

A peer-reviewed version of this preprint was published in PeerJ on 29 October 2013.

[View the peer-reviewed version](https://peerj.com/articles/190) (peerj.com/articles/190), which is the preferred citable publication unless you specifically need to cite this preprint.

Page RDM. 2013. BioNames: linking taxonomy, texts, and trees. PeerJ 1:e190 <https://doi.org/10.7717/peerj.190>

BioNames: linking taxonomy, texts, and trees

BioNames is a web database of taxonomic names for animals, linked to the primary literature and, wherever possible, to phylogenetic trees. It aims to provide a taxonomic "dashboard" where at a glance we can see a summary of the taxonomic and phylogenetic information we have for a given taxon and hence provide a quick answer to the basic question "what is this taxon?" BioNames combines classifications from the Global Biodiversity Information Facility (GBIF) and GenBank, imagery from the Encyclopedia of Life (EOL), animal names from the Index of Organism Names (ION), and bibliographic data from multiple sources including the Biodiversity Heritage Library (BHL) and CrossRef. The user interface includes display of full text articles, interactive timelines of taxonomic publications, and zoomable phylogenies. It is available at <http://bionames.org>.

- 1 Roderic D. M. Page
- 2 Institute of Biodiversity, Animal Health and Comparative Medicine
- 3 College of Medical, Veterinary and Life Sciences
- 4 Graham Kerr Building
- 5 University of Glasgow
- 6 Glasgow G12 8QQ, UK

- 7 Telephone: +44 141 330 4778
- 8 Email: Roderic.Page@glasgow.ac.uk

9 Introduction

10 Large-scale digitisation of biodiversity data is underway on at least three broad fronts. The first,
11 and perhaps the only category that is genuinely "born digital" is DNA sequencing (Benson et al.
12 2012). DNA barcoding (Hebert 2003) and, more recently, "metabarcoding" (Taberlet et al. 2012)
13 is generating a flood of sequence data, much of it tied to a specific place and time. The contents
14 of natural history collections are being digitised (Baird 2000), both the specimens themselves
15 (Blagoderov et al. 2012) and metadata about those specimens. The latter is being aggregated by
16 the Global Biodiversity Information Facility (GBIF; <http://data.gbif.org>) to provide an overview
17 of the spatial distribution of life on Earth. Much of the biological literature is similarly being
18 converted from physical to digital form, most notably by the Biodiversity Heritage Library (BHL;
19 <http://www.biodiversitylibrary.org>). Taxonomic publication is becoming increasingly digital
20 through rise of "mega" journals such as *Zootaxa* (<http://www.mapress.com/zootaxa/>), and
21 semantically enriched journals such as *ZooKeys* (<http://www.pensoft.net/journals/zookeys/>).

22 The increasing use of sequence data has made taxonomic relationships readily computable (e.g.,
23 by building phylogenetic trees). Yet many DNA sequences are disconnected from classical
24 taxonomy because they lack formal taxonomic names (Page 2011c; Parr et al. 2012). Barcoding
25 has been responsible for a massive influx of these "dark taxa" into the sequence databases (Page
26 2011c). Many of these unnamed barcode taxa have since been suppressed by GenBank. But even
27 without the barcoding sequences, dark taxa have been steadily increasing in number in recent
28 years. Names may have a special place in the hearts of taxonomists (Patterson et al. 2010), but
29 the pace of biodiversity discovery is outstripping our ability to put names on taxa, as evidenced
30 by the rise of dark taxa in GenBank. There are increasing calls to adopt less formal taxonomic

31 naming schemes (Schindel and Miller 2010), or to focus on describing biodiversity without
32 necessarily naming it (Deans et al. 2012; Maddison et al. 2012). A significant challenge will be
33 determining whether these dark taxa represent newly discovered taxa, or come from known taxa
34 but have not been identified as such (Hibbett and Glotzer 2011; Nagy et al. 2011).

35 The vision of "Biodiversity Information on Every Desktop" (Edwards 2000) (perhaps updated to
36 "biodiversity on every device") rests on our ability to not only digitise life (and the documents we
37 have generated during centuries of cataloguing and studying biodiversity) but also to integrate the
38 wealth of data emerging from sequencing machines and optical scanners. There are numerous
39 points of contact between these different efforts, such as specimen codes, bibliographic
40 identifiers, and GenBank accession numbers (Page 2008a; 2010). Figure 1 shows a simplified
41 model of the core entities that make up taxonomy and related disciplines (e.g., systematics). The
42 diagram is not meant to be exhaustive, nor does it attempt to rigorously define relationships in
43 terms of one or more available ontologies. Instead, it simply serves as a way to visualise the links
44 between taxon names, the publications (and authors and journals) where those names first appear,
45 the application of those names to taxa, and data associated with those taxa (e.g., DNA sequence-
46 based phylogenies).

47 Despite the wealth of possible connections between biodiversity data objects, the most commonly
48 shared identifier that spans sequences, specimens, and publications remains the taxonomic name
49 (Sarkar 2007; Patterson et al. 2010). We rely on names to integrate data, despite the potential
50 ambiguity in what a given taxonomic name "means" (Kennedy et al. 2005; Franz and Cardona-
51 Duque 2013). Unfortunately, it is often difficult to obtain information on a taxonomic name,
52 either to track its origins and subsequent use, or to verify that it has been correctly used. Typically
53 when taxonomic literature is cited in databases, it is typically as a text string with no link to the

54 growing corpus of digitised literature. Hence taxonomic databases are little more than online
55 collections of 5×3 index cards, technology taxonomy's founding father Linnaeus himself
56 pioneered (Müller-Wille & Charmantier 2012). Ideally, for any given taxon name we should be
57 able to see the original description, track the fate of that name through successive revisions, and
58 see other related literature. At present this is almost impossible to do, even in well studied taxa.

59 **EOL Challenge**

60 In response to the Encyclopedia of Life (EOL) Computational Data Challenge
61 (<http://eol.org/info/323>) I constructed BioNames (<http://bionames.org>) (Page 2012). Its goal is to
62 create a database of taxonomic names linked to the primary literature and, wherever possible, to
63 phylogenetic trees. Using existing globally unique identifiers for taxonomic names, concepts,
64 publications, and sequences rather than cryptic text strings (for example, abbreviated
65 bibliographic citations) simplifies the task of linking — we can rely on exact matching of
66 identifiers rather than approximate matching between names for what may or may not be the
67 same entity. This is particularly relevant once we start to aggregate information from different
68 databases, where the same information (e.g., a publication) may be represented by different
69 strings. Furthermore, if we use existing identifiers we increase the potential to connect to other
70 databases (Page 2008a). This paper outlines how BioNames was built, describes the user
71 interface, and discusses future plans.

72 **Materials & Methods**

73 BioNames integrates data on taxonomic names and classifications, literature, and phylogenies
74 from a variety of sources. Given the inevitable differences in how different databases treat the
75 same data (as well as internal inconsistencies within individual databases), considerable effort
76 must be spent cleaning and reconciling data. Much of this process involves mapping "strings" to
77 "things" (Bollacker et al. 2008), or more precisely, mapping strings to identifiers for things.

78 **Taxon names**

79 At present the taxonomic scope of BioNames is restricted to names covered by the International
80 Code of Zoological Nomenclature (animals and those eukaryotes not covered by the International
81 Code of Nomenclature for algae, fungi, and plants). Taxonomic names were obtained from the
82 Index of Organism Names (ION; <http://www.organismnames.com>). Each name in ION has a Life
83 Science Identifier (LSID) (Martin et al. 2005) which uniquely identifies that name. LSIDs can be
84 dereferenced to return metadata in Resource Description Framework format (RDF) (Page 2008b).
85 ION LSIDs provide basic information on a taxonomic name using the TDWG Taxon Name LSID
86 Ontology (<http://rs.tdwg.org/ontology/voc/TaxonName>), in many cases including bibliographic
87 details for the publication where the name first appeared (Fig. 2).

88 The publication in which the name first appeared is listed in the contents of the "PublishedIn"
89 property. In the example in Figure 2 this is the string "Description of a new species of
90 Pinnotheres, and redescription of *P. novaezealandiae* (Brachyura: Pinnotheridae). New Zealand
91 Journal of Zoology, 10(2) 1983: 151-162. 158 (Zoological Record Volume 120)". I used regular
92 expressions to parse citation strings into their component parts (e.g., article title, journal, volume,
93 pagination), and then attempted to locate the corresponding reference in an external database (see
94 below).

