, notably the katepisternum without any setae, about three to four irregular rows of marginal setae of supraalar area moderately thickened. Postalar wall with a strong, three-sided pyramidal tubercle. Scutoscutular suture deep and narrow, without any trace of a

prescutellum. Anterodorsal margin of anepisternum, at the border with postpronotum, with a dorsally directed peg (Fig. 9). Katatergite smooth, with only a small posterodorsal part bearing slightly elongated microtrichia, which seem to be arranged as oblique dorsoventral stripes (Hippa & Ståhls, 2005: character 50, state 1; requires verification). Katepimeron convex. Metathoracic spiracle directed posteriorly, occupying nearly half the distance from the ventral margin of the katatergite to the base of the hind coxa, membranous cleft of metapleural sclerites posterior to spiracle narrow, the sclerite ventral to the cleft posterodorsally incised. Pometacoxal bridge well-developed, its posteromedian apex simple. Posterior basale short (Fig. 10), pleural wing process with a short membranous cleft, subalar sclerite with a moderately inflated anterodorsal apex, plumule small, dark brown, apparently with simple setae (requires verification with SEM). A small deep pit at the posterodorsal corner of the meron. Posteroventral area of anepimeron flat, not sunk, thus forming a unified, flat surface together with the anterior part of the katatergite. Ventral surface of scutellum concave. Subscutellum small. Anterior part of anepisternum convex, only slightly differentiated from the posterior part, the posterior part with a rudimentary posterodorsal swelling. Katepisternum anteriorly and at the anapleural suture unusually deeply invaginated (possibly an artefact from being dried out from alcohol). Proepimeron strongly protruding ventrally (Fig. 8). Proepimeron laterally slightly bulging (Fig. 8), its ventral part bare, its dorsal part with sparse short more or less spinose setae. The arms of the anteropronotal collar touching dorsally, the cleft between them only slightly extended posteriorly. Haltere dark brown, except for the median part of the stalk, which is yellow.

**Wing.** (Fig. 5). Infuscated brown, entirely covered by unusually dense microtrichia. Costa ending far from apex, with a short stump distal to R<sub>4+5</sub>, M<sub>1</sub> bent towards wing margin, then recurrent and joining R<sub>4+5</sub> at an obtuse angle, the recurrent section without external spur. M<sub>2</sub> present as a very short spur, R<sub>4+5</sub> deeply curved into cell r<sub>4+5</sub>, with an additional cross-vein connecting the curve of R<sub>4+5</sub> with M<sub>1</sub> at the level of maximum proximity, cross-vein Sc-R present, R<sub>1</sub> long and reaching C at about two-thirds the distance from Sc to R<sub>2+3</sub>. Tegula densely setose, costagium densely setose, except for a narrow bare stripe along its posterior margin, setae all black (including ventral ones), C microtrichose, with two rows of setae, except for the basal part from the costagium to about one-third of the second costal section, where there are several irregular rows ventrally, the ventral setae being elongated (up to about three times costal width), and two to three moderately irregular rows dorsally from apical half of first costal section to first third of second section. Humeral cross-vein slightly shifted apically in relation to the 'scar' near the base of the radial stem vein. Cross-vein R-M placed at about two-thirds of cell d<sub>m</sub>, posterior angle of cell d<sub>m</sub> angular, with a short veinlet (CuA<sub>1</sub>). Vena spuria reaching R-M as a pigmented (brown) wing fold, depigmented proximal to the slight kink

in the middle. A<sub>1</sub>+CuA<sub>2</sub> in proximal half straight and strong, then weakened and smoothly curved towards wing margin. Lower calyptery dark brown, without long setae on its disk, marginal fringe of upper part white, composed of rather dense, apparently simple setae (observed with stereomicroscope at 80×), marginal fringe of lower part light brown and composed of multifurcate trichiae.

**Legs.** (Figs 1, 6). All parts black or reddish brown, with ill-defined borders between the colours, setosity dense, black to dark brown and generally rather short, except for moderately long setae posterolaterally on fore and mid femora. Fore femur anterobasally with a patch of very dense, short setae. Apex of fore tibia anterolaterally with dense, irregularly placed setae. Fore coxa short (Fig. 7), about as long as its median margin as it is wide apically, microtrichose (except for basal rim), anteriorly with black setae. Outer part of ventral margin of anterior mesocoxite rather smoothly folded (compared with *Merodon*, where the fold is sharp), posterior mesocoxite flat, microtrichose, bare. Hind femur (Fig. 6) moderately thickened, at its widest point (except for process) 0.28 times as wide as long, lower side slightly concave, with an anterolateral apical triangular flange, which bears slightly thickened setae at its posterior margin, without posterolateral subapical spinose setae, ventrally with a bare stripe from base to apex, with a spot of dense, short, thick setae anterolaterally near base, almost entirely microtrichose, except for small areas at the subapical process, the bare ventral area posterior to the subapical flange slightly depressed. Hind tibia with a blunt, anteroventral, sparsely setose carina for about proximal two-fifths, the scar placed in middle of the tibia, where the tibia is thickened, apex ventrally with a sharp transverse edge across nearly full width, almost entirely setose, except for a spot at the dorsal apex, which is bare and bright orange, the setae entirely black, the posteroapical setae of the dorsal half elongated (about three-quarters apical width of hind tibia), nearly entirely microtrichose, except for the anteroventral carina, the apicoventral fifth, and the bare dorsopalpal spot. Hind trochanter with a moderate hump, covered by dense, short setae, distally bare. All tarsi simple. Claws large (one claw of hind tarsus 0.61 mm long), base dark red.

**Abdomen.** (Figs 1, 12). Antetergite almost fused with tergite 1, bare. Tergite 1 light reddish brown, darkened towards the sides, covered by short, black (semi-) adpressed setae, except for the posterior margin, which is bare, but densely microtrichose, moderately folded, and about as long as the width of the first tarsomere of mid leg. Lateral part of tergite 4 extended medioventrally, meeting sternite 4 (Fig. 12). Tergites 2–4 densely covered by bright golden fur, except for a small patch of black setae at the anteromedian part of tergite 2 and the ventral extension of tergite 4, each with a pair of oblique impressions, those on tergite 3 curved (as in *Eumerus* spp.), the cuticle entirely matt, the setae placed on small black tubercles. First abdominal spiracle placed in the pleural membrane next to the metepimeron (Fig. 11), pleural membrane dark brown.

Sternite 1 entirely sclerotized, its anterior corners broadly fused with the metepimeron (Fig. 11), with a pair of large, setose, lateral margins and median line broadly bare. Sternites 2 and 3 much wider than long, covered with short, black setae, leaving the median line bare, sternite 4 (Fig. 12) nearly twice as wide as long laterally, lateral margins thickened, posterior margin excavated, the posterolateral margins folded towards the body (i.e. upwards), covered by black setae, which are very short (much shorter than on sternite 3), but long at the lateral margins, median line with a fine sharp line. Terminalia enormously enlarged (Fig. 1), black setose.

