OXFORD UNIVERSITY PRESS | Systematic Biology

For peer review only. Do not cite.

How to Render Species Comparable Taxonomic Units Through Deep Time: a Case Study on Intraspecific Osteological Variability in Extant and Extinct Lacertid Lizards

Journal:	*Systematic Biology*
Manuscript ID	USYB-2021-099.R1
Manuscript Type:	Regular Manuscript
Date Submitted by the Author:	n/a
Complete List of Authors:	Tschopp, Emanuel; Universität Hamburg Centrum für Naturkunde; American Museum of Natural History, Division of Paleontology; Università degli Studi di Torino, Dipartimento di Scienze della Terra; Universidade Nova de Lisboa GeoBioTec Geobiociências Geotecnologias e Geo-engenharias Napoli, James; American Museum of Natural History, Division of Paleontology; Richard Gilder Graduate School Division of Paleontology Wencker, Lukardis; Università degli Studi di Torino, Dipartimento di Scienze della Terra Delfino, Massimo; Università degli Studi di Torino, Dipartimento di Scienze della Terra; Institut Català de Paleontologia Miquel Crusafont Upchurch, Paul; University College London, Earth Sciences
Keywords:	morphological disparity, osteology, intraspecific variation, Lacertidae, taxonomic bias, species, species delimitation

SCHOLARONE™ Manuscripts

How to Render Species Comparable Taxonomic Units Through Deep Time: a Case Study on Intraspecific Osteological Variability in Extant and Extinct Lacertid Lizards

Emanuel Tschopp[1,2,3,4], James G Napoli[2,45], Lukardis CM Wencker[3], Massimo Delfino[3,56], Paul Upchurch[76]

1- Centrum für Naturkunde, Universität Hamburg, Hamburg, Germany

2- Division of Paleontology, American Museum of Natural History, New York, USA

3- Dipartimento di Scienze della Terra, Università di Torino, Italy

4- GeoBioTec, NOVA School of Science and Technology, NOVA University Lisbon, Caparica, Portugal

5- Richard Gilder Graduate School, American Museum of Natural History, New York, USA

65- Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Barcelona, Spain.

76- Department of Earth Sciences, University College London, London, UK

Corresponding author: E. Tschopp, emanuel.tschopp@uni-hamburg.de, +49 1525-140-5200

Abstract

Generally, the species is considered to be the only naturally occurring taxon. However, species recognised and defined using different species delimitation criteria cannot readily be compared, impacting studies of biodiversity through Deep Time. This comparability issue is particularly marked when comparing extant with extinct species, because the only available data

For Peer Review Only

for species delimitation in fossils is derived from their preserved morphology, which is generally restricted to osteology in vertebrates. Here, we quantify intraspecific, intrageneric, and intergeneric osteological variability in extant species of lacertid lizards using pairwise dissimilarity scores based on a dataset of 253 discrete osteological characters for 99 specimens referred to 24 species. Variability is always significantly lower intraspecifically than between individuals belonging to distinct species of a single genus, which is in turn significantly lower than intergeneric variability. Average values of intraspecific variability and associated standard deviations are consistent (with few exceptions), with an overall average within a species of 0.208 changes per character scored. Application of the same methods to six extinct lacertid species (represented by 40 fossil specimens) revealed that intraspecific osteological variability is inconsistent, which can at least in part be attributed to different researchers having unequal expectations of the skeletal dissimilarity within species units. Such a divergent interpretation of intraspecific and interspecific variability among extant and extinct species reinforces the incomparability of the species unit. Lacertidae is an example where extant species recognised and defined based on a number of delimitation criteria show comparable and consistent intraspecific osteological variability. Here, as well as in equivalent cases, application of those skeletal dissimilarity values to palaeontological species delimitation potentially provides a way to ameliorate inconsistencies created by the use of morphology to define species.

Running head: ~~MORPHOLOGICAL SPECIES DELIMITATION IN LACERTIDAE~~SPECIES COMPARABILITY IN BIOLOGY AND PALAEONTOLOGY

Keywords: species delimitation, morphological disparity, osteology, intraspecific variation, Lacertidae, taxonomic bias

For Peer Review Only

Species are the fundamental biological ~~units, and~~units and are often considered the only naturally occurring taxa (e.g., Simpson 1940; Dunbar 1950; Gingerich 1985; Haffer 1986; Baum 1998; Harrison 1998; Wiens and Penkrot 2002; Hey et al. 2003; Queiroz 2005, 2007; Rieppel 2009; Hausdorf and Hennig 2010). However, the observable nature of a species is difficult to ~~grasp, and~~grasp and may vary from species to species. This difficulty of recognizing species and describing them in a consistent way (the "species problem"; Trueman 1924) is among the oldest problems in biology (Queiroz 2005; Allmon 2013), and has culminated in the formulation of nearly 40 species concepts, ~~since~~ most famously Mayr's (1942) ~~coined the famous~~ Biological Species Concept (Zachos 2016, 2018). However, most of these ~~proposed~~ species concepts have the same underlying assumption, namely that species are independently evolving lineages. This communality was recognized by Simpson (1951), who noted that earlier species concepts mainly diverged in the operational criteria they suggested to delimit species. This view was further developed by Wiley (1978) and later by ~~de~~ Queiroz (1998, 2005, 2007), who proposed a general or unified species concept, solely based on this communality of independent evolution. Consequently, the issue of describing species in a consistent way across all biological sciences has since been recognized to be of an operational nature and should thus be called the species delimitation problem (Queiroz 2005).

