Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2021 doi:10.20944/preprints202102.0623.v1

Phylogenetic signal in characters from Aristotle's History of Animals

Michel Laurin [1*] and Marcel Humar [2]

1 Centre de Recherches en Paléontologie – Paris (CR2P), UMR 7207, CNRS/MNHN/SU/EPHE, Muséum national d'Histoire naturelle, Paris, France.

2 Humboldt-Universität zu Berlin, Institut für Klassische Philologie, Berlin, Germany.

*** Corresponding author. E-mail: michel.laurin@mnhn.fr**

Abstract: The great Greek philosopher Aristotle (384–322 BCE) is almost unanimously acclaimed as the founder of zoology. There is a consensus that he was interested in attributes of animals, but whether or not he tried to develop a zoological taxonomy remains controversial. Fürst von Lieven and Humar compiled a data matrix and showed, through a parsimony analysis published in 2008, that these data produced a hierarchy that matched several taxa recognized by Aristotle. However, their analysis leaves some questions unanswered because random data can sometimes yield fairly resolved trees. In this study, we update the scores of many cells and add four new characters to the data matrix (147 taxa scored for 161 characters) and quote passages from Aristotle's *Historia animalium* to justify these changes. We confirm the presence of a phylogenetic signal in these data through a test using skewness in length distribution of a million random trees, which shows that many of the characters discussed by Aristotle were systematically relevant. Our parsimony analyses on the updated matrix recover far more trees than reported by Fürst von Lieven and Humar, but their consensus includes many taxa that Aristotle recognized and apparently named for the first time, such as *selachē* (selachians) and *dithyra* (Bivalvia). This study suggests that even though taxonomy was clearly not Aristotle's chief interest in *Historia animalium*, it was probably among his secondary interests. These results may pave the way for further taxonomic studies in Aristotle's zoological writings in general. Despite being almost peripheral to Aristotle's writings, his taxonomic contributions are clearly major achievements.

Keywords: history of biology; history of zoology; taxonomy; biological nomenclature; metazoans

Introduction

The great Greek philosopher Aristotle (384–322 BCE) is almost unanimously acclaimed as the founder of zoology. Mayr (1982: 149) even stated "The history of taxonomy starts with Aristotle (384–322 BC)", and also that "On the whole, in spite of some incongruous combinations and unclassified residues, Aristotle's higher taxa of animals were distinctly superior to those of Linnaeus, whose primary interest was in plants." (Mayr 1982: 152). Similarly, Darwin stated, in a letter (Gotthelf 1999) to the physician and classicist William Ogle who has sent Darwin a copy of his translation of Aristotle's works on the parts of animals, that "Linnaeus and Cuvier have been my two gods, though in very different ways, but they were mere school-boys to old Aristotle." Cuvier (1841: 148–149) also praised Aristotle's achievements in zoology:

© 2021 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2021 doi:10.20944/preprints202102.0623.v1

"Aristote, dès son introduction, expose aussi une classification zoologique qui n'a laissé que bien peu de choses à faire aux siècles qui sont venus après lui. Ses grandes divisions et subdivisions du règne animal sont étonnantes de précision, et ont presque toutes résisté aux acquisitions postérieures de la science."

"Aristotle, from his introduction [in *History of Animals*], also exposes a zoological classification which left very little to do in the subsequent centuries. His great divisions and subdivisions of the animal kingdom are astonishingly precise, and almost all of them have withstood the test of time." (Our translation)

Much of Aristotle's zoological work describes biodiversity in the broad sense: organs of animals, their habits, and groups of organisms and their characteristics. Especially regarding the last aspect of his writings there has been an extensive discussion. Most authors agree that Aristotle's zoological work was not meant to be primarily taxonomic (Pellegrin 1986, Gotthelf 1988, Lennox 2001, Sandford 2019, but see Lloyd 1961 and Carraro 2019), which would explain why he never summarized his classification in a compact form, contrary to most more recent systematists. In systematics, a taxonomy is a hierarchical classification with no partial overlap between the sets, except in rare cases created by hybridization; for instance, *Primates*[1] includes *Strepsirhini* (lemurs) and *Haplorhini* (*Tarsius*, monkeys and apes), and no member of *Strepsirhini* can be a member of *Haplorhini*, and vice versa, and there is no other way of dividing Primates in the context of a taxonomy. On the contrary, there are many alternative non-taxonomic (not strictly hierarchical) classifications of primates, for instance by habitat (arboreal vs terrestrial), diet (insectivorous, frugivorous, folivorous), daily activity pattern (diurnal vs nocturnal), and the sets created by these classifications can overlap (e.g. some arboreal primates are diurnal, whereas others are nocturnal). Aristotle discussed both logical, not strictly hierarchical classifications, and hierarchical classifications reminiscent of taxonomies, so the main question is: Did Aristotle view these as different but equally valid and interesting ways of classifying animals, or did he have a preferred classification scheme for animals that is akin to a taxonomy?

Assessing Aristotle's intentions in animal classification is complicated by the fact that the texts explaining this topic or containing an overall view might be lost (Pellegrin 1986, Hall 1991: 111–112). Aristotle's focus was apparently on explaining animals' design (structure), which was supplemented by illustrations obtained from dissections in his lost work *Anatomai* (Fürst von Lieven et al. 2020), and lifestyle (Pratt 1984: 272) and showing character linkage (Fürst von Lieven and Humar 2008: 244). Lennox (2006) discussed the debate about whether Aristotle's aim was more at defining taxa or only their attributes; in any case, both are linked because Aristotle found groups of animals that were produced by grouping according to several correlated characters. Aristotle did not follow the Platonic method of classical (dichotomous) division based on a single character and a twofold subdivision (called the "Platonic division method" below, for short) to classify animals. This is logical because the main purpose of Platonic division is to define entities, rather than to classify (Balme 1987: 70). Instead, Aristotle grouped animals based on several co-occurring characters (Mayr 1982: 151; Carraro 2019: 157). Thus, we avoid the term

[1] Note that in conformity with most rank-based codes of biological nomenclature, except for the Zoological Code, and as also recommended by the PhyloCode (Cantino and de Queiroz 2020), we italicize all taxon names. We capitalize taxon names, except for specific epithets and names used by Aristotle, given that modern editions of ancient texts do not capitalize these names.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2021 doi:10.20944/preprints202102.0623.v1

"division", even though Falcon (1977: 136) considers that this is a second form of division. The problems which arise when using dichotomy in classification are exemplified by Aristotle: "For if one uses dichotomy, it is either altogether impossible to grasp something (since the same thing falls into many divisions and opposed things into the same division), or there will be only one difference, and this one, whether it is simple or the result of interweaving, will be the final form." (PA I 3, 463b13–16, Translation after Lennox 2001). On Aristotle's method of division in biology, see Balme (1987), Falcon (1997) and Kullmann (2014: 145–147). Similarly, Stoyles (2013: 5) argued that Aristotle's classification aimed at finding "the widest classes possessing the various animal features", and that this avoided repetition in Aristotle's descriptions.

This can be illustrated by the following example. Aristotle noticed that a small group of *ichthyes* (the paraphyletic group[2] that became Pisces or "fishes" later on) do not possess a covering of the gills (HA II 13, 504b35); further the taxa in this group that Aristotle knew about are viviparous (contrary to teleosts, which included the greatest biodiversity in Aristotle's *ichthyes*) and have a cartilaginous endoskeleton (on the last two features, see PA II 9, 655a24 and IV 1, 676b1–3). These *ichthyes* are known as *selachē* (gr. σελάχη), a term probably coined by Aristotle because it is undocumented in earlier sources and Aristotle does not indicate that it had previously been recognized (here, and in the following, we use an overbar over some letters to specify Greek letters in our transcription which facilitates the detection of names; therefore, the letter ē indicates the η, while ō stands for ω); also Pliny in hist. nat. 9, 90 assumes that Aristotle invented the term. There is no need to list all attributes of *selachē* separately for all species included in that group. Note that many selachians are oviparous (Dulvy and Reynolds 1997), but Aristotle apparently was unaware of this, although in one passage he writes about the fishing-frog, which he wrongly classified as a selachian, that "one alone lays a complete egg outside" (GA III, 754a23–26).

Aristotle apparently coined other new terms referring to animal groups. The group called *dithyra* (from the adjective δίθυρος = with two doors or entrances) referring to the *Bivalvia* is not mentioned before Aristotle; the same holds true for the term *strombōdē* (gr. στρομβώδη, snails; also sometimes στρομβοειδῆ) which occurs the first time in the writings of Aristotle (e.g. in HA IV 4, 528b8, PA IV 9, 684b34). In other cases, Aristotle re-delimited previously-named groups, such as *kētē* (κήτη), which originally referred to any huge sea-creature like seals (Hom. Od. 4, 446) and sharks (Hom. Od. 12, 96–97). Aristotle, presumably for systematic reasons, confines it exclusively to the spouting whales and dolphins (todays Cetacea).

