

Running Title: Semantic Approach to Comparative Homology

1
2
3
4
5
6
7
8
9
10
11

Assessing similarity: a semantic approach to non- evolutionary comparative homology

Lars Vogt

Institut für Evolutionsbiologie und Ökologie, Universität Bonn,

An der Immenburg 1, D-53121 Bonn, Germany

[*lars.m.vogt@gmail.com*](mailto:lars.m.vogt@gmail.com)

Telephone: +49 228 73 5122

Fax: +49 228 73 5129

Abstract

13 The concept of phylogenetic homology has been criticized of involving circular argumentation
14 resulting from a methodological gap between its ontological definition and its empirical
15 recognition criteria. Based on the role of similarity for the recognition of phylogenetic
16 homologues I argue that phylogenetic homology presupposes non-evolutionary comparative
17 homology. Due to their use of Platonic ideals, archetypes and the requirement of the *a priori*
18 assumption of a stable positional reference system, pre-Darwinian notions of homology cannot be
19 used in this context. Based on Young's concept of comparative homology, I suggest a similarity-
20 based procedure for recognizing comparative homologues that utilizes a semantic framework.
21 This approach circumvents the problems of the pre-Darwinian notions of comparative homology
22 and provides the so far lacking conceptual basis for the application of recognition criteria of
23 phylogenetic homology, adding the missing link to the logical chain for assessing phylogenetic
24 homology and closing the methodological gap between its ontological definition and its
25 epistemological recognition criteria. This approach to comparative homology provides an
26 essential methodological framework for generalizing over individual organisms and identifying
27 and demarcating anatomical structural kinds. It thus represents an important contribution to the
28 theoretical and methodological foundation of both morphology and comparative biology.

29
30 **Keywords:** comparative homology, phylogenetic homology, homology concept, ontology,
31 semantic instance anatomy, semantic graph

32

33 **1 Introduction**

34 Homology is considered to be one of the most important concepts in biology (e.g., Jardine
35 and Jardine, 1969; de Beer, 1971; Bock, 1974; Riedl, 1979; Brigandt and Griffiths, 2007; Minelli
36 and Fusco, 2013). It is essentially a comparative concept, whose relationship to systematics
37 always has been very close.

38 By attributing homology to a transformation within a common ancestor, Hennig (1950)
39 linked homology explicitly and unambiguously to the phylogeny of species, therewith
40 establishing the methodological basis for using phylogenetic character statements as evidence for
41 phylogeny reconstruction. Whereas homology takes in a central role in phylogeny reconstruction,
42 it is important to specify which interpretation of the homology concept is relevant in
43 phylogenetics, because various evolutionary and non-evolutionary interpretations exist (for a
44 general discussion of homology see e.g. Rieppel, 1993; Panchen, 1994; Hall, 1994, 2012; Bock
45 and Cardew, 1999).

46 However, before phylogeneticists can analyze character matrices through various
47 algorithms of numerical tree inference, they first have to conduct some other steps of inquiry (e.g.
48 Vogt *et al.*, 2010). When using morphology as a source of evidence, the anatomical organization
49 of several organisms first has to be studied and compared. On this basis, hypotheses of
50 phylogenetic homology are inferred and coded into morphological phylogenetic character
51 statements. Not only is the coding of morphological characters and character states far from being
52 trivial (e.g., Hauser and Presch, 1989; Platnick *et al.*, 1991; Maddison, 1993; Pleijel, 1995;
53 Wilkinson, 1995; Rieppel, 1996; Hawkins *et al.*, 1997; Lee and Bryant, 1999; Strong and
54 Lipscomb, 1999; Sereno, 2007; Ramírez, 2007), but it also seems to be difficult to develop a
55 consistent theoretical and methodological framework for the preceding comparative study.

56 Morphological character analysis has been more or less ignored in the theoretical
57 literature so far, with exceptions usually referring to primary homology assessment (see e.g.
58 Bock, 1977; Neff, 1986; Rieppel and Kearney, 2002; Jenner, 2002, 2004; Vogt, 2004; Ramírez,
59 2007). Especially when comparing to the step of numerical tree inference it becomes obvious that
60 there has been little discussion of methods and problems regarding the step of the morphological
61 comparative study and the individuation of morphological phylogenetic characters (Pogue and
62 Mickevich, 1990). Wilkinson (1995) claimed that biology still needs a unified theory of character
63 construction. de Pinna comments, when discussing primary homology, that "*the proposition of*

Running Title: Semantic Approach to Comparative Homology

64 *primary homologies has never become satisfactorily objective (cf. Jardine, 1969: 329), and one*
 65 *must agree with Patterson (1982: 58) that "hypotheses of homology are conjectures whose*
 66 *source is immaterial to their status", and with Woodger (1937: 137) that "[t]here is a primary*
 67 *sense of 'homology' which we all use intuitively and upon which all the more sophisticated senses*
 68 *of the word depend."* (de Pinna, 1991: p. 373). de Pinna concludes that *"the recognition of*
 69 *similarities stands as an essentially imprecise and subjective process"* (de Pinna, 1991: p. 377).

70 In the following I will introduce a semantic approach to a similarity-based comparative
 71 homology concept that is based on ideas of Young (1993). This approach has the potential to
 72 substantially increase the much needed semantic transparency and computer parsability for
 73 similarity propositions in morphology, consequently subjecting them to criticism and corrections.
 74 The approach not only contributes an important methodological link for morphological character
 75 analysis in phylogenetics, but also an important methodological framework to the comparative
 76 method in general.

77 **2 Phylogenetic Homology**

78 **2.1 Phylogenetic Characters and the Phylogenetic Homology Concept**

79 In an influential publication, de Pinna (1991) provided a straight forward interpretation of
 80 homology within the framework of the cladistic methodology, with homology being synonymous
 81 with synapomorphy, a position that has been advanced before (e.g., Wiley, 1975; Bock, 1977;
 82 Rieppel, 1980, 1988; Patterson, 1982; Stevens, 1984), but also criticized (e.g., de Queiroz, 1985;
 83 Roth, 1988; Wagner, 1989). de Pinna (1991) distinguished primary from secondary homology as
 84 two interdependent and complementary notions of homology that are applied in two consecutive
 85 steps during phylogeny reconstruction. **Topographical correspondence** is applied as a criterion
 86 for postulating a **conjecture of homology** (= primary homology). The underlying test has been
 87 referred to as the similarity or identity test in cladistic literature (Bock, 1977; Cracraft, 1981;
 88 Patterson, 1982, 1988; Vogt, 2002, 2004, 2008a; Agnarsson and Coddington, 2008; for an
 89 argument against similarity constituting a test see de Pinna, 1991). In a subsequent step, primary
 90 homology can be supported or rejected as **phylogenetic homology** (= secondary homology)
 91 based on additional criteria that involve the **congruence test** (Patterson, 1988; de Pinna, 1991;

Running Title: Semantic Approach to Comparative Homology

92 Vogt, 2002, 2004, 2008a; Richter, 2005; Agnarsson and Coddington, 2008). de Pinna (1991)
 93 argued that whereas the congruence test may refute particular primary homology statements,
 94 these statements will nevertheless be of significance for phylogeny reconstruction and contain
 95 homology statements at a lower level of generality (e.g., the primary homology '*presence of*
 96 '*wings in vertebrates*' coded in a matrix will result in the secondary homologies '*wings of bats*' and
 97 '*wings of birds*' on a most parsimonious tree). This is a consequence of primary homology
 98 statements being closer to the empirical evidence that secondary homology statements attempt to
 99 explain. If a given primary homology cannot be explained by a single transformation event that
 100 would imply that it represents a synapomorphy, a given tree still has to explain it somehow, even
 101 when requiring the assumption of additional transformation events that translate the primary
 102 homology hypothesis to several secondary homology hypotheses.

103 Phylogenetic character statements are used for inferring the phylogeny of a given group of
 104 taxa. Character statements are statements about putative homologies (e.g., Freudenstein, 2005), in
 105 which (i) a character state refers to a putative **taxic homology**, which is epistemologically based
 106 on topographical identity, and in which (ii) the relation between different character states of a
 107 character refers to a putative **transformational homology**, which is epistemologically based on
 108 topographical correspondence (Brower and Schawaroch, 1996; for a discussion of other
 109 meanings of *taxic* and *transformational homology* see Brower, 2014). Phylogenetically
 110 informative characters are therefore indicative of transformation events (Hawkins *et al.*, 1997;
 111 Lee and Bryant, 1999), and at least one of the character states of a character is considered to
 112 represent a putative synapomorphy (Vogt, 2002).

113 2.2 Methodological Problems with the Phylogenetic Homology Concept

114 2.2.1 *Difference between Ontological Definition and Empirical Recognition*

115 Synapomorphies, or phylogenetic homologues, represent theoretical entities whose
 116 ontological definition refers to historical essences or **historical identity** (identity as a result of
 117 common evolutionary origin). As a consequence, because historical identity cannot be recognized
 118 directly, the existence of phylogenetic homologues cannot be identified and demarcated directly
 119 on a purely observational basis, but, instead, only indirectly, mediated through directly

120 observable entities that function as diagnostic characters. The recognition criteria of phylogenetic
121 homology thus lack clearly defined necessary and sufficient criteria.

122 The epistemological status of phylogenetic homologues is thus comparable to that of
123 biological species and monophyla, and assertions about their existence are necessarily always
124 **hypothetical**. As a consequence, a morphological character statement never directly represents
125 an anatomical entity and its properties—phylogenetic character statements are not descriptive
126 statements about the anatomical organization of a specimen, they are not morphological data in a
127 strict sense, but instead statements about hypothetical entities. Whereas "*similarity or*
128 *topographical correspondence is factual*" (de Pinna, 1991, p. 373) and can be documented in
129 morphological descriptions, character statements go beyond a factual description of what exists,
130 as they include a hypothetical and explanatory component (Vogt *et al.*, 2010).

131 According to the ontological definition of phylogenetic homology, phylogenetic
132 homologues do *not necessarily* have to be similar in appearance (e.g., Cracraft, 1967; Gans,
133 1985). As a consequence, although each phylogenetic homology possesses clearly defined
134 necessary and sufficient **ontological** criteria for particular anatomical entities to be member of it
135 (i.e. to be phylogenetically homologous), their **epistemological** recognition criteria lack them.
136 However, without underlying similarity, the recognition of phylogenetic homologues is
137 practically impossible. Therefore, it is epistemologically necessary to distinguish phylogenetic
138 homologues that share similar features from phylogenetic homologues that do not share similar
139 features, with only the former being epistemologically relevant to phylogenetics (see Vogt,
140 2002). As a consequence, similarity represents a necessary but not a sufficient condition for
141 phylogenetically relevant homologues (Vogt *et al.*, 2010).

142 Morphological character statements relate distribution patterns of morphologically similar
143 inheritable anatomical entities of different operational taxonomic units (OTUs) to evolutionary
144 events. Establishing this relation requires the interpretation of observable evidence that is
145 documented in morphological descriptive data in the context of the theory of evolution, assuming
146 that a given distribution pattern is the result of a series of particular evolutionary events,
147 consisting of at least one transformation event, followed by speciation events the number of
148 which depends on the number of OTUs sharing the similarity (given that the OTUs represent
149 separate species). Similarity, in this context, means that anatomical entities must be equivalent or
150 similar in at least one heritable aspect across several OTUs. And since such

151 equivalence/similarity represents only a necessary but not a sufficient condition for phylogenetic
152 homology (Vogt *et al.*, 2010), phylogeneticists consider similarity to be indicative for **putative**
153 **phylogenetic homology** (Rieppel, 2005). Thus, distribution patterns of similar inheritable
154 anatomical entities are used as empirical scientific arguments for the reconstruction of phylogeny
155 (Vogt, 2002, 2004).