95 **Bibliographic identifiers**

96 When populating BioNames every effort has been made to map each bibliographic string to a
97 corresponding identifier, such as a Digital Object identifier (DOI). While DOIs are the best-
98 known bibliographic identifier, there are several others that are relevant to the taxonomic
99 literature (Page 2009). DOIs are themselves based on Handles (<http://hdl.handle.net>), an identifier
100 widely used by digital repositories such as DSpace (Smith et al. 2003). A number of journals,
101 such as the *Bulletins* and *Novitates* of the American Museum of Natural History are available in
102 DSpace repositories and consequently have Handles. Other major archives such as JSTOR
103 (<http://www.jstor.org/>) and the Japanese National Institute of Informatics (CiNii;
104 <http://ci.nii.ac.jp/>) have their own unique identifiers (typically integer numbers that are part of a
105 URL). Having a variety of identifiers can complicate the task of finding existing identifiers for a
106 particular publication. Whereas for some identifiers, such as DOIs and CiNii NAIDs (National
107 Institute of Informatics Article IDs) there are OpenURL resolvers for this task (Van de Sompel &
108 Beit-Arie 2001), for other identifiers there may be no obvious way to find the identifier other than
109 by using a search engine.

110 For the example in Figure 2, the citation string "Description of a new species of Pinnotheres, and
111 redescription of *P. novaezelandiae* (Brachyura: Pinnotheridae). New Zealand Journal of Zoology,
112 10(2) 1983: 151-162. 158 (Zoological Record Volume 120)" corresponds to the article with the
113 DOI 10.1080/03014223.1983.10423904. Once we have a DOI, we can then use services such as
114 those provided by CrossRef (<http://www.crossref.org>) to retrieve author and publisher
115 information for an article (see Fig. 11 below for one use of publisher information).

116 Identifiers also exist for aggregations of publications, such as journals. The historical practice of
117 abbreviating journal titles in citations has led to a plethora of ways to refer to the same journal.
118 For example, the BioStor database (<http://biostor.org>; Page 2011b) has accumulated more than
119 ten variations on the name of the journal *Bulletin of Zoological Nomenclature* (such as "Bull Zool
120 Nomen", "Bull Zool Nom.", "Bull. Zool. Nomencl.", etc.). This practice, presumably motivated
121 by the desire to conserve space on the printed page, complicates efforts to match citations to
122 identifiers. One approach to tackling this problem is to map abbreviations to journal-level
123 globally unique identifiers, such as International Standard Serial Numbers (ISSNs) (for the
124 *Bulletin of Zoological Nomenclature* the ISSN is 0007-5167). In addition to reducing ambiguity,
125 there are web services such as that provided by WorldCat (<http://www.worldcat.org>) that take
126 ISSNs and return the history of name changes for a journal, which in turn can help clarify the
127 (often complicated) history of long-lived journals.

128 Documents

129 Taxonomic publications are available under a variety of licenses, ranging from explicitly open
130 access licenses (MacCallum 2007) to articles that are "free", to articles that are behind a paywall.
131 Archives such as JSTOR and CiNii have a mixture of free and subscription-based content. Many
132 smaller journals, often published by scientific societies, are providing their content online for
133 free, if not explicitly under an open license. The Biodiversity Heritage Library (the single largest
134 source of taxonomic articles in BioNames, Fig. 11) makes its content available under a Creative
135 Commons license. Where PDFs were available online either "for free" or under open access,
136 these were downloaded and locally cached. Pages were extracted and converted into bitmap
137 images for subsequent display in a web browser.

138 Closed-access publications that are available online are linked to by their identifier (e.g., DOI).
139 Access to some of these publications may be available for short-term "rent" by services such as
140 DeepDyve (<http://www.deepdyve.com>): where possible BioNames includes a link those services.

141 **Clustering taxonomic names**

142 Taxonomic names comprise a "canonical" name and the name's authorship, for example *Homo*
143 *sapiens* Linnaeus comprises the canonical name "Homo sapiens" and the authorship string
144 "Linnaeus". Names in taxonomic databases such as ION display numerous variations in spelling
145 of authors, and instances of the same canonical name published by different authors (e.g.,
146 homonyms), so the names were clustered before populating BioNames. For each set of taxon
147 names with the same canonical name the authorship was compared. If one name lacked an author
148 and the other had an author, the names were automatically merged into a cluster. Given more than
149 two names a graph was constructed where the nodes are the authorship strings, and a pair of
150 nodes is connected if their corresponding strings were sufficiently similar. String similarity was
151 computed by converting the strings to a "finger print" comprising lower case letters with all
152 accented characters replaced by non-accented equivalents, and all punctuation removed, then
153 finding the longest common subsequence of the two strings. If the length of the subsequence
154 relative to the input strings was longer than a specified threshold (by default, 0.8, where identical
155 strings have a similarity of 1.0) then the two author strings were connected by an edge in the
156 graph. The components of the graph correspond to clusters of names with similar authorship
157 strings, and were treated as being the same name. Figure 3 shows a graph for the different names
158 that all have "Rhacophorus" as the canonical name.

159 **Mapping names to taxa**

160 BioNames includes two taxonomic classifications, sourced from GBIF
161 (<http://uat.gbif.org/dataset/d7dddbf4-2cf0-4f39-9b2a-bb099caae36c>) and NCBI
162 (<ftp://ftp.ncbi.nih.gov/pub/taxonomy>), respectively. These provide the user with a way to navigate
163 through taxonomic names, as well as view data associated with each classification (e.g.,
164 phylogenies).

165 Ideally there would be a one-to-one mapping between a taxonomic name and a taxon, but
166 complications often arise. In addition to the well-known problems of synonymy (more than one
167 name for the same taxon) and homonymy (the same name used for different taxa), name and
168 taxon databases may store slightly different representations of the same name. For example, ION
169 has four records for the name "Nystactes" (each name is followed by its LSID):

170	<i>Nystactes</i>	urn:lsid:organismnames.com:name:2787598
171	<i>Nystactes</i> Bohlke	urn:lsid:organismnames.com:name:2735131
172	<i>Nystactes</i> Gloger 1827	urn:lsid:organismnames.com:name:4888093
173	<i>Nystactes</i> Kaup 1829	urn:lsid:organismnames.com:name:4888094

174 GBIF has three taxa with this name (the number is the GBIF species id):

175	<i>Nystactes</i> Böhlke, 1957	2403398
176	<i>Nystactes</i> Gloger, 1827	2475109
177	<i>Nystactes</i> Kaup, 1829	3239722

178 Note the differences in the name string ("o" versus "ö" in "Böhlke", presence or absence of years
179 and commas). To automate the mapping of names to concepts in cases like this I constructed a
180 bipartite graph where the nodes are taxon names, divided into two sets based upon which
181 database they came from (e.g., one set of names from ION, the other from GBIF). I then connect
182 the nodes of the graph by edges whose weights are the similarity of the two strings computed
183 using the longest common subsequence that the two strings share. For example, Figure 4 shows
184 the graph for "Nystactes". Computing the maximum weighted bipartite matching of this graph
185 creates a map between the two sets of names. Ideally GBIF should have only one entry for
186 *Nystactes* because each animal name (with a few exceptions) must be unique. If a newer name
187 has already been published before, then it should be replaced by a new name. In this case,
188 *Nystactes* (Böhlke 1957) has since been replaced by *Nystactichthys* (Böhlke 1958), and *Nystactes*
189 (Kaup 1829) by *Paramyotis* (Bianchi 1916). Unfortunately these changes have not yet percolated
190 their way from the primary literature into the GBIF taxonomy.

191 **Images**

192 To help the user recognise the taxa being displayed images for as many taxa as possible were
193 obtained using EOL's API, which provides access to both the images, and a mapping between
194 GBIF and NCBI taxon concept identifiers and the corresponding record in EOL.

195 **Phylogenies**

196 Phylogenies were obtained from the PhyLoTA database (<http://phylota.net>) (Sanderson et al.
197 2008). This database contains eukaryote phylogenies constructed from automatically assembled
198 clusters of nucleotide sequences (loosely corresponding to "genes"). A MySQL data dump was

199 downloaded (version 184, corresponding to the GenBank release of the same version number)
200 and used to populate a local MySQL database. Metadata for the sequences in each phylogeny was
201 obtained from the European Bioinformatics Institute (EBI; <http://www.ebi.ac.uk>), and used to
202 populate the MySQL database with basic information such as taxon and locality information, as
203 well as bibliographic details for the sources of the sequences.

204 Database

205 Once aggregated, cleaned, and reconciled, the data was converted to JSON (JavaScript Object
206 Notation) and stored in a CouchDB database. CouchDB is a "NoSQL" document database that
207 stores objects in JSON format. Unlike typical SQL databases, CouchDB does not have a database
208 schema and does not support ad hoc queries. Instead CouchDB accepts semi-structured
209 documents, and the user defines fixed queries or "views" (Anderson et al. 2010).

210 Results

211 BioNames comprises a CouchDB database and a web interface. Key features of the interface are
212 outlined below.

213 Search

214 BioNames features a simple search interface that takes a scientific name and returns matching
215 taxonomic names and concepts, together with any publications and phylogenies that contain the
216 name. Figure 5 shows an example search result.