A caveat about these taxonomic and nomenclatural innovations is that previous knowledge on animal systematics that Aristotle could have relied on is poorly documented. Thus, Simon Byl (Byl 1980: 331; see also Mayr 1982, 149) argued that previous works on animals and the experience of professionals ("gens de métier") influenced or facilitated Aristotle's work on animal classification. Similarly, Meyer (2015, 36–58) mentioned previous classifications of animals in Homer and Hesiodus, which might have had an impact on Aristotle's work. Perhaps, folk

[2] A paraphyletic group includes only some of the descendants of the last common ancestor of its members. Since Hennig (1965), such groups have been dismantled by systematists. For an explanation of paraphyly, with comparisons with monophyly and polyphyly, see Fürst von Lieven and Humar 2008, 231–232.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2021 doi:10.20944/preprints202102.0623.v1

taxonomies (attested in Homer and other authors) were Aristotle's starting point. But given Aristotle's extensive work on animal classification, he probably improved substantially previous taxonomies, which becomes evident in several taxa that he apparently named first to fill a taxonomic (and verbal) gap. A perhaps less likely alternative is that several names attested for the first time in Aristotle's works were already evoked in now-lost previous works by earlier Greeks, or even, scholars from Mesopotamia or Egypt whose writings Aristotle may have had access to. The following quote from Aristotle suggests both that he is interested in developing or improving (pre-existing) systematics, and that relevant previous sources existed. In PA I 1, 639b4–6 he wrote: "I mean the question of whether one should study things in common according to kind [genos] first, and then later their distinctive characteristics, or whether one should study them one by one straight away. At present this matter has not been determined […]".

Systematists now consider that the best classification, the one that best explains how characters are distributed in biological organisms and why they appear to be correlated, is a taxonomy reflecting the phylogeny (Hennig 1965). Aristotle's zoological classification, at its most basic level, apparently used genealogical criteria by recognizing that parents give birth to offspring of the same kind. Note that this is not a phylogenetic criterion because Aristotle was not evolutionist, and he used the genealogical criterion only for the trivial task of assigning organisms to low-level taxa. At a higher level, he used similarity and functional criteria because structure and function of organs were both important for him (Pratt 1984: 274). In this context, we can wonder if within Aristotle's classification, all individuals or more inclusive groups of a given group necessarily possess all features of the higher-ranking group that includes them. This question is difficult to address since Aristotle did not stipulate if members of a group (*genos*) have to display all features peculiar to that group. However, we find some passages that deal with this question in Aristotle. Thus, he wrote in PA I 5, 645b1–3: "It is necessary first to divide the attributes associated with each kind [*genos*] that belong in themselves [*essential attributes*] to all the animals, and next to try to divide their causes." [Translation after Lennox]. This passage could be read as a statement that all attributes of one kind are visible in every species belonging to it. Nevertheless, in PA I 3, 643b1–2, he states that "there are certain kinds [*genē*] to which both differences belong and that are flyers and wingless, just like the ant kind [*genos*]."

In other words, how did Aristotle deal with exceptions, a pervasive phenomenon in biology? Much later, the concept of Homeostatis Property Clusters (HPCs) was proposed to deal with the fact that taxa lack defining, eternal, necessary and sufficient intrinsic properties (Boyd 1999). Did Aristotle's zoological practice anticipate on that? We leave this for future investigations.

Aristotle already might have distinguished between analogy and homology, which are important evolutionary concepts (Balme 1962:89; Fürst von Lieven and Humar 2008). His concept of 'the more and the less' in organs reflects the concept of homology (e.g., PA I 4, 644b11–15; Balme 1962: 89; Lennox 1987).

Pellegrin (1986; cited in Romeyer-Dherbey 1986) argued strongly that Aristotle did not do taxonomy as we intend it, as shown by the fact that the words *genos* and *eidos* are used at various levels in Aristotle's work. Pellegrin (1985: 95) even commented: "[...] the Aristotelian concepts of γένος and εἶδος, far from being prefigurations of our notions of genus and species, do not have a biological sense",

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2021 doi:10.20944/preprints202102.0623.v1

which seems a little excessive to the extent that Aristotle clearly designated animal groups by these terms, and that he did not use the Platonic division method to define them (Mayr 1982: 151). Furthermore, we can hardly criticize Aristotle for not having expressed a clear definition of biological species, given that this concept is still vague, with 146 meanings of 'biological species' documented in a recent compilation (Lherminier and Solignac 2005: 111–123). In an *eidos-genos* relationship of two groups of animals the former is included in the latter (Carraro 2019). Thus, the *eidos* of a given level can, according to Pellegrin, become *genos* at a lower level and be subdivided into *eidē*. However, Kullmann has argued in several works (a short summary is given in Kullmann 2014: 137–141) that the term *eidos* has to be understood as the smallest taxonomical unit and that *eidos* and *genos* in Aristotle are not used interchangeably in different levels; this misunderstanding might be a result of a missing reflection on the different meaning of *eidos* (form, type, but also appearance). Kullmann (2014:139) argued: "Die Nichtbeachtung dieser sprachgeschichtlichen Tatsache, daß das Wort eidos eine unterschiedliche Bedeutung hat, führt häufig zu der falschen Auffassung, daß ein und dieselbe Gruppe von Lebewesen auf fast jeder Stufe der Allgemeinheit sowohl Genos (kind) als auch Eidos (form) bezeichnet werden kann." [Neglecting this linguistic fact that the word *eidos* has different meanings often leads to the wrong assumption that one and the same group of living beings can be denoted both *genos* (kind) and *eidos* (form) at almost every level of generality.] Nevertheless, the different terms suggest that *genos* and *eidos* imply a hierarchy, as is found in modern taxonomy, but that they cannot be equated with the fixed levels in this hierarchy that genus and species represent in rank-based (Linnaean) nomenclature. This flexible use of the words *genos* and *eidos* and the absence of Linnaean (absolute) categories in Aristotle's classification has sometimes been used to argue that the latter does not represent a taxonomy (Pellegrin 1986 cited in Romeyer-Dherbey 1986). However, this does not follow because starting with Hennig (1969, 1981), an increasing number of authors has argued that Linnaean categories should be abandoned (de Queiroz and Gauthier 1990; Cantino et al. 1997; Laurin 2005a), and the word "taxon" can be used at any hierarchical level, just like Aristotle's *genos* and *eidos*. Indeed, Henry (2011: 200) already noted "I defend the claim that Aristotle's biology should remain of interest to philosophers and biologists alike insofar as it combines pluralism and realism with a rank-free approach to classification, which some philosophers [and many systematists] see as the way forward in systematics." This apparently extends to the taxa that appear to be single species (in the modern connotation of the word); "Aristotle was clearly aware that what are, prima facie, single species may in fact be divisible into more than one, as he points out in several instances." (Hall 1991: 132).

Pellegrin's (1986) claim that Aristotle's biological classification reflected only partly what we would now call phylogeny is unsurprising because the idea of biological evolution came much later, and the idea that evolutionary relationships could be depicted by trees originated with Lamarck, de Barbançois and Darwin in the 19th century (Tassy 2011). Balme (1962: 85) even stated that "there is no classification scheme in the background, and all attempts to construct one for Aristotle have failed".

Aristotle's writing *De partibus animalium* (On the Parts of Animals), which can be treated as the first introduction into the method of biological research (Düring 1943 :31 speaks of it as "a general introduction to the biological course"), includes in book I a lengthy attack on the Platonic division method and Aristotle explained that each animal kind should be defined by several characters, none of which is sufficient to provide an essential definition (Sloan 1972: 6; Mayr 1982: 151). While collecting characters of certain species, it is also important to take characters which are peculiar of a certain *genos* (*koinē kata genos*) into account (Aristotle also states that one should not include characters which are only accidentally found in animals such as sleep; on this, see PA I 1, 639a20–21 and Cat. V, 2b30–33). This suggests that for Aristotle, animals could not simply be grouped according to the then-prevailing logical principles. These considerations are compatible with current concepts of taxa; given that they evolve, any character may in theory be lost, so no intrinsic character should be considered essential. Aristotle was apparently misinterpreted by Popper and Hull on this point; these two influential authors apparently thought that "Aristotle considered the term to be defined as a name of the essence of the thing, and the defining formula as the description of the essence. And he insists that the defining formula must give an exhaustive description of the essence or the essential properties of the thing in question." Winsor (2003: 390).

More recent studies by biologists (rather than by philosophers) seriously question this last claim (e.g., Carraro 2019: 155–156). By concentrating for the first time on characters and their taxonomic distribution rather than only on the taxa and their names, Fürst von Lieven and Humar (2008) convincingly argued that Aristotle's work produced a hierarchy, which can be a taxonomy without absolute ranks (also see Moser 2013: 56). Like modern taxonomists, Aristotle often grouped animals using several characters for each taxon, rather than dividing them using single characters, as his predecessors had done, and he indicated that each taxon could occur only once in a classification (Fürst von Lieven and Humar 2008: 243).

A phylogenetic (parsimony) analysis of 147 terminal taxa included in Aristotle's HA, books II–V, scored from these same works for 161 characters attributed to various taxa by Aristotle, produced 58 groups, 29 of which have equivalents (similarly delimited) in Aristotle's work, and a further 12 have equivalents in modern works but not in Aristotle's (Fürst von Lieven and Humar 2008). Of 47 groups recognized by Aristotle and considered in their study, Fürst von Lieven and Humar (2008: 249) stated that 25 were still valid. The tree resulting from a parsimony analysis of these data matches only partly the currently accepted phylogeny (Fürst von Lieven and Humar 2008: fig. 4). Thus, insects, crustaceans and teleosts form mutually exclusive clades as they should, but tunicates are located very far from vertebrates, and echinoderms form a clade with gastropods and bivalves, rather than with chordates, to mention only two of the many unorthodox results contained in the tree.

The results of Fürst von Lieven and Humar (2008) suggest that Aristotle's work reflects an underlying taxonomy (which was never neatly summarized by Aristotle, contrary to what would be done two thousand years later by his successors), but doubts remain, partly because no statistical assessment of the similarities between the tree and Aristotle's classification was made. Are the similarities between the tree obtained by Fürst von Lieven and Humar (2008) from Aristotle's data and the current taxonomy merely coincidental? A visual inspection of their tree suggests that this is not very likely, but assessing this in a statistical framework would greatly improve our confidence on the conclusions that can be drawn from there results.