**Terminalia.** (Figs 13–17). Epandrium dorsally short, the depressed area proximal to the cerci almost reaching the proximal margin of epandrium, laterally with an extensive membranous area, but without a triangular membranous incision (as generally found in the Eumerini). Cerci largely sclerotized, fused laterally with epandrium for about two-thirds their length, the sclerite basally with a narrow, membranous cleft. Subepandrial sclerite about twice as long as wide, its posterior third heavily sclerotized and with a pair of medially directed processes, at the margins, close to the attachment of surstyli, with a pair of smaller processes, the median line membranous as is the most part of the proximal two-thirds, which are strongly sclerotized laterally and at the anterior margins only. Surstyli firmly fused with epandrium, with well-developed suture between them, posterior lobe (i.e. posterior to the attachment to epandrium) small, its apex curved ventrally, and with a sharp transverse edge, except for the posterior end rather densely clothed with short, black, somewhat spine-like setae, the remaining part of surstylus smooth, bare, except for a small area at its posterior end medially, which has an extension of the

setosity of the posterior surstyle. Hypandrium with an extraordinarily voluminous base, the anterodorsal corners, which are attached to the anteroventral corners of the epandrium, long and slender, as is the dorsomedian process, with which the hypandrium is attached to the subepandrial sclerite, its ventral margin forming an even curve of ca. 140°C, except for the slender apical part without any ventral or lateral processes, ridges, invaginations, etc. (as is generally found in the Eumerini), with a pair of medially directed triangular extensions of the dorsal margins, its dorsal surface otherwise membranous from just posterior to the dorsomedian process for entire length, the proximal half dorsolaterally with a fine granulose sculpture, with fine, oblique transverse striae present laterally and ventrally, but most parts of hypandrium more or less smooth, no trace of a suture discernible where the superior lobes are fused with the hypandrium, the narrow apical part [i.e. caudal to the (apparently bare) lingual region] with a subapical, well-developed stenidion, apical half deeply invaginated ventrally, its median wall with a densely pilose, triangular ventral process, otherwise simple. Phallus extremely long, fine and hair-like from just posterior to its base to the apex, the two ducts separated a short distance from the phallic base, all the other sclerotized structures present in Eumerini (as in almost all Syrphidae) reduced to a completely unsclerotized, apparently featureless

membrane, phallic apodeme (Fig. 17) small, pump small, very heavily sclerotized. The hamus is entirely absent.

**Diagnosis.** Distinct from all other known Syrphidae by the presence of at least the following four unique autapomorphies: an additional, complete cross-vein between R<sub>4+5</sub> and M<sub>1</sub> (Fig. 5); trochanteral process of the mesocoxite (coxal prong) shortened; posteroverentral part of epimeron not sunk, together with anterior part of katatergite forming a unified flat surface; anterior corners of sternite 1 broadly fused with the metepimeron (Fig. 11).

Differing from all other Eumerini by the following two character states: katepisternum completely without setae; hamus missing. [The 'hamus' is a hook-like paired structure attached to the base of the phallus and the interior wall of the hypandrium (Doczkal, 1996), and a characteristic feature of the male terminalia of Eumerini.]