217 **Document display**

218 BioNames uses the DocumentCloud (<https://github.com/documentcloud/document-viewer>)
219 viewer to display both PDFs, and page images from digital archives such as BioStor and Gallica
220 (<http://gallica.bnf.fr/>) (Fig. 6).

221 **Journals**

222 Much of the work in populating BioNames comprises mapping citation to string to bibliographic
223 identifiers and, where possible, linking those citations to full text. For each journal that has a
224 ISSN, BioNames has a corresponding web page that lists all the articles from that journal that are
225 in the database, and provides a graphical summary of how many of those articles have been
226 located online (Fig. 7).

227 **Timeline**

228 BioNames can display timelines of the numbers of taxonomic names published in higher
229 taxonomic groups, inspired by Taxatoy (Sarkar et al. 2008) (Fig. 8). For a given node in the
230 taxonomic hierarchy the children of that node are displayed as a treemap where the size of each
231 cell is proportional to the log of the number of taxa in the subtree rooted on that child taxon. The
232 number of names in that taxon published in each year is displayed as an interactive chart.
233 Clicking on an individual year will list the corresponding publications for that year.

234 **Taxa**

235 Each GBIF or NCBI taxon in BioNames has a corresponding web page that lists the associated
236 taxonomic names, publications linked to those names, and other relevant data (e.g., Fig. 9).

237 **Phylogenies**

238 Phylogenies from PhyLOTA are rendered in an interactive viewer using the Scalable Vector
239 Graphics (SVG) format. The user can zoom in and out, and change the drawing style. Terminal
240 taxa with the same label have the same colour (Fig. 10). This makes it easier to recognise clusters
241 of sequences from the same taxon (e.g., conspecific samples), as well as highlight possible errors
242 (e.g., mislabelled or misidentified sequences). At present the colours are arbitrarily chosen, other
243 schemes could be added in future (Lespinats and Fertil 2011).

244 **Dashboard**

245 The BioNames web site features a "dashboard" which displays various summaries of the data it
246 contains. For example, Fig. 11 shows a bubble chart of the number of articles different publishers
247 have made available online. "Publisher" in this context is broadly defined to include digital
248 archives such as BioStor and JSTOR, repositories using DSpace, and commercial publishers such
249 as Elsevier, Informa UK, Magnolia Press, Springer, and Wiley.

250 Discussion

251 The EOL Computational Data Challenge imposed a deadline on the first release of BioNames,
252 however development of both the database and web interface is ongoing. Below I discuss some
253 potential applications and future directions.

254 Links

255 BioNames makes extensive use of identifiers to clean and link data, but the real value of
256 identifiers becomes apparent when they are shared, that is, when different databases use the same
257 identifiers for the same entities, instead of minting their own. Reusing identifiers can enable
258 unexpected connections between databases. For example, the PubMed biomedical literature
259 database has a record (PMID:948206) for the paper "Monograph on '*Lithoglyphopsis' aperta*, the
260 snail host of Mekong River Schistosomiasis" (Davis et al. 1976). The PubMed record contains
261 the abstract for the paper, but not a link to where the user can obtain a digital version of the paper.
262 However, this reference is in a volume that has been scanned by the Biodiversity Heritage
263 Library, and the article has been extracted by BioStor (<http://biostor.org/reference/102054>). If
264 PubMed was linked to BHL, users of PubMed could go straight to the content of the article. But
265 this is just the start. The Davis et al. paper also mentions museum specimens in the collection of
266 the Academy of Natural Sciences of Drexel University, Philadelphia. Metadata for these
267 specimens has been aggregated by GBIF, and the BioStor page for this article displays those links
268 (<http://biostor.org/reference/102054>). In an ideal world we should be able seamlessly to traverse
269 the path PubMed → BioStor → GBIF. Likewise, we should be able to traverse the path in the
270 other direction. At present, a user of GBIF simply sees metadata for these specimens and a

271 locality map. They are unaware that these specimens have been cited in a paper (Davis et al.
272 1976) which demonstrates that the snails host the Mekong River schistosome. This connection
273 would be trivial to make if the reciprocal link was made: GBIF → BioStor. Furthermore, a link
274 BioStor → PubMed would give us access to Medical Subject Headings (MeSH) for the
275 schistosome paper. Hence we could imagine ultimately searching a database of museum
276 specimens (GBIF) using queries from a controlled vocabulary of biomedical terms (MeSH).

277 Making these connections requires not only that we have digital identifiers, but also that where
278 ever possible we reuse existing identifiers. In practice forging these links can be hard work (Page
279 2011a), and many links may be missing from existing databases (Miller et al. 2009). However, if
280 we restrict ourselves to project-specific identifiers then we stymie attempts to create a network of
281 connected biodiversity data.

282 **Text mining**

283 Much of the value of a scientific publication lies dormant unless it is accessible to text mining,
284 which requires access to full text. Where possible BioNames stores information on the publisher
285 of each article (Fig. 11), which could then be used to prioritise discussions with publishers on
286 gaining access to full text (Van Noorden 2012). Fortunately, the single largest "publisher" of
287 content in BioNames is BioStor (Page 2011b), which contains scans and OCR text from the
288 Biodiversity Heritage Library. BHL makes its content available under a Creative Commons
289 license, and so can be readily mined. Indeed, the text has already been indexed by tools that can
290 recognise taxonomic names (Akella et al. 2012).

291 **Impact of taxonomic literature**

292 The taxonomic community has long felt disadvantaged by the role of citation-based "impact
293 factor" in assessing the importance of taxonomic research (Garfield 2001; Krell 2000; Werner
294 2006) especially as much of the taxonomic literature appears in relatively low-impact journals. A
295 common proposal is to include citations to the taxonomic authority for every name mentioned in
296 a scientific paper (Wägele et al. 2011). Regardless of the merits of this idea, in practice these
297 citations are often hard to locate, which is another motivation for BioNames.

298 There is additional value in surfacing identifiers for the taxonomic literature. In addition to
299 helping construct citation networks, global identifiers can facilitate computing other measures of
300 the value of a taxonomic paper. There is a growing interest in additional measures of post-
301 publication impact of a publication in terms of activity such as social bookmarking, and
302 commentary on web sites ("alt-metrics") (Yan and Gerstein 2011). Gathering these metrics is
303 greatly facilitated by using standard bibliographic identifiers (otherwise, how do we know
304 whether two commentators are discussing the same article or not?). If taxonomic literature is be
305 part of this burgeoning conversation then it needs to be able to be identified unambiguously.

306 **Dark taxa**

307 One of the original motivations for constructing BioNames is the rise of "dark taxa" in genomics
308 databases (Page 2011c). It is clear that some dark taxa do, in fact, have names. For example,
309 consider the frog "*Gephyromantis* aff. *blanci* MV-2005" (NCBI taxonomy id 321743), which has
310 a single DNA sequence AY848308 associated with it. This sequence was published as part of a

311 DNA barcoding study (Vences et al. 2005). If we enter the accession number AY848308 into
312 Google we find two documents, one the supplementary table for (Vences et al. 2005), the other a
313 subsequent paper (Vences and Riva 2007) that describes the frog with this sequence as a new
314 species, *Gephyromantis runewsweeki*. This example is relatively straightforward, but it still
315 required significant time to track down the species description. A key question facing attempts to
316 find names for dark taxa is whether the methods available can be scaled to handle the magnitude
317 of the problem.

318 Alternatively, one could argue that newer technologies such as DNA barcoding make classical
319 taxonomy less relevant, and perhaps the effort in digitising older literature and exposing the
320 taxonomic names it contains is misplaced. A counter argument would be that the taxonomic
321 literature potentially contains a wealth of information on ecology, morphology and behaviour,
322 often for taxa in areas that have been subsequently altered by human activity. Given the rarity of
323 many taxa (Lim et al. 2011), and the uneven taxonomic and geographic distribution of taxonomic
324 expertise (May 1998; Gaston and May 1992), for many species the only significant data on their
325 biology may reside in the legacy literature (possibly under a different name (Solow et al. 1995)).
326 As this legacy becomes more accessible through projects such as BHL (and services that build
327 upon that project; Page 2011a) there will be considerable opportunities to mine that literature for
328 basic biological data (Thessen et al. 2012).

329 **Publishing platform**

330 Recently some taxonomic journals have begun to mark up taxonomic names and descriptions
331 (Penev et al. 2010), which is a precursor to linking names and data together. But these
332 developments leave open the problem of what these links will point to. If we have a database of

333 all taxonomic names and the associated literature (such as BioNames aims to be for zoological
334 names), then such a database would provide an obvious destination for those links. Indeed,
335 ultimately, we could envisage publishing new taxonomic publications within such a database, so
336 that each new publication becomes simply another document within the database (Gerstein and
337 Junker 2002). In the same way, we could use automated methods to extend the process of tagging
338 names, specimens and literature cited to the legacy literature (Page 2010), so that the entire body
339 of taxonomic knowledge becomes a single interwoven web of names, citations, publications, and
340 data.