The doubts raised above are reinforced by the fact that our preliminary search based on the same data matrix hit the limit of 200 000 trees (rather than the 1000 trees

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2021 doi:10.20944/preprints202102.0623.v1

reported by Fürst von Lieven and Humar 2008) and that their strict consensus (a tree that includes only the clades found in all the source trees; Day 1985) includes several large polytomies. Furthermore, Hillis and Huelsenbeck (1992: 189) reported that "Analysis of random data often yields a single most-parsimonious tree, especially if the number of characters examined is large [which is the case in the matrix of Fürst von Lieven and Humar 2008] and the number of taxa examined is small [which is not the case]". One might wonder if the similarities between that consensus tree (and the groups recognized by Aristotle found on that tree) and the currently accepted phylogeny are merely coincidental, or if these data include a reliable phylogenetic signal. The notion of phylogenetic signal, which is of course foreign to Aristotle, can be defined as the tendency for closely related taxa to resemble each other more than distantly-related taxa (these relationships have to be established on the basis of other evidence, to avoid circularity); in other words, phylogenetic signal is inherited resemblance (Revell et al. 2008).

Methods

To reassess the implications of the taxonomic data included in Aristotle's works, we first updated the data matrix compiled by Fürst von Lieven and Humar (2008), using Mesquite (Maddison and Maddison 2019) to visually scan all characters for anomalous distributions, which were then checked again in Aristotle's writings to verify if these justified the current scores. Several mismatches were detected that way, and we report below many passages in Aristotle's HA that justify updating scores.

For the ordered analysis, we ordered seven characters 27, 33, 39, 42, 64, 131, and 159 (a new character). For some characters, such as 27, we had to reorder the states first because the initial order (Eyelashes: 0, upper eyelashes; 1, upper and lower eyelashes; 2, no eyelashes) did not reflect the possible cline. The new state order (0, no eyelashes; 1, upper eyelashes; 2, upper and lower eyelashes) reflects the logical hypothesis that it is easier for a taxon that already has upper eyelashes to acquire lower ones, than for a form without any eyelashes to acquire upper and lower eyelashes. It also reflects the highly probable hypothesis that the primitive condition is the absence of eyelashes (which occur only in some tetrapods). The primitive condition has conventionally been designated by the first symbol of a logical series, such a, a', a'' at least since Hennig (1965), but later, with the advent of computer-assisted phylogenetic analysis, numerical symbols became the norm and the state 0 has been most frequently used for the primitive state of phenotypic characters (e.g., Swofford and Maddison 1987).

We split the character for the number of gills (character 64) into two because the states were initially: 0, one (simple); 1, two (one duplicate, one simple); 2, four (simple); 3, four (three duplicate, one simple); 4, five (duplicate); 5, eight (duplicate); 6, no gills. States 0 to 5 seem to form a cline with an increasing number of gills, assuming that the word "simple" designates a hemibranch (a gill arch covered by gills only on one side) and that "duplicate" designates a holobranch (gill arch covered by gills on both sides), in standard anatomical nomenclature (Romer and Parsons 1977). However, state 6 (no gills) is at the wrong end of the cline. It would have been possible to reorder the states by moving "no gills" to state 0 and shifting all other states upward by one position. However, this would have implied that when gills appeared, there was initially a single one, and that the number increased over time. On the contrary, the presence of numerous gill slits in urochordates, cephalochordates, cyclostomes (Romer and Parsons 1977) and early jawless vertebrates (Janvier 1996a)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2021 doi:10.20944/preprints202102.0623.v1

suggests that gills were numerous when they first appeared and that their number became subsequently reduced. Thus, we preferred removing state 6 (no gills) and making a new character (158, presence of gills), with the states absent (0) and present (1). Note that the four new characters recognized in this study (158–161) have been added after the 157 characters included in Fürst von Lieven and Humar (2008) to facilitate comparisons between both studies. In character 64, we rescored the taxa that had state 6 (no gills) to inapplicable (-). This solution has the benefit of allowing the primitive condition to be anywhere in the cline (including in the middle), even though gills (the kind present in vertebrates) were obviously primitively absent in animals (as Aristotle called them; this taxon is now called Metazoa). For character 95 we redefined the states to better capture variations in the position of the embryos in the wombs of viviparous animals.

For character 39 (breasts), the initial states were: 0, two; 1, four; 2, many; 3, none. Again, there seems to be a cline (two, four, many), but the state "none" is at the wrong end of it, and we did not wish to assume how many breasts were present when they first appeared. We could have split the character as we did for number 64, but given the smaller number of states, we developed a step matrix (Maddison and Maddison 1992:58) instead. It allows transitions between absence of breasts and any number of breasts in a single step, but going from two to many breasts requires two steps because we assume that "four breasts" must be an intermediate state. This step matrix was also used for two other characters presenting a similar configuration (42, testicles, and 131, shell surfaces of bivalves).

We also changed the scores of several cells in the matrix. For character 33 (set of teeth) we corrected the matrix because a passage in Aristotle (HA II 1, 501a23–24) states clearly that most of the *ichthyes* have curved teeth. Regarding character 60 (number of fins), we changed the scores for the *myraina* (no fins) and the *gongros* (two fins) as well as the scores for the sting-rays (*leiobatos* and *narkē*: no fins) after a revision of the relevant passage in HA I 5, 489b26–32. For character 86 (gut-appendages or caeca), we corrected the scores to account for the fact that *perdix* and *alektryōn* have gut-appendages as stated in HA II 17, 509a17–22. For character 107 (proportion of body-size and length of feet), we revised the crucial passage in HA IV 1, 524a20–24 and changed scores for all *polypoda* named there (*polypous*, *heledōnē*, *nautilos*). For character 121 (number of fins on the tail = uropods) we changed the score for the *karkinion* (no fins on tail) because in HA IV 4, 529b20–26 Aristotle states that the *karkinion* resembles spiders in shape. For character 138, we changed the score in all mussels named *lepas* (*lepas* and *lepas agria*), which have the quasi-liver (*mēkōn*) deep in the shell (see HA IV 4, 529a29–31).

We also erected a new character (159: possession of sinews and their relative development) that we found in Aristotle's texts that we studied to improve the scores in the data matrix. We split up a character into new characters (54, initially "way of closing the eyes, with four states, into three binary characters: new 54, possession of lower eyelid, 160, presence of a medial corner eyelid, and 161, possession of upper eyelid) and changed the scores for several animals. In HA II 4, 502a31–32 Aristotle states that "[the ape] has very thin eyelashes [on both eyelids] while all other quadrupeds do not have them on both [eyelids]" which led us to state for the quadrupeds that they have both eyelids (a lower and a upper eyelid); they were initially scored as unknown (?).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2021 doi:10.20944/preprints202102.0623.v1

For a single character, the presence of lungs (character 70), we had to widen the search because the lungs of the dolphin are described shortly in HA VII (VIII) 2, 589a31–b6 but are not mentioned in the books II–IV (our main corpus). Hence, we scored the dolphin as having lungs.

We then assess the phylogenetic signal in the data, which Aristotle must have interpreted simply as similarity between groups of organisms. This allows us to determine if the characters that Aristotle used to describe his taxa contain a phylogenetic signal in the context of a classification of the taxa included in our data matrix, which are discussed extensively in HA. The presence of such a phylogenetic signal would imply that Aristotle discussed relevant diagnostic characters and, hence, that he had a good eye for important and relevant characters. This is a minimal requirement to suggest that Aristotle's groups were formed using good characters, but the presence of such a signal yields no unambiguous clues about Aristotle's intentions.

To study the phylogenetic signal in Aristotle's data (as compiled by Fürst von Lieven and Humar 2008), we use a method based on tree length distribution skewness, a method that has been shown to yield reliable data about the presence of phylogenetic signal (Le Quesne 1989; Huelsenbeck 1991; Hillis 1991; Hillis and Huelsenbeck 1992). This statistic is based on the fact that random data generates trees that have an approximately normal length distribution. This has been checked by examination of all possible trees for datasets of a few taxa, and examination of a population of random trees for larger datasets. On the contrary, datasets that contain a phylogenetic signal produce tree length distributions that are skewed to the left because there are few optimal trees, but many trees are much longer. Tables have been produced to determine significance thresholds for trees containing variable numbers of taxa and characters (Hillis 1991; Hillis and Huelsenbeck 1992: table 1 extended). We thus examined skewness in the length of samples of 1 000 000 trees in PAUP* (Swofford 2002): one sample with all characters unordered, and three samples with some characters ordered, to determine if this number of trees was sufficient to obtain a reliable estimate of skewness.