natural language

natural language  
analog

natural language  
analog  
1000s of journals

**consequences?**

**impossible** to query



# **impossible** to query



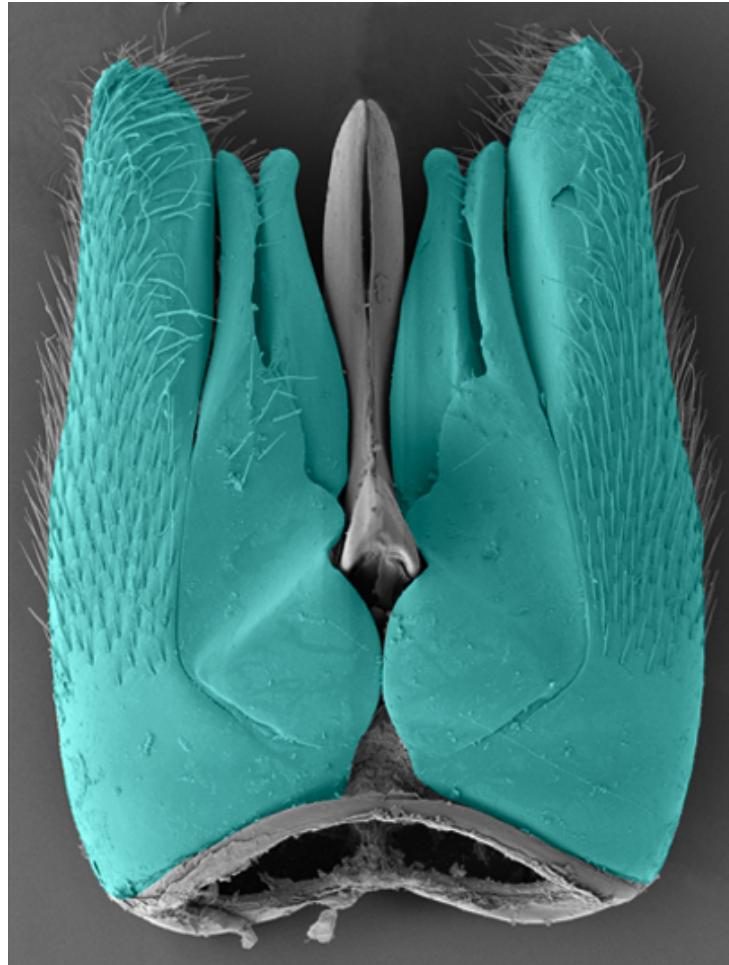
color of abdomen: metallic red

abdomen ruby-colored

reddish are the segments of the metasoma

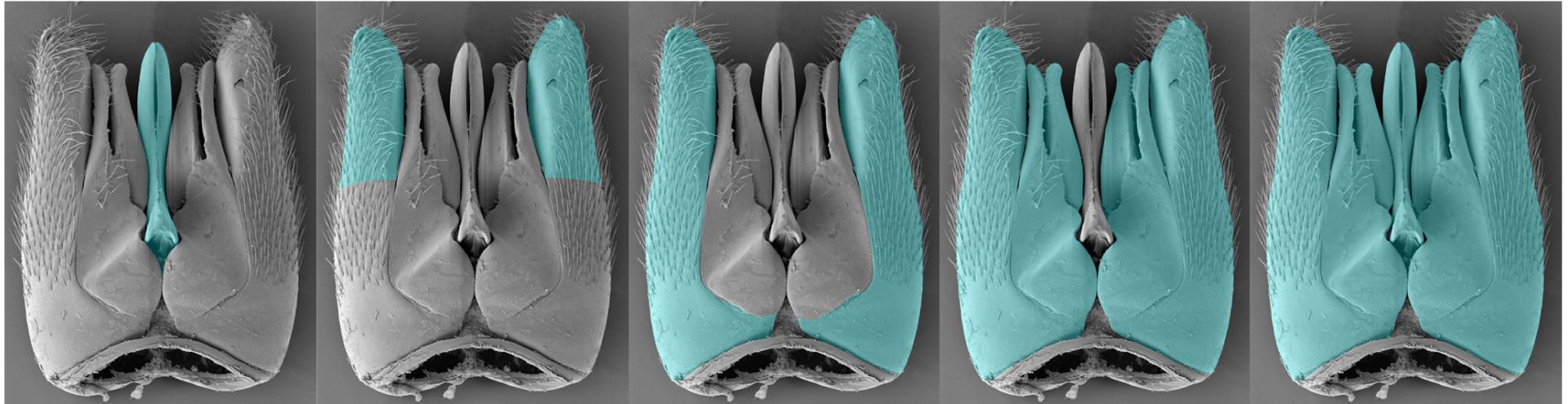
# concept drift is rampant

paramere in the 1800s



# concept drift is rampant

paramere *now*

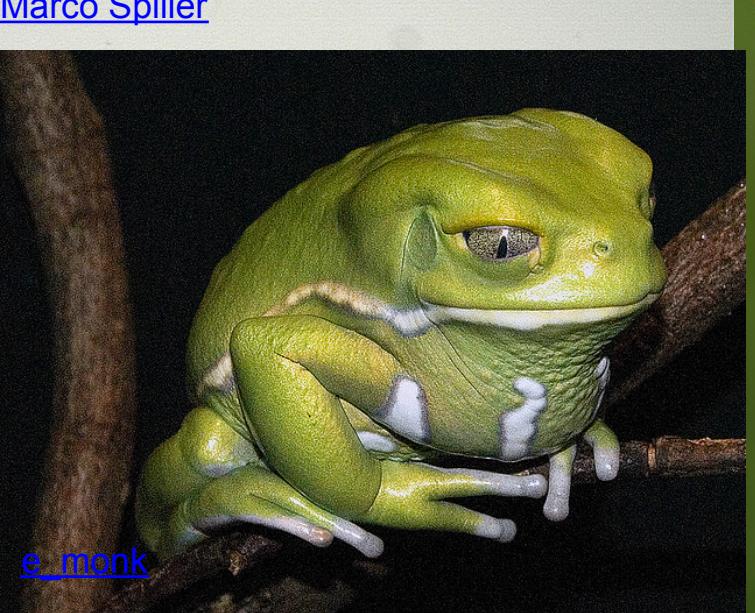
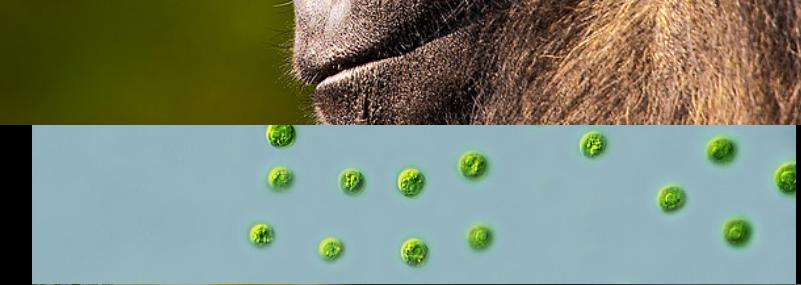
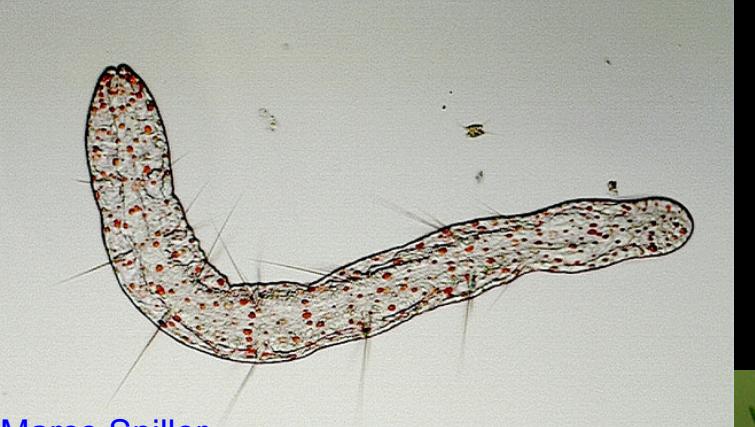


[doi:10.1371/journal.pone.0015991](https://doi.org/10.1371/journal.pone.0015991)



[ilovepics11](#)

[davidandbecky](#)

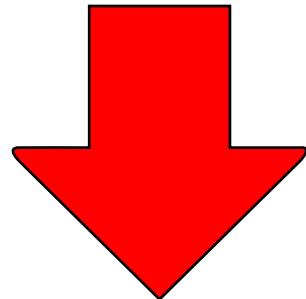


**Millions** of analog (re)descriptions,  
containing **tens of millions** of  
natural language phenotype  
annotations, distributed across  
thousands of journals

**Millions** of analog (re)descriptions,  
containing **tens of millions** of  
natural language phenotype  
annotations, distributed across  
thousands of journals

(largely)  
wasted data

# New paradigm



make these phenotype annotations  
(*i.e.*, our descriptive statements)  
more broadly available

# **Semantic phenotype annotations**

wasp



ant





wasp

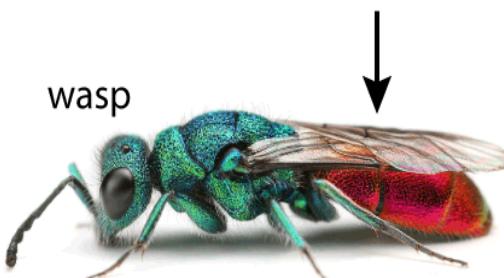
Specimen data  
(GBIF, ZooBank, etc.)



ant

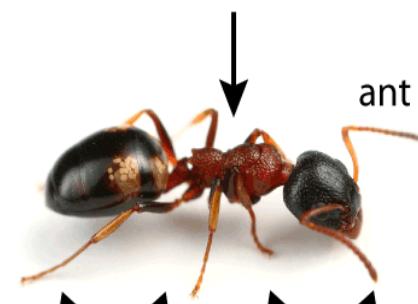
Specimen data  
(GBIF, ZooBank, etc.)

- 1) posterior tagma metallic red
- 2) abdomen, red shiny
- 3) color of metasomal segments: red



Specimen data  
(GBIF, ZooBank, etc.)

- 1) alitrunk dark red
- 2) thorax red with low brightness
- 3) color of mesosomal segments: red

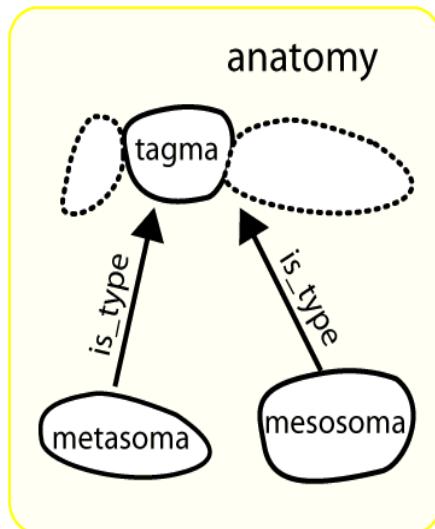


Specimen data  
(GBIF, ZooBank, etc.)

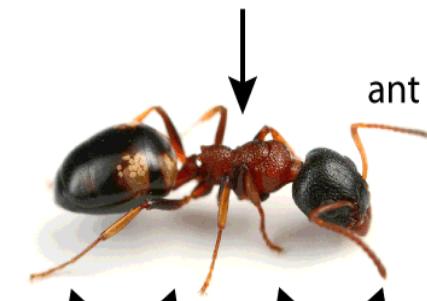
- 1) posterior tagma metallic red
- 2) abdomen, red shiny
- 3) color of metasomal segments: red



Specimen data  
(GBIF, ZooBank, etc.)