The disparate operational criteria proposed in ~~conflicting~~ different ~~earlier~~ species concepts ~~resulted in~~led researchers to develop various approaches to delimit species~~. Diverging species delimitation methods~~, which ~~may in turn~~ often lead to conflicting species counts when applied to a single dataset ~~(Haffer 1986; Wiens and Penkrot 2002; Doan and Castoe 2003; Agapow et al. 2004; Sites and Marshall 2004; Marshall et al. 2006; Knowles and Carstens 2007;~~

For Peer Review Only

Queiroz 2007; Hausdorf and Hennig 2010; Hausdorf 2011; Carstens et al. 2013), because some methods may recognize several units as distinct species(Haffer 1986; Wiens and Penkrot 2002; Doan and Castoe 2003; Agapow et al. 2004; Sites and Marshall 2004; Marshall et al. 2006; Knowles and Carstens 2007; Queiroz 2007; Hausdorf and Hennig 2010; Hausdorf 2011; Carstens et al. 2013), while other approaches would group them into a single species. Any such recognized "species" unit (taxonomic species, sensu Simpson 1940) is an estimate of the naturally occurring species (real species, sensu Simpson 1940). As such, the resulting taxonomic species units and will approach the real species units to differing degrees depending on the applied methodology and the inherent biological properties of the real species. Hence, although different methodologies createthese diverging units that are all called "species", they, but these are not necessarily comparable (Cracraft 1987), and should not be used in a comparative context (e.g., when studying changes in biodiversity). For instance, application of different species delimitation methods (based on molecular or morphological data) to a clade of the extant phrynosomatid lizard *Sceloporus*, resulted in recognition of five species by all approaches, but only two of the species were the same (Wiens and Penkrot 2002). The same delimitation methods applied to the gymnophthalmid lizard *Proctoporus bolivianus* suggested that this taxon actually represents either two or three distinct species (Doan and Castoe 2003). Delimiting species based on a non-phylogenetic and a phylogenetic species criterion (Agapow et al. 2004) found that the latter approach usually resulted in higher species counts, be it in plants, fungi, invertebrates, or vertebrates. Different approaches applied to salamanders recognized between one and eleven species in the *Ensatina eschscholtzii* complex (Sites and Marshall 2004). In trapdoor spiders, six molecular delimitation approaches yielded species counts ranging from three to 18, and not a single one was recognized as the same species by all approaches (Carstens

For Peer Review Only

et al. 2013). Given that species are generally used as fundamental units in a variety of biological studies, these issues have wide-reaching implications~~, be it in population-level analyses, biogeography, ecology, macroevolution, or conservation biology~~ (Sites and Marshall 2003, 2004; Balakrishnan 2005). We herein call this issue the "species comparability problem".

The species comparability problem is ~~likely to be~~ especially pronounced when comparing extinct and extant species, where not only the methodology to delimit species is often different, but there are also fewer available data upon which species delimitation can be based (Simpson 1951; Benton and Pearson 2001; Bruner 2004; Allmon and Smith 2011; Barnosky et al. 2011; Carrasco 2013; Miller III 2016). The~~is~~ restricted amount of data results ~~both~~ from having few specimens ~~available~~ per species, as well as having a limited range of data preserved in each specimen. ~~As a c~~Consequen~~ce~~tly, even if palaeontologists ~~may~~ agree with biologists on a particular species criterion (e.g., reproductive isolation), the available data in the fossil record may not allow accurate application of that criterion (Benton and Pearson 2001), given that fossil specimens of extinct species "caught in the act" of reproducing have been found but are exceedingly rare (see Joyce et al. 2012). In fact, within palaeontology, the "species problem" has been recognized as comprising three distinct, ~~but~~ interdependent issues: 1) the "species nature problem" (what constitutes a species in living organisms?); 2) the "species recognition problem" (can extant species be recognized in the fossil record?); and 3) the "species study problem" (can extinct species be studied as are modern species?) (Allmon 2013). ~~In the context of these palaeontological "species problems", we add t~~The species comparability problem ~~as~~ can be added as a fourth ~~aspect~~issue~~. It~~, ~~results~~ resulting from the "species nature problem" and the "species recognition problem" – as long as we delimit extant species using methodologies that

For Peer Review Only

cannot be applied to fossils, we cannot assume that those taxonomic units, created based on disparate delimitation criteria, are comparable, even if we all call them "species".

Most species of fossil taxa are delimited based on some understanding of "significant" morphological differences, either in a strict comparative context, or based on a phylogenetic analysis and resulting apomorphic features (Wood 1931; Queiroz 2007; Reichenbacher et al. 2007; Bernardi and Minelli 2011; Carrasco 2013; Allmon 2016; Kimura et al. 2016; Miller III 2016; Brochu and Sumrall 2020). These morphological differences can be calculated in two ways: 1) in direct comparison to the holotype specimen (a typological interpretation of the species), or 2) in comparison to observed intraspecific variability of a "type population", where the holotype may or may not represent the arithmetic mean (a polytypic or population interpretation of the species; Simpson 1940; Mayr 1942; Dzik 1985). In vertebrate palaeontology, assessing these differences is hampered by low sample sizes and limited availability of preserved data.