156 2.2.2 Circular Methodological Argumentation

157 Phylogenetic homology has been criticized of involving a **circular methodological**
158 **argument**, because in order to identify phylogenetic homologues the underlying phylogeny must
159 be known, which, however, can only be inferred through the identification of homologous
160 characters (e.g. Boyden, 1947; Eckhardt, 1964; Ghiselin, 1966, 1976; Sokal and Sneath, 1973;
161 Sattler, 1984; Stevens, 1984). Young (1993) argues that the circularity problem largely stems
162 from the **methodological gap** between the ontological definition of phylogenetic homology,
163 which refers to historical identity, and the empirical recognition criteria for phylogenetic
164 homologues, which rely primarily on similarity. This discrepancy cannot be solved by providing
165 an ontological definition of phylogenetic homology that does not refer to similarity, implying that
166 similarity is not part of the logical chain of assessing phylogenetic homology, as a specific notion
167 of similarity is still employed as the recognition criterion (Young, 1993).

168 Epistemologically, the phylogenetic homology concept is rooted in a non-evolutionary
169 homology concept for recognizing similarity that is used as an index to phylogenetic
170 homologues, which at their turn are used for reconstructing phylogeny and evolution. This
171 duality has been recognized before (see *homology and homogeny*, Lankester, 1870; *observational*
172 *and theoretical components*, Jardine, 1969; *topographical and phylogenetic homology* Rieppel,
173 1980; *preliminary and final testing*, Kluge and Strauss, 1985; *topological correspondence and*
174 *homology*, Rieppel, 1988) and also underlies de Pinna's distinction between primary and
175 secondary homology.

176 2.3 The Historical Relation between Non-Evolutionary Homology Concepts and 177 the Recognition of Phylogenetic Homologues

178 Whereas the underlying ontological definition of de Pinna's homology concept is based on
179 the theory of evolution, its criteria of identification, which go back to Remane's (1952) homology
180 criteria and beyond, rest to a large degree on former empirical experience with the phenomenon
181 of 'homology'. The underlying historical continuity becomes even more obvious when comparing
182 Remane's (1952) *criteria of position* and of *special quality* with Etienne Geoffroy Saint-Hilaire's
183 (1830) **principle of connections** and his **unity of composition**. According to the *principle of*
184 *connections* homologues share the same spatial relationship within the organism and according to
185 the *unity of composition* they are composed of the same substances and constituent parts (c.f.
186 Russell, 1916; Rieppel, 1988; Young, 1993; Hall, 2012). Some of these ideas can even be found
187 in Linnaeus' categories of quantity of observed entities, their basic geometrical form, their
188 spatiotemporal distribution, and their relative size—categories which Linnaeus applied in his
189 sexual system for the classification of plants (Linnaeus, 1735, 1751; for a discussion see Vogt,
190 2008b). The historical roots of recognizing homologues can be traced back even further in time,
191 as Belon's illustration of 1555 indicates (reproduced in Panchen, 1994). The earliest English
192 reference to homology is attributed to William Sharp MacLeay in 1821 (Strickland, 1846;
193 Panchen, 1994). Some notion of homology therefore obviously arose prior to the mid-nineteenth
194 century and thus even prior to Owen's (1843) widely cited non-evolutionary definition of
195 homology (Staton, 2011).

196 A theoretically elaborate non-evolutionary concept of homology has been first developed
197 in the nineteenth century, during which the idea arose that two categories of similarity between
198 observable anatomical entities can be distinguished. On the one hand, anatomical entities that are
199 'misleadingly' identical and that occur isolated, which were called **analogues**, and their sameness
200 was traced back to equivalence of form and function (Rieppel, 1993). On the other hand,
201 anatomical entities that appear to be 'truly' identical, because they occur in reciprocally
202 corroborating aggregates, which were at that time traced back to the idea of essential identity
203 (Rieppel, 1993) and were called **homologues** (Owen, 1843, 1848, 1849; see also *affinities*,
204 Strickland, 1840a,b). Interestingly, the demarcation criterion of homologues against analogues
205 was based on a non-cladistic notion of congruence, i.e. the expectation that classifications based

206 on different homologues tend to confirm each other, because the distribution patterns of several
207 properties coincide.

208 Despite the problems with Owen's homology concept being based on Platonic ideals and
209 his bottom-up methodology of using perceivably simplest structures as archetypes (see Spencer,
210 1867), the criterion for distinguishing analogues from homologues is nevertheless comparable to
211 de Pinna's demarcation criterion for secondary homology based on primary homology (de Pinna,
212 1991), as both rely on some notion of congruence. However, they differ in that respect that the
213 former distinction rests on the untestable assumption that a natural order of organisms would
214 have to stand out because of a natural affinity of the corresponding organisms (Whewell, 1847),
215 which results from an innate property that is associated with a specific archetype or idealized
216 form (Owen, 1843).

217 2.4 The Application of the Phylogenetic Homology Concept presupposes a Non- 218 Evolutionary Homology Concept

219 Any use of general terms in morphology is linked to the idea of identifiable and re-
220 identifiable **structural kinds** (Mahner and Bunge, 1997; Rieppel, 2007) and thus types/classes of
221 anatomical entities. This implies that particular anatomical entities can be identified as instances
222 of a structural kind. Positional relationships, meaning Etienne Geoffroy Saint-Hilaire's *principle*
223 *of connections*, have traditionally been one of the main criteria used for the identification of
224 anatomical entities (Piaget, 1971; Young, 1993). Irrespective of whether applied for identifying
225 instances of structural kinds or for the recognition of non-evolutionary or phylogenetic
226 homologues, Young (1993) already argued that the application of the *principle of connections*
227 presupposes the ***a priori* assumption of a stable positional reference system**. As a consequence,
228 the application of the *principle of connections* will result in a circular chain if no **landmark**
229 **entities** (i.e., Young's disjointed intrinsically identifiable anatomical entities; Young, 1993) are
230 present that build a **stable pattern of spatial arrangement** that can be identified independent of
231 the reference system itself. These landmark entities are required for establishing a stable
232 reference system of relative positions for comparing the anatomy of different organisms.

233 Because most anatomical entities are not landmark entities, the *a priori* assumption of the
234 existence of a stable positional pattern is unavoidable in morphological research (e.g., Rieppel,

Running Title: Semantic Approach to Comparative Homology

235 1988). The importance of this basic assumption as a prerequisite for the applicability of the
236 *principle of connections* can be found throughout the biological literature (see in Young, 1993).
237 The recognition of the necessity of such an assumption can be seen as the methodological
238 justification for the necessity of the application of an alternative, non-evolutionary, similarity-
239 based **comparative homology concept** that does not rest on Platonic ideals and Owen's bottom-
240 up methodology. Such a comparative homology concept is required for descriptive and
241 comparative purposes and represents a methodological prerequisite for the applicability of the
242 phylogenetic homology concept—*"without some similarity, we should not even dream of*
243 *homology"* (Stevens, 1984, p. 403).

244 Conceptually, comparative homology is very close to what de Pinna characterized as the
245 basic form of homology meaning *"equivalence of parts"* that serves *"as a sorting procedure to*
246 *investigate the validity of comparative information"* (de Pinna, 1991, p. 368). A central question
247 regarding the comparative homology concept, however, is the question of how to recognize
248 comparative homologues, which implies the more fundamental question of *"how do we recognize*
249 *similarity?"* (Young, 1993, p. 233).

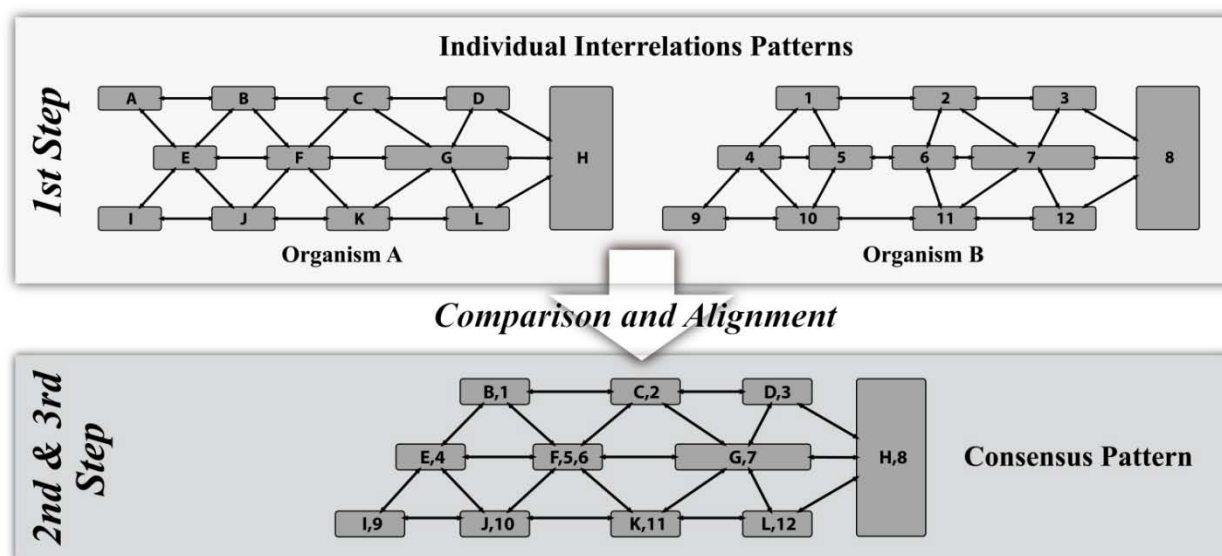
250 **3 A Semantic Approach to a Non-Evolutionary, Similarity-Based** 251 **Comparative Homology Concept**

252 Despite their obvious conceptual connections, none of the non-evolutionary homology
253 concepts of the nineteenth century can function as a similarity-based comparative homology
254 concept in present day research, because of the reasons discussed above (i.e. Platonic idealism,
255 Owen's archetype, unjustifiable presupposition of a stable positional reference system). Young
256 (1993), however, suggested an alternative and very promising approach following a three-step
257 procedure that emphasizes the interrelations between anatomical entities instead of their intrinsic
258 qualities for identifying units of comparison. This approach is independent of an *a priori*
259 assumption of a stable positional reference system.

260 Unfortunately, Young's approach has not been adopted by many morphologists so far. A possible
261 reason for the lack of acceptance might be the fact that it is not easily applicable, because it
262 requires morphologists to execute additional steps of translating morphological descriptions into

Running Title: Semantic Approach to Comparative Homology

263 a highly formalized network-like format, and Young only provided a very basic idea of how this
 264 network must be structured and how the different steps must be practically conducted (see Fig. 1)
 265 As a consequence, Young's approach poses considerable problems when one attempts to actually
 266 apply it to real data. However, by implementing semantic techniques and using ontology terms
 267 and a standardized and highly formalized syntax for morphological descriptions that can be
 268 visualized as a network-like semantic graph, Young's (1993) approach can be modified and made
 269 applicable.



270
 271 **Fig. 1.** Schematic diagram of the three steps of Young's (1993) approach. **1st Step:** *individual interrelations patterns*
 272 of anatomical entities of two organisms (A and B) are shown. The nodes represent anatomical entities and the arcs
 273 the relations between them. **2nd & 3rd Step:** the two *individual interrelations patterns* have been compared and
 274 aligned and a *consensus pattern* has been build. The combination of letters and numbers in each node denotes the
 275 comparative homologues present in the two organisms (A and B). (Figure adapted from figure 1, Young, 1993)

276 In the following, I provide a brief description of each step of this approach and argue how
 277 current semantic techniques could be applied to make each step actually operable by
 278 morphologists.

279 3.1 1st Step: Semantic Representations of Patterns of Interrelations of Anatomical 280 Entities

281 The first step concerns the description of **patterns of interrelations of anatomical**
282 **entities**, visualized as network diagrams, in which anatomical entities and their properties are
283 represented as nodes and the interrelations between them as arcs (see 1st step in Fig. 1). This can
284 be conducted using the Web Ontology Language (OWL) that can be serialized to the Resource
285 Description Framework (RDF) syntax.