- 1) alitrunk dark red
- 2) thorax red with low brightness
- 3) color of mesosomal segments: red

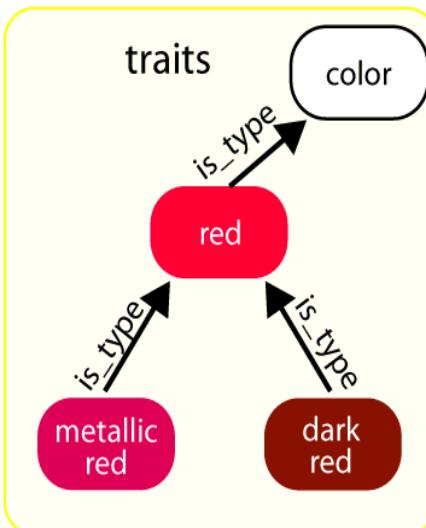
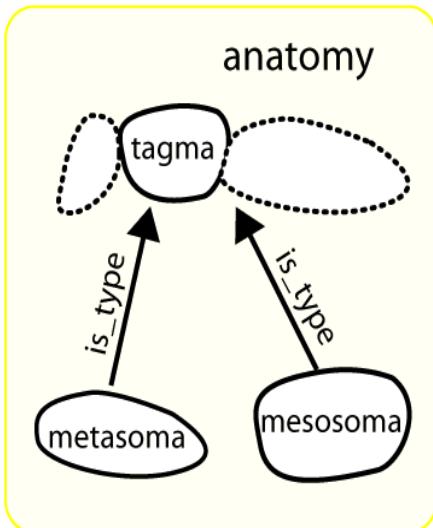


Specimen data  
(GBIF, ZooBank, etc.)

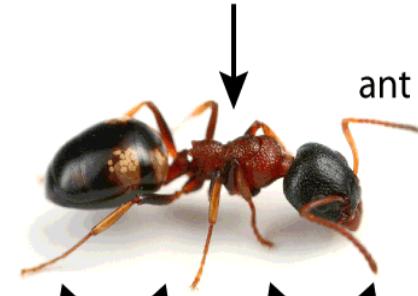
- 1) posterior tagma metallic red
- 2) abdomen, red shiny
- 3) color of metasomal segments: red



Specimen data  
(GBIF, ZooBank, etc.)



- 1) alitrunk dark red
- 2) thorax red with low brightness
- 3) color of mesosomal segments: red