Many species of extinct vertebrates are known from single specimens (Watanabe 2016; Tschopp and Upchurch 2019), rendering any morphological comparison necessarily typological. Additionally, possible comparisons are mainly restricted to hard tissues, given that the morphology of soft tissues only preserve in exceptional circumstances (e.g., Christiansen and Tschopp 2010; Rauhut et al. 2012; Zheng et al. 2017; Fabbri et al. 2020; Bell and Hendrickx 2021). Even the preserved fossil hard parts we do have are often incomplete, hampering comparison both among fossil taxa and between fossil and extant taxa (e.g., Mannion and Upchurch 2010; Cleary et al. 2015; Brown et al. 2019). At the same time, osteology plays a minor role in species diagnoses or in identification keys of extant vertebrates (see Villa et al. 2018, 2019; Čerňanský and Syromyatnikova 2019; Villa and Delfino 2019; for notable

For Peer Review Only

exceptions in lizards). Taxonomists and systematists generally identify specimens of extant taxa based on external (soft tissue) morphology, while species delimitation methods are almost entirely based on molecular approaches (~~e.g., Carstens et al. 2013~~Wiens 2007). ~~Despite extensive discussion, it remains to be seen if species units delimited based on such disparate criteria are actually comparable. In fact~~Therefore, it has been proposed that at least some palaeontological species are more inclusive than neontological species (Trueman 1924; Cope and Lacy 1992; Brochu and Sumrall 2020), meaning that they may rather correspond to neontological genera or other higher-level taxonomic units. This may result from the fact that fixed diagnostic morphological traits do not necessarily exist (e.g., in cryptic species; Wiley 1978; Wiens and Servedio 2000; Wiens and Penkrot 2002; Allmon and Smith 2011; Brochu and Sumrall 2020), or they only occur in soft tissues. On the other hand, sexual dimorphisms may not be recognized as such in fossils and could instead be interpreted as diagnostic features of two distinct extinct species, ~~effectively~~ which would erroneously ~~doubling~~double the ~~real~~ species count for sexually reproducing species (Wiley 1978).

~~Although particularly pronounced between palaeontology and neontology, t~~The species comparability problem may also affect entirely palaeontological datasets. The application of different values of morphological disparity to delimit species sometimes results in diverging interpretations of diversity. Possible examples are Cambrian versus Ordovician trilobites (Foote 1990), and the Dmanisi hominins in Georgia (Arsdale and Wolpoff 2013; Lordkipanidze et al. 2013; Schwartz et al. 2014; Zollikofer et al. 2014; Rightmire et al. 2019). These issues impact ~~especially macroevolutionary~~any ~~studies~~study using “species” as its basic unit, including analyses of biodiversity through Deep Time (e.g., Carrasco 2013), ~~as well as assessments of current~~ extinction rates ~~compared to the Big Five Mass Extinctions of the geological past, which~~

For Peer Review Only

~~mostly rely on changes in species numbers over time~~, and evolutionary tempo or mode. Indeed, this challenge was the main reason for unexpected results in species diversity curves of small North American mammals (Carrasco 2013), and was recognised as one of the major "severe data comparison problems" by Barnosky et al. (2011: box 1) when trying to understand the extent of any current extinction.

Although several species delimitation methods are known and regularly applied in molecular phylogenetics and phylogenomics (e.g., Sites and Marshall 2003, 2004; Marshall et al. 2006; Carstens et al. 2013), only a few of these approaches are applicable to morphological data, and a very limited number of species delimitation methodologies has been explicitly used to define species based on morphological data in the past (see Tschopp and Upchurch (2019), and references therein) – although using intraspecific variation in extant species to guide delimitation of extinct species was first proposed by Matthew (1930).

Numerous methodological approaches to mathematically quantify variability have been developed ~~since then~~ and were applied for taxonomic purposes in extant taxa (e.g., Anderson and Abbe 1934; Cain and Harrison 1958), culminating in the development of "Numerical Taxonomy" (Sneath and Sokal 1973), which was mostly applied at higher taxonomic levels than the species. Although numerical taxonomy as a field has since been abandoned in favour of phylogenetic approaches, these methodologies continue to be used to quantify morphological disparity, including intraspecific variability (e.g., Anderson and Whitaker 1934; Zarapkin 1939; Wood et al. 1991; Dayan et al. 2002; Reichenbacher et al. 2007; Bever 2009; Foth et al. 2015). There has also been continuous support for the idea that morphological intraspecific variability may be used as a proxy for the presence of other operational species delimitation criteria (e.g., Hull 1965; Brochu and Sumrall 2020)~~, further reinforcing Matthew's proposal from 1930~~.

For Peer Review Only

However, only in the study of fossil mammals has an explicit application of extant variability scores, to delimit extinct species or assess their validity, been relatively widespread (e.g., Simpson 1941; Gingerich 1981; Kay 1982; Kelley 1986; Roth 1992).

In lacertid lizards – the focus of our study – knowledge of morphological intraspecific variability is mostly limited to external or soft tissue features. Several previous studies have analysed or discussed intraspecific variability in lacertid lizards, but mostly focused on single species or particular character complexes (e.g., Mateo 1988; Bruner et al. 2005; Brecko et al. 2008; Bruner and Costantini 2009; Kirchhof et al. 2012; Borczyk et al. 2014; Tayhan et al. 2016). Few studies have quantified variability among extant taxa on a larger scale (Barahona and Barbadillo 1998), or assessed the validity of extinct species based on osteological intraspecific variability from extant relatives (Mateo 1988; Barahona et al. 2000). These latter studies focused on the particular traits that have been suggested as diagnostic for certain putatively extinct species. ~~We are not aware of any study attempting to quantify overall intraspecific variability, and use the values observed in extant species to delimit extinct relatives. Hence, o~~Our study is the first to quantify intraspecific variability across a number of extant and extinct lacertid species ~~and~~, based on a large sample of osteological characters. ~~As such, it significantly increases both character and taxon sampling within Lacertidae to assess intraspecific variability in the entire clade, and to test if values obtained from extant species can be applied to fossil samples.~~

MATERIALS & METHODS

Our study comprises three analytical steps. First, we characterised intraspecific (comparing two specimens assigned to one species), intrageneric (comparing two specimens assigned to two distinct species of a single genus), and intergeneric (comparing two specimens assigned to distinct genera) osteological variability of lacertid lizards based on a dataset of 253 osteological character statements and

For Peer Review Only

99 individual specimens from 24 extant species. Second, we added 40 fossil specimens of six different species to the same dataset, to test for diverging species delimitation in neontological versus palaeontological understandings of lacertid species. Third, for the extant species, we simulated the impact of missing data and limited anatomical overlap (as observed in our sample of extinct species), to study how this affects our morphological dissimilarity analyses.