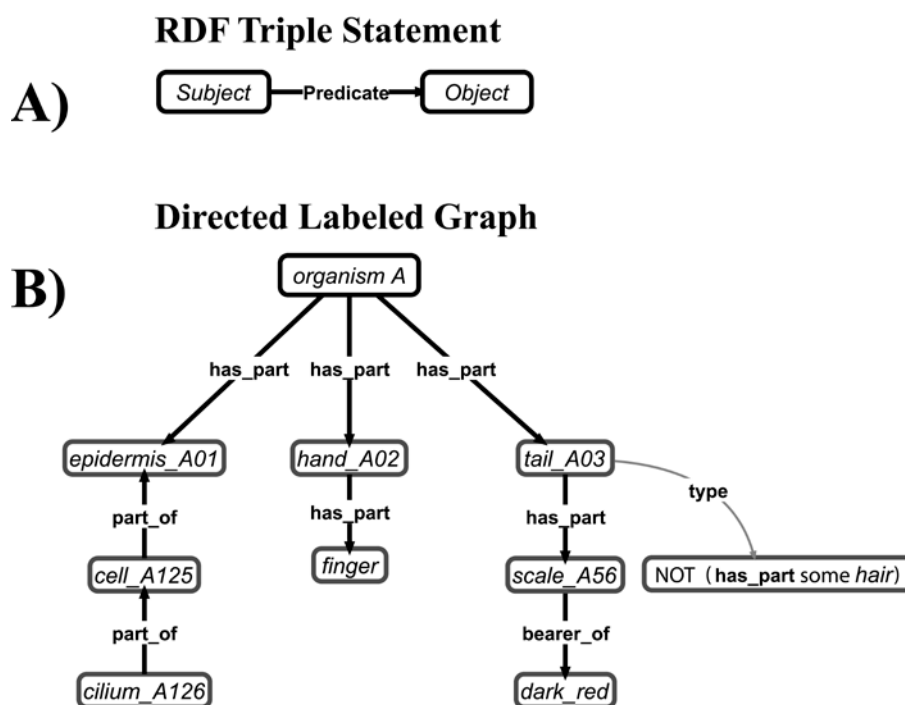
286 *3.1.1 Using RDF and OWL for describing the anatomical Organization of individual* 287 *Organisms*

288 In RDF, each proposition is structured as a triple statement consisting of a *Subject*, a
289 *Predicate* and an *Object*. The *Subject* represents a resource in the form of a Uniform Resource
290 Identifier (URI). The *Predicate* represents a relation in which this resource stands to another
291 resource or a literal. The *Predicate* itself is another resource that is expressed in the form of a
292 URI. The *Object* represents a resource (URI) or some value expressed as a literal. A resource
293 always refers to a real thing or a piece of data (e.g. a Web page, a class concept) and the literal
294 can be a unique ID, a numerical value, an arbitrary label, a proper name, a general term, or a
295 natural language text. A triple statement therefore describes a given resource by relating its URI
296 to another URI or some value via a URI that defines a predicate. In other words, the *Predicate*
297 determines how *Subject* and *Object* relate to each other (Fig. 2A).

298 The RDF triple statement can be modeled as a graph, with *Subject* and *Object* each
299 forming nodes that are connected through a labeled directed arc (edge), which is the *Predicate*
300 (Fig. 2A). A collection of such graphs can jointly form a semantic network called **directed**
301 **labeled graph** or **semantic graph** (Fig. 2B). Directed labeled graphs are very useful tools for
302 modeling domain knowledge and for analyses and inferences using graphs logic (e.g., Sowa,
303 1984, 2008).

304 The RDF syntax can be used for describing the relation between particular anatomical
305 entities of a particular specimen, as for instance parthood relations (e.g., '*scale_A56 part_of*
306 *tail_A03*'), or for describing specific properties of a particular anatomical entity (e.g., '*scale_A56*
307 **bearer_of dark_red**'). Triple statements can be interlinked by using the *Object* of one statement

308 as the *Subject* of another statement (e.g., '*cilium_A126 part_of cell_A125*' and '*cell_A125*
 309 *part_of epidermis_A01*'), resulting in chains or networks of triple statements (Fig. 2B). The
 310 resulting semantic graph describes an instance anatomy, which is why I refer to it as a *semantic*
 311 *instance anatomy*.



312

313 **Fig. 2. A) RDF Triple Statement:** the syntax of the Resource Description Framework (RDF) consists of three
 314 elements, i.e. a *Subject* and an *Object* that are linked through a *Predicate*. This triple statement can be visualized as a
 315 graph, in which *Subject* and *Object* are represented as nodes that are linked through a directed labeled arc (edge),
 316 which represents the *Predicate*. **B) Directed Labeled Graph:** The resource taking the *Object* position in one triple
 317 statement can take the *Subject* position in another, linking triple statements into chains and networks. One can use
 318 the RDF syntax to describe and visualize the anatomical organization of complex individual anatomical entities or
 319 individual organisms. Due to the open world assumption, negations are important for being able to express the
 320 absence of a specific anatomical entity and its properties. In this instance-based approach to description, negation is
 321 expressed by classifying the instance that does not have a specific type of anatomical entity as its part as an instance
 322 of the class that does not have this type of anatomical entity as its part (e.g., the triple statement *tail_A03 type* 'NOT
 323 (*has_part* some *hair*)' expresses that the tail of *organism A* does not possess any *hair*). For reasons of clarity, the
 324 *Subjects*, *Predicates* and *Objects* are not represented with their URIs but with their labels, with the letter-number
 325 codes indicating that they represent particular anatomical entities.

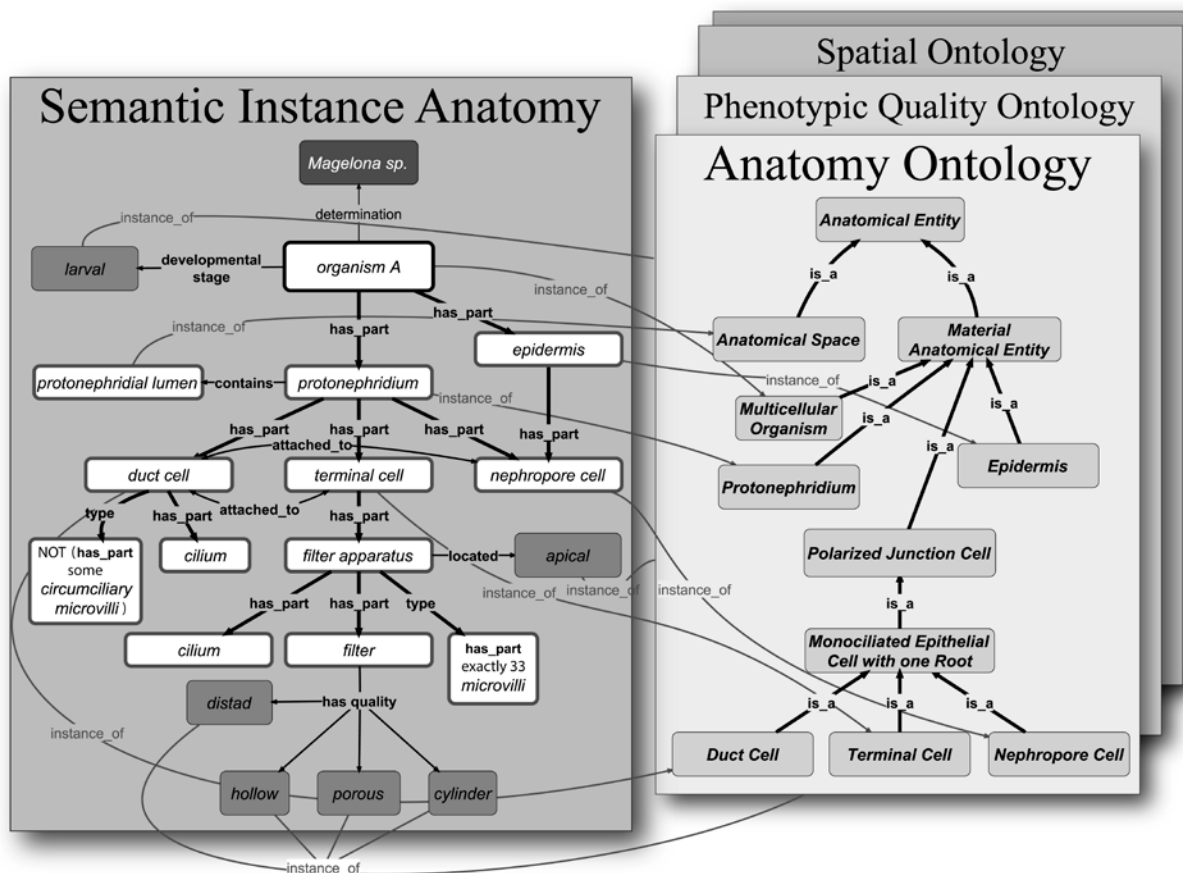
326 The Web Ontology Language (OWL) represents the language that is commonly used to
 327 describe ontologies. OWL is a semantically transparent knowledge representation language and a
 328 popular ontology language that can be serialized to RDF. OWL allows the specification of

329 restrictions to the applicability of *Predicates*, formally limiting what types of resources can be
330 linked to a *Predicate* as its *Subject* (domain of the *Predicate*) and its *Object* (range of the
331 *Predicate*). In addition to such global restrictions, OWL also allows assigning property
332 characteristics to *Predicates*, as for instance **transitivity** (if 'A is_a B' and 'B is_a C', then 'A is_a
333 C'), **reflexivity** ('A is_a A') and **antisymmetry** (if 'A is_a B' and 'B is_a A', then A and B are
334 identical), which are important for querying, information retrieval, reasoning, and knowledge
335 inference. Furthermore, OWL allows the specification of quantified relations (e.g., 'some', 'only'),
336 of cardinality (e.g., 'exactly 1'), and of necessary and sufficient conditions and it enables the
337 expression of negated statements.

338 Negated statements are important due to the **open world assumption** that necessarily
339 underlies every description of a specimen's anatomy, and thus every *semantic instance anatomy*.
340 It is impossible to provide a comprehensive description of any real world object—there is always
341 something missing in a description. As a consequence, we cannot infer the absence of a property
342 or a specific type of anatomical entity from the lack of information about it in a given *semantic*
343 *instance anatomy*. In other words, if something is not represented in a *semantic instance*
344 *anatomy*, we cannot conclude that it is necessarily absent, but only that information about it is
345 missing. Therefore it is important to be able to express negated statements in formalized
346 anatomical descriptions (Fig. 2B, Fig. 3; cf. Balhoff *et al.*, 2014).

347
348 The practical implementation of ontology-based morphological descriptions has been
349 shown to be feasible (e.g., Vogt, 2008b; Beck *et al.*, 2009; Washington *et al.*, 2009; Vogt *et al.*,
350 2010; Deans *et al.*, 2012; Mikó *et al.*, 2012; Balhoff *et al.*, 2013). When describing the
351 anatomical organization of complex anatomical entities or individual organisms, *Subjects* and
352 *Objects* can reference to terms of anatomy ontologies (e.g., *Foundational Model of Anatomy* for
353 *Homo sapiens*, Rosse and Mejino, 2007; *Fly Anatomy Ontology* for *Drosophila melanogaster*,
354 Grumbling and Strelets, 2006; *Hymenoptera Anatomy Ontology*, Mikó *et al.*, 2012), indicating
355 that a particular anatomical entity is an instance of a specific general type or class of anatomical
356 entity. When describing properties of anatomical entities, *Subjects* and *Objects* can also refer to
357 terms of quality ontologies (e.g., *Phenotypic Quality Ontology*, Mungall *et al.*, 2007). The
358 *Predicates*, on the other hand, refer to relations defined in various ontologies, including specific
359 relations ontologies (e.g., *OBO Relations Ontology*, Smith *et al.*, 2005, 2007) (Fig. 3). In this

360 way, the resulting descriptions can make use of the semantic machinery provided by these
 361 ontologies, which will become relevant in the second step (see below).