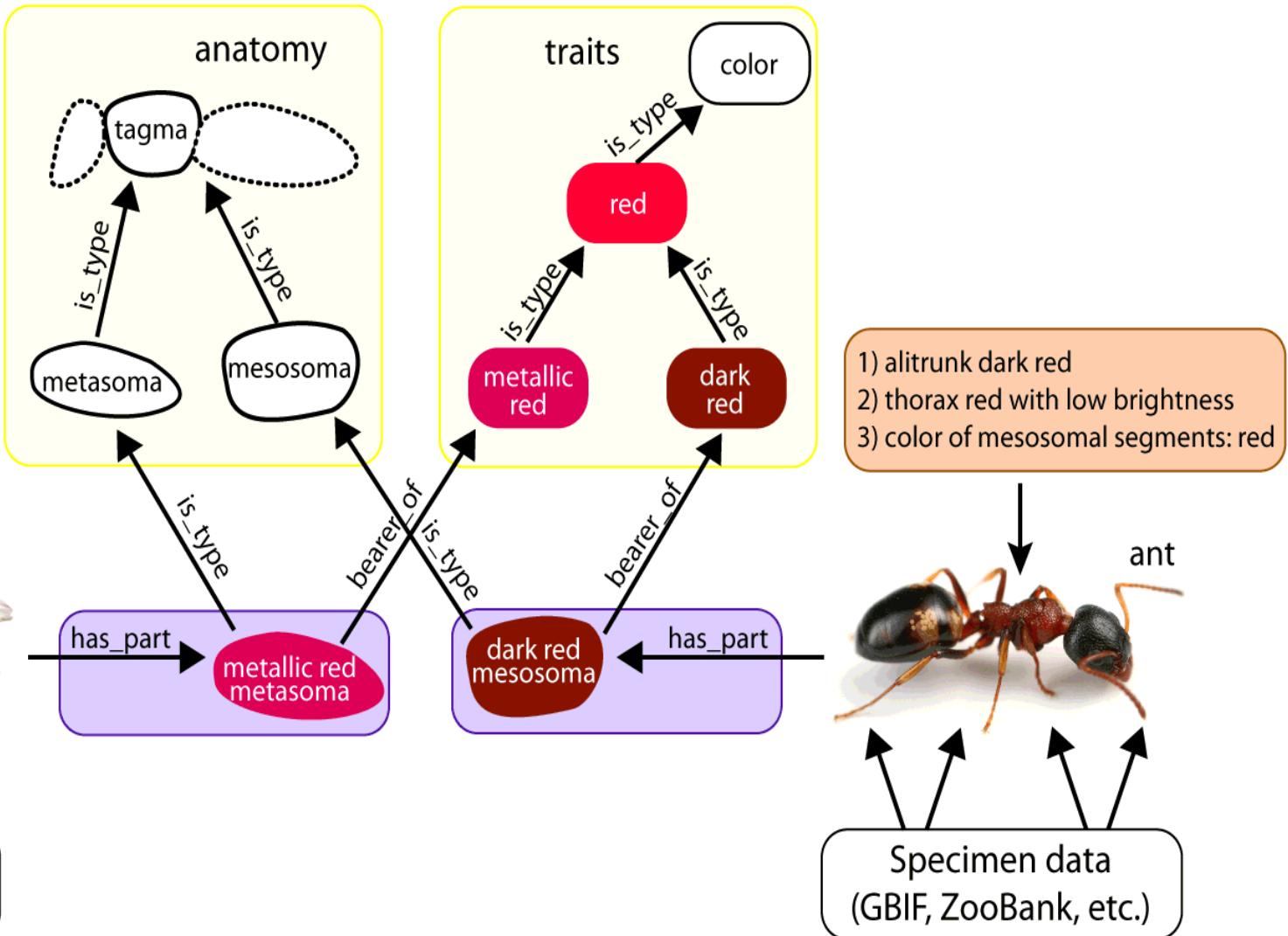


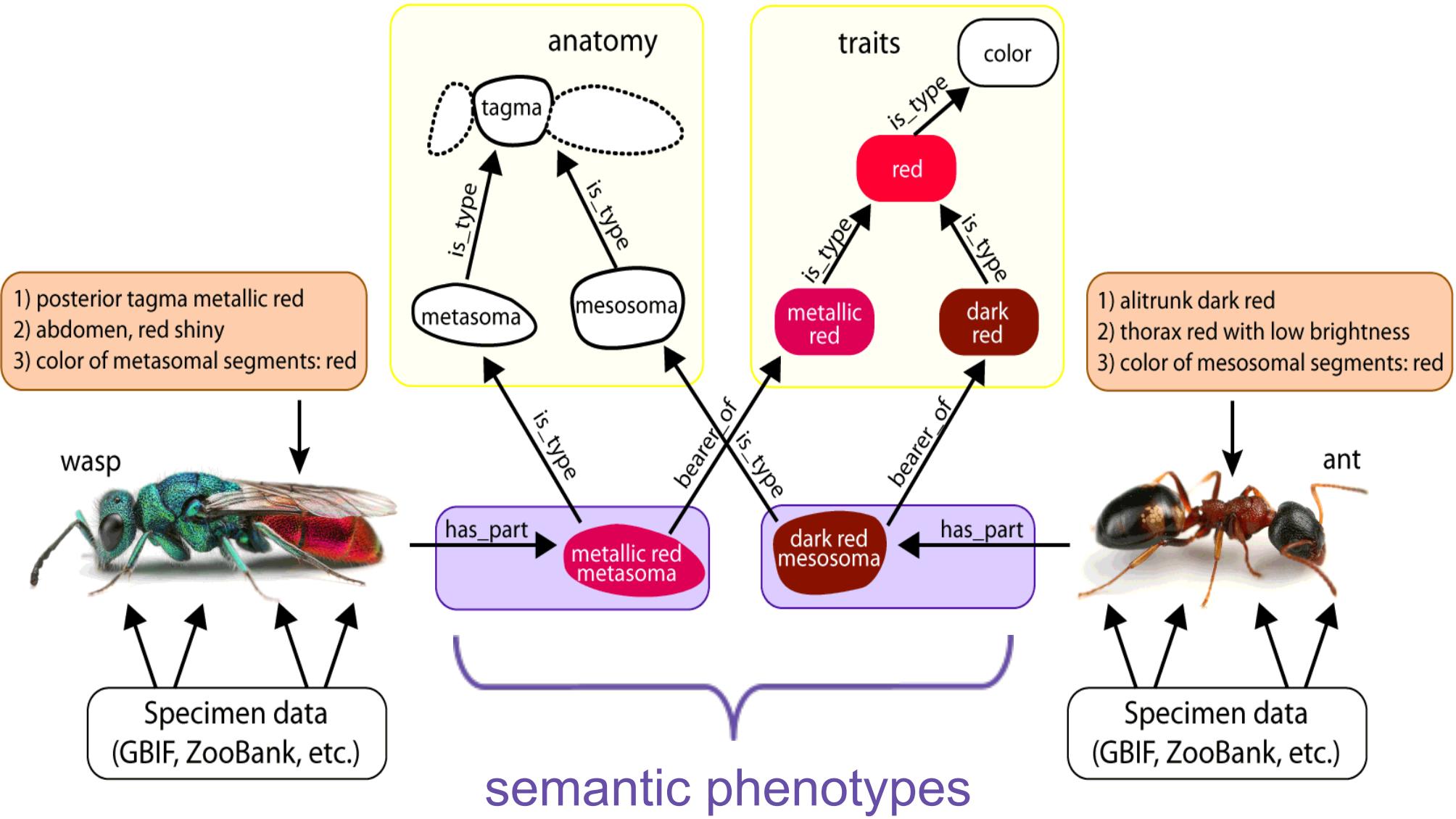
Specimen data  
(GBIF, ZooBank, etc.)

- 1) posterior tagma metallic red  
 2) abdomen, red shiny  
 3) color of metasomal segments: red



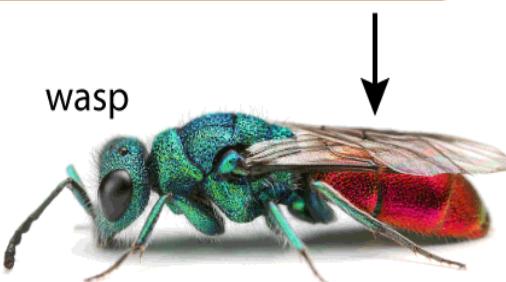
Specimen data  
 (GBIF, ZooBank, etc.)



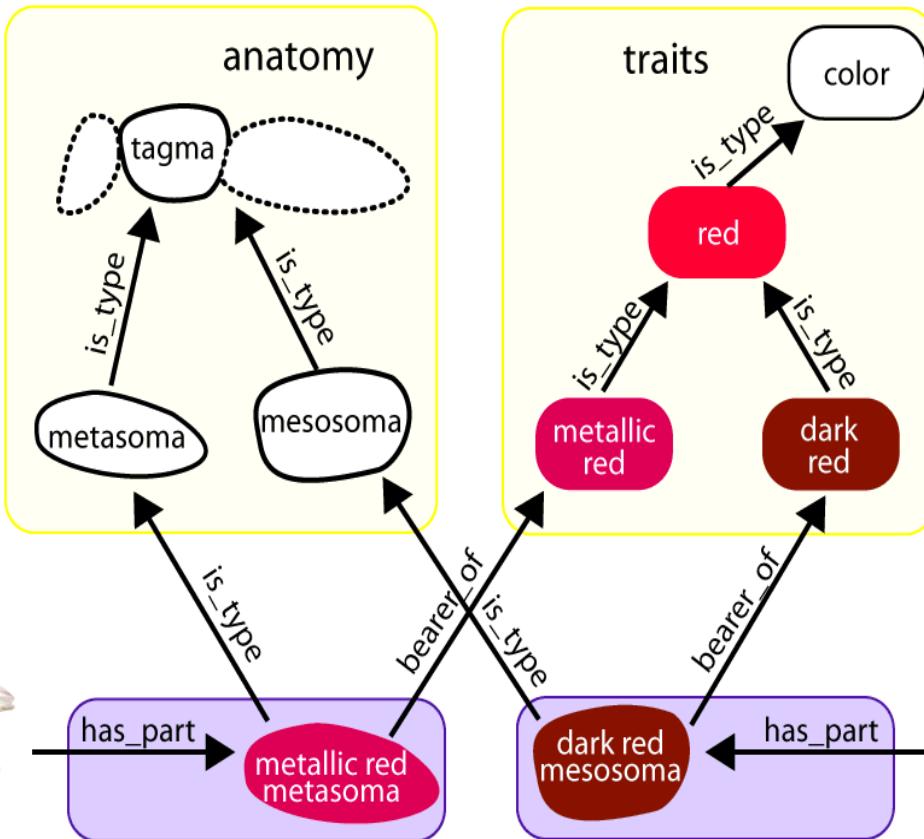


"Show me all the taxa with at least one red tagma."

- 1) posterior tagma metallic red
- 2) abdomen, red shiny
- 3) color of metasomal segments: red



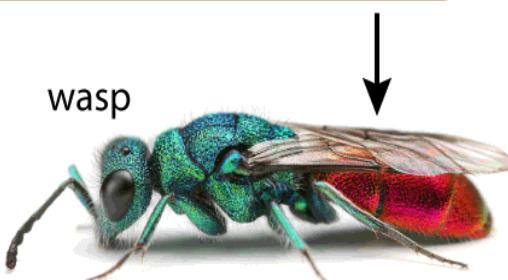
Specimen data  
(GBIF, ZooBank, etc.)



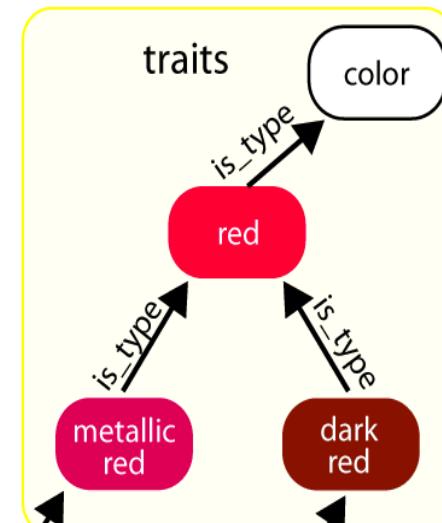
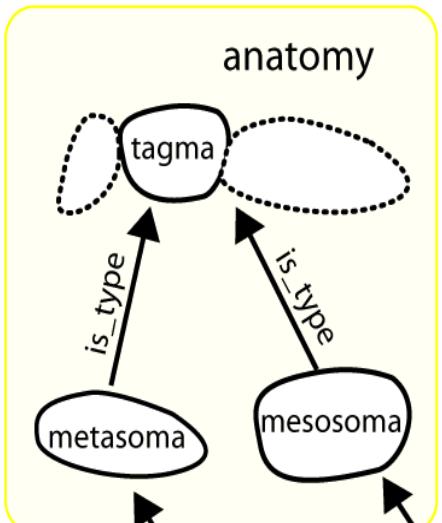
- 1) alitrunk dark red
- 2) thorax red with low brightness
- 3) color of mesosomal segments: red

"Show me all the taxa with at least one red tagma."