Dataset

The dataset of lacertid lizards used herein is a modified version of the phylogenetic matrix initially published ~~in~~ by Villa et al. (2017) and extended and modified by Tschopp et al. (2018b). ~~Their~~ These datasets ~~was~~ were initially imported into, and modified in, Mesquite (v. 3.6; Maddison and Maddison 2017), and subsequently transferred to, and managed on MorphoBank (O'Leary and Kaufman 2012). The modified matrix includes 30 additional characters in respect to Tschopp et al. (2018b), whereas the taxon sampling follows Villa et al. (2017) in including specimen-level operational taxonomic units (OTUs), but more than triples their sample of 37 extant OTUs by adding 62 extant and 40 extinct OTUs. The final matrix is available on MorphoBank (http://morphobank.org/permalink/?P4084), and among the supplementary material on Dryad (add doi).

Character sampling.—Disparity analyses do not depend on characters being phylogenetically significant (i.e., invariable within a certain clade, so it carries a clear phylogenetic signal), because variability is assessed on a pairwise basis, independent of any phylogenetic context (Gerber 2019). In fact, case studies have shown that disparate character coding strategies in discrete datasets do not have any significant impact on the outcome of disparity studies in caecilian amphibians (Hetherington et al. 2015). Hence, inclusion of as many characters as possible, irrespective of their variability within and among species, should yield more accurate estimates of overall intraspecific osteological disparity.

For Peer Review Only

~~Osteological characters are often the only preserved anatomical data in fossils of extant and extinct species, whereas extant species are generally recognised through molecular phylogenetics and/or a combination of external and (only marginally) internal morphology, behaviour, and geographic provenance of living specimens. Hence, when comparing extant and extinct species, we can only directly compare osteological variability in most cases, which is often only partially known in extant species (for European lacertids see Barahona and Barbadillo 1997, 1998; Barahona et al. 2000; Villa et al. 2017; Čerňanský and Syromyatnikova 2019; Villa and Delfino 2019). Therefore, we restricted our dataset to osteological characters.~~

Several characters were added based on existing literature (Queiroz 1987; Estes et al. 1988; Denton and O'Neill 1995; Scanlon 1996; Lee 1998; Conrad 2008; Brizuela 2010; Gauthier et al. 2012; Bailon et al. 2014; Čerňanský et al. 2016b; Quadros et al. 2018) and personal observations. Because we were interested in morphological disparity in general, and intraspecific variability more specifically, we did not restrict the character sampling to phylogenetically significant characters, but explicitly also included characters that ranged from high to no variability among the scored specimens (even within species). Whereas this may be problematic for phylogenetic analysis (Wilkinson 1997; Gerber 2019), it is the preferred approach for morphological disparity analyses, which effectively represent a phenetic approach to measure morphological diversity (Lloyd 2016). The final dataset included 253 characters, 219 of which are qualitative, and 34 quantitative (all of them discretized). Cranial characters constitute the majority of the dataset (167), followed by postcranial (69), and dental features (17). The character list is provided as Supplementary Data 2.

Extant Taxon and Specimen sampling.—Pairwise dissimilarity is calculated between two specimens, so two specimens per taxa are sufficient to obtain a score for variability within that

For Peer Review Only

taxon. Because we were interested in intraspecific, intrageneric, and intergeneric variability, we included all specimens of any genus represented by three or more specimens in total (up to 59 in *Lacerta*). By doing so, some included species are represented by a single specimen, which, consequently, only contributed to the calculations of intrageneric and intergeneric variability. The choice of these species and genera was mostly determined by the availability of skeletal specimens in scientific collections. The final species sampling amounts to 24 extant species belonging to seven genera of all three main subclades of Lacertidae (Gallotiinae, Eremiadini, Lacertini; Supplementary Table 1).

~~Table 1: Species sampling for disparity analyses of osteological variability in extant lacertids.~~

~~Species~~	~~Subclade~~	~~Number of specimens~~
~~*Gallotia caesaris*~~	~~Gallotiinae~~	~~1~~
~~*Gallotia simonyi*~~	~~Gallotiinae~~	~~1~~
~~*Gallotia stehlini*~~	~~Gallotiinae~~	~~1~~
~~*Iberolacerta cyreni*~~	~~Lacertini~~	~~1~~
~~*Iberolacerta monticola*~~	~~Lacertini~~	~~2~~
~~*Lacerta agilis*~~	~~Lacertini~~	~~12~~
~~*Lacerta bilineata*~~	~~Lacertini~~	~~12~~
~~*Lacerta media*~~	~~Lacertini~~	~~4~~
~~*Lacerta pamphylica*~~	~~Lacertini~~	~~3~~
~~*Lacerta schreiberi*~~	~~Lacertini~~	~~6~~
~~*Lacerta strigata*~~	~~Lacertini~~	~~3~~
~~*Lacerta trilineata*~~	~~Lacertini~~	~~8~~