341 **Availability**

342 BioNames is accessible at <http://bionames.org>. The source code used to build the web site is
343 available on GitHub <http://github.com/rmpage/bionames>. Scripts used to fetch, clean, and
344 reconcile the data are archived in <http://github.com/rmpage/bionames-data>

345 **Acknowledgements**

346 I thank Ryan Schenk for his work on the BioNames, and Cyndy Parr (EOL) for managing the
347 EOL Computational Challenge and providing helpful feedback on the development of BioNames.
348 Some of the ideas in this manuscript were first explored in a talk at the "Anchoring Biodiversity
349 Information: From Sherborn to the 21st century and beyond" symposium held at The Natural
350 History Museum, London, October 28th 2011. I thank Ellinor Michel for the invitation to speak
351 at that meeting.

352 **References**

353 Akella, L., Norton, C. N., & Miller, H. (2012). NetiNeti: discovery of scientific names from text
354 using machine learning methods. *BMC Bioinformatics*, 13(1), 211. doi:10.1186/1471-
355 2105-13-211

356 Anderson, J. Chris, Jan Lehnardt and Noah Slater (2010). *CouchDB: The Definitive Guide*.
357 O'Reilly, ISBN: 978-0-596-15589-6

358 Baird, R. (2010). Leveraging the fullest potential of scientific collections through
359 digitisation. *Biodiversity Informatics*, 7(2).
360 <https://journals.ku.edu/index.php/jbi/article/view/3987>

361

362 Benson, Dennis A., Ilene Karsch-Mizrachi, Karen Clark, David J. Lipman, James Ostell, and Eric
363 W. Sayers (2012). GenBank. *Nucl. Acids Res.* (2012) 40 (D1): D48-D53.
364 doi:10.1093/nar/gkr1202

365 Bianchi 1916. *Annuaire du Musee Zoologique de l'Academie d. Sciences de St. Petersburg*
366 21:xxiii-xxxii (not seen)

367 Blagoderov, V., Kitching, I., Livermore, L., Simonsen, T., & Smith, V. (2012). No specimen left
368 behind: industrial scale digitization of natural history collections. *ZooKeys*, 209(0), 133–
369 146. doi:10.3897/zookeys.209.3178

- 370 Bollacker, K., Evans, C., Paritosh, P., Sturge, T., & Taylor, J. (2008). Freebase. Proceedings of the
371 2008 ACM SIGMOD international conference on Management of data - SIGMOD '08 (p.
372 1247). Association for Computing Machinery. doi:10.1145/1376616.1376746
- 373 Böhlke, J. E. (1957). On the Occurrence of Garden Eels in the Western Atlantic, with a Synopsis
374 of the Heterocongrinae. Proceedings of the Academy of Natural Sciences of Philadelphia,
375 109: 59-79. <http://www.jstor.org/stable/4064494>
- 376 Böhlke, J. E. (1958). Substitute Names for *Nystactes* Bohlke and *Lucaya* Bohlke, Preoccupied.
377 Copeia, 1958(1), 59. doi:10.2307/1439557
- 378 Conle, Oskar V, and Frank H Hennemann (2002) Revision of neotropic Phasmatodea: The tribe
379 Anisomorphini sensu Bradley & Galil 1977: (Insecta, Phasmatodea, Pseudophasmatidae).
380 Spixiana Supplement 28: 1–141. <http://biostor.org/reference/118220>
- 381 Davis GM, Kitikoon V, Temcharoen P (1976) Monograph on "*Lithoglyphopsis*" *aperta*, the snail
382 host of Mekong River schistosomiasis. Malacologia 15(2): 241-87.
383 <http://biostor.org/reference/102054>
- 384 Deans, A. R., Yoder, M. J., & Balhoff, J. P. (2012). Time to change how we describe biodiversity.
385 Trends in Ecology & Evolution, 27(2), 78–84. doi:10.1016/j.tree.2011.11.007
- 386 Edwards, J. L. (2000). Interoperability of Biodiversity Databases: Biodiversity Information on
387 Every Desktop. Science, 289(5488), 2312–2314. doi:10.1126/science.289.5488.2312)

- 388 Faulkes, C. G., Bennett, N. C., Cotterill, F. P. D., Stanley, W., Mgone, G. F., & Verheyen, E.
389 (2011). Phylogeography and cryptic diversity of the solitary-dwelling silvery mole-rat,
390 genus *Heliophobius* (family: Bathyergidae). (A. Kitchener, Ed.) *Journal of Zoology*,
391 285(4), 324–338. doi:10.1111/j.1469-7998.2011.00863.x
392
- 393 Franz, N. M., & Cardona-Duque, J. (2013). Description of two new species and phylogenetic
394 reassessment of *Perellesschus* O'Brien & Wibmer, 1986 (Coleoptera: Curculionidae), with
395 a complete taxonomic concept history of *Perellesschus* sec. Franz & Cardona-Duque, 2013
396 . *Systematics and Biodiversity*, 11(2), 209–236. doi:10.1080/14772000.2013.806371
- 397 Garfield, E. (2001). *Nature*, 413(6852), 107–107. doi:10.1038/35093267
- 398 Gaston, K. J., & May, R. M. (1992). Taxonomy of taxonomists. *Nature*, 356(6367), 281–282.
399 doi:10.1038/356281a0
- 400 Gerstein, M. and Jochen Junker (2002). Blurring the boundaries between scientific 'papers' and
401 biological databases. *Nature* [http://www.nature.com/nature/debates/e-](http://www.nature.com/nature/debates/e-access/Articles/gerstein.html)
402 [access/Articles/gerstein.html](http://www.nature.com/nature/debates/e-access/Articles/gerstein.html)
- 403 Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications
404 through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*,
405 270(1512), 313–321. doi:10.1098/rspb.2002.2218
- 406 Hibbett, D., & Glotzer, D. (2011). Where are all the undocumented fungal species? A study of
407 *Mortierella* demonstrates the need for sequence-based classification. *New Phytologist*,
408 191(3), 592–596. doi:10.1111/j.1469-8137.2011.03819.x

- 409 Kaup, J. J., & Stejneger, L. (1829). Skizzirte Entwicklungs-Geschichte und natürliches System
410 der europäischen Thierwelt : Erster Theil welcher die Vogelsäuethiere und Vögel nebst
411 Andeutung der Entstehung der letzteren aus Amphibien enthält /. Smithsonian Institution
412 Biodiversity Heritage Library. doi:10.5962/bhl.title.63915
- 413 Kennedy, J. B., Kukla, R., & Paterson, T. (2005). Scientific Names Are Ambiguous as Identifiers
414 for Biological Taxa: Their Context and Definition Are Required for Accurate Data
415 Integration (pp. 80–95). Springer-Verlag. doi:10.1007/11530084_8
- 416 Krell, F.-T. (2000). *Nature*, 405(6786), 507–508. doi:10.1038/35014664
- 417 Lespinats, Sylvain, & Bernard Fertil. (2011). ColorPhylo: A Color Code to Accurately Display
418 Taxonomic Classifications. *Evolutionary Bioinformatics*, 257. doi:10.4137/EBO.S7565
- 419 Lim, G. S., Balke, M., & Meier, R. (2011). Determining Species Boundaries in a World Full of
420 Rarity: Singletons, Species Delimitation Methods. *Systematic Biology*, 61(1), 165–169.
421 doi:10.1093/sysbio/syr030
- 422 MacCallum, C. J. (2007). When Is Open Access Not Open Access? *PLoS Biology*, 5(10), e285.
423 doi:10.1371/journal.pbio.0050285
- 424 Maddison, D. R., Guralnick, R., Hill, A., Reysenbach, A.-L., & McDade, L. A. (2012). Ramping
425 up biodiversity discovery via online quantum contributions. *Trends in Ecology &
426 Evolution*, 27(2), 72–77. doi:10.1016/j.tree.2011.10.010

- 427 Martin, S., Hohman, M. M., & Liefeld, T. (2005). The impact of Life Science Identifier on
428 informatics data. *Drug Discovery Today*, 10(22), 1566–1572. doi:10.1016/S1359-
429 6446(05)03651-2
- 430 MAY, R. M. (1988). How Many Species Are There on Earth? *Science*, 241(4872), 1441–1449.
431 doi:10.1126/science.241.4872.1441
- 432 Miller, H., Norton, C. N., & Sarkar, I. N. (2009). GenBank and PubMed: How connected are
433 they? *BMC Research Notes*, 2(1), 101. doi:10.1186/1756-0500-2-101
- 434 Müller-Wille, S., & Charmantier, I. (2012). Natural history and information overload: The case of
435 Linnaeus. *Studies in History and Philosophy of Science Part C: Studies in History and
436 Philosophy of Biological and Biomedical Sciences*, 43(1), 4–15.
437 doi:10.1016/j.shpsc.2011.10.021
- 438 Nagy, L. G., Petkovits, T., Kovács, G. M., Voigt, K., Vágvölgyi, C., & Papp, T. (2011). Where is
439 the unseen fungal diversity hidden? A study of *Mortierella* reveals a large contribution of
440 reference collections to the identification of fungal environmental sequences. *New
441 Phytologist*, 191(3), 789–794. doi:10.1111/j.1469-8137.2011.03707.x
- 442 Page, R. D. M. (1983). Description of a new species of *Pinnotheres* , and redescription of *P.*
443 *novaezelandiae* (Brachyura: Pinnotheridae) . *New Zealand Journal of Zoology*, 10(2),
444 151–162. doi:10.1080/03014223.1983.10423904