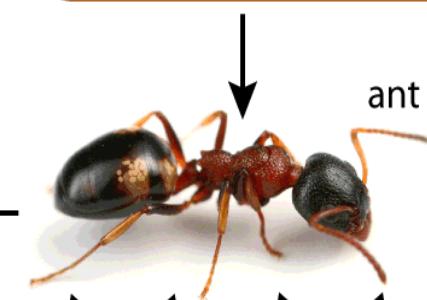
- 1) posterior tagma metallic red
- 2) abdomen, red shiny
- 3) color of metasomal segments: red



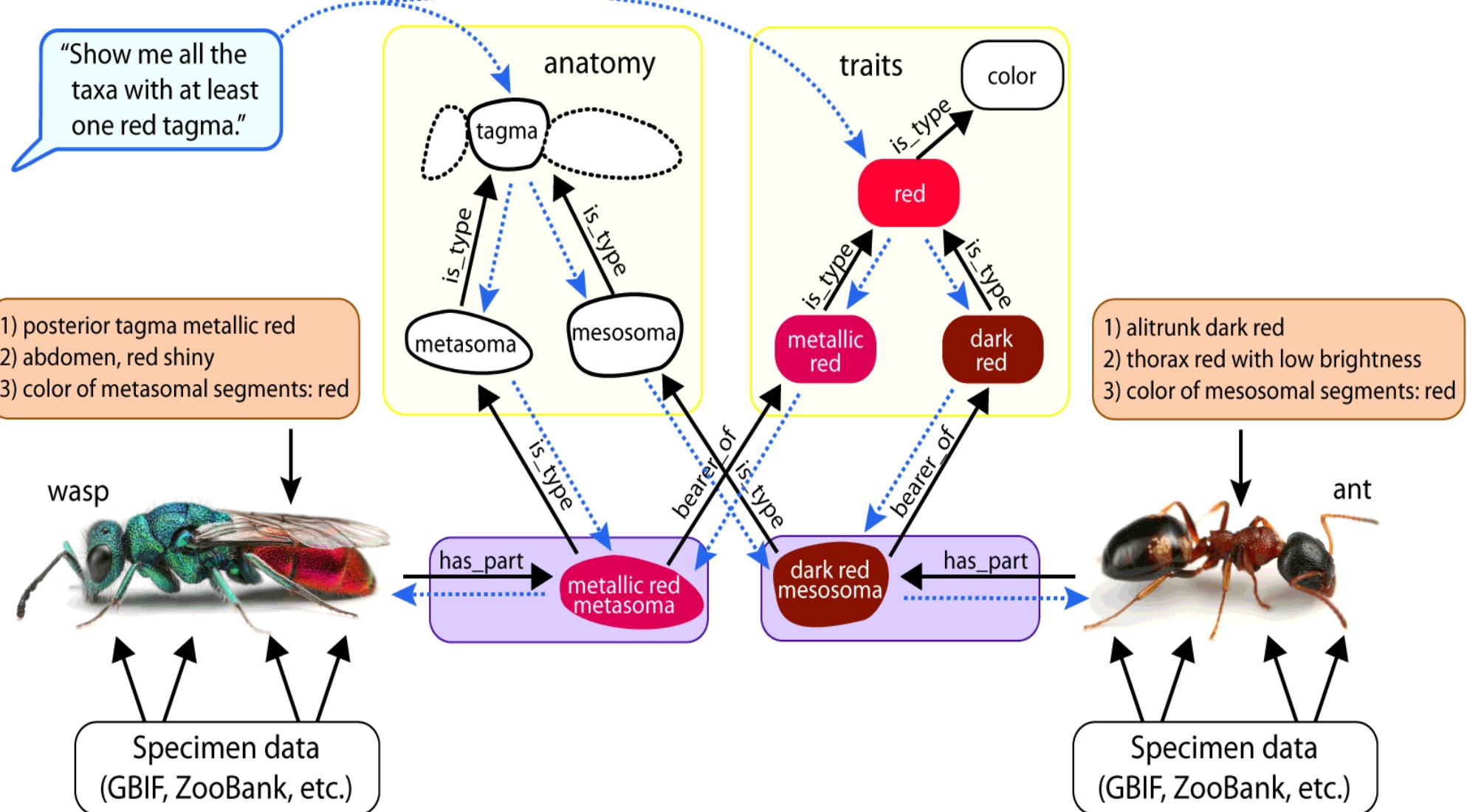
Specimen data  
(GBIF, ZooBank, etc.)



- 1) alitrunk dark red
- 2) thorax red with low brightness
- 3) color of mesosomal segments: red



Specimen data  
(GBIF, ZooBank, etc.)





I have a crazy-looking insect  
with a red butt.

What *is* this thing?!



[Alan English](#)

# red posterior tagma



[Daniel Jimenez](#)



[tehadl](#)



[Finger Food](#)



$H_0$ : insects with at least one red tagma are more common in tropical rainforests





[Alan English](#)



[InSectHunte  
L](#)



[tehACL](#)



[Finger Food](#)