362
 363 **Fig. 3.** Semantic instance anatomy of a protonephridium of a larval specimen of *Magelona*. **Semantic Instance**
 364 **Anatomy:** Simplified description of a particular *protonephridium* of *organism A*. All nodes either refer to particular
 365 anatomical entities, types of properties, relations or to numerical values, and in case of negations and quantity
 366 statements also to class expressions. All arcs (edges) refer to *Predicates*. The anatomical entities connected through
 367 the *Predicate* 'has_part' build the organizational backbone of the semantic instance anatomy, forming a tree-like
 368 hierarchy (partonomy). The root node ('*organism A*') represents the entity that is partitioned and all nodes connected
 369 through 'has_part' *Predicates* its constituent parts. The parthood relation between *duct cell* and *circumciliary*
 370 *microvilli* is negated, implying that there exists no particular entity that is at the same time an instance of the class
 371 '*circumciliary microvilli*' and a part of *organism A*'s duct cell—the duct cell of the described organism does not
 372 possess any circumciliary microvilli. Additional nodes and arcs specify some of the properties of some anatomical
 373 entities, like for instance their *developmental stage* or they specify topological relations, as for instance their *location*
 374 or what they are *attached to*. Semantic instance anatomies use class-terms from ontologies. Therefore, all nodes
 375 referring to particular anatomical entities are linked to the corresponding terms of respective ontologies through an
 376 '*instance_of*' relation (shown only for some anatomical entities). The '*instance_of*' predicate relates a particular to a
 377 class, asserting that it is an instance of that class. *Description of protonephridium based on Bartolomaeus, 1998.*
 378 **Anatomy Ontology:** Hierarchical graph-based representation of a taxonomy of types of anatomical entities.
 379 Respective terms denote classes that are organized in a tree-like hierarchy of classes and their subclasses through the

380 'is_a' predicate. This taxonomy implies that any *terminal cell* is a special type of *monociliated epithelial cell with one*
381 *root*, and thus also a special type of *polarized junction cell*.

382 3.1.2 Semantic Representations of Individual Interrelations Patterns

383 For the first step of the semantic approach to comparative homology that is based on
384 Young's (1993) procedure, the method of description introduced above is applied for describing
385 **patterns of interrelations of anatomical entities** of individual organisms (= *individual*
386 *interrelations patterns*). It is important that all particular anatomical entities described must be
387 individuated and identified exclusively based on their intrinsic properties (i.e. Young's disjointed
388 intrinsically identifiable anatomical entities; Young, 1993), taking interior physical discontinuity
389 or qualitative heterogeneity as their recognition criteria, like for instance a sharp gradient of
390 material constitution, color, texture, or electric charge (cf. *bona fide objects*, Smith, 1994, 1995,
391 2001; Smith and Varzi, 1997, 2000; for a general discussion of boundaries of anatomical entities
392 see Vogt *et al.*, 212b) that is defining for a specific type of anatomical entity. These **landmark**
393 **entities** are identified solely based on their intrinsic qualities and independent of any spatial
394 relations, meaning independent of their position within the organism and independent of their
395 position relative to other anatomical entities. The respective classification is documented by
396 linking the node representing a particular anatomical entity to the corresponding term of an
397 anatomy ontology through the *Predicate instance_of* (cf. Fig. 3).

398 In many cases, this results in a somewhat unspecific classification of the anatomical
399 entities to be described. A photoreceptor cell, for instance, would not be represented as a
400 'photoreceptor cell' in an *individual interrelations pattern*, because no preparation and
401 visualization technique produces evidence that unambiguously indicates that a cell is a
402 photoreceptor cell, at least if solely based on its observable intrinsic properties. Instead, one
403 would represent the entity as a 'cell' (evidence: histology → nucleus; ultrastructure → cell
404 membrane) that has a 'portion of opsin' as its part (evidence: immunohistology → opsin) (see Fig.
405 4).

406 After having identified the landmark entities, all anatomical entities are described that can
407 be unambiguously individuated and identified in reference to such landmark entities (i.e. Young's
408 adjoining intrinsically identifiable anatomical entities; Young, 1993). This can lead to a more
409 specific representation of some of the anatomical entities in an *individual interrelations pattern*.

Running Title: Semantic Approach to Comparative Homology

410 For instance, opsin, which has been identified in the previous step, can be further specified based
411 on the anatomical entity in which it has been detected. If opsin has been detected in a microvilli
412 group (evidence: ultrastructure → cell membrane pattern), it must be rhabdomeric opsin and if it
413 has been detected in a branched cilium (evidence: ultrastructure → cell membrane pattern), it
414 must be ciliary opsin (see Fig. 4).

415 Unfortunately, however, because positional relationships represent one of the main
416 criteria for the identification of anatomical entities (Saint-Hilaire's *principle of connections*,
417 Piaget, 1971; Young, 1993), and the identification of anatomical entities also depends on the
418 preparation of the object of study and the visualization techniques applied, many anatomical
419 entities described that way will only be classified to a very general level, as for instance as a *cell*,
420 a *cell part*, a *cell group*, or a *cell cluster* (see Vogt, 2010; Vogt *et al.*, 2011). As a consequence,
421 the identification of the individual correlate of such a broadly classified anatomical entity in
422 another organism is not possible. Therefore, it is important that the observed positional and
423 topographical relations between the anatomical entities are also represented in the resulting
424 *individual interrelations patterns*. This spatial information, however, is not documented in the
425 nodes of the graph that refer to anatomical entities, but represented through the arcs (= *Predicates*)
426 connecting these nodes and through the nodes that refer to property values belonging
427 to these anatomical entities.

428 The generation of *individual interrelations patterns* does not presuppose any *a priori*
429 (=not empirically supported) stable positional reference system. In this context it is important to
430 note that, contrary to previous approaches to homology assessment that emphasize the intrinsic
431 properties of anatomical entities, the *individual interrelations patterns* instead emphasize the
432 **interrelationships between** anatomical entities.

433
434 What actually can be identified as landmark entities directly depends on what techniques
435 of preparation and visualization have been applied when studying the object of interest. Classical
436 histology allows the identification of cells (through cell nuclei) but not necessarily also of
437 organelles. Depending on information provided about the techniques and methods applied, an
438 adequate **domain granularity framework** could be applied that could interpret any underlying
439 anatomy ontology and filter out all those ontology terms that refer to entities that cannot be
440 unambiguously identified in the given circumstances (Keet, 2006, 2008; for domain granularity

Running Title: Semantic Approach to Comparative Homology

441 framework for morphology see Vogt, 2010; Vogt *et al.*, 2012a,b). This can lead to a very
 442 simplified ontology consisting only of a few different types of anatomical entities (in case of
 443 classical histology, e.g. 'cell', 'fiat cell part', 'cell cluster' and 'cell group'). Using such a *domain*
 444 *granularity framework* and taking the information about the preparation methods and
 445 visualization techniques applied into account, one could even take a semantic instance anatomy
 446 with fully classified anatomical entities and automatically translate it into an *individual*
 447 *interrelations pattern* in which only the landmark entities are fully classified.

448
 449 In addition to the here discussed *semantic instance anatomy* approach, an alternative
 450 approach to documenting the anatomical organization of a given specimen has been proposed,
 451 and the resulting representations have been called *Semantic Phenotypes* (e.g. Mungall *et al.*,
 452 2010; Deans *et al.*, 2012; Balhoff *et al.*, 2013). Contrary to the *semantic instance anatomy*
 453 approach, which follows an instance-based framework, the *Semantic Phenotypes* approach
 454 follows a class-based framework that defines *Semantic Phenotypes* as class expressions that are
 455 formally described following an Entity-Quality (EQ) expression (Mungall *et al.*, 2007; Dahdul *et*
 456 *al.*, 2010; Balhoff *et al.*, 2010). The syntax of the EQ statements is similar to that of phylogenetic
 457 character statements. In order to prevent circularity of reasoning, the *Semantic Phenotypes*
 458 approach cannot be applied to the here described semantic approach to identifying comparative
 459 homologues, because it presupposes the identification of anatomical structural kinds and the
 460 identification of units of comparison and thus of comparative homologues in the first place.

461 3.2 2nd Step: Comparing several Individual Interrelations Patterns and Inferring a 462 Consensus Pattern

463 The second step in the semantic approach to comparative homology involves the
 464 comparison of several *individual interrelations patterns* of different organisms for recognizing,
 465 extracting, and documenting parts within the patterns that are shared by all organisms that have
 466 been compared, resulting in a **consensus pattern** for these organisms (Fig. 1; cf. Young, 1993).
 467 This requires aligning nodes referring to anatomical entities across different graphs. The
 468 alignment is primarily based on identifying shared positional and topographical relations through
 469 respective spatial *Predicates*, but all other types of *Predicates*, as for instance those referring to

Running Title: Semantic Approach to Comparative Homology

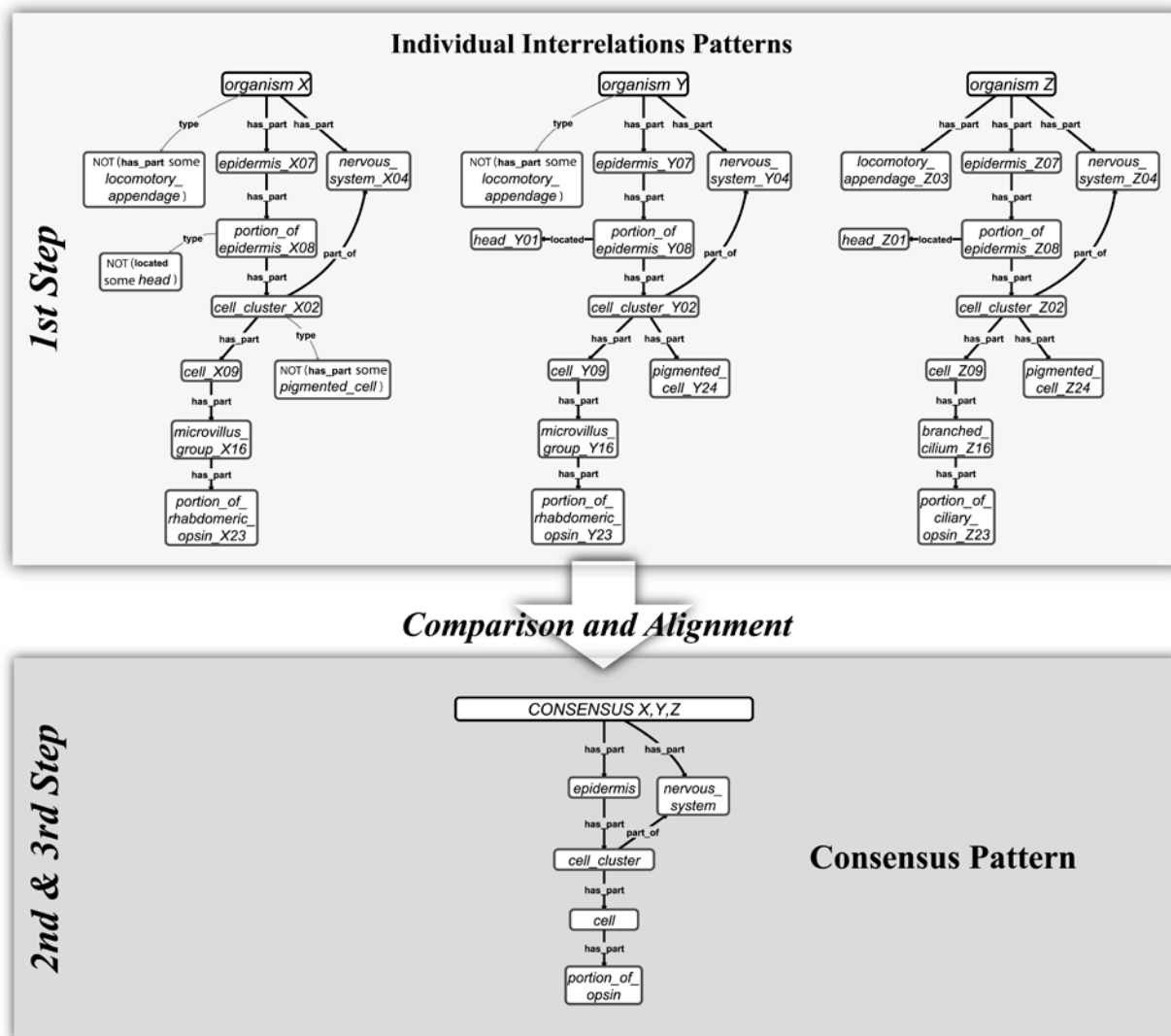
470 relative size, composition, association with other anatomical entities, and development, as well as
471 *Objects* referring to property values can also be considered.