- 445 Page, R. D. M. (2008a). Biodiversity informatics: the challenge of linking data and the role of
446 shared identifiers. *Briefings in Bioinformatics*, 9(5), 345–354. doi:10.1093/bib/bbn022
- 447 Page, R. D. (2008b). LSID Tester, a tool for testing Life Science Identifier resolution services.
448 *Source Code for Biology and Medicine*, 3(1), 2. doi:10.1186/1751-0473-3-2
- 449 Page, R. D. (2009). bioGUID: resolving, discovering, and minting identifiers for biodiversity
450 informatics. *BMC Bioinformatics*, 10(Suppl 14), S5. doi:10.1186/1471-2105-10-S14-S5
- 451 Page, R. D. M. (2010). Enhanced display of scientific articles using extended metadata. *Web*
452 *Semantics: Science, Services and Agents on the World Wide Web*, 8(2-3), 190–195.
453 doi:10.1016/j.websem.2010.03.004
- 454 Page, R. D. M. (2011a). Linking NCBI to Wikipedia: a wiki-based approach. *PLoS Currents*, 3,
455 RRN1228. doi:10.1371/currents.RRN1228
- 456 Page, R. D. (2011b). Extracting scientific articles from a large digital archive: BioStor and the
457 Biodiversity Heritage Library. *BMC Bioinformatics*, 12(1), 187. doi:10.1186/1471-2105-
458 12-187
- 459 Page, R. D. M. (2011c). Dark taxa: GenBank in a post-taxonomic world.
460 <http://iphylo.blogspot.co.uk/2011/04/dark-taxa-genbank-in-post-taxonomic.html>
- 461 Page, R. D. M. 2012. EOL Computable Data Challenge. doi 10.6084/m9.figshare.92091

- 462 Parr, C. S., Guralnick, R., Cellinese, N., & Page, R. D. M. (2012). Evolutionary informatics:
463 unifying knowledge about the diversity of life. *Trends in Ecology & Evolution*, 27(2), 94–
464 103. doi:10.1016/j.tree.2011.11.001
- 465 Patterson, D. J., Cooper, J., Kirk, P. M., Pyle, R. L., & Remsen, D. P. (2010). Names are key to
466 the big new biology. *Trends in Ecology & Evolution*, 25(12), 686–691.
467 doi:10.1016/j.tree.2010.09.004
- 468 Penev, L., Agosti, D., Georgiev, T., Catapano, T., Miller, J., Blagoderov, V., Roberts, D., et al.
469 (2010). Semantic tagging of and semantic enhancements to systematics papers: ZooKeys
470 working examples. *ZooKeys*, 50(0). doi:10.3897/zookeys.50.538
- 471 Sanderson, M., Boss, D., Chen, D., Cranston, K., & Wehe, A. (2008). The PhyLoTA Browser:
472 Processing GenBank for Molecular Phylogenetics Research. *Systematic Biology*, 57(3),
473 335–346. doi:10.1080/10635150802158688
- 474 Sarkar, I. N. (2007). Biodiversity informatics: organizing and linking information across the
475 spectrum of life. *Briefings in Bioinformatics*, 8(5), 347–357. doi:10.1093/bib/bbm037
- 476 Sarkar, I., Schenk, R., & Norton, C. N. (2008). Exploring historical trends using taxonomic name
477 metadata. *BMC Evolutionary Biology*, 8(1), 144. doi:10.1186/1471-2148-8-144
- 478 Schindel DE, Miller SE (2010) Provisional nomenclature: the on-ramp to taxonomic names. In:
479 Polaszek A (Ed) *Systema Naturae 250 - The Linnaean Ark*. CRC Press, 109-115 pp.

- 480 Smith, M., Barton, M., Branschofsky, M., McClellan, G., Walker, J. H., Bass, M., Stuve, D., et al.
481 (2003). DSpace. D-Lib Magazine, 9(1). doi:10.1045/january2003-smith
- 482 Solow, A. R., Mound, L. A., & Gaston, K. J. (1995). Estimating the Rate of Synonymy.
483 Systematic Biology, 44(1), 93–96. doi:10.1093/sysbio/44.1.93
- 484 TABERLET, P., COISSAC, E., POMPANON, F., BROCHMANN, C., & WILLERSLEV, E.
485 (2012). Towards next-generation biodiversity assessment using DNA metabarcoding.
486 Molecular Ecology, 21(8), 2045–2050. doi:10.1111/j.1365-294X.2012.05470.x
- 487 Thessen, A. E., Cui, H., & Mozzherin, D. (2012). Applications of Natural Language Processing in
488 Biodiversity Science. Advances in Bioinformatics, 2012, 1–17. doi:10.1155/2012/391574
- 489 Van de Sompel, H., & Beit-Arie, O. (2001). Open Linking in the Scholarly Information
490 Environment Using the OpenURL Framework. D-Lib Magazine, 7(3).
491 doi:10.1045/march2001-vandesompel
- 492 Van Noorden, R. (2012). Trouble at the text mine. Nature, 483(7388), 134–135.
493 doi:10.1038/483134a
- 494 Vences M, Riva IDL (2007) A new species of *Gephyromantis* from Ranomafana National Park,
495 south-eastern Madagascar (Amphibia, Anura, Mantellidae). Spixiana 30(1): 135-143.
- 496 Vences, M., Thomas, M., van der Meijden, A., Chiari, Y., & Vieites, D. R. (2005).Frontiers in
497 Zoology, 2(1), 5. doi:10.1186/1742-9994-2-5

498 Wägele, H., Klusmann-Kolb, A., Kuhlmann, M., Haszprunar, G., Lindberg, D., Koch, A., &
499 Wägele, J. W. (2011). The taxonomist - an endangered race. A practical proposal for its
500 survival. *Frontiers in Zoology*, 8(1), 25. doi:10.1186/1742-9994-8-25

501 Werner, Y. L. (2006). The case of impact factor versus taxonomy: a proposal. *Journal of Natural*
502 *History*, 40(21-22), 1285–1286. doi:10.1080/00222930600903660

503 Yan, K.-K., & Gerstein, M. (2011). The Spread of Scientific Information: Insights from the Web
504 Usage Statistics in PLoS Article-Level Metrics. (A. Vespignani, Ed.) *PLoS ONE*, 6(5),
505 e19917. doi:10.1371/journal.pone.0019917

Figure 1

Taxonomy data model

Simplified diagram of the relationships between the core entities that make up taxonomy, such as authors, publications, taxon names, and taxa. Relationships between entities are represented by lines, those in black are the focus of BioNames.