The fact that Aristotle did not summarize his taxonomy (possibly, the texts are now lost or he explained orally a detailed taxonomy in his lectures) excludes the use of tests that rely on comparisons between a reference tree (which should reflect Aristotle's hypothetical taxonomy) and a population of random trees (e.g., Laurin 2005b) using a tree statistic (typically, parsimony character steps, for discrete morphological data). Similarly, it would be inappropriate to compare the length of the most parsimonious tree with the length of trees derived from randomized data (Archie 1989) because Aristotle produced no such tree and his classification matched the strict consensus of the most parsimonious trees only partly (Fürst von Lieven and Humar 2008).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2021 doi:10.20944/preprints202102.0623.v1

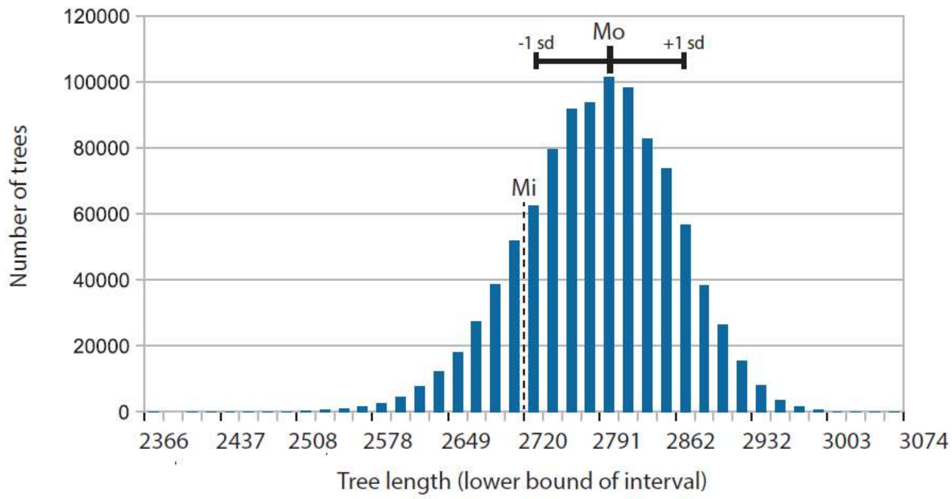


Fig. 1. Phylogenetic signal in the matrix produced by Fürst von Lieven and Humar (2008) from books II-V of Aristotle's HA. This is assessed by tree length distribution skewness on a population of one million random trees obtained from PAUP 4 (Swofford 2002) with some characters ordered. Skewness ($g1 = -0.248034$; we here illustrate the least skewed of the tree samples that we evaluated) is smaller than the threshold value (-0.12; $p < 0.01$) for 25 taxa and 100 characters (these values increase with number of taxa and characters; see Hillis and Huelsenbeck 1992). Note that the lowermost and uppermost of the 40 bins of the histograms contains one and two tree each, respectively. To better visualize the skewness, the middle of the range (Mi = 2720 steps) is shown. Note the noticeable shift to the right of the mode (Mo = 2800 steps) compared to the middle of the length range (Mi), which is narrowly excluded from the 1 sd (standard deviation) interval (±70,50) from the mode. The shortest of these random trees (2375 steps; much longer than the shortest trees of 306 steps) is farther from the mean length (2788.74 steps) than the longest tree (3065 steps); these distances are 413.74 and 276.26 steps, respectively.

We also re-analyzed the resulting matrix in PAUP* (Swofford 2002) to find the most parsimonious trees and their strict and majority-rule consensus (a tree that includes the clades most frequently encountered in the source trees; Day 1985). We performed this both with all characters unordered, as done by Fürst von Lieven and Humar (2008), and with characters that form clines ordered, given that mathematical principles and simulations show that this gives better results (Rineau et al., 2015, 2018). To increase the probability of finding consensus trees that correctly reflect the data, we performed a search with Maxtrees set to 200 000, with all characters unordered. With some characters ordered (see above for details), we performed two such searches. In addition, still with some ordered characters, we performed 10 additional searches with maxtrees set to 10 000. In all cases, random addition sequences were used, holding 2 trees at each step, and varying the random seed number for every search. Other search settings are: tree-bisection-reconnection (TBR) with reconnection limit = 8 and steepest descent option not in effect.

Results

With all characters unordered, we obtain 200 000 trees (there were undoubtedly many more, but this limit was set because of memory limitations) of 301 steps. These have a consistency index of 0.7076, a homoplasy index of 0.2924; when excluding uninformative characters, these stats become 0.6966 and 0.3034,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2021 doi:10.20944/preprints202102.0623.v1

respectively. The retention index is 0.9697 and the rescaled consistency index is 0.6862.

With some characters ordered, we still obtain 200 000 trees (again, there were certainly many more) of 306 steps. The fact that 8 out of the 10 searches with maxtrees set at 10 000 recovered trees with identical scores and yielded identical strict consensus and very similar majority-rule consensus trees suggests that we have recovered the most parsimonious trees and that our consensus trees adequately summarize the phylogenetic information. The most parsimonious trees have a consistency index of 0.6961, a homoplasy index of 0.3039; when excluding uninformative characters, these stats become 0.6847 and 0.3153, respectively. The retention index is 0.9685 and the rescaled consistency index is 0.6742.

Skewness of a population of one million random trees was obtained from PAUP 4 (Swofford 2002) with all characters unordered. Its skewness index (g1) is smaller (g1 = -0.253218; $p < 0.01$) than the threshold value (-0.12) for 25 taxa and 100 characters established (Hillis and Huelsenbeck 1992), which indicates significant skew **(Fig. 1**). With some characters ordered, the skewness index (g1) varied slightly between three samples of 1 000 000 trees (from -0.258032 in run 2 to -0.248034 in run 3), but was always strongly negative. Given that these values increase with number of taxa and characters, our test is conservative, and the close values obtained in the three samples with some characters ordered indicates that our sample is large enough to provide a reliable estimate of tree length distribution skewness. Most importantly, these results show that the data drawn from Aristotle's HA contain a strong phylogenetic signal, which shows that Aristotle was a great observer of animals.

The majority-rule consensus of the most parsimonious trees with ordered characters will be described briefly, to highlight similarities and differences between groups recognized by Aristotle and recovered as clades here, and the current consensus.

Our tree, like the one obtained by Fürst von Lieven and Humar (2008), contains many clades still recognized today: a clade equivalent to modern *Arthropoda* (not named by Aristotle), *malakostraka* (*Malacostraca*)*, karkinoi* (*Brachyura*)*, karides* (*Natantia* and *Stomatopoda*)*, entoma* (land arthropods), two mutually exclusive clades of *Polypoda* (*Myriapoda* and *Octopoda*[3]), *Hexapoda* (not named by Aristotle), *diptera, echinoi* (*Echinodermata*)*, monothyra* (*Archaeogastropoda*)*, strombōdē* (Caenogastropoda), *kochloi* (sea snails), *dithyra* (*Bivalvia*)*, malakia* (*Cephalopoda*)*, enhaima* (*Gnathostomata*)*, selachē* (*Euselachii*)*, platea selachōn* (*Batoidei*)*, kalymmata* [*sc. echonta*] (*Teleostei*)*,* a clade equivalent to modern *Tetrapoda* (not named by Aristotle)*,opheis* (*Serpentes*), *ornithes* (*Aves*)*, bareis* (*Galliformes*)*, gampsōnyches* (*Accipitriformes*)*, Mammalia* (not named by Aristotle), *pithēkoi* (*Catarrhini*)*, mōnyches* (*Equidae*)*, dichala* (*Artiodactyla*)*,* and *keratophora* (*Ruminantia*)*.* All these clades were also identified by Fürst von Lieven and Humar (2008).

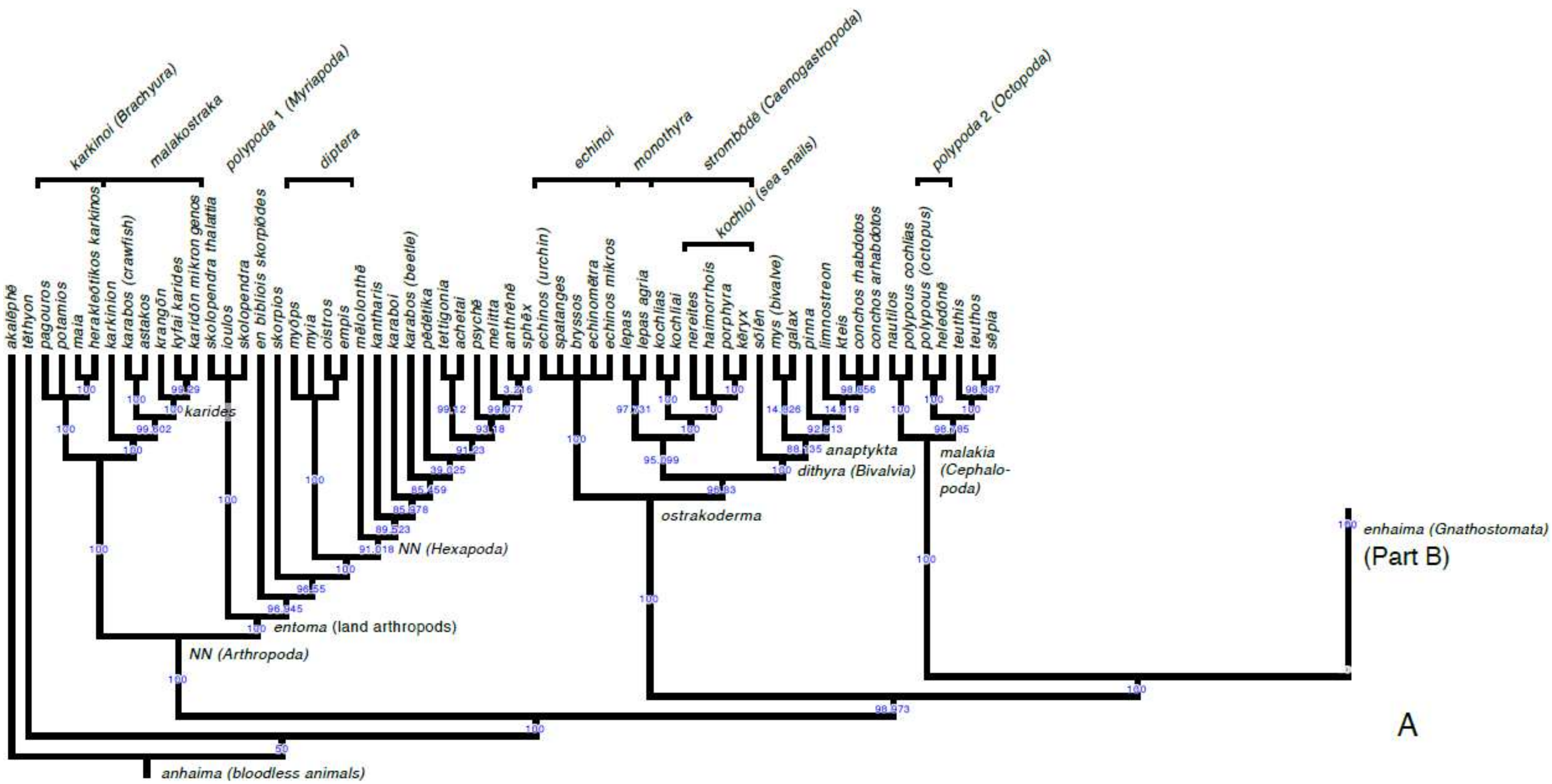
[3] Some animal's names are used homonymously in Aristotle's writing. One case is the name *polypous* ('many-footed') which refers to the octopus on one hand, and on the other to the centipede. Similar cases can be found in the terms *echinos* (the sea-urchin and the hedgehog), the *batrachos* (literally frog) which refers to the common frog and the fishing-frog (*Lophius piscatorius*) and the *karabos* (crawfish and a kind of beetle). To avoid confusion we indicate this in figure 2.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2021 doi:10.20944/preprints202102.0623.v1