472 The more detailed an *individual interrelations pattern* covers the anatomical organization
473 of an organism, the more valuable is the resulting *consensus pattern* for understanding the
474 morphological system (cf. Young, 1993). However, if specific anatomical entities exhibit
475 extensive variation across different organisms, many nodes representing the organisms' properties
476 will not be represented in the resulting *consensus pattern* and the entities themselves, if at all, are
477 only represented in a less specific classification, reflecting their limited degree of similarity (see
478 Fig. 4). A *consensus pattern* therefore usually contains less information than its underlying set of
479 *individual interrelations patterns*, and the degree of matching between them can be understood as
480 an indicator of the extent of morphological evolution or morphological variability within the
481 group of organisms compared (cf. Young, 1993).

482 The extraction of the *consensus pattern* from a collection of *individual interrelations*
483 *patterns* does not depend on an *a priori* frame of reference, but rather is determined by the
484 maximum congruence possible between the relationships among the anatomical entities, which
485 can determine the similarity or even equivalence of two anatomical entities (Young, 1993). The
486 comparison of *individual interrelations patterns* requires the mapping of nodes across all
487 compared networks. This alignment of nodes is facilitated by the landmark entities that serve as
488 unambiguous reference points for the alignment.

489 The semantic framework allows utilizing existing algorithms for aligning directed labeled
490 graphs for the alignment of sets of *individual interrelations patterns*. These algorithms have been
491 specifically developed for the mapping of correspondences (=logical equivalences) and the
492 matching of similarities between concepts of different ontologies (see, e.g., Noy *et al.*, 2005;
493 Zhang and Bodenreider, 2007a,b; Parmentier *et al.*, 2010; Travillian *et al.*, 2010; Bertone *et al.*,
494 2013) and would have to be adapted to the needs and requirements of the here proposed semantic
495 approach. In order for these algorithms to be most efficiently applicable, the *individual*
496 *interrelations patterns* to be compared must use a common set of *Predicates*. Only if the
497 networks use the same set of *Predicates*, their nodes will be comparable and thus alignable.

498



499
 500 **Fig. 4.** Diagram of the three step procedure of the semantic approach to comparative homology with an example. **1st**
 501 **Step:** *individual interrelations patterns* of anatomical entities of three organisms (X, Y, and Z) are shown (for the
 502 sake of clarity, their **instance_of** relations to ontology terms are not shown). The nodes represent particular
 503 anatomical entities and the arcs the relations between them. **2nd & 3rd Step:** the three *individual interrelations*
 504 *patterns* have been compared and aligned and a *consensus pattern* has been build (because the **has_part** property is
 505 *transitive*, not only the microvillus group or the branched cilium has a portion of opsin as its part, but also the cell it
 506 is part of).

507 Since the primary focus lies on the identification of topographical correspondences, the
 508 algorithms should, at least initially, consider only those triple statements that are (i) connected to
 509 some anatomical entity as their *Subject* and/or *Object* and that (ii) possess *Predicates* that refer to
 510 topographical relations (e.g., parthood, connectivity) and relations of relative position,
 511 entailment, and adjacency. The algorithms should start with mapping or matching the landmark

Running Title: Semantic Approach to Comparative Homology

512 entities. In doing so, the algorithms can make use of *Predicates* that are transitive, antisymmetric
513 and reflexive binary relations (*partial ordering relations*; Bittner and Donnelly, 2007) that give
514 rise to *granular partitions* (e.g., Vogt, 2010; Vogt *et al.*, 2012b). Such granular partitions are
515 used in the hierarchical organization of terms in ontologies (=taxonomies; *taxonomic inclusion*,
516 Bittner *et al.*, 2004) and can be found also in the *individual interrelations patterns* (e.g.,
517 paronomies, which are based on parthood relations) (Fig. 3, Fig. 4). The algorithms can use both
518 ordering systems (i.e. taxonomies and paronomies) for further modifying the network-structure
519 of *individual interrelations patterns*, in order to identify instances of topographical
520 correspondence or similarity between specific nodes or even between entire branches within the
521 networks.

522 These modifications include traversing the taxonomy of ontology terms towards more
523 general terms in order to identify the taxonomic granularity level at which two given anatomical
524 entities are equivalent. For example, if a portion of rhabdomeric opsin and a portion of ciliary
525 opsin must be aligned, the respective anatomical entity would have to be represented as a portion
526 of opsin in the *consensus pattern*, because this is the least common taxonomic level at which they
527 are equivalent (Fig. 4). For this purpose, the algorithms should not only use the class-subclass
528 relationships that are explicitly stated in the ontology, but also all class-subclass relationships
529 implied in the ontology that can be inferred using a reasoner.

530 The modifications also include traversing the hierarchy of topographical granular
531 partitions (and other types of parthood-based granular partitions) in *individual interrelations*
532 *patterns* towards coarser granularity levels in order to identify the mereological or resolution-
533 dependent level of granularity at which two given anatomical entities are equivalent. For
534 example, if in one organism a cell cluster is part of a portion of the epidermis that is located at the
535 head and in another organism it is part of a portion of the epidermis that is located at some other
536 body region, but not the head, and the two cell clusters must be aligned, in the *consensus pattern*
537 their relative position would have to be represented less specific as being part of the epidermis,
538 because independent on where within the epidermis the cell cluster is located, in both cases it is
539 part of the epidermis (Fig. 4). This example also shows how a difference in position within the
540 organism of otherwise equivalent anatomical entities poses no problem for this procedure. Their
541 equivalence will be identified and their positional characteristics generalized to the degree
542 necessary to capture their level of similarity.

Running Title: Semantic Approach to Comparative Homology

543 This approach must be applied to all nodes within the network. As long as parts of the
544 network show similarities across all organisms compared, these parts will be represented in the
545 resulting *consensus pattern*, whereas their connection to the rest of the network will most likely
546 be modified to represent less specific information. This first alignment cycle of landmark entities
547 across *individual interrelations patterns* must be further differentiated in following cycles, in
548 which the algorithms attempt to align all other anatomical entities and their properties using the
549 same methods of traverse and modification.

550 If one or more anatomical entities of at least one of the organisms compared shares
551 similarities to more than one anatomical entity of another organism, the alignment of *individual*
552 *interrelations patterns* will result in more than one *consensus pattern*. This is not problematic as
553 such, because it only reflects a complex network of similarity relations that cannot be
554 unambiguously represented in a single anatomical organization scheme. Depending on the
555 purpose of the comparative study, one can either use all the resulting *consensus patterns* in
556 subsequent analyses steps or infer the *overall consensus pattern* that will necessarily contain less
557 information than the set of *consensus patterns* on which it is based. Alternatively, one could
558 conduct a comparison for all possible combinations of *individual interrelations patterns*,
559 resulting in a set of various *consensus patterns*, each of which reflects the similarity relations
560 between the particular *individual interrelations patterns* compared. This approach is
561 computationally challenging and will likely require heuristic approaches comparable to those that
562 we know from numerical tree inferences.

563 3.3 3rd Step: Comparative Homologues

564 The final step of the procedure involves the identification and definition of all those
565 anatomical entities that are referenced in a *consensus pattern* to be **comparative homologues**
566 (Fig. 1, Fig. 4).

567 By defining only those anatomical entities as comparative homologues that exhibit the
568 same interrelations pattern in a given *consensus pattern*, the semantic approach uses the same
569 basic definition for homology as Owen, without, however, referring to notions of archetypes and
570 Platonic ideals and without applying Owen's bottom-up methodology. Moreover, whereas
571 traditional phylogenetic homology assessment can only be conducted against the background of a

Running Title: Semantic Approach to Comparative Homology

572 specific positional frame of reference, whose *a priori* selection cannot be justified (Rieppel,
573 1988), the semantic approach is different, because it does not assume that comparative homology
574 is based on similarity of structure of specific anatomical entities but rather on **similarity of their**
575 **relationships** (Young, 1993).

576 When applying semantic techniques to the comparison of *individual interrelations*
577 *patterns* as well as to the identification and definition of anatomical entities as comparative
578 homologues, a more general approach can be followed than Young suggested. Instead of
579 restricting comparative homology to attributes of position and connection and thus to the
580 traditional criteria of Etienne Geoffroy Saint-Hilaire and of the Owenian non-evolutionary
581 homology concept, the here proposed semantic approach can evaluate all kinds of similarity
582 relations, based on structural, functional and/or developmental properties and relations.
583 Consequently, the resulting comparative homology concept is more general than Young's
584 concept. Moreover, due to its computer-parsability and formal structure, the semantic approach to
585 comparative homology can be applied very flexible as well, restricting similarity to positional
586 and connective attributes where required or to other attributes, as for instance specific functions
587 or developmental relations, whenever necessary. This flexibility can also be used within a
588 phylogenetic framework, where one could start with a notion of similarity that is based on
589 topographical *Predicates*, and in subsequent steps one could test the effect that the inclusion of
590 additional *Predicates* has on the resolution of the *consensus pattern* and thus on the recognition
591 of putative phylogenetic homologues. In this way, one can evaluate the influence of function-
592 related properties and dispositions, intrinsic structural qualities of anatomical entities and
593 developmental relations on the assessment of comparative homology and, indirectly, also on the
594 assessment of phylogenetic homology (see discussion below).

595 This flexibility can be achieved by taking advantage of the fact that ontology *Predicates*,
596 which are used in the *individual interrelations patterns* to specify the different relations between
597 anatomical entities and their properties, are already classified into different categories. Moreover,
598 each *Predicate* does not only possess a textual definition, but oftentimes also specific property
599 characteristics and a specific domain and range restriction (see above). The application of a
600 *domain granularity framework* that enables the specification of pre-defined **granularity**
601 **perspectives** that can be applied to any *individual interrelations pattern* for filtering out specific
602 triple statements based on their *Predicates* and their entity types can facilitate in practically

Running Title: Semantic Approach to Comparative Homology

603 managing this flexibility by, for instance, reducing any given graph to triple statements about
604 those entity types and relations that are of interest for a given purpose (e.g. Vogt, 2010; Vogt *et*
605 *al.*, 2012a,b). Moreover, by utilizing granular partitions of ontologies and of *individual*
606 *interrelations patterns*, algorithms can always traverse not only the explicit hierarchies but also
607 all relevant implicit hierarchies that are present in these representations when attempting to find
608 logically equivalent terms for mapping or sufficiently similar terms for matching between
609 *individual interrelations patterns*. Thus, if for a given node of a specific *individual interrelations*
610 *pattern* no comparable node can be found in another *individual interrelations pattern*, the
611 algorithms would search again on a more general level of taxonomic granularity or on a coarser
612 level of parthood granularity based on the granularity perspectives defined in the granularity
613 framework. In other words, using an adequate domain granularity framework for morphology can
614 substantially facilitate the guided traversing of various hierarchical systems and the filtering of
615 irrelevant *Predicates* of the network structure of *individual interrelations patterns*.