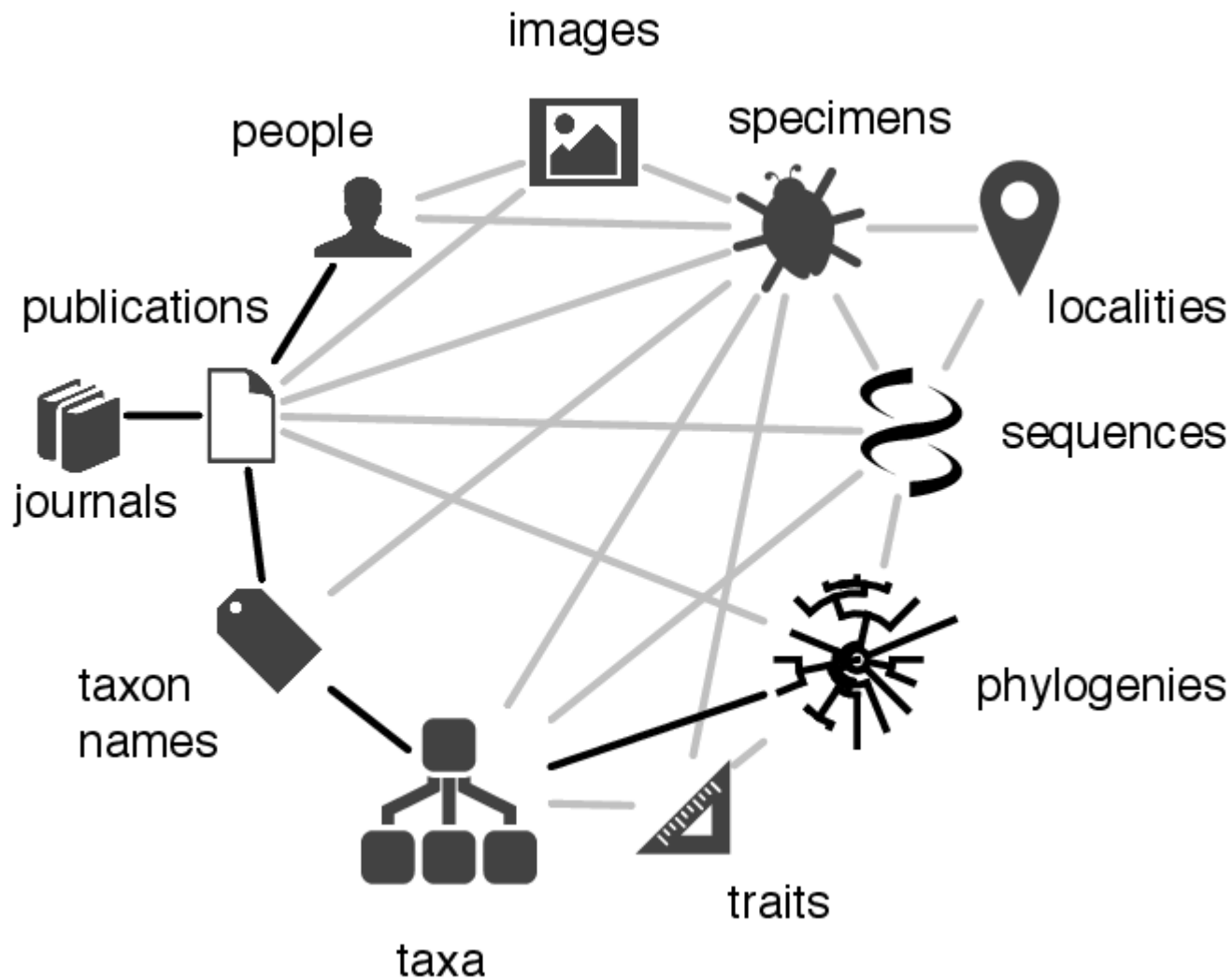


Figure 2

RDF for taxon name

The RDF retrieved by dereferencing the LSID urn:lsid:organismnames.com:name:371873, which identifies the taxonomic name *Pinnotheres atrinicola*.

```
<?xml version="1.0"?>
<rdf:RDF xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:dcterms="http://purl.org/dc/terms/" xmlns:rdf="http://
www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:tdwg_co="http://rs.tdwg.org/ontology/voc/Common#" xmlns:tdwg_pc="http://rs.tdwg.org/ontology/voc/
PublicationCitation#" xmlns:tdwg_tn="http://rs.tdwg.org/ontology/voc/TaxonName#">
  <tdwg_tn:TaxonName rdf:about="371873">
    <dc:identifier>371873</dc:identifier>
    <dc:creator rdf:resource="http://www.organismnames.com"/>
    <dc:Title>Pinnotheres atrinicola</dc:Title>
    <tdwg_tn:nameComplete>Pinnotheres atrinicola</tdwg_tn:nameComplete>
    <tdwg_tn:nomenclaturalCode rdf:resource="http://rs.tdwg.org/ontology/voc/TaxonName#ICZN"/>
    <tdwg_co:PublishedIn>Description of a new species of Pinnotheres, and redescription of P.
novaezealandiae (Brachyura: Pinnotheridae). New Zealand Journal of Zoology, 10(2) 1983: 151-162. 158
[Zoological Record Volume 120]</tdwg_co:PublishedIn>
    <tdwg_co:microreference>158</tdwg_co:microreference>
    <rdfs:seeAlso rdf:resource="http://www.organismnames.com/namedetails.htm?lsid=371873"/>
  </tdwg_tn:TaxonName>
</rdf:RDF>
```

Figure 3

Clustering taxonomic names

Graph depicting similarity between different authorship strings associated with the name "Rhacophorus". The components of this graph correspond to the name clusters recognised by BioNames.

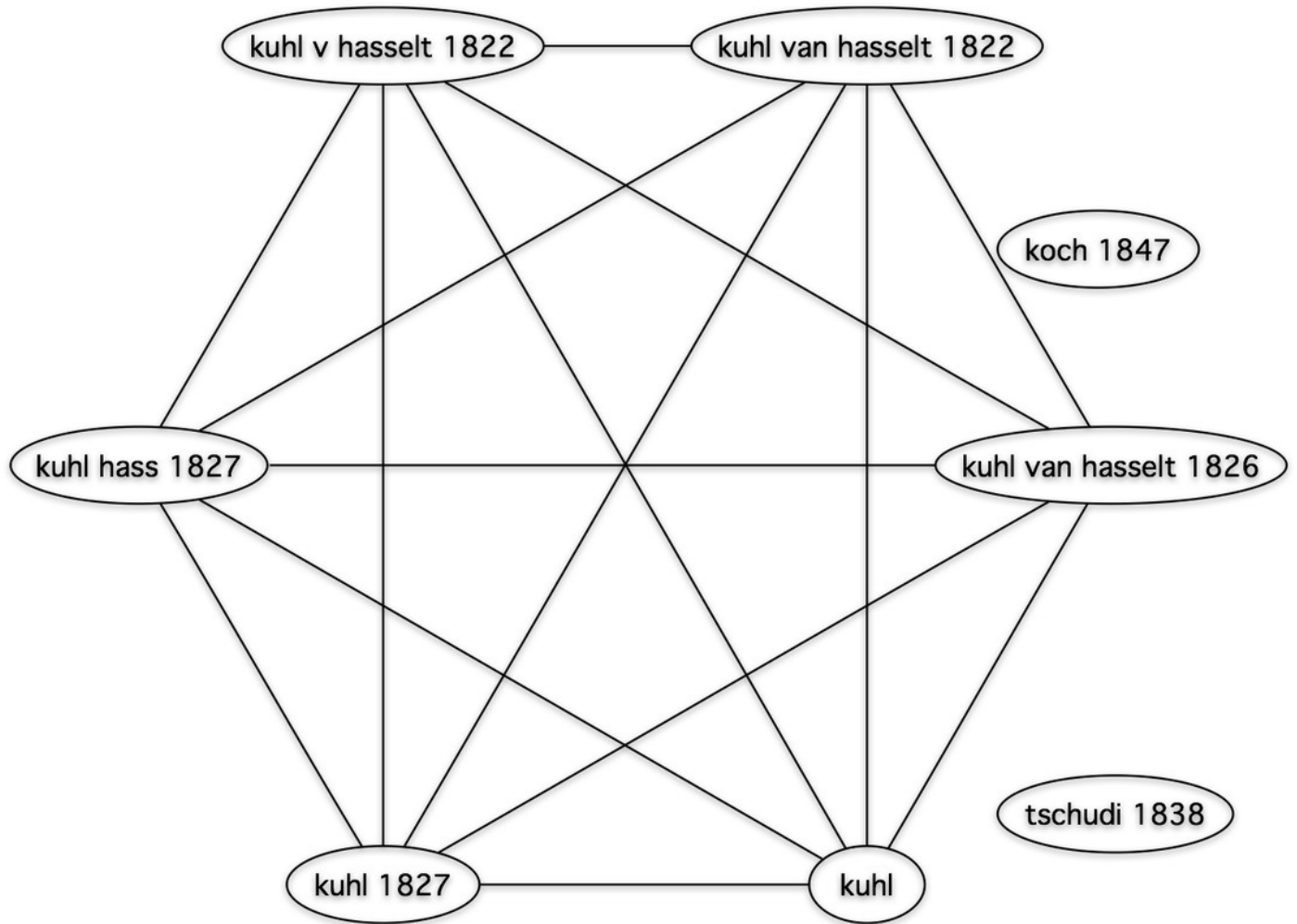


Figure 4

Matching taxonomic names to taxa

Bipartite graph of string similarities between taxonomic names containing the string "Nystactes" in the ION and GBIF databases. Solid edges in the graph represent the maximum weighted bipartite matching, and define the mapping between ION and GBIF names.