For Peer Review Only

~~*Lacerta viridis*~~	~~Lacertini~~	~~11~~
~~*Ophisops elegans*~~	~~Eremiadini~~	~~4~~
~~*Podarcis hispanicus*~~	~~Lacertini~~	~~1~~
~~*Podarcis muralis*~~	~~Lacertini~~	~~4~~
~~*Podarcis siculus*~~	~~Lacertini~~	~~3~~
~~*Podarcis tiliguerta*~~	~~Lacertini~~	~~2~~
~~*Podarcis waglerianus*~~	~~Lacertini~~	~~1~~
~~*Psammodromus algirus*~~	~~Gallotiinae~~	~~4~~
~~*Timon kurdistanicus*~~	~~Lacertini~~	~~1~~
~~*Timon lepidus*~~	~~Lacertini~~	~~10~~
~~*Timon pater*~~	~~Lacertini~~	~~3~~
~~*Timon princeps*~~	~~Lacertini~~	~~1~~

The specimen sampling of the matrix of Villa et al. (2017) was considerably increased through scoring of additional lacertid specimens in European collections we could study first-hand, and of specimens that were extensively figured in recent literature (e.g., Čerňanský and Syromyatnikova 2019). This approach limited the number of specimens that could be included. However, we specifically targeted certain collections to capture as much variability as possible, be it geographical, ontogenetic, or sexual variability. ~~The total number of~~We included 99 extant specimens ~~included~~ in the dataset for the dissimilarity analyses~~ is 99~~. Of the 24 sampled species, 16 were represented by two or more specimens (up to twelve; ~~Table 1;~~ Supplementary Table 1)~~,~~ ~~amounting to~~; a total of 91 specimens ~~that~~ were used for our calculations of intraspecific osteological variability. These include all eight sampled species of *Lacerta* (*L. agilis, L.*

For Peer Review Only

bilineata, L. media, L. pamphylica, L. schreiberi, L. strigata, L. trilineata, L. viridis), three species of *Podarcis* (*P. muralis*, *P. siculus*, *P. tiliguerta*), two species of *Timon* (*T. lepidus* and *T. pater*), *Iberolacerta monticola*, *Ophisops elegans*, and *Psammodromus algirus*. The remaining eight specimens of the other ten eight species solely contributed to the calculation of intrageneric and intergeneric variability. All seven genera represented in our dataset include at least three specimens (up to 59 in *Lacerta*).

Extinct Taxon and Specimen sampling.—In order to test to what degree our approaches can be applied to the fossil record, we sampled 40 OTUs belonging to six extinct species of lacertids. These are *Dracaenosaurus croizeti*, "*Lacerta*" *filholi* and "*L.*" *siculimelitensis*, *Mediolacerta roceki*, *Plesiolacerta lydekkeri*, and *Pseudeumeces cadurcensis* (Supplementary Table 2).

Dracaenosaurus croizeti is here represented by seven specimens including three partial, semi-articulated skulls and skeletons from Cournon (France), and four disarticulated, tooth-bearing bones from Coderet (France). Our sample of "*Lacerta*" *filholi* includes four specimens: two dentaries (including the holotype) and a maxilla from Pech du Fraysse (France), and a third dentary from Coderet (France). It would have been possible to include other material based on published figures (e.g., Augé and Smith 2009), but these are all single, disarticulated bones, so the utility of their inclusion is limited. "*Lacerta*" *siculimelitensis* is also solely known from disarticulated material. Here, we use locality-level OTUs instead of specimen-level OTUs so we could score more characters per OTU. These are from five different sites: 1) Wied Incita Quarry (Malta), 2) Contrada Fusco (Italy), 3) Spinagallo (Italy), 4) Gargano (Italy), and 5) Monte Tuttavista (Italy). Using locality-level OTUs instead of single specimens increases the number of characters available for pairwise comparison, which would be very low or non-existent in fossil

For Peer Review Only

specimens that only preserve bones from disparate skeletal regions. However, this approach also increases the amount of potentially polymorphic features, equivalent to the use of a species- or any other higher-level OTU (Wiens 1995, 2000; Prendini 2001; Brusatte 2010; Tschopp and Upchurch 2019). We adopted a frequency scoring approach if a feature was observed to be polymorphic among the recovered material from a single locality, following recommendations of Wiens (1995, 2000). Thus, the calculated intraspecific variability in "*L.*" *siculimelitensis* does not represent differences between individuals, but rather differences between potentially distinct populations in time and space. Our sample of *Mediolacerta roceki* includes four specimens: the most complete fossil of the species, a nearly complete lower jaw; the holotypic dentary; and two disarticulated tooth-bearing bones from France and Germany. No articulated specimen is known from *Plesiolacerta lydekkeri*. We included 12 specimens of *P. lydekkeri* in our dataset, many based on figures by Čerňanský and Augé (2013). The included specimens comprise cranial and postcranial material from several sites in France. However, a combination of these into locality-level OTUs as implemented for "*L.*" *siculimelitensis* was not justifiable because most of the material is from historic collections from a single locality (Quercy, France), where the respective stratigraphic levels were not recorded, so that considerable time could be represented in the sample. Hence, we also used this sample to test the impact of the absence of anatomical overlap between specimens on disparity analyses. *Pseudeumeces cadurcensis* is here represented by eight specimens: an articulated lower jaw (the most complete individual specimen to our knowledge), and seven disarticulated cranial bones from a number of localities in France.