616
617 As a side note: the here discussed procedure can also be used for identifying and
618 classifying anatomical entities. By aligning an *individual interrelations pattern* with *semantic*
619 *instance anatomies* that have been validated by experts and therefore comprise fully classified
620 anatomical entities, all anatomical entities within the *individual interrelations pattern* can be
621 classified in reference to their comparative homologues in the *semantic instance anatomies*. This
622 approach to the classification of individuals does not require that the corresponding classes of
623 anatomical entity types of the relevant anatomy ontologies possess formalized property
624 descriptions, but utilizes the information relevant to classification that is implicitly contained in
625 the *semantic instance anatomies* that are used for comparison.

626 **4 The General Role of Comparative Homology in Biology**

627 4.1 The Epistemic Relation between Comparative Homology and Phylogenetic 628 Homology

629 As already discussed further above, primary homology covers two notions of
630 phylogenetic homology, meaning taxic and transformational homology. With the comparative

631 homology concept proposed above one can individuate units of comparison, some of which can
632 be interpreted in a phylogenetic framework as **initial hypotheses of transformational homology**
633 and others as **initial hypothesis of taxic homology** (=putative synapomorphy). Establishing a
634 relation of comparability that is based on spatio-structural similarity between particular
635 anatomical entities of different OTUs sets the frame for a phylogenetic character. Observable
636 differences between these comparable anatomical entities can be interpreted to have been caused
637 by transformation events in a common ancestor, at least if these difference refer to inheritable
638 properties and relations.

639 Putative transformational homologues are identified as those comparable homologues that
640 are represented in the consensus pattern either by a taxonomically more general or mereologically
641 coarser entity than its corresponding entities from the underlying individual interrelations
642 patterns (e.g., see Fig.4: the *portion of rhabdomeric opsin* in *organisms X and Y* and the *portion*
643 *of ciliary opsin* in *organism Z* are putatively transformational homologues as *portion of opsin*).
644 Any unit of comparison that shares properties and/or relations across only some of the studied
645 OTUs gives rise to a hypothesis of taxic homology, because according to the phylogenetic
646 homology concept, equivalence in a single property can be sufficient for assuming phylogenetic
647 homology, as long as no other evidence supports the opposite (e.g. see Fig.4: the *portion of*
648 *rhabdomeric opsin* in *organisms X and Y* are taxic homologues).

649 As a consequence, one can conclude that because comparative homologues are based on
650 similarity, any comparative homologue is a putative phylogenetic homologue and **comparative**
651 **homology is indicative of primary homology**. Further analyses and considerations are of course
652 required for primary homology assessment, especially when considering that not every data unit
653 (=triple statement of a *semantic instance anatomy*) and not every anatomical unit (=anatomical
654 entity) necessarily also represents an evidence unit (=phylogenetic homologue) for reconstructing
655 phylogeny; primary homology usually involves several data units that often refer to more than
656 one anatomical unit (see *delimitation problem* in Vogt *et al.*, 2010).

657 That homologues identified through similarity may reflect historic identity and thus
658 phylogenetic homology has been discussed before (e.g., Wagner, 1980; Cracraft, 1981; Brady,
659 1985; Wilkinson, 1988), and the philosophical and methodological possibility of such a transition
660 already has been demonstrated (Rieppel, 1988; Young, 1993). The comparative homology
661 concept provides the conceptual basis necessary for the application of the recognition criteria of

662 phylogenetic homology independent of an *a priori* assumption of the existence of a stable
663 positional pattern. The comparative homology concept thus adds the missing link to the logical
664 chain for the method of assessing phylogenetic homology and closes the methodological gap
665 between the ontological definition and the epistemological recognition criteria of phylogenetic
666 homology. It thus provides the conceptual answer to the criticism that phylogenetic homology
667 would involve a circular methodological argument: we can assess phylogenetic homology on the
668 basis of similarity and without having to have prior knowledge about its underlying phylogeny.

669 4.2 Comparative Homology and Morphological Terminology

670 Any use of general terms in morphology and any attempt to compare and identify
671 similarities and equivalences between particular anatomical entities is linked to the idea of
672 identifiable and re-identifiable structural kinds (Mahner and Bunge, 1997; Rieppel, 2007) and
673 thus to classes or types of anatomical entities. In reverse one can paraphrase: '*no similarity*' is
674 tantamount to '*no comparative method*' in biology and thus to no justification for the applicability
675 of general terms in morphology. The comparative homology concept provides a conceptually and
676 methodologically consistent method for identifying and documenting instances of similarity
677 between the parts of particular organisms. As a consequence, it represents an essential
678 methodological component for generalizing over individual organisms and identifying and
679 demarcating structural kinds in morphology. In a nutshell one could say that identifying
680 comparative homologues in morphology equals identifying anatomical structural kinds.

681 As a consequence, the process of identifying anatomical structural kinds could be
682 reasonably automated by the three-step procedure discussed above. The consensus pattern
683 retrieved from the second step not only specifies comparative homologues, but also their
684 consensus properties, which represent putative defining properties for their corresponding
685 anatomical structural kinds. Morphologists would have to provide adequate general terms for
686 reference to these structural kinds and evaluate which consensus properties are best suited for
687 defining the kind.

688 Contrary to the phylogenetic homology concept, the here proposed comparative
689 homology concept focuses on the factual aspects of similarity and is completely agnostic about
690 similarity's causal underpinning. It is therefore used in descriptive contexts, whereas the

691 phylogenetic homology concept is used in explanatory contexts. However, the comparative
692 homology concept does not only take in an essential function for describing anatomical entities,
693 as it enables the identification of types of anatomical entities, but also provides the conceptual
694 foundation for comparative biology. Therefore, comparative homologues are entities that are
695 important to be named, described and characterized, because they are of relevance for biology in
696 general. Respective general terms should belong to any comprehensive anatomical terminology.

697 **5 Conclusion**

698 One reason for the lack of theoretical analyses and conceptual understanding of the
699 morphological character analysis might be found in the perception that the assessment of
700 phylogenetic homology presupposes a specific frame of reference, for whose selection no *a priori*
701 justification exists (see in Rieppel, 1988). Young (1993) contrasted this point of view with an
702 approach that questions the assumption that homology is exclusively based on similarity of
703 intrinsic properties, and he instead suggested that it is based on a combination of observable
704 intrinsic qualities and similarity of relationships.

705 Instead of defining homology in reference to a common historical identity, as the
706 phylogenetic homology concept does, the definition of Young's comparative homology concept is
707 based on the notion of a shared set of specific properties and relations, an alternative notion of
708 homology that has been discussed before (e.g., Ghiselin, 1969, 1987; Cracraft, 1981; Sattler,
709 1984; Aboitiz, 1987, 1988). Above, I have described how this comparative notion of homology
710 can be implemented in a semantic approach and how comparative homologues can be recognized
711 and individuated on the basis of semantic representations of *individual interrelations patterns*,
712 without the requirement of an *a priori* assumption of a positional frame of reference. This
713 approach uses the interrelations between anatomical entities themselves as reference in addition
714 to the identification of landmark entities and compares them across different organisms to
715 identify similarities in their pattern that are documented as *consensus patterns*. These *consensus*
716 *patterns* designate which anatomical entities can be compared across all the organisms that have
717 been investigated, and these comparable entities are considered to be comparative homologues.

718 With the theoretical and methodological framework provided by semantics and
719 ontologies, Young's approach for the recognition of comparative homologues can be turned into a

Running Title: Semantic Approach to Comparative Homology

720 straight forward procedure that can be applied to real data. Its application is transparent and can
721 be implemented using adequate algorithms. As a consequence, comparative research could
722 become more reproducible, allowing the procedures and methods of comparative studies and of
723 the character analysis step of phylogenetics to become communicable and subjected to criticism
724 and corrections. In other words, by applying a formal semantic framework, the analytical
725 accessibility of comparative studies in general and the comparative part of morphological
726 phylogenetic studies in particular (the *alignment* of morphology; Stevens, 2000) will be increased
727 significantly, and comparative studies will become less subjective and more open to critical
728 analysis. The formal semantic framework and the semantic approach to comparative homology
729 thus contribute the basis for developing the unified theory of character construction that
730 Wikinson (1995) claimed that biology needs. It also contributes to increasing the objectivity of
731 propositions of primary homology (cf. Jardine, 1969) and, although the recognition of similarities
732 might still stand “*as an essentially imprecise and subjective process*” (de Pinna, 1991, p. 377),
733 the semantic framework contributes the much needed semantic transparency and computer-
734 parsability for documenting, communicating and analyzing similarity propositions, all of which
735 represent essential preconditions for subjecting similarity propositions to criticism and to
736 corrections and with them the comparative method and the character analysis step.

737 Against this background and evaluated in this context, the semantic approach to
738 comparative homology represents an important theoretical and methodological contribution to
739 biology in general and morphology in particular, because it provides a highly formalized
740 methodological basis for individuating anatomical structural kinds, which at their turn provide
741 the basic reference system for all supra-molecular biological entities, providing the **descriptive**
742 **framework for the supra-molecular domain of biology**. It thus contributes to one of the core
743 competencies of morphology as a covering discipline (Grobe and Vogt, 2014). The semantic
744 approach to comparative homology is also essential for comparative anatomical studies and thus
745 provides a **key method for morphology**. Moreover, it bridges the methodological gap of
746 phylogenetic homology assessment in morphology and thus contributes an important
747 methodological link for morphological character analysis in phylogenetics.

748 By replacing phylogenetic homology with comparative homology as the central concept
749 in morphology, morphology can free itself from the iron grip of having to take phylogenetic
750 homology as its paradigm theoretical and methodological framework. After all, morphology is

Running Title: Semantic Approach to Comparative Homology

751 more than a basis for taxonomy and phylogenetics—nobody would restrict molecular biology to
 752 phylogenetics and taxonomic research! Comparative homology is the most basic methodological
 753 concept in morphology, offering interesting starting points for various research programs. Like
 754 molecular biology, morphology is primarily a way to produce data and thus to describe a certain
 755 aspect of reality. Its explanatory impact is diverse and not restricted to only a couple of research
 756 areas. The semantic approach to comparative homology can help re-opening morphology to all
 757 kinds of biological disciplines and various research topics.

758 **Acknowledgements**

759 I thank Thomas Bartolomaeus, Peter Grobe and Chris Mungall for helping with the semantic
 760 description of a protonephridium and Björn Quast and Christoph Bleidorn for commenting on an
 761 earlier draft of this MS. It goes without saying, however, that I am solely responsible for all the
 762 arguments and statements in this paper. This study was supported by the German Research
 763 Foundation DFG (VO 1244/7-1). I am also grateful to the taxpayers of Germany.

764 **References**

- 765 Aboitiz F. 1987. “Homology” in anatomy and molecular biology. *Cell* 51, 515–516.
- 766 Aboitiz F. 1988. Homology: A comparative or a historical concept? *Acta Biotheoretica* 37, 27–
 767 29.
- 768 Agnarsson I., Coddington J.A. 2008. Quantitative tests of primary homology. *Cladistics* 24, 51–
 769 61.
- 770 Balhoff J.P., Dahdul W.M., Kothari C.R., Lapp H., Lundberg J.G., Mabee P., Midford P.E.,
 771 Westerfield M., Vision T.J. 2010. Phenex: ontological annotation of phenotypic
 772 diversity. *PLoS One* 5, e10500.
- 773 Balhoff J.P., Dececchi T.A., Mabee P.M., Lapp H. 2014. Presence-absence reasoning for
 774 evolutionary phenotypes. In: International Society for Computational Biology, pp. 1–4.
- 775 Balhoff J.P., Mikó I., Yoder M.J., Mullins P.L., Deans A.R. 2013. A semantic model for species
 776 description, applied to the ensign wasps (Hymenoptera: Evaniidae) of New Caledonia.
 777 *Systematic Biology* 62, 639–659.