We also recovered taxa recognized by Aristotle that are no longer recognized today, like *ostrakoderma, ichthyes* (*Pisces*)*, ōotoka tetrapoda* (oviparous tetrapods, including frogs and most squamates, but not snakes), and *amphodonta* (solenodont mammals) as well as the taxon named *anaptykta* (not identified in Fürst von Lieven and Humar, 2008) which refers to the mussels that can open, which is a curiosity because all bivalves (Aristotle's *dithyra*) can open their shelves but he allegedly identified mussels which are not able to open (the so-called *synkekleismena*, see below).

We also recovered clades that do not reflect the current consensus and that were not named by Aristotle, but that are implied by his *Scala naturae* (HA VIII 1, 588b4 sqq. and GA II 1, 732b15 sqq.) which comprises a sliding transition from inanimate bodies to plants and from those to the sessile living beings (especially the shellfish is emphasized), then on to "segmented animals" (mostly arthropods), cephalopods, and finally, the vertebrates, from the aquatic ones (ichthyes), through oviparous tetrapods, and finally, viviparous tetrapods (mammals), with Man on the very top. In GA Aristotle erects a grouping of animals according to the perfection of the mode of reproduction of the animals which is made dependent on their body warmth and their body moisture (on the *Scala naturae* in Aristotle, see Kullmann 2014, 147). Our tree replicates much of this scale. For instance, our tree places *malakia* (cephalopods) closer to gnathostomes than to other bloodless animals. In Aristotle's *Scala naturae, malakia* was indeed located just below oviparous vertebrates. Similarly, the position of the dolphin outside other mammals in our tree reflects the position of cetaceans below that of other mammals in Aristotle's *Scala naturae.* The only notable exception is that our tree places *ostrakoderma* closer to vertebrates than to arthropods, whereas the *Scalae natura* implies the reverse.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2021 doi:10.20944/preprints202102.0623.v1



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2021 doi:10.20944/preprints202102.0623.v1

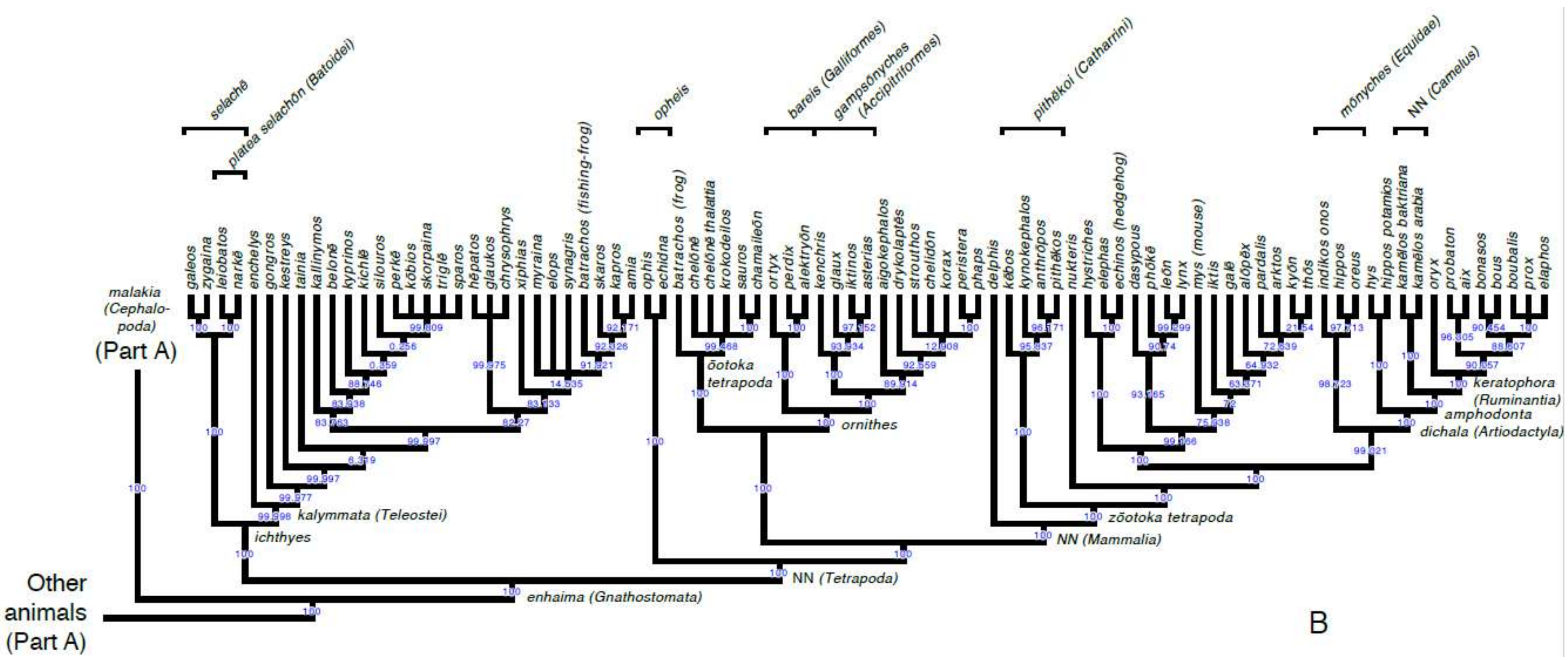


Fig. 2. Majority-rule consensus of the 200 000 trees produced by a search on the matrix produced by Fürst von Lieven and Humar (2008) with some characters ordered. The proportion of source trees that incorporate the various clades is indicated by blue numbers on each branch. See text for details and statistics of the source trees. The majority-rule consensus tree with all characters ordered differs only very slightly. The clades that match taxa named by Aristotle are labeled as such with the current name of these clades in parentheses, when Aristotle's terminology is not intelligible to most readers. Thus, *ichthyes* is obviously equivalent to *Pisces*, just like *selachē* matches *Euselachii* and *echinoi* is equivalent to *Echinodermata*; the modern name is not indicated on the figure for such taxa. Part A corresponds to Aristotle's *anhaima*; part B is *enhaima* (gnathostomes). Statistics of this majority-rule consensus tree (with some characters ordered): Component information (consensus fork) = 125 (normalized = 0.862); Nelson-Platnick term information = 1511; Nelson-Platnick total information = 1636; Mickevich's consensus information = 0.202; Colless weighted consensus fork (proportion max. information) = 0.152; Schuh-Farris levels sum = 45739 (normalized = 0.088); Rohlf's CI(1) = 0.972; Rohlf's -ln CI(2) = 656.931 (CI(2) = 4.99e-286).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2021 doi:10.20944/preprints202102.0623.v1

Discussion

Our results indicate the presence of phylogenetic signal in the data. This suggests that similarities between the strict consensus of the most parsimonious trees (which includes many taxa recognized by Aristotle) and the currently-accepted taxonomy established by systematists in the 20th and 21st centuries are not merely coincidental. This reinforces either Fürst von Lieven and Humar's (2008) suggestion that Aristotle's work rested on an underlying taxonomy, or more likely that using the characters that he had selected, Aristotle was able to discover many valid animal groups while describing characters of animals. Aristotle appears to "divide Nature by the joints", as suggested by Wiener (2015). Obviously, this activity led him to propose new names for some of the taxa that he discovered (like *selachē, dithyra,* and *strombōdē),* or to re-delimit some taxa, like *kētē,* and all this suggests that Aristotle was interested in taxonomy, even though this was obviously not his primary focus in zoology.