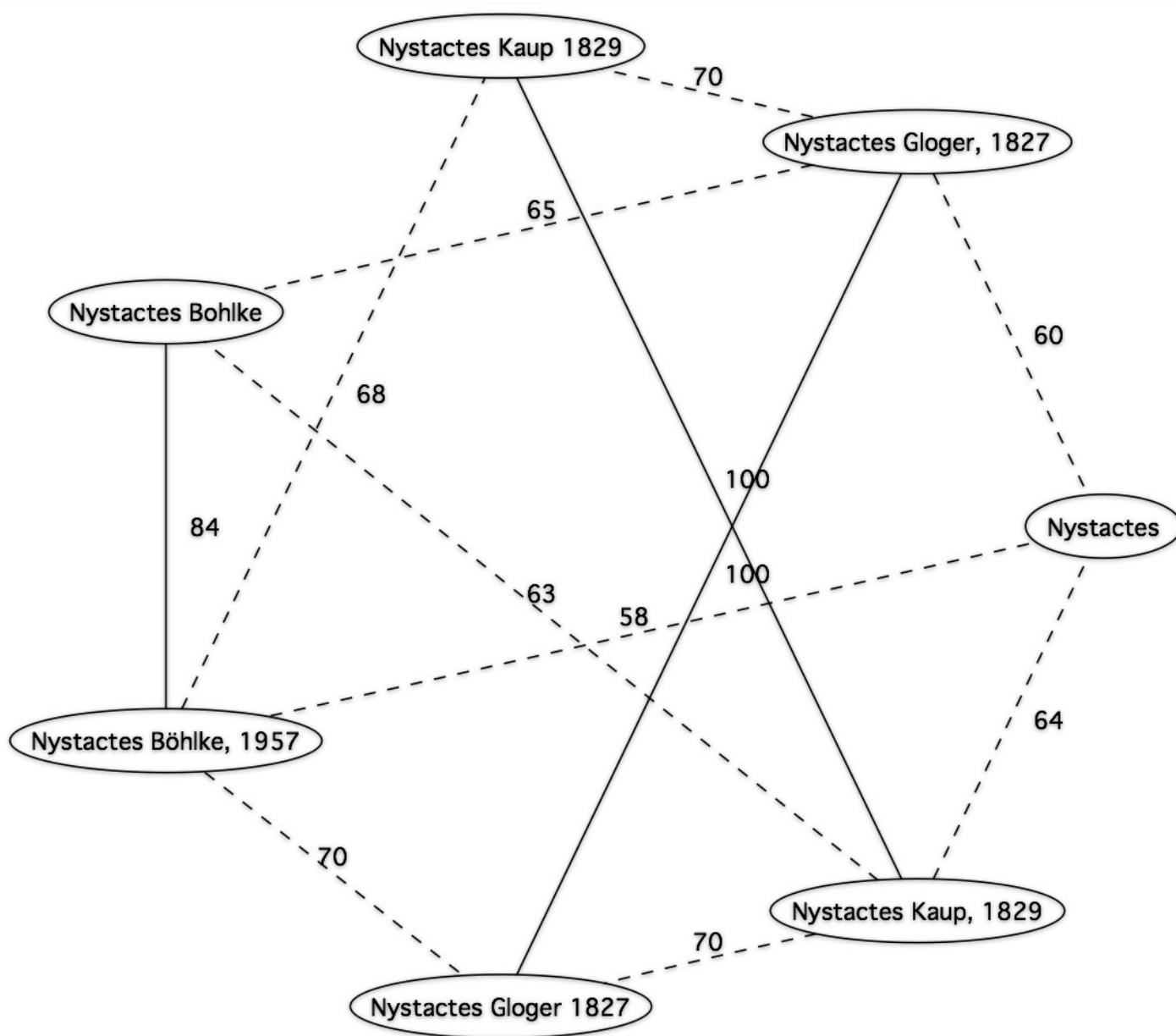


Figure 5

Search results


Screenshot of the search results for a query BioNames. The results include names that match the query, taxon concepts from GBIF and NCBI with thumbnail images from EOL, phylogenies containing members of the genus, and relevant taxonomic publications.


BioNames Dashboard

Names

Pristimantis Pristimantis moro Pristimantis yukpa Pristimantis orcus Pristimantis chimu






Taxa

 **Pristimantis Jiménez**
According to GBIF 2424987


 **Pristimantis**
According to NCBI 449102


GBIF NCBI


Phylogenies


 NADH dehy...  COI  12S ribosom...  16S ribosom...  trnV


Articles


 **Fauna neotropicals spe...**
by M Jiménez De La Espada; *Jornal de Sciencias Lisboa* ix (1870)
BioStor 100700


 **Batrachios Argentinos. En...**
by Carlos Berg; *Anales del Museo de Buenos Aires* v pages 147--226 (1896)
BioStor 103833


 **Evolutionary relationshi...**
by J D Lynch; *Misc Publ Mus Nat Hist Univ Kansas* 53 pages 1--238 (1971)
BioStor 59550


 **Frogs of the genus Eleut...**
by William E Duellman; Jennifer B Pramuk; *Scientific Papers Natural History Museum the University of Kansas* 13 pages 1--78 (1999)


 **The Eleutherodactylus o...**
by William Edward Duellman; John D Lynch; *University of Kansas Natural History Museum Miscellaneous Publication* 69 pages 1--86 (1980)

 **Fauna neotropicals spe...**
by M Jiménez De La Espada; *Jornal de Sciencias Lisboa* ix (1870)
BioStor 100700

 **A new species of the Pri...**
by Edgar Lehr; Gunther Kohler; *Zootaxa* 1621 pages 45--54 (2007)

 **A diminutive new specie...**
Salamandra 43(3) pages 165--171 (2007)

 **Three new malodorous r...**
by D Bruce Means; J A Y M Savage; *Zootaxa* 1658 pages 39--55 (2007)

 **A new peculiar frog spe...**
by Giovanni Boano; Stefano Mazzotti; Roberto Sindaco; *Zootaxa* 1674 pages 51--57 (2008)

Did you mean

[Pristimantinae](#)
[Pristimantis moro](#)
[Pristimantis royi](#)
[Pristimantis mars](#)
[Pristimantis pecki](#)
[Pristimantis palsa](#)
[Pristimantis orcus](#)
[Pristimantis rozei](#)
[Pristimantis uisae](#)
[Pristimantis vidua](#)
[Pristimantis myops](#)
[Pristimantis turik](#)
[Pristimantis stipa](#)
[Pristimantis yukpa](#)
[Pristimantis avius](#)
[Pristimantis altae](#)
[Pristimantis adnus](#)
[Pristimantis leoni](#)
[Pristimantis cacao](#)
[Pristimantis bambu](#)
[Pristimantis lemur](#)
[Pristimantis danae](#)
[Pristimantis galdi](#)
[Pristimantis chimu](#)
[Pristimantis ridens](#)
[Pristimantis rivasi](#)
[Pristimantis riveti](#)
[Pristimantis repens](#)
[Pristimantis onorei](#)
[Pristimantis roseus](#)
[Pristimantis myersi](#)
[Pristimantis orcesi](#)
[Pristimantis ortizi](#)
[Pristimantis piceus](#)
[Pristimantis pugnax](#)
[Pristimantis toftae](#)
[Pristimantis xestus](#)
[Pristimantis zoliae](#)
[Pristimantis zophus](#)
[Pristimantis mendax](#)
[Pristimantis wiensi](#)
[Pristimantis viejas](#)

Names: 5 Taxa: 2 Phylogenies: 5 Articles: 42

Figure 6

Displaying an article

Screenshot of BioNames displaying a document from BioStor (Conle and Hennemann 2002).

The document viewer can display page images, thumbnails, and (where available) text.

The screenshot displays the BioNames web interface. At the top, there is a search bar and a 'Dashboard' link. Below this, navigation tabs for 'View', 'Details', 'Names' (with a count of 11), and 'Grid' are visible. The main content area is a document viewer with tabs for 'DOCUMENT', 'PAGES', and 'TEXT'. The 'PAGES' tab is active, showing a grid of 24 page thumbnails labeled from p. 109 to p. 139. The thumbnails show various types of content, including text pages and pages with scientific illustrations of insects. A zoom control and a search bar are located at the top of the viewer. To the right of the viewer is a sidebar containing the following information:

- Revision of neotropic Phasmatodea: The tribe Anisomorphini sensu Bradley & Galii 1977: (Insecta, Phasmatodea, Pseudophasmatidae)**
- NAMES**: 11
- Authors**: Oskar V Conle, Frank H Hennemann
- [View on BioStor](#)
- 0 comments**
- Message input field: "Leave a message..."
- Sort options: Best, My Disqus, Share, Settings
- Text: "No one has commented yet."
- ALSO ON BIONAMES**:
 - [Rhachotropis Smith 1883](#) (1 comment • 16 days ago)
 - [Cyclopodia horsfieldi de Meij.](#) (1 comment • 15 days ago)
 - [Histiostrongylus parnelli Webster 1971](#) (1 comment • 14 days ago)
 - [L'état actuel de nos connaissances sur les Chiroptères fossiles-\(Note préliminaire\)](#) (1 comment • 14 days ago)
- DISQUS**: Comment feed, Subscribe via email

Figure 7

Displaying a journal

Screenshot of the page in BioNames for the journal *Proceedings of the Entomological Society of Washington* (ISSN 0013-8797). The centre column lists the articles in a volume selected by the user using the index on the left. The right hand column displays basic data about the journal, and a graphical display of how many articles have been mapped to a globally unique identifier.