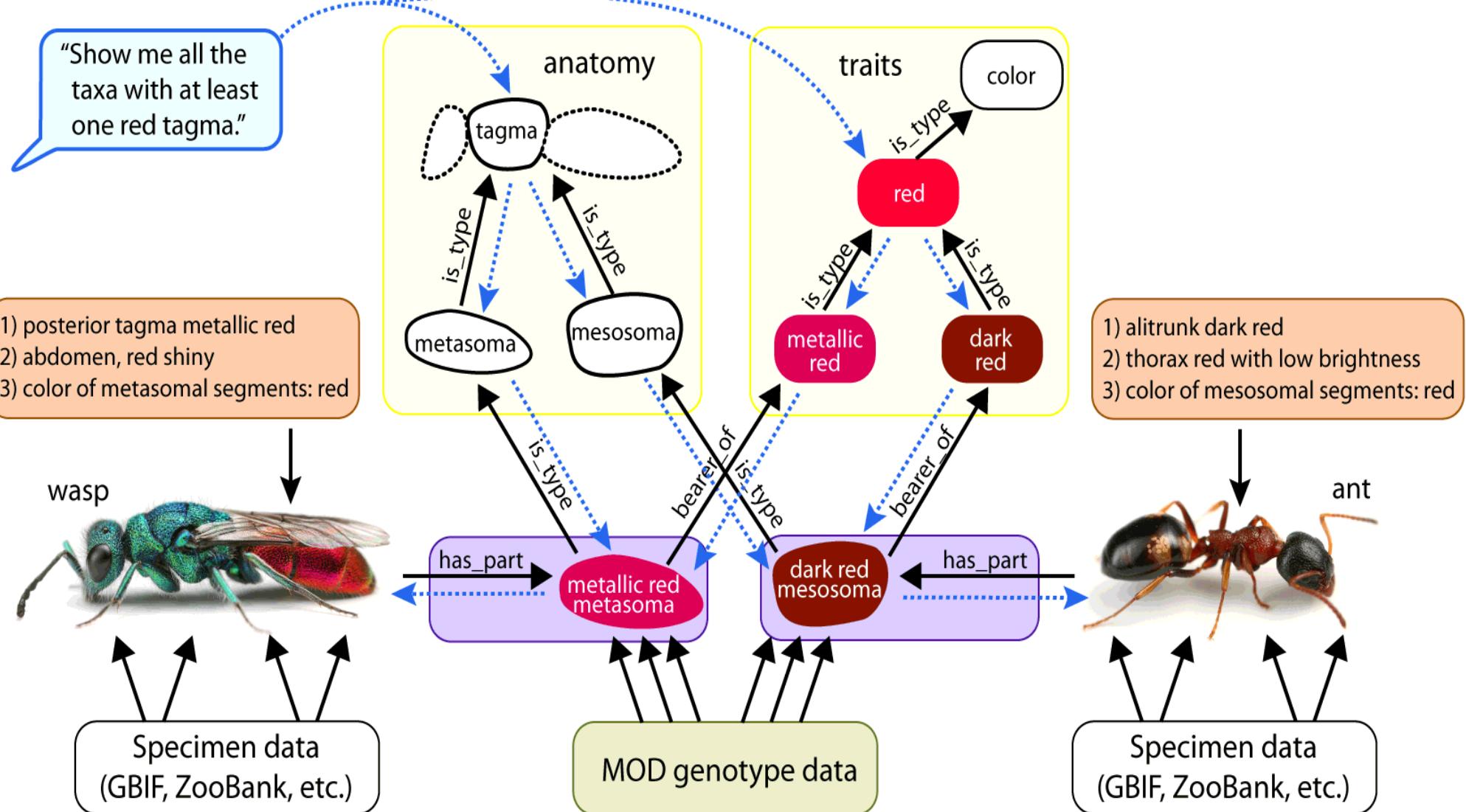


[Daniel Jimenez](#)



[Marco Gaiani](#)

red  
tagmata





$H_0$ : red coloration is governed by the same pigment pathway in all organisms



[bayucca](#)



[Wood Spectre](#)

# red major body region



[Gossamer1013](#)



[mad plumerian](#)



[geoffreyhpierson](#)



Our data become broadly available



**Sounds promising!  
Are we there yet?**



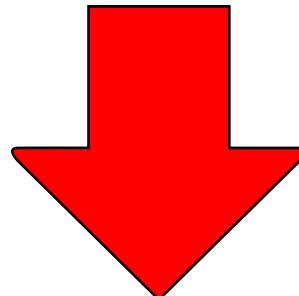
No.

## We have:

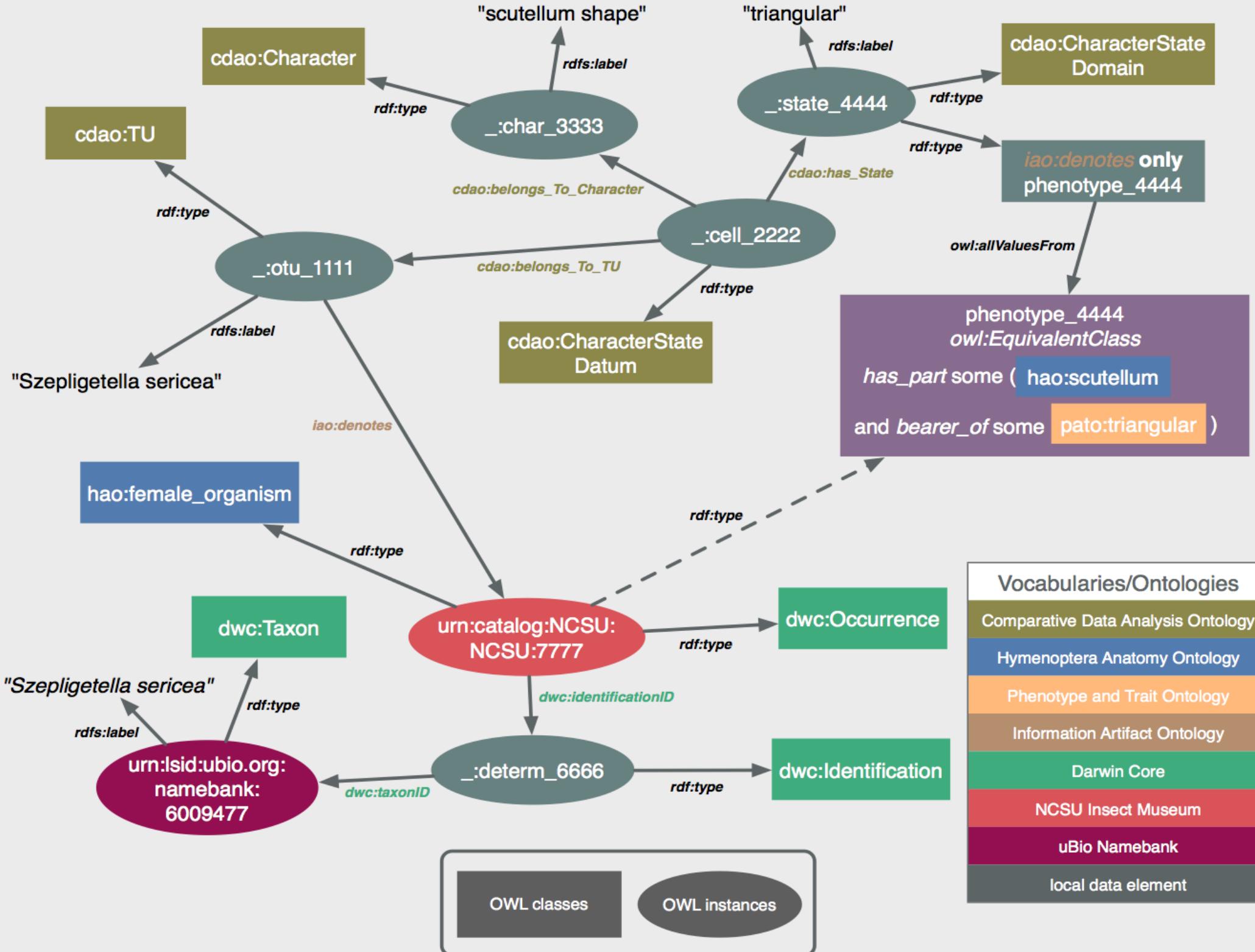
- multi-species anatomy ontologies
- phenotype ontologies
- knowledge representation standards (OWL, RDF, SWRL, XML, etc.)

## We have:

- multi-species anatomy ontologies
- phenotype ontologies
- knowledge representation standards (OWL, RDF, SWRL, XML, etc.)



model for representing knowledge  
about phenotypes at a fine scale  
and in a semantic way



## We need:

- *more* multi-species anatomy ontologies
- *refined* phenotype ontologies
- *better* reasoners
- to address expressivity issues
- applications for taxonomists

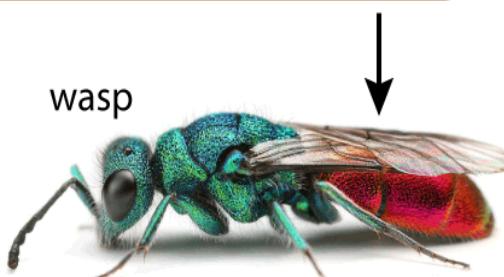
1,800,000 *known* species × 25 annotations = 45,000,000 statements

1,800,000 *known* species × 25 annotations = 45,000,000 statements  
16,000 × 40 annotations = 640,000 statements annually

1,800,000 *known* species × 25 annotations = 45,000,000 statements  
16,000 × 40 annotations = 640,000 statements annually  
\$263 BILLION to finish the process