Dracaenosaurus croizeti was first described by Gervais (1848-1852) based on material from the Oligocene of Marcoin in southern France. It can be recognized by a strongly

For Peer Review Only

amblyodont dentition with very large posterior teeth in the maxilla and dentary. Additional material has since been referred to the species from other localities in France and in Germany, all of Oligocene age (Hoffstetter 1944; Müller 2004; Augé 2005; Čerňanský et al. 2016a, 2017). The species is here represented by seven specimens including three partial, semi-articulated skulls and skeletons from Cournon (France), and four disarticulated, tooth-bearing bones from Coderet (France).

"*Lacerta*" *filholi* was described as a new species of lacertid by Augé (1988), being diagnosed by a dentary dentition of monocuspid anterior teeth, bicuspid central teeth, and markedly tricuspid posterior teeth; a distinct coronoid facet on the labial surface of the dentary, and the absence of zygosphenes in the vertebrae. The species is known from the Oligocene of France and Belgium and the early Miocene of France (Augé and Smith 2009). Our sample includes four specimens: the holotypic dentary from Pech di Fraysse (France), a second dentary and a maxilla from the same site, and a third dentary from Coderet (France). It would have been possible to include other material based on published figures (e.g., Augé and Smith 2009), but these are all single, disarticulated bones, so the utility of their inclusion is limited (see also *Plesiolacerta lydekkeri* below).

"*Lacerta*" *siculimelitensis* was initially described by Böhme and Zammit-Maempel (1982), based on a peculiar heterodont dentition in a large-sized dentary. Additional material has since been referred to it (Caloi et al. 1986; Esu et al. 1986; Kotsakis 1996), or described as having a similar dentition and size (Delfino and Bailon 2000; Delfino 2001, 2002; Tschopp et al. 2018b). However, no articulated specimen has been found to date, hampering a precise systematic assessment of the material. Here, we use locality-level OTUs instead of specimen-level OTUs so we could score more characters per OTU.

For Peer Review Only

We included five locality-level OTUs representing the combined material referred to or mentioned as similar to "*L.*" *siculimelitensis* based on morphology and size (following the reported referrals in the literature) from five different sites from the Pleistocene: 1) Wied Incita Quarry, Malta (type locality), 2) Contrada Fusco, Sicily (Italy), 3) Spinagallo, Sicily (Italy), 4) Gargano, Apulia (Italy), and 5) Monte Tuttavista, Orosei, Sardinia (Italy). The material from Sardinia was referred to *Timon* sp. by Tschopp et al. (2018b), who also noted similarities with "*L.*" *siculimelitensis*. Because a detailed revision of "*L.*" *siculimelitensis* is lacking, and because Tschopp et al. (2018b) recovered the material as the sister taxon to all extant species of *Timon*, the Sardinian taxon was referred to *Timon* sp. Böhme and Zammit-Maempel (1982) initially referred "*L.*" *siculimelitensis* to *Lacerta* sensu latu, which at the time also included *Timon*, so that an inclusion of the Sardinian material as part of "*L.*" *siculimelitensis* in our dataset is justifiable.

Using locality-level OTUs instead of single specimens increases the number of characters available for pairwise comparison, which would be very low or non-existent in fossil specimens that only preserve bones from disparate skeletal regions. However, this approach also increases the amount of potentially polymorphic features, equivalent to the use of a species- or any other higher-level OTU (Wiens 1995, 2000; Prendini 2001; Brusatte 2010; Tschopp and Upchurch 2019). As in Tschopp et al. (2018b) for the Sardinian material, we adopted a frequency scoring approach if a feature was observed to be polymorphic among the recovered material from the other localities of "*L.*" *siculimelitensis*, following recommendations of Wiens (1995, 2000). Thus, the calculated intraspecific variability in "*L.*" *siculimelitensis* does not represent differences between individuals, but rather differences between potentially distinct populations in time and space.

For Peer Review Only

Mediolacerta roceki was described as a lacertid with slightly amblyodont posterior teeth (Augé 2005). It is known from the Oligocene of France (Augé 2005; Augé and Hervet 2009) and Germany (Čerňanský et al. 2016a). Our sample includes the most complete fossil of *M. roceki*, a nearly complete lower jaw, the holotypic dentary, and two other disarticulated tooth-bearing bones from France and Germany.

Plesiolacerta lydekkeri was named by Hoffstetter (1942) based on dorsal vertebrae with a distinct zygosphene-zygantrum articulation. Although no articulated specimen is known, numerous other fossils have subsequently been referred to this species, including cranial material (see Čerňanský and Augé (2013) for a summary). The referred material comes from the middle Eocene to the early Oligocene of England and France (Čerňanský and Augé 2013). As with "*Lacerta*" *filholi*, no articulated specimen is known from this species. In order to highlight the impact of the absence of anatomical overlap between specimens on the disparity analyses, and because several referred specimens were extensively figured by Čerňanský and Augé (2013), we opted to include 12 specimens of *P. lydekkeri* in our dataset. A combination of these into locality-level OTUs as implemented for "*Lacerta*" *siculimelitensis* was not justifiable because most of the material is from historic collections from a single locality (Quercy, France), where the respective stratigraphic levels were not recorded, so that considerable time could be represented in the sample. The included specimens comprise cranial and postcranial material from several sites in France.