Running Title: Semantic Approach to Comparative Homology

- 778 Bartolomaeus T. 1998. Head Kidneys in Hatchlings of *Scoloplos armiger* (Annelida: Orbiniida):
779 Implications for the Occurrence of Protonephridia in Lecithotrophic Larvae. *Journal of*
780 *the Marine Biological Association of the United Kingdom* 78, 183–192.
- 781 Beck T., Morgan H., Blake A., Wells S., Hancock J.M., Mallon A.-M. 2009. Practical application
782 of ontologies to annotate and analyse large scale raw mouse phenotype data. *BMC*
783 *Bioinformatics* 9, S2.
- 784 Belon P. 1555. *L’histoire de la nature des oyseaux*. Paris: Guillaume Cavellet.
- 785 Bertone M.A., Mikó I., Yoder M.J., Seltmann K.C., Balhoff J.P., Deans A.R. 2013. Matching
786 arthropod anatomy ontologies to the Hymenoptera Anatomy Ontology: results from a
787 manual alignment. *Database* 2013, 1–9.
- 788 Bittner T., Donnelly M. 2007. Logical properties of foundational relations in bio-ontologies.
789 *Artificial Intelligence in Medicine* 39, 197–216.
- 790 Bittner T., Smith B., Donnelly M. 2004. Individuals, universals, collections: on the foundational
791 relations of ontology. In: Varzi A., Vieu L., (Eds), Proceedings of the International
792 Conference on Formal Ontology in Information Systems. IOS Press, Amsterdam, pp.
793 37–48.
- 794 Bock G.R., Cardew G. 1999. *Homology*. Novartis F. Chichester: John Wiley & Sons.
- 795 Bock W.J. 1974. Philosophical foundations of classical evolutionary classification. *Systematic*
796 *Zoology* 22, 375–392.
- 797 Bock W.J. 1977. Foundations and methods of evolutionary classification. In: Hecht M.K., Goody
798 P.C., Hecht B.M., (Eds.), Major patterns in vertebrate evolution. Plenum Press, New
799 York, pp. 851–895.
- 800 Boyden A. 1947. Homology and analogy. *American Midland Naturalist* 37, 648–669.
- 801 Brady R.H. 1985. On the independence of systematics. *Cladistics* 1, 113–126.
- 802 Brigandt I., Griffiths P.E. 2007. The importance of homology for biology and philosophy.
803 *Biology & Philosophy* 22, 633–641.
- 804 Brower A.V.Z. 2014. Transformational and taxic homology revisited. *Cladistics* 31(2), 197–201.
- 805 Brower A.V.Z., Schawaroch, V. 1996. Three steps of homology assessment. *Cladistics* 12, 265–
806 272.
- 807 Cracraft J. 1967. Comments on homology and analogy. *Systematic Zoology* 16, 355–359.

Running Title: Semantic Approach to Comparative Homology

- 808 Cracraft J. 1981. The use of functional and adaptive criteria in phylogenetic systematics.
809 *American Zoologist* 21, 21–36.
- 810 Dahdul W.M., Balhoff J.P., Engeman J., Grande T., Hilton E.J., Kothari C., Lapp H., Lundberg
811 J.G., Midford P.E., Vision T.J., Westerfield M., Mabee P.M. 2010. Evolutionary
812 Characters, Phenotypes and Ontologies: Curating Data from the Systematic Biology
813 Literature. *PLoS One* 5, e10708.
- 814 de Beer G.R. 1971. *Homology, an Unsolved Problem*. Glasgow: Oxford University Press.
- 815 de Pinna M.G.G. 1991. Concepts and tests of homology in the cladistic paradigm. *Cladistics* 7,
816 367–394.
- 817 de Queiroz K. 1985. The ontogenetic method for determining character polarity and its relevance
818 to phylogenetic systematics. *Systematic Zoology* 34, 280–299.
- 819 Deans A.R., Yoder M.J., Balhoff J.P. 2012. Time to change how we describe biodiversity. *Trends*
820 *in Ecology & Evolution* 27, 78–84.
- 821 Eckhardt T. 1964. Das Homologieproblem und Fälle strittiger Homologien. *Phytomorphology* 14,
822 79–92.
- 823 Freudenstein J.V. 2005. Characters, states, and homology. *Systematic Biology* 54, 965–973.
- 824 Gans C. 1985. Differences and similarities: comparative methods in mastication. *American*
825 *Zoologist* 25, 291–301.
- 826 Geoffroy Saint-Hilaire E. 1830. *Principes de Philosophie Zoologique, Discutés en Mars 1830, au*
827 *sein de L'Académie Royale des Sciences, par M. Geoffroy Saint-Hilaire*. Paris: Pichon et
828 Didier.
- 829 Ghiselin M.T. 1966. An application of the theory of definitions to systematic principles.
830 *Systematic Zoology* 15, 127–130.
- 831 Ghiselin M.T. 1969. The distinction between similarity and homology. *Systematic Zoology* 18,
832 148–149.
- 833 Ghiselin M.T. 1976. The nomenclature of correspondence: a new look at “homology” and
834 “analogy.” In: Hodos W., Jerison H., (Eds.), *Evolution, Brain, and Behavior: Persistent*
835 *Problems*. John Wiley & Sons, New York, pp. 129–142.
- 836 Ghiselin M.T. 1987. Species concepts, individuality, and objectivity. *Biology & Philosophy* 2,
837 122–143.

Running Title: Semantic Approach to Comparative Homology

- 838 Grobe P., Vogt L. 2014. Documenting Morphology: Morph·D·Base. In: Wägele J.W.,
839 Bartolomaeus T., (Eds.), *Deep Metazoan Phylogeny: The Backbone of the Tree of Life -*
840 *New Insights from Analyses of Molecules, Morphology, and Theory of Data Analysis.*
841 de Gruyter, Berlin/Boston, pp. 475–503.
- 842 Grumbling G., Strelets V. 2006. FlyBase: Anatomical data, images and queries. *Nucleic Acids*
843 *Research* 34, D484–D488.
- 844 Hall B.K. 1994. *Homology - The hierarchical basis of comparative biology.* New York, London:
845 Academic Press.
- 846 Hall B.K. 2012. Homology, homoplasy, novelty, and behavior. *Developmental Psychobiology* 55,
847 4–12.
- 848 Hauser D.L., Presch W. 1989. The effect of ordered characters on phylogeny reconstruction.
849 *Cladistics* 7, 243–265.
- 850 Hawkins J.A., Hughes C.E., Scotland R.W. 1997. Primary homology assessment, characters and
851 character states. *Cladistics* 13, 275–283.
- 852 Hennig W. 1950. *Grundzüge einer Theorie der Phylogenetischen Systematik.* Berlin:
853 Zentralverlag.
- 854 Jardine N. 1969. The observational and theoretical components of homology: a study based on
855 the morphology of the dermal skull-roofs of rhipidistian fishes. *Biological Journal of the*
856 *Linnean Society* 1, 327–361.
- 857 Jardine N., Jardine C.J. 1969. Is there a concept of homology common to several sciences.
858 *Classification Society Bulletin* 2, 12–18.
- 859 Jenner R.A. 2002. Boolean logic and character state identity: pitfalls of character coding in
860 metazoan cladistics. *Contributions to Zoology* 71, 67–91.
- 861 Jenner R.A. 2004. The scientific status of metazoan cladistics: why current research practice must
862 change. *Zoologica Scripta* 33, 293–310.
- 863 Kluge A.G., Strauss R.E. 1985. Ontogeny and systematics. *Annual Review of Ecology and*
864 *Systematics* 16, 247–268.
- 865 Keet C.M. 2006. A taxonomy of types of granularity. IEEE Conference in Granular Computing
866 (GrC2006), 10–12 May 2006, Atlanta, USA. Available at
867 http://www.meteck.org/files/GrCGranTypes_CMK.pdf (accessed 04 July 2015).

- 868 Keet C.M. 2008. *A Formal Theory of Granularity - Toward enhancing biological and applied life*
869 *sciences information system with granularity*. Bozen: Free University of Bozen -
870 Bolzano. Available at
871 http://www.meteck.org/files/AFormalTheoryOfGranularity_CMK08.pdf (accessed 04
872 July 2015).
- 873 Lankester E.R. 1870. On the use of the term homology in modern zoology, and the distinction
874 between homogenetic and homoplastic agreements. *Ann Mag Nat Hist* 6, 34–43.
- 875 Lee D.-C., Bryant H.N. 1999. A reconsideration of the coding of inapplicable characters:
876 Assumptions and problems. *Cladistics* 15, 373–378.
- 877 Linnaeus C. 1735. *Systema naturae sive Regna tria naturae. Systematicae proposita per Classes,*
878 *Ordines, Genera, & Species*. Leiden: Theodor Haak.
- 879 Linnaeus C. 1751. *Philosophia botanica*. Stockholm: Kiesewetter.
- 880 Maddison W.P. 1993. Missing data versus missing characters in phylogenetic analysis.
881 *Systematic Biology* 42, 576-581.
- 882 Mahner M., Bunge M. 1997. *Foundations of biophilosophy*. Berlin, Heidelberg: Springer.
- 883 Mikó I., Friedrich F., Yoder M.J., Hines H.M., Deitz L.L., Bertone M.A., Seltmann K.C.,
884 Wallace M.S., Deans A.R. 2012. On Dorsal Prothoracic Appendages in Treehoppers
885 (Hemiptera: Membracidae) and the Nature of Morphological Evidence. *PLoS One* 7,
886 e30137
- 887 Minelli A., Fusco G. 2013. Homology. In: Kampourakis K., (Ed.), *The Philosophy of Biology: A*
888 *Companion for Educators. History, Philosophy and Theory of the Life Sciences* 1.
889 *History, Philosophy and Theory of the Life Sciences*. Springer Netherlands, Dordrecht.
- 890 Mungall C.J., Gkoutos G.V., Smith C.L., Haendel M.A., Lewis S.E., Ashburner M. 2010.
891 Integrating phenotype ontologies across multiple species. *Genome Biology* 11, R2.
- 892 Mungall C.J., Gkoutos G., Washington N., Lewis S. 2007. Representing Phenotypes in OWL.
893 Proceedings of the OWLED 2007 Workshop on OWL: Experience and Directions: June
894 6-7, 2007; Innsbruck, Austria 2007.
- 895 Neff N.A. 1986. A rational basis for a priori character weighting. *Systematic Zoology* 35, 110–
896 123.
- 897 Noy N.F., Doan A., Halevy A.Y. 2005. Semantic Integration. *AI Magazine* 26, 7–10

Running Title: Semantic Approach to Comparative Homology

- 898 Owen R. 1843. *Lectures on the comparative anatomy and physiology of the invertebrate animals:*
899 *delivered at the Royal college of surgeons.* London: Longman, Brown, Greens &
900 Longmans.
- 901 Owen R. 1848. *On the archetype and homologies of the vertebrate skeleton.* London: John van
902 Voorst.
- 903 Owen R. 1849. *On the nature of limbs.* London: John van Voorst.
- 904 Panchen A.L. 1994. Richard Owen and the concept of homology. In: Hall B.K., (Ed.), *Homology*
905 *- The hierarchical basis for comparative biology.* Academic Press, New York, London,
906 pp. 22–62.
- 907 Parmentier G., Bastian F.B., Robinson-Rechavi M. 2010. Homolonto: generating homology
908 relationships by pairwise alignment of ontologies and application to vertebrate anatomy.
909 *Bioinformatics* 26, 1766–1771.
- 910 Patterson C. 1982. Morphological characters and homology. In: Joysey K.A., Friday A.E., (Eds.),
911 *Problems of Phylogenetic Reconstruction.* Academic Press, London, pp. 21–74.
- 912 Patterson C. 1988. Homology in classical and molecular biology. *Molecular Biology and*
913 *Evolution* 5, 603–625.
- 914 Piaget J. 1971. *Biology and knowledge: an essay on the relations between organic regulations*
915 *and cognitive processes.* Chicago: University of Chicago Press.
- 916 Platnick N.I., Griswold C.E., Coddington J.A. 1991. On missing entries in cladistic analysis.
917 *Cladistics* 7, 337–343.
- 918 Pleijel F. 1995. On character coding for phylogeny reconstruction. *Cladistics* 11, 309–315.
- 919 Pogue M.G., Mickevich, M.F. 1990. Character definitions and character state delineation: The
920 bête noire of phylogenetic inference. *Cladistics* 6, 319–361.
- 921 Ramírez M.J. 2007. Homology as a parsimony problem: a dynamic homology approach for
922 morphological data. *Cladistics* 23, 588–612.
- 923 Remane A. 1952. Die Grundlagen des natürlichen Systems, der vergleichenden Anatomie und der
924 Phylogenetik. Theoretische Morphologie und Systematik I. Geest & Portig, Leipzig.
- 925 Richter S. 2005. Homologies in phylogenetic analyses—concept and tests. *Theory in Biosciences*
926 124, 105–120.
- 927 Riedl R. 1979. *Order in Living Organisms.* Chichester: Wiley.