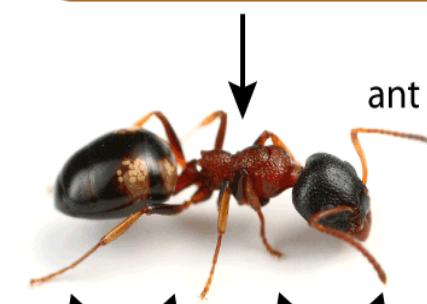
1,800,000 known species  $\times$  25 annotations = 45,000,000 statements  
16,000  $\times$  40 annotations = 640,000 statements annually  
\$263 BILLION to finish the process

- 1) posterior tagma metallic red
- 2) abdomen, red shiny
- 3) color of metasomal segments: red



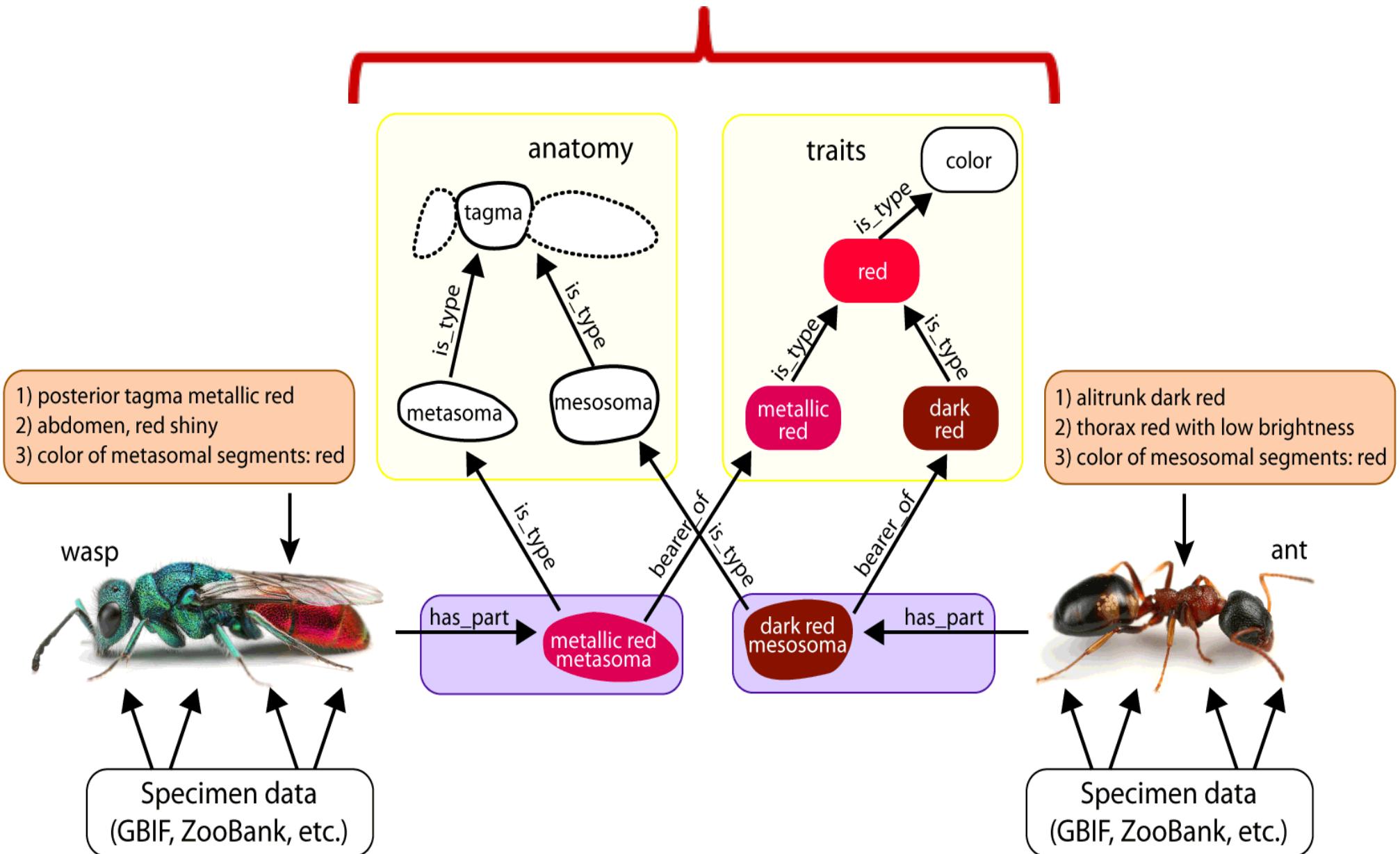
Specimen data  
(GBIF, ZooBank, etc.)

- 1) alitrunk dark red
- 2) thorax red with low brightness
- 3) color of mesosomal segments: red



Specimen data  
(GBIF, ZooBank, etc.)

# invest





## Welcome

The Phenotype Ontology Research Coordination Network (RCN) was funded by NSF to establish a network of scientists who are interested in comparing phenotypes across species and in developing the tools and methods needed in making this possible. [Read more...](#)

## Announcements

[Funding for Meetings](#)

[Collaborative Exchange Opportunities](#)

[Annual Meeting Feb. 23-25, 2012](#)

## EOL Integration

POSTED ON AUGUST 1, 2011 | COMMENTS OFF

I registered as a content partner for EOL. The description for the Channel Catfish has been harvested and now I can preview it on the specie's page. It worked great. Upon the administrator's decision, the newly added information will be publicly available. Until then, I will post here a screen of how it looks. Maybe some html formatting will help for a better presentation, but at this point it's a relieve that everything went smooth with obtaining this preview.

TABLE OF CONTENTS	MORPHOLOGY	CONTRIBUTE
<ul style="list-style-type: none"><li><a href="#">Overview</a></li><li><a href="#">Comprehensive Description</a></li><li><a href="#">Distribution</a></li><li><a href="#">Physical Description</a></li><li><a href="#">Morphology</a></li><li><a href="#">Size</a></li><li><a href="#">Diagnostic Description</a></li><li><a href="#">Ecology</a></li><li><a href="#">Habitat</a></li><li><a href="#">Trophic Strategy</a></li><li><a href="#">Associations</a></li><li><a href="#">Diseases and Parasites</a></li><li><a href="#">Life History and Behavior</a></li><li><a href="#">Behavior</a></li><li><a href="#">Life Cycle</a></li></ul>	<p>The information highlighted in yellow below has not been reviewed.</p> <h3>Morphology</h3> <p>SOURCE AND ADDITIONAL INFORMATION</p> <p>SUPPLIER: Ontophenotype-EOL</p> <p>LICENSE: No rights reserved</p> <p>INDEXED: July 31, 2011</p> <p>Has a coronomeckelian. Presents a bigger maxilla. Presents an urohyal median process. Presents a posterior cranial fontanel. The caudal fin is bifurcated. Has a posterior dentation of pectoral fin spine. Has a mesocoracoid bone. Presents a larger claustrum bone. Presents a hypobranchial 2 bone distal cartilage. Has a gill raker. Has an interhyal bone. Has an ossified basibranchial element. Has an ossified hypobranchial 2 element. Presents a</p>	<ul style="list-style-type: none"><li><a href="#">Submit an image</a></li><li><a href="#">Submit text</a></li><li><a href="#">More information on how to help</a></li><li><a href="#">Latest Changes</a></li></ul> <p>CURATION</p> <p><a href="#">Who can curate this page?</a></p> <p>EXPLORE</p> <p> <a href="#"><i>Lythrypnus albigena</i></a> Bussing, 1990 Goby</p>

# Changing the way we describe biodiversity

<http://bitly.com/DeansESA2011>



[John Hallmén](#)