Pseudeumeces cadurcensis was initially described as *Plestiodon cadurcense* by Filhol (1877). It possesses an amblyodont dentition that is intermediate between *Mediolacerta* and *Dracaenosaurus*. Fossils from different localities from the early to late Oligocene of France have been referred to *P. cadurcensis* in the past (Augé and Hervet 2009). Tentative referrals also

For Peer Review Only

~~include material from the late Oligocene of Germany (Čerňanský et al. 2016a) and the earliest Miocene of France (Augé and Hervet 2009). An articulated lower jaw is the most complete individual specimen to our knowledge and is here included in the dataset. Additional material sampled comprises disarticulated cranial bones from a number of localities in France.~~

Specimen Identification.—Correct species identifications of the sampled specimens is ~~obviously~~ paramount to ~~any~~ stud~~y~~ies of intraspecific variability. Here, 28 of the 99 specimens of extant species were collected, identified based on external morphological features and locality data, and then prepared by one of us (MD). The other identifications were mostly adopted from the collection catalogues, which were assumed to have been compiled by other expert herpetological taxonomists. Exceptions to this were made when we encountered identifications that appeared highly dubious based on the associated collection data and/or strongly aberrant size or morphology of the specimen, and where responsible collection ~~managers and/or curators~~staff urged caution. ~~These~~ All specimens with dubious identification were excluded from scoring. ~~Additionally, m~~Many of the studied specimens were referred to a species and accessioned in collections ~~of the above-mentioned institutions~~ before important revisions of those respective species or genera were published, and the ID associated with the specimens we studied has not been updated since. These include specimens identified as "*Lacerta ocellata*" and *Lacerta viridis*. The populations formerly ascribed to the first taxon are now referred to several different species included in the genus *Timon*. The species *L. viridis* is still a valid species within the genus *Lacerta*, but populations previously referred to the subspecies *L. viridis bilineata* were raised to species rank in the 1990s (see, among others, Arnold et al. 2007). All the species currently recognized as valid have distinct geographical distributions, and therefore museum

For Peer Review Only

specimens, skeletal preparations included, catalogued as "*Lacerta ocellata*" and *Lacerta viridis* with associated locality information could still be attributed to their respective species.

The identification of the fossil specimens was taken entirely from literature and museum catalogues for analytical reasons. Because we wanted to test if extinct species as recognized by palaeontologists had disparate intraspecific variability compared to extant species, we had to resort to those earlier referrals by default.

Phylogenetic Framework

The phylogenetic framework we followed ~~for the extant and extinct species~~ is based on earlier works (Carranza et al. 2004; Arnold et al. 2007; Kapli et al. 2011; Pyron et al. 2013; Mendes et al. 2016; Čerňanský et al. 2016b, 2017; Cruzado-Caballero et al. 2019). Given that the use of our compiled morphological matrix ~~includes highly variable character statements that may not be phylogenetically informative, and that, therefore, its feasibility~~ for phylogenetic inference may be limited (see Dataset – Character sampling), we ~~here~~ refrain from performing an independent analysis based on our own dataset. However, the main importance for this study is that all included species belong to Lacertidae, so we can assess if osteological intraspecific variability is consistent among the extant species in this particular clade, and could reasonably be used as a guideline to delimit extinct lacertid species, as well.

Molecular, morphological, and total-evidence phylogenetic analyses all recover the extant species in our dataset as members of Lacertidae. ~~Lacertidae can be subdivided into the two subclades Gallotiinae and Lacertinae (Carranza et al. 2004; Arnold et al. 2007; Pyron et al. 2013), and the latter further splits into Lacertini and Eremiadini (Arnold et al. 2007; Kapli et al. 2011; Pyron et al. 2013).~~ All three major lacertid clades are represented in our dataset: *Gallotia* and *Psammodromus* are gallotiine lacertids (Carranza et al. 2004; Arnold et al. 2007; Pyron et al.

For Peer Review Only

2013; Mendes et al. 2016; Čerňanský et al. 2016b, 2017; Cruzado-Caballero et al. 2019), *Ophisops elegans* is an eremiadin~~i~~ lacertine (Kapli et al. 2011; Pyron et al. 2013), and the remaining species belong to Lacertini (Carranza et al. 2004; Arnold et al. 2007; Kapli et al. 2011; Pyron et al. 2013; Mendes et al. 2016).

The extinct species ~~analysed in the second part of our analysis~~ were identified as lacertids based on particular diagnostic characters (mostly in the jaw; Supplementary Data 3). Some were later confirmed to be lacertids in phylogenetic analyses, although their exact position within Lacertidae often remains uncertain (Čerňanský et al. 2016b, 2017; Tschopp et al. 2018b; Cruzado-Caballero et al. 2019; Wencker et al. ~~in review~~2021).

~~Dracaenosaurus croizeti was referred to Lacertidae by Hoffstetter (1944), based on the open Meckelian canal in the dentary, which is covered by the splenial almost up to the symphysis, the triangular retroarticular process, and other features on the mandible. Phylogenetic analysis recovered it as a member of Gallotiinae (Čerňanský et al. 2017; Cruzado-Caballero et al. 2019).~~

~~“Lacerta” filholi was described as a new species of lacertid by Augé (1988), based on the open Meckelian canal, an arched subdental ridge (his “lame horizontale”; Augé 1988: p. 468), and a rounded ventral margin of the dentary. Augé (1988) mentioned close affinities with extant Timon lepidus (then Lacerta lepida) and extinct Plesiolacerta lydekkeri, while highlighting that there were insufficient diagnostic osteological characters to distinguish Lacerta (including Timon) from Gallotia and Podarcis, which were previously treated as subgenera of Lacerta and had just recently been identified as distinct genera (Arnold 1973). An attribution to Lacertidae seems well-supported; this is also corroborated by a recent phylogenetic study that found “L.” filholi as part of Gallotiinae (Wencker et al., in review).~~

For Peer Review Only

Although validity of "L." siculimelitensis has been questioned by several authors (e.g., Mateo 1988; Barahona and Barbadillo 1997), the lacertid affinity of the holotypic and referred material remained widely accepted (e.g., Estes 1983; Holman 1998; Delfino and Bailon 2000; Delfino 2001, 2002), and is supported by the widely open Meckelian canal, its tooth attachment, and tooth crown morphology. Böhme and Zammit-Maempel (1982) further mentioned that the referred caudal vertebrae are distinct from those of Gallotia, and that the biogeography of Gallotia and Lacerta (incl. Timon at the time, see above) further supports a referral to the latter. The phylogenetic analysis of the Sardinian fossils included here as "L." siculimelitensis by Tschopp et al. (2018b) recovered it as the sister taxon to Timon spp., further confirming its lacertid affinity and, more specifically, a referral to Lacertini.