An important question regarding Aristotle's sources arises when one compares his observations with the zoological facts now known to us: In some cases, where the score is surprising, one could suggest that Aristotle never saw the animals he describes in his works. For example, his description of the *indikos onos*, which is the rhinoceros, is false in many respects. The same holds true for his description of the bending of the legs in the elephant (according to Kullmann 2014: 131–132, Aristotle has seen the elephant at first hand) and the descriptions of the *Oryx* as having one horn (it is not clear what animal Aristotle's *Oryx* actually is); this idea might have its origin in people looking at some representations of Antelopes in side view where the two horns seem to be only one horn; such illustrations are illustrated in pharaonic art, some of which Aristotle, or one of his assistants (see below), might have seen (Lones 1907). Classical scholarship includes discussions of what authors Aristotle consulted albeit he rarely named his sources by name (with few exceptions such as Ctesias and Herodotus). Kullmann (2014: 129–134) also suggested that Aristotle gave a kind of a 'checklist' to other students who should collect data for him in foreign regions he did not visit (Aristotle certainly never was in Africa, for instance, to see the *Oryx*). It is likely that the students made mistakes during their studies. Further, it is commonly accepted that Aristotle mingles facts acquired by dissections with second-hand reports from other sources. It is unclear that Aristotle did to verify reports about animals. All these points should be kept in mind when readers encounter sometimes surprising, false descriptions of several animals. The data matrix that we provide (as Online Resource 1), which includes scores and character-state definitions, should enable further progress in this field by facilitating expansion of our investigation to further books of Aristotle, or for further scrutiny of HA books I–IV, in which some relevant passages might have escaped our scrutiny.

This necessity of inventing names in pioneering works in zoology is also palpable regarding the names for structures Aristotle describes. Aristotle, aware of the fact that some structures or organs he describes in animals do not have proper names, sometimes uses metaphors to address this problem. By inventing metaphorical terms for unnamed structures or organs, Aristotle did what every scientist – modern as well as ancient – sometimes has to do when coining a new name for the yet unnamed (see Fürst von Lieven and Humar 2017; on metaphor in Aristotle, see Coughlin 2013 and Driscoll 2012).

This turns out to be a crucial problem in analyzing the characters of animals. It is difficult to determine if a metaphor can be treated as a character equivalent to non-metaphorical terms. For instance, the teeth of gnathostomes and the 'teeth' of the cephalopods (HA IV 1, 524b2) which refer to the horny beak are apparently treated as one character in Aristotle's work (since the horny beaks serve as 'teeth'), whereas no modern zoologist would homologize these structures composed of different tissues; in this case it is not even clear if Aristotle used a metaphor here or simply misinterpreted the beak as two teeth.

But there are also obvious discrepancies: E.g. the metaphorical 'horns' (*keraia*) of the *strombōdē* (HA IV 4, 528b24), a group of snails. These structures on their fore-head (the *antennae*) cannot be taken as the character 'having horns' in ruminants. The same holds true for the 'horns' of the crustaceans and the 'horns' of insects (both a kind of antennae or 'feelers'). Similarly, the feet [*podes*] of the cephalopods cannot be treated as the character 'having feet' because this term is clearly metaphorical (see HA IV 1, 523b21–22). The scores that we have entered in the matrix reflect these interpretations of Aristotle's characters and of their taxonomic distribution.

Another problem regarding Aristotle's terminology can be found in the term *phrēn* (φρήν), in plural *phrenes*. This term probably refers to the midriff or diaphragm (*diazōma*) in some animals (Fritts 1976). In some passages Aristotle introduces this term as a synonym; e.g. in HA II 15, 506a5–6: "All animals which have blood, have the heart and the diazōma, which is called phrenes"; see also HA I 15, 496b11 and PA III 10, 672b11. But it is not clear what the *phrenes* actually is. It can't be the diaphragm or midriff because *ichthyes* do not have this organ (but Aristotle in the passage quoted above clearly states that all animals with blood have this structure). To interpret the *phrenes* as the pericardium would make sense because a true heart is unique to vertebrates. And this is much more plausible than the interpretation of *phrenes* as a diaphragm. This interpretation has been introduced by Körner (1929) for the use of *phrenes* in Homer. It is also possible to interpret this structure as simply the set of membranes that cover the organs in the thorax and that divide it from the abdomen (Ireland 1975). As such, it would include diaphragm, pericardium, and also the pulmonary pleurae (and possibly more). Since we are primarily interested in the taxonomic implications of Aristotle's characters with the taxonomic distribution that he indicated, we do not need to determine exactly what this structure is in modern terminology.

Our analyses, like those of Fürst von Lieven and Humar's (2008), suggest that Aristotle recognized more than two ranks of named taxa; this interpretation is also explicit in Voultsiadou et al. (2017: fig. 1), which shows a classification of animals with three main levels and a fourth one for a few taxa. From most to least inclusive, these taxa of four levels include *enhaima, ichthyes, selachē*, and the last level includes *galeōdeis* and *platea*. This is coherent with Pellegrin's (1986) interpretations of the multi-level taxonomic usage of *genos* and *eidos* in Aristotle's zoological works. However, Hall (1991: 133) wrote: "The crucial question for my enquiry is, did Aristotle recognize gene intermediate between the major *genos*, Birds, and individual bird species? He obviously did not have a scheme of **intermediate *genē*** already worked out, which he does not happen to expound in any surviving work: his references to groups of birds are too haphazard and inconsistent." [Emphasis ours] On the contrary, we find evidence, at least for several taxa, of taxa of intermediate ranks between Aristotle's most inclusive and least inclusive taxa. Thus, in his writings, Aristotle's

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2021 doi:10.20944/preprints202102.0623.v1

anhaima (bloodless animals) designates a group that includes *ostrakoderma* (hard-shelled bloodless animals), which itself includes *strombōdē* (roughly Caenogastropoda), which includes *kochloi* (sea snails), which itself includes four taxa (Fig. 2), as far as we can tell. Part of this hierarchy was already noticed by Voultsiadou et al. (2017: fig. 1). Our analyses provide several more such examples (Fig. 2), especially among arthropods, *malakia* (cephalopods) and *enhaima* (vertebrates; see above). That Aristotle's classification of animals includes more than two ranks can also be seen at the beginning of book four (HA IV 1, 523a31–12, with omissions): "We now treat the animals devoid of blood [*anhaima*], […] There are several groups [genē], one consists of the so-called 'softs' [*malakia*]. […] Another [genus] is that of the malacostraca. […] An [other genus] is that of the ostracoderms. […] The fourth [genus] is that of insects, […]." After reporting about the peculiar structures of the some insects and cephalopods, Aristotle turns in chapter four to the ostracoderms (HA IV 4, 528a11–18: "Of other [ostracoderms] some are bivalved [*dithyra*] and some univalved [*monothyra*]; and by 'bivalves' I mean such as are enclosed within two shells, and by 'univalved' such as are enclosed within a single shell […]. Of the bivalves, some can open out [*anaptykta*], like the scallop and the mussel; […]. Other bivalves are closed on both sides together [*synkekleismena*], like the *sōlēn*."

This hierarchy with several taxonomic levels is obvious despite the fact that Aristotle's taxa do not match perfectly those that we recognize today. Aristotle's *ostrakoderma* is such an example. It includes all animals that have a hard outer shell. The sub-groups of *ostrakoderma* match fairly well taxa still recognized today, but they are not all closely related. These groups include the *dithyra,* which are equivalent to the modern taxon of Bivalvia and the *monothyra,* which refer to 'mussels' with only one shell. Aristotle obviously has the limpets in mind, and those animals belong to the *Gastropoda*. All his *monothyra* are aquatic snails with a conical shell that often live on stones on the shore. Hence, Aristotle wrongly considered them as mussels (or had a much broader concept of mussel than us). The bivalves that 'can open out' are mussels whose two shells are articulated via a hinge; hence, Aristotle describes them as 'the one that can open out' (all common mussels we know today). The other group mentioned (mussels with two shells 'closed on both sides alike') within the *dithyra*, are, on the contrary, mussels that allegedly do not open via a single juncture and, hence, seem to be closed. These animals are probably mussels from the genus *Pharidae*. Another group within the *ostrakoderma* are the *strombōdē* (roughly Caenogastropoda), which are characterized and differentiated from all other members of the *ostrakoderma* by having a spiral-shell (HA IV 4, 528b6–7). The most obvious anomaly in *ostrakoderma* for modern zoologists is the inclusion of *echinoi* (our echinoids, sea urchins), which are echinoderms, hence deuterostomians very distantly related to mollusks; the starfish is only mentioned once in the HA (V 15, 548a7) and Aristotle never specifies to what group it belongs. Nevertheless, the subgroups of *ostrakoderma* defined by Aristotle have an equivalent in modern taxonomy.

Our tree reveals a few more striking mismatches compared to contemporary taxonomies. Thus, snakes (*opheis*) are placed as the sister-group of other tetrapods, which probably reflects the absence of limbs. Similarly, the dolphin (*delphis*) is the sister-group of *zōotoka tetrapoda* (viviparous quadrupeds), which include all other mammals. Again, their absence of digited limbs, along with their aquatic lifestyle, may have misled Aristotle. The position of *pithēkoi* (catarrhine primates), including *anthrōpos* (*Man*) as the sister-group of all other terrestrial mammals in our tree is more surprising; it reflects mostly the presence of a narrow chest in other mammals (ch. 13),

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2021 doi:10.20944/preprints202102.0623.v1

four legs (Aristotle did not consider primate arms as legs), and other similar limb characters, in addition to the presence of two (upper and lower) eyelids (ch. 27).