BioNames Search Dashboard

1880's

1890's

1900's

1910's

1920's

1930's

1940's

1950's

1960's

1970's

1980's

- 1980
 - vol. 82 **37**
- 1981
 - vol. 83 **41**
- 1982
 - vol. 84 **42**
- 1983
 - vol. 85 **63**
- 1984
 - vol. 86 **55**
- 1985
 - vol. 87 **41**
- 1986
 - vol. 88 **39**
- 1987
 - vol. 89 **47**
- 1988
 - vol. 90 **22**
- 1989
 - vol. 91 **34**

1990's

2000's

A new subgenus for Forcipomyia, with descriptions of eight new species (Diptera: Ceratopogonidae)

by B De Meillon; W W Wirth;
Proceedings of the Entomological Society of Washington 82(1) pages 9--24 (1980)

- BioStor 59619

The Geomydoecus oregonus complex (Mallophaga: Trichodectidae) of the Western United States pocket gophers (Rodentia: Geomyidae)

by Roger DeForrest Price; Ronald A Hellenthal;
Proceedings of the Entomological Society of Washington 82(1) pages 25--38 (1980)

- BioStor 65011

New Species Of The Riffle Beetle Genus Portelmis From Ecuador

by P J Spangler;
Proceedings of the Entomological Society of Washington 82(1) pages 63--68 (1980)

- BioStor 75920

Two new species of Chloroperlidae (Plecoptera) from Mississippi

by Surdick ; Stark ;
Proceedings of the Entomological Society of Washington 82(1) pages 69--73 (1980)

- BioStor 70149

Notes Of American Aradinae (Hemiptera, Aradidae)

by N A Kormilev;
Proceedings of the Entomological Society of Washington 82(1) pages 99--107 (1980)

- BioStor 75880

New Species Of Midge Of The Genus Forcipomyia Melgen (Diptera, Ceratopogonidae) From North America

by P G Bystrak; D H Messersmith;
Proceedings of the Entomological Society of Washington 82(1) pages 108--116 (1980)

Proceedings of the Entomological Society of Washington

ISSN-L **0013-8797**

ISSN **0013-8797**

Latest articles **RSS**

Vol. 1, no. 1 (Feb. 29, 1884 to Dec. 3, 1885)-

Identifier coverage

DOI, Handle, BioStor, JSTOR, CINII, PMID, PMC

Feedback

Figure 8

Timeline of taxonomic names for birds

Screenshot of the distribution overtime of publications of new names for birds (Aves). The treemap on the left displays taxa below Aves in the taxonomic hierarchy, the chart on the right displays the number of publications in each year that publish a new bird name. The user has clicked on "2012", resulting in a list of the papers published in that year appearing below the timeline.

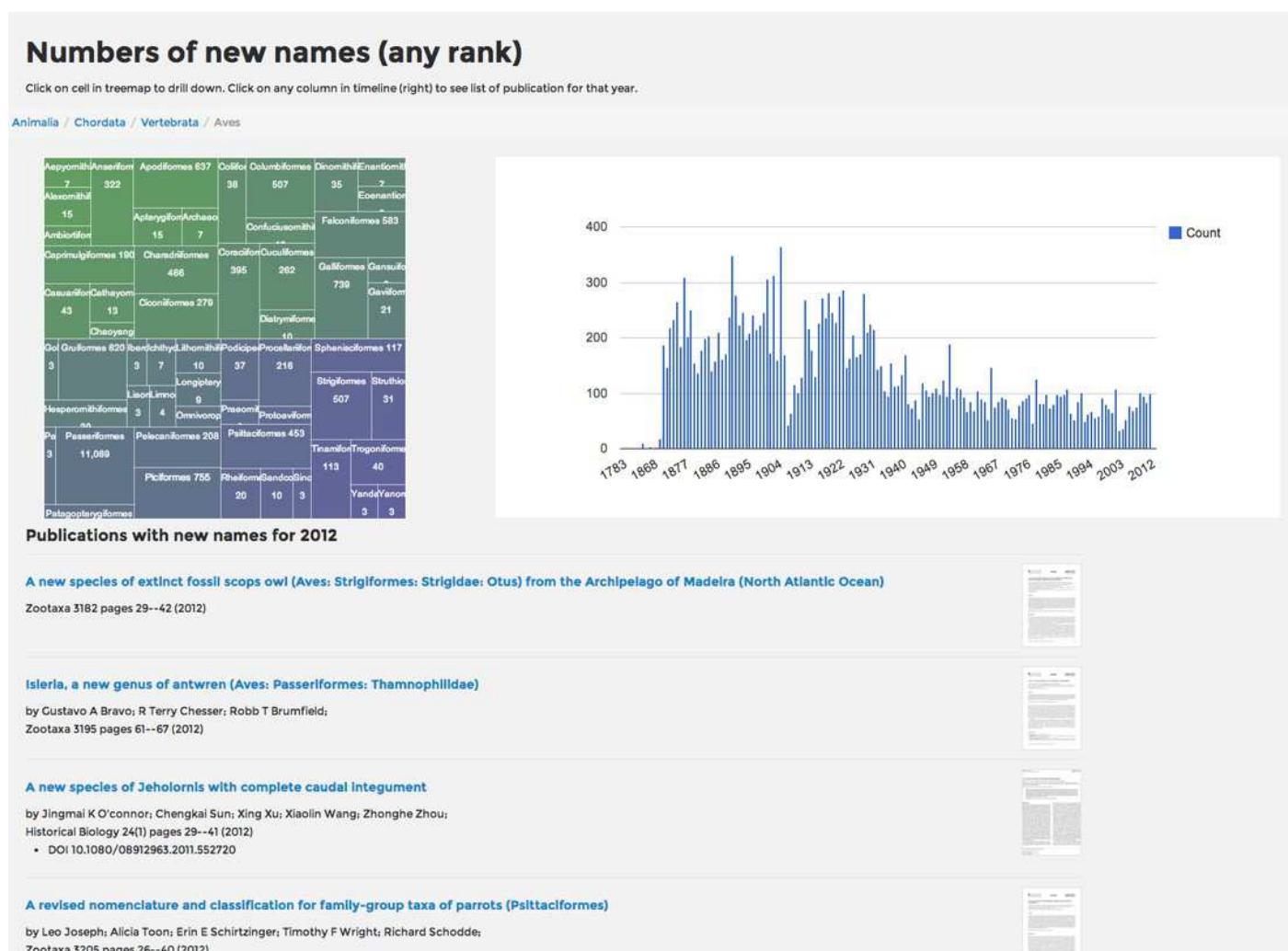


Figure 9

Bibliography for a taxon

Screenshot of the bibliography tab on a taxon page in BioNames. This example shows the publications relevant to the bat genus *Rousettus*, including those for synonyms. The user can select publications from a given time slice and/or combination of synonyms.

The screenshot displays the BioNames interface for the taxon *Rousettus*. At the top, there is a search bar and a dashboard link. The main navigation includes 'Name', 'Bibliography' (with a count of 31), 'Map', and 'About'. The 'Bibliography' tab is active, showing a timeline chart from 1880 to 1990. The chart lists synonyms: Cynonycteris, Eleutherura, Rousettus, Rousettus, Senonycteris, Stenonycteris, and Xantharpyia. Below the chart, a list of publications is shown, including 'The families and genera of bats' by Gerrit S. Miller (1907), 'On Pterocyon, Rousettus and Myonycteris' by Knud Andersen (1907), 'V. On a collection of Mammals made by Mr. S. A. Neave in Rhodesia...' by R. C. Wroughton (1907), 'Catalogue of the Chiroptera in the collection of the British Museum. Volume I: Megachiroptera' by Knud Andersen and George Edward Dobson (1912), and 'Dos nuevos murciélagos frugívoros' by Angel Cabrera (1920). On the right side, the taxon name 'Rousettus Gray, 1821' is displayed, along with its rank (genus), number of names (1), and number of publications (31). There are sections for 'Images' (from EOL), 'Classification' (showing the hierarchy from Pteropodidae to Rousettus), and '0 comments'.

Figure 10

Phylogeny viewer

Screenshot of phylogeny from PhyLoTA as displayed in BioNames. The user can zoom in and out and pan, as well as change the layout of the tree.

The screenshot displays the BioNames Phylogeny viewer interface. At the top, there is a search bar and a dashboard link. The main area shows a circular phylogenetic tree with various taxa labeled, including *Cyprina* and *Bathymodiolinae*. A green bar at the bottom of the tree area indicates "Parsed OK (use mouse to zoom and pan)". To the right, a sidebar provides details for the selected taxon, *Bathymodiolinae*, including the number of sequences (153), taxa (58), and sources (7). Below this, there is a map showing the localities of sequences and a section for related trees, including COI, ND4, and various ribosomal RNA genes. The interface also includes navigation icons at the top of the tree area and a comment section at the bottom right.

Figure 11

Relative importance of different publishers of taxonomic literature

Bubble chart showing relative numbers of taxonomic articles made available online by different publishers.