Running Title: Semantic Approach to Comparative Homology

- 928 Rieppel O. 1980. Homology, a deductive concept? *Zeitschrift für zoologische Systematik und*
929 *Evolutionsforschung* 18, 315–319.
- 930 Rieppel O. 1988. *Fundamentals of Comparative Biology*. Basel: Birkhäuser Verlag.
- 931 Rieppel O. 1993. The conceptual relationship of ontogeny, phylogeny, and classification - The
932 taxic approach. *Evolutionary Biology* 27, 1–32.
- 933 Rieppel O. 1996. Testing homology by congruence: the pectoral girdle of turtles. *Proceedings of*
934 *the Royal Society of London Series B Biological Sciences* 263, 1395–1398.
- 935 Rieppel O. 2005. Modules, kinds, and homology. *Journal of Experimental Zoology Part B:*
936 *Molecular and Developmental Evolution* 304B, 18–27.
- 937 Rieppel O. 2007. The performance of morphological characters in broad-scale phylogenetic
938 analyses. *Biological Journal of the Linnean Society* 92, 297–308.
- 939 Rieppel O., Kearney M. 2002. Similarity. *Biological Journal of the Linnean Society* 75, 59–82.
- 940 Rosse C., Mejino Jr. J.L. V 2007. The Foundational Model of Anatomy Ontology. In: Burger A.,
941 Davidson D., Baldock R., (Eds.), *Anatomy Ontologies for Bioinformatics: Principles*
942 *and Practice*. Springer, New York, pp. 63–117.
- 943 Roth V.L. 1988. The biological basis of homology. In: Humphries C.J., (Ed.), *Ontogeny and*
944 *Systematics*. Columbia University Press, New York, pp. 1–26.
- 945 Russell E.S. 1916. *Form and Function - A Contribution to the History of Animal Morphology*.
946 London: John Murray.
- 947 Sattler R. 1984. Homology - a continuing challenge. *Systematic Botany* 9, 382–394.
- 948 Sereno P.C. 2007. Logical basis for morphological characters in phylogenetics. *Cladistics* 23,
949 565–587.
- 950 Smith B. 1994. Fiat objects. In: Guarino L.V., Pribbenow S., (Eds.), *Parts and Wholes:*
951 *Conceptual Part-Whole Relations and Formal Mereology*, 11th European Conference on
952 *Artificial Intelligence*. European Coordinating Committee for Artificial Intelligence,
953 Amsterdam, pp. 15–23.
- 954 Smith B. 1995. On Drawing Lines on a Map. In: Frank A.U., Kuhn W., Mark D.M., (Eds.),
955 *Spatial Information Theory: Proceedings in COSIT '95*. Springer,
956 Berlin/Heidelberg/Vienna/New York/London/Tokyo, pp. 475–484.
- 957 Smith B. 2001. Fiat Objects. *Topoi* 20, 131–148.

Running Title: Semantic Approach to Comparative Homology

- 958 Smith B., Ashburner M., Rosse C., Bard J., Bug W., Ceuster W., Goldberg L.J., Eilbeck K.,
 959 Ireland A., Mungall C.J., Leontis N., Rocca-Serra P., Ruttenger A., Sansone S.-A.,
 960 Scheuermann R.H., Shah N., Whetzel P.L., Lewis S. 2007. The OBO Foundry:
 961 coordinated evolution of ontologies to support biomedical data integration. *Nature*
 962 *Biotechnology* 25, 1251–1255.
- 963 Smith B., Ceusters W., Klagges B., Köhler J., Kumar A., Lomax J., Mungall C., Neuhaus F.,
 964 Rector A.L., Rosse C. 2005. Relations in biomedical ontologies. *Genome Biology* 6,
 965 R46.
- 966 Smith B., Varzi A.C. 1997. Fiat and Bona Fide Boundaries: Towards an Ontology of Spatially
 967 Extended Objects. *Spatial Information Theory: A Theoretical Basis for GIS*. Springer,
 968 Berlin, Heidelberg, pp. 103–119.
- 969 Smith B., Varzi A.C. 2000. Fiat and Bona Fide Boundaries. *Philosophy and Phenomenological*
 970 *Research* 60, 401–420.
- 971 Sokal R.R., Sneath P.H.A. 1973. *Numerical taxonomy*. San Francisco: Freeman.
- 972 Sowa J.F. 1984. *Conceptual structures: information processing in mind and machine*. Reading
 973 (Massachusetts): Addison-Wesley Publishing Company, Inc.
- 974 Sowa J.F. 2008. Conceptual Graphs. In: van Harmelen F., Lifschitz V., Porter B., (Eds.),
 975 *Handbook of Knowledge Representation*. Elsevier. pp. 213–237.
- 976 Spencer H. 1867. *Principles in Biology* (Vol. 2). New York: Appleton.
- 977 Staton J.L. 2011. Homology in Character Evolution. *eLS*.
 978 doi:10.1002/9780470015902.a0001776.pub2.
- 979 Stevens P.F. 1984. Homology and phylogeny: morphology and systematics. *Systematic Botany*
 980 9(4), 395–409.
- 981 Stevens P.F. 2000. On characters and character states: Do overlapping and non-overlapping
 982 variation, morphology and molecules, all yield data the same value? In: Scotland R.,
 983 Pennington R.T., (Eds.), *Homology and systematics: Coding characters for phylogenetic*
 984 *analysis*. Taylor & Francis, London, pp. 81–105.
- 985 Strickland H.E. 1840a. On the true method of discovering the natural system in zoology and
 986 botany. *Ann Mag Nat Hist* 6, 184–186.
- 987 Strickland H.E. 1840b. Observations upon the affinities and analogies of organized beings. *Ann.*
 988 *Mag Nat Hist* 4, 219–226.

Running Title: Semantic Approach to Comparative Homology

- 989 Strickland H.E. 1846. On the structural relations of organized beings. *Philosophical Magazine*
990 *and Journal of Science* 3, 354–364.
- 991 Strong E.E., Lipscomb D. 1999. Character coding and inapplicable data. *Cladistics* 15, 363–371.
- 992 Travillian R.S., Adamusiak T., Burdett T., Gruenberger M., Hancock J., Mallon A.-M., Malone
993 J., Schofield P., Parkinson H. 2010. Anatomy Ontologies and Potential Users: Bridging
994 the Gap. In: Herre H., Hoehndorf R., Kelso J., Schulz S., (Eds.), IMISE-Report 2/2010:
995 OBML 2010 Workshop Proceedings, Mannheim, September 9-10, 2010. Mannheim:
996 Universität Leipzig, IMISE-Reports, Prof. Dr. Markus Löffler, pp. 47–50.
- 997 Vogt L. 2002. Testing and weighting characters. *Organisms, Diversity and Evolution* 2, 319–333.
- 998 Vogt L. 2004. Signs and phylogeny: a semiotic approach to systematics. *Semiotica* 149, 125–159.
- 999 Vogt L. 2008a. The unfalsifiability of cladograms and its consequences. *Cladistics* 24, 62–73.
- 1000 Vogt L. 2008b. Learning from Linnaeus: towards developing the foundation for a general
1001 structure concept for morphology. *Zootaxa* 1950, 123–152.
- 1002 Vogt L. 2010. Spatio-structural granularity of biological material entities. *BMC Bioinformatics*
1003 11, 289.
- 1004 Vogt L., Bartolomaeus T., Giribet G. 2010. The linguistic problem of morphology: structure
1005 versus homology and the standardization of morphological data. *Cladistics* 26, 301–325.
- 1006 Vogt L., Grobe P., Quast B., Bartolomaeus T. 2011. Top-Level Categories of Constitutively
1007 Organized Material Entities—Suggestions for a Formal Top-Level Ontology. *PLoS One*
1008 6(4), e18794.
- 1009 Vogt L., Grobe P., Quast B., Bartolomaeus T. 2012a. Accommodating Ontologies to Biological
1010 Reality—Top-Level Categories of Cumulative-Constitutively Organized Material
1011 Entities. *PLoS One* 7, e30004.
- 1012 Vogt L., Grobe P., Quast B., Bartolomaeus T. 2012b. Fiat or Bona Fide Boundary—A Matter of
1013 Granular Perspective. *PLoS One* 7, e48603.
- 1014 Wagner G.P. 1989. The origin of morphological characters and the biological basis of homology.
1015 *Evolution* (N Y) 43, 1157–1171.
- 1016 Wagner W.H. 1980. Origin and philosophy of the groundplan-divergence method of cladistics.
1017 *Systematic Botany* 5, 173–193.

Running Title: Semantic Approach to Comparative Homology

- 1018 Washington N.L., Haendel M.A., Mungall C.J., Ashburner M., Westerfield M., Lewis S.E. 2009.
1019 Linking human diseases to animal models using ontology-based phenotype annotation.
1020 *PLoS Biol* 7, e1000247.
- 1021 Whewell W. 1847. *The Philosophy of the Inductive Sciences, Founded upon their History - Vol.*
1022 *I.* 2nd ed. John W. West Strand: Parker.
- 1023 Wiley E.O. 1975. Karl R . Popper , Systematics , and Classification : A Reply to Walter Bock and
1024 Other Evolutionary Taxonomists. *Systematic Zoology* 24, 233–243.
- 1025 Wilkinson M. 1988. Evolutionary and classical concepts of homology: a reply to Aboitiz. *Acta*
1026 *Biotheoretica* 37, 315–319.
- 1027 Wilkinson M. 1995. A comparison of two methods of character construction. *Cladistics* 11, 297–
1028 308.
- 1029 Woodger J.H. 1937. *The Axiomatic Method in Biology*. Cambridge University Press, London.
- 1030 Young B.A. 1993. On the necessity of an archetypal concept in morphology: With special
1031 reference to the concepts of “structure” and “homology.” *Biology & Philosophy* 8, 225–
1032 248.
- 1033 Zhang S., Bodenreider O. 2007a. Lessons learned from cross-validating alignments between large
1034 anatomical ontologies. *Studies in health technology and informatics* 129, 822–826.
- 1035 Zhang S., Bodenreider O. 2007b. Experience in Aligning Anatomical Ontologies. *International*
1036 *Journal on Semantic Web and Information Systems* 3, 1–26.