Many of the taxa that Aristotle recognized and that have been refuted by recent phylogenetic studies have long been accepted by other zoologists. Thus, *anhaima* (bloodless animals, now called invertebrates) has long been known to be paraphyletic because they do not include all descendants of their last common ancestor; in this case, some invertebrates, such as echinoderms, are now known to be more closely related to vertebrates than to arthropods, annelids and mollusks, among others. Yet, this group was accepted by zoologists well into the 20th century as *Invertebrata*, and many textbooks still include "invertebrate zoology" in their title. In fact, a Google Scholar search (performed on December 27, 2020) found 15 700 results for "invertebrate zoology" for publications dating from 2001 to the present. Thus, Aristotle's *anhaima* is still with us, under another name. Our data matrix does not allow testing whether Aristotle's data would recover this taxon as a clade because the outgroup used to root the tree is *akalēphē*, interestingly introduced by Aristotle as "own group" (*idion genos*) in HA IV 6, 531a31, (a sea anemone; a cnidarian; on the anatomical description of the *akalēphē*, see HA IV 6, 531a31–b17), which is part of *anhaima*. Thus, paraphyly of *anhaima* is constrained by this choice of outgroup. However, there is no better alternative among the taxa described by Aristotle because sponges are too different to meaningfully polarize the characters and only briefly described by Aristotle, and ctenphorans were not described by Aristotle as far as we know. Furthermore, this choice has the advantage of being the correct outgroup, among the taxa included in our matrix, according to recent phylogenies (Pandey and Braun 2000).

Ichthyes is another noteworthy example. It has been accepted well into the 20th century as *Pisces* and indeed, there are still many ichthyology departments in contemporary academic institutions, but contrary to Pliny the Elder (Moser 2013: 30) and many of his successors, Aristotle did not include other marine metazoans in this taxon (a tradition that explains the names of animals such as starfish, cuttlefish, and jellyfish, among others).

The taxa that Aristotle recognized but that were subsequently dismantled also include, most notably, the *ostrakoderma* (see above). This taxon long remained in use, though under the name *Testacea*. This term was implicitly coined by Pliny who speaks of animals "enclosed by hard shells" (*testis conclusa duris*, hist. nat. 9,83) or "stony shells" (*silicea testa*, hist. nat 1,19); it was dismantled by Cuvier (1795), who transferred much of its contents to the mollusks. Of course, many taxa now considered as mollusks had been included in ostrakoderma/Testacea all along (the Bivalves and Gastropods), but Cuvier (1817) also included among mollusks organisms without an outer shell, such as the cephalopods (Aristotle's *malakia*). But, interestingly, the taxon *Ostracodermi* was erected by Cope (1889) for a group of Paleozoic armored jawless vertebrates. Given that the paraphyly of this taxon was suspected long ago and demonstrated by phylogenetic analyses (e.g. Janvier 1996b), it is no longer used formally, but the vernacular form "ostracoderm" still appears in recent scientific papers.

Other important mismatches compared to recent taxonomies include the position of bivalves and gasteropods in *ostrakoderma* and the distant placement of cephalopods as *malakia,* which we recovered (Fig. 2).

The relationships between the large clades found in our tree largely agree with Aristotle's *Scala naturae,* with the sole notable exception of the position of *ostrakoderma*. This suggests that Aristotle had a relatively good intuition for the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2021 doi:10.20944/preprints202102.0623.v1

relationships between taxa that the implied by the characters that he recognized. Of course, he must have conceptualized these relationships in terms of similarity rather than kinship (phylogeny).

Voultsiadou et al. (2017) similarly argued that Aristotle made tremendous progress in taxonomy and that his attempts at classifying animals according to other (i.e., not phylogenetic) criteria, such as reproductive mode and lifestyle, were misinterpreted by subsequent authors. These were not combined into a fuzzy taxonomy with overlapping taxa because Aristotle distinguished groups of animals that take a certain name, which we call taxa, from groups based on other criteria, such as lifestyle, which he arguably did not intend to be considered as taxa (Voultsiadou et al. 2017: 477). He called many of these groups "anonymous" or "nameless", but some of these groups may have been considered valid by Aristotle because they include the blooded and bloodless animals (Wiener 2015). Thus, even though Aristotle's primary goal in his biological work does not appear to have been to produce a taxonomy, he may have relied on a taxonomy to organize his zoological knowledge and present it to his readers. Alternatively, he may have written and classified simultaneously, in which case what looks like a taxonomy arose spontaneously and coincidentally. Our analyses cannot discriminate between these alternatives, though they do show that Aristotle's classification was remarkably good, given the tools and data available to him.

The works by Fürst von Lieven and Humar (2008) and Voultsiadou et al. (2017) could thus be used to support the view that in his classification of animals, Aristotle was, to an extent, a monistic realist, because some ways of separating animals into sets appeared to be more valid than others to Aristotle (Henry 2011: 205), even though the works cited here make no claim about monism in Aristotle. Henry (2011: 206) argues that Aristotle was a pluralist realist because, in his view, Aristotle's "Great Kinds" (major taxa, such as birds and fish) "do not enjoy a privileged status as ontological groupings", but this question deserves to be re-examined in light of these recent analyses. For instance, Henry (2011: 207) points out that "For Aristotle, natural kinds are limited to those groups whose shared similarities are underwritten by common causes." The main common cause of shared similarities between taxa in biological systematics is now considered to be the phylogeny (though convergence, often reflecting a similar ecology, can also play a role), and this, along with the evidence provided by Fürst von Lieven and Humar (2008) and Voultsiadou et al. (2017) logically leads to the conclusion that there is indeed one best way to classify biodiversity (into clades), which is compatible with monistic realism and with modern taxonomic practice, given that we seek to uncover the Tree of Life (TOL).

What remains to be determined is whether Aristotle's writings, taken globally and accounting for their inconsistencies, suggest that Aristotle had in mind a classification akin to a taxonomy, if he was trying to discover one, or if the taxonomy that is palpable in his writings and which we recovered to a large extent through our analyses of his data emerged spontaneously (and perhaps unintentionally) as he wrote the HA. We hope to have made some progress on this front, but it would be worth re-reading Aristotle to determine if his writings were globally consistent with the idea that he had a rankless taxonomy (without the equivalent of Linnaean categories, such as genera, families and orders, which were introduced much later) given that most recent philosophers and historians of science who studied this topic seem to have assumed, after Pellegrin (1985, 1986), that a taxonomy has to feature such fixed ranks. Yet, recent developments in biological nomenclature show that such absolute

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2021 doi:10.20944/preprints202102.0623.v1

ranks are not necessary, and indeed, that it might even be better to drop them (Hennig 1969, 1981; de Queiroz and Gauthier 1990; Cantino et al. 1997; Laurin 2005a). At least some zoologists who recent studied Aristotle concluded that his works suggest an underlying taxonomy (Voultsiadou and Vafidis 2007; Ganias et al. 2017; Voultsiadou et al. 2017), and our findings add support to this hypothesis.

Intention is usually difficult to assess, and we may never know what Aristotle's intentions were. This is unfortunate because it has been argued that "nothing is more important than intentions" (Hodge 1972: 129). However, Aristotle's search for coherent sets of characters that yield appropriate divisions between animal classes simultaneously yields a classification of animals that looks like a taxonomy, as our results show. Of course, our data are derived from Aristotle's work, but our analytical treatment (computer-assisted parsimony analysis) is very different. Despite this, our tree is remarkably congruent with Aristotle's classification which demonstrates the accuracy of his descriptions of animals. The only taxon recovered by our search that was not named by Aristotle is Arthropoda. The many similarities between the clades that we recovered and animal groups recognized by Aristotle are suggestive of a taxonomy. How did Aristotle uncover so many groups that have been validated by recent phylogenetic investigations, despite being fixist? To what extent does this reflect previous folk classifications? Does this reflect his choice of relevant systematic characters? These points deserve further investigation.

Author contributions

Both authors contributed to the study conception and design. Data collection, mostly through searches in Aristotle's original writings, were performed by Marcel Humar. Data analyses were performed by Michel Laurin, who also drafted the figures. The first draft (much shorter than this version) of the manuscript was written by Michel Laurin and both authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Acknowledgments

We thank Sylvain Charbonnier (CR2P, Paris) for information about crustacean morphology and Alexander Fürst von Lieven (Institut für Zoologie, Freie Universität, Berlin) for comments on an earlier draft of this paper. This research was funded by the CR2P's recurring grant (ML).

Online Resources

Online Resource 1. Data matrix and consensus trees in Mesquite format. Cells highlighted in yellow have updated scores, compared to those of Fürst von Lieven and Humar (2008). Cells highlighted in green have surprising scores that have been confirmed by a new examination of Aristotle's writings.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2021 doi:10.20944/preprints202102.0623.v1

References

Archie JW (1989) A randomization test for phylogenetic information in systematic data. Syst Zool 38:239–252

Balme DM (1962) ΓΕΝΟΣ and ΕΙΔΟΣ in Aristotle's Biology. The Classical Quarterly 12:81–98

Balme DM (1987) Aristotle's use of division and differentiae. In: Gotthelf A, Lennox JG (eds) Phylosophical Issues in Aristotle's Biology. Cambridge University Press, Cambridge, pp 69–89

Boyd R (1999) Homeostasis, species, and higher taxa. In: Wilson RA (ed) Species: New interdisciplinary Essays. MIT Press, Cambridge, Massachusetts, pp 141–185

Byl S (1980) Recherches sur les grands traités biologiques d'Aristote: sources écrites et préjugés. vol 64. Memoires de la Classe des Lettres, 2ème Série, vol 3. Academie Royale de Belgique, Bruxelles

Cantino PD, de Queiroz K (2020) International Code of Phylogenetic Nomenclature (PhyloCode): A Phylogenetic Code of Biological Nomenclature. CRC Press, Boca Raton, Florida

Cantino PD, Olmstead RG, Wagstaff SJ (1997) A comparison of phylogenetic nomenclature with the current system: a botanical case study. Syst Biol 46:313–331

Carraro N (2019) Dualisers in Aristotle's Biology. apeiron 52:137–165 doi:10.1515/apeiron-2018-0004

Cope E (1889) Synopsis of the families of Vertebrata. Am Nat 23:849–877

Cuvier G (1795) Mémoire sur la structure interne et externe, et sur les affinités des animaux auxquels on a donné le nom de vers; lu à la société d'Histoire-Naturelle, le 21 floréal de l'an 3. La Décade philosophique, littéraire et politique 5:385–396

Cuvier G (1841) Histoire des sciences naturelles depuis leur origine jusqu'à nos jours, chez tous les peuples connus. Première partie, comprenant les siècles antérieurs au 16[e] de notre ère. vol 1. Fortin, Masson et Cie, Paris

Day WH (1985) Optimal algorithms for comparing trees with labeled leaves. Journal of classification 2:7–28

de Queiroz K, Gauthier J (1990) Phylogeny as a central principle in taxonomy: Phylogenetic definitions of taxon names. Syst Zool 39:307–322

Dulvy NK, Reynolds JD (1997) Evolutionary transitions among egg-laying, live-bearing and maternal inputs in sharks and rays. Proc R Soc Lond B 264:1309–1315

Düring I (1943) Aristotle's De Partibus Animalium: Critical and Literary Commentaries. Wittergren & Kerbers Förlag, Göteborg

Falcon A (1997) Aristotle's Theory of Division. Bulletin of the Institute of Classical Studies Supplement 68:127–146

Fritts HW (1976) On the nature of the diaphragm; the evolution of three viewpoints. Transactions of the American Clinical and Climatological Association 87:16

Fürst von Lieven A, Humar M (2008) A Cladistic Analysis of Aristotle's Animal Groups in the "Historia animalium". Hist Philos Life Sci 30:227–262

Fürst von Lieven A, Humar M (2017) Die Bilder hinter den zoologischen Schriften des Aristoteles. Antike Naturwissenschaft und ihre Rezeption 27:71–140

Fürst von Lieven A, Humar M, Scholtz G (2020) Aristotle's lobster: the image in the text. Theory in Biosciences:1–15 doi:10.1007/s12064-020-00322-6

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2021 doi:10.20944/preprints202102.0623.v1

Ganias K, Mezarli C, Voultsiadou E (2017) Aristotle as an ichthyologist: Exploring Aegean fish diversity 2,400 years ago. Fish Fish 18:1038–1055

Gotthelf A (1988) Historiae I: plantarum et animalium: On Natural Science, Physics and Metaphysics, Ethics, Religion and Rhetoric. In: Fortenbaugh WW, Sharples RW (eds) Theophrastean Studies, vol 3. New Brunswick-Oxford, pp 100–135

Gotthelf A (1999) Darwin on Aristotle. Journal of the History of Biology 32:3–30

Hall J (1991) The Classification of Birds, in Aristotle and Early Modern Naturalists (I). History of science 29:111-151 doi:10.1177/007327539102900201

Hennig W (1965) Phylogenetic systematics. Annual Review of Entomology 10:97–116

Hennig W (1969) Die Stammesgeschichte der Insekten. Kramer, Frankfurt am Main

Hennig W (1981) Insect Phylogeny. John Wiley & Sons, Chichester

Henry D (2011) Aristotle's pluralistic realism. The Monist 94:197–220

Hillis DM (1991) Discriminating between phylogenetic signal and random noise in DNA sequences. In: Miyamoto MM, Cracraft J (eds) Phylogenetic analyses of DNA sequences. Oxford University Press, Oxford, pp 278–294

Hillis DM, Huelsenbeck JP (1992) Signal, noise, and reliability in molecular phylogenetic analyses. The Journal of Heredity 83:189–195

Hodge MJS (1972) The Universal Gestation of Nature: Chambers' "Vestiges" and "Explanations". Journal of the History of Biology 5:127–151

Huelsenbeck JP (1991) Tree-length distribution skewness: an indicator of phylogenetic information. Syst Zool 40:257–270

Ireland S (1975) Φρένες as an anatomical organ in the works of Homer. Glotta 53:183–195

Janvier P (1996a) The dawn of the vertebrates: characters versus common ascent in the rise of current vertebrate phylogenies. Palaeontology 39:259–287

Janvier P (1996b) Early vertebrates. 1 edn. Oxford University Press, Oxford

Körner O (1929) Die Ärztlichen Kenntnisse in Ilias und Odyssee. J. F. Bergmann, Münich

Kullmann W (2014) Aristoteles als Naturwissenschaftler. vol 38. Walter de Gruyter GmbH & Co KG, Berlin

Laurin M (2005a) The advantages of phylogenetic nomenclature over Linnean nomenclature. In: Minelli A, Ortalli G, Sanga G (eds), vol 1. Instituto Veneto di Scienze, Lettere ed Arti, Venice, pp 67–97

Laurin M (2005b) Embryo retention, character optimization, and the origin of the extra-embryonic membranes of the amniotic egg. Journal of Natural History 39:3151–3161

Le Quesne WJ (1989) Frequency distributions of lengths of possible networks from a data matrix. Cladistics 5:395–407

Lennox JG (1987) Kinds, forms of kinds, and the more and the less in Aristotle's biology. In: Gotthelf A, Lennox JG (eds) Philosophical Issues in Aristotle's Biology. Cambridge University Press., Cambridge, pp 339–359

Lennox JG (2001) Aristotle's philosophy of biology: studies in the origins of life science. Cambridge University Press, Cambridge

Lennox JG (2006) Aristotle's biology and Aristotle's philosophy. In: Gill ML, Pellegrin P (eds) A companion to ancient philosophy, vol 25. Blackwell, Malden, MA (USA), pp 292–315

Lherminier P, Solignac M (2005) De l'espèce. Syllepse, Paris

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2021 doi:10.20944/preprints202102.0623.v1

Lloyd GER (1961) The development of Aristotle's theory of the classification of animals. Phronesis 6:59–81
Maddison WP, Maddison DR (1992) MacClade: Analysis of phylogeny and character evolution. Version 3.0. Sinauer Associates, Sunderland, Massachusetts
Maddison WP, Maddison DR (2019) Mesquite: a modular system for evolutionary analysis. Version 3.61.
Mayr E (1982) The growth of biological thought—Diversity, evolution, and inheritance. vol 1. 1 edn. The Belknap Press of Harvard University Press, Cambridge
Meyer MF (2015) Aristoteles und die Geburt der biologischen Wissenschaft. Springer Spektrum, Wiesbaden
Moser B (2013) The Roman Ethnozoological Tradition: Identifying Exotic Animals in Pliny's Natural History. MA, University of Western Ontario
Pandey A, Braun EL (2020) Phylogenetic Analyses of Sites in Different Protein Structural Environments Result in Distinct Placements of the Metazoan Root. Biology 9:64
Pellegrin P (1985) Aristotle: a zoology without species. In: Gotthelf A (ed) Aristotle on Nature and Living Things. Mathesis Publications/Bristol Classical Press, Bristol, pp 95–116
Pellegrin P (1986) Aristotle's classification of animals: biology and the conceptual unity of the Aristotelian corpus. University of California Press, Berkeley
Pratt V (1984) The Essence of Ari[s]totle's Zoology. Phronesis 29:267–278
Revell LJ, Harmon LJ, Collar DC (2008) Phylogenetic signal, evolutionary process, and rate. Syst Biol 57:591–601
Rineau V, Grand A, Zaragüeta R, Laurin M (2015) Experimental systematics: sensitivity of cladistic methods to polarization and character ordering schemes. Contributions to Zoology 84:129–148
Rineau V, Zaragüeta I Bagils R, Laurin M (2018) Impact of errors on cladistic inference: simulation-based comparison between parsimony and three-taxon analysis. Contributions to Zoology 87:25–40
Romer AS, Parsons TS (1977) The Vertebrate Body. 5 edn. W. B. Saunders, Philadelphia
Romeyer-Dherbey G (1986) La classification des animaux chez Aristote. Statut de la Biologie et unité de l'aristotélisme. Revue de Métaphysique et de Morale 91:428–430
Sandford S (2019) From Aristotle to contemporary biological classification: what kind of category is "Sex"? Redescriptions 22:4–17
Sloan PR (1972) John Locke, John Ray, and the problem of the natural system. Journal of the History of Biology 5:1–53
Stoyles BJ (2013) Μέγιστα Γένη and Division in Aristotle's Generation of Animals. apeiron 46:1–25
Swofford DL (2002) PAUP* Phylogenetic Analysis Using Parsimony (*and other methods). Version 4.0a, build 167 edn. Sinauer Associates, Sunderland, Massachusetts
Swofford DL, Maddison WP (1987) Reconstructing Ancestral Character States Under Wagner Parsimony. Math Biosci 87:199–229
Tassy P (2011) Trees before and after Darwin. J Zoo Syst Evol Research 49:89–101
Voultsiadou E, Vafidis D (2007) Marine invertebrate diversity in Aristotle's zoology. Contributions to Zoology 76:103–120

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 February 2021
doi:10.20944/preprints202102.0623.v1

Voultsiadou E, Gerovasileiou V, Vandepitte L, Ganias K, Arvanitidis C (2017) Aristotle's scientific contributions to the classification, nomenclature and distribution of marine organisms. Mediterranean Marine Science 18:468–478

Wiener C (2015) Dividing Nature by the Joints. Apeiron 48:285–326

Winsor MP (2003) Non-essentialist methods in pre-Darwinian taxonomy. Biology and Philosophy 18:387–400