Mediolacerta roceki was described as a lacertid by Augé (2005). No particular trait was mentioned that would support the referral to Lacertidae, but as with Dracaenosaurus croizeti, "Lacerta" filholi, and "L." siculimelitensis, the Meckelian canal is widely open, and reaches the symphysis (Augé 2005; ET & LCMW, pers. obs.), as is typical for lacertids. Furthermore, Augé (2005) proposed a very close relationship with "L." filholi. Phylogenetic analysis confirmed M. roceki as lacertid, but different approaches recovered it in conflicting positions within the clade (Wencker et al., in review).

The referred dentaries of Plesiolacerta lydekkeri have a widely open Meckelian canal, supporting lacertid affinities. In their detailed review of the species, Čerňanský and Augé (2013) stated that P. lydekkeri is close to or within crown Lacertidae, but they did not include it in a phylogenetic analysis. Recently, Čerňanský and Syromyatnikova (2019) listed 17 features shared by Timon lepidus and P. lydekkeri, some of which are otherwise unique among lacertids. Also, Wencker et al. (in review) noticed a close phylogenetic relationship between P. lydekkeri and the

For Peer Review Only

genus Timon. Thus, a lacertid affinity, and probably a referral to Lacertini, seems well-supported for P. lydekkeri.

Pseudeumeces cadurcensis was recognized as a lacertid by Hoffstetter (1944), based on the same diagnostic features used to identify D. croizeti. Phylogenetic analyses later confirmed the lacertid affinity, and recovered it as a member of Gallotiinae (Čerňanský et al. 2016b, 2017; Tschopp et al. 2018b; Cruzado-Caballero et al. 201Based on these works, the species can be tentatively referred to the subcla9).des Gallotiinae (*Dracaenosaurus croizeti*, "*Lacerta*" *filholi*, *Pseudeumeces cadurcensis*, and possibly *Mediolacerta roceki*) and Lacertini ("*L.*" *siculimelitensis*, *Plesiolacerta lydekkeri*).

Pairwise Dissimilarity

Pairwise dissimilarity as well as other disparity measures based on discrete morphological characters have long been used in palaeontology to study variability and/or morphospace occupation over time (e.g., Foote 1990, 1992a, 1993; Briggs et al. 1992; Lupia 1999; Bever 2009; Foth et al. 2015). The numerous proposed analytical approaches have various properties; the choice of methodology strongly depends on the kind of disparity one plans to study, and the type of data you haveone has available (Ciampaglio et al. 2001). Pairwise dissimilarity intuitively fits the purpose of quantifying intraspecific variability, and it also has been shown to be relatively insensitive to sample size, especially when using averages (Foote 1992b, 1993; Ciampaglio et al. 2001), rendering this methodology useful for morphological datasets of fossils. To our knowledge, pPairwise dissimilarity based on a discrete character matrix was first applied to delimit taxonomic units by Benson et al. (2012;) in plesiosaurs, who used using mean values of species-level OTUs to delimit genera). The same approach was applied and by Tschopp et al. (2015) ; in sauropod dinosaurs, where both species and genus

For Peer Review Only

delimitation were partially based on weighted pairwise dissimilarity scores). However, we are not aware of any previous study that has explored intraspecific osteological variability by means of pairwise dissimilarity in extant species to test its applicability for delimitation of closely related extinct species. Our analysis provides a nearly ideal test case because the taxonomy of the included specimens of extant lacertid species is known a priori and was probably not based on osteological features in most cases. Hence, we can test to what degree osteological intraspecific variability varies within extant species and assess if these data may be of use to delimit extinct species, which would render extant and extinct lacertid species comparable taxonomic units.

We used a custom R script (R Core Team 2019) to conduct our analyses (~~available upon request~~Supplementary Data 4). The script computes pairwise dissimilarity between all specimens, categorising them by species and classifying them as intraspecific (comparing two specimens assigned to one species), interspecific and intrageneric (comparing two specimens assigned to two distinct species of a single genus), or intergeneric (comparing two specimens assigned to distinct genera). With regard to multistate characters, disparity was calculated as the numerical difference between character scorings (e.g., a comparison between state 0 and state 2 is regarded as a disparity of 2), because all 30 multistate characters form morphoclines that are treated as ordered in a phylogenetic analysis (Foote 1992a; Brazeau 2011). Because we discretized all quantitative characters, the number of states in the multistate characters in our dataset amounts to three (23 characters), four (6 characters), to a maximum of five (1 character), so their impact on the entire analysis is not expected to be considerably strong. Polymorphisms were treated as the average of their scored states (e.g., 0&1 was treated as 0.5) because polymorphic characters capture informative details and should not be ignored (Wiens 1995, 1998; Watanabe 2016; Tschopp and Upchurch 2019). When a character was not scored in one of

For Peer Review Only

the individuals, dissimilarity for that character was not computed. The total dissimilarity across all characters for each pairwise comparison was then divided by the total number of computed dissimilarities (i.e., the number of characters scored in both individuals) to calculate weighted pairwise dissimilarity, representing disparity in units of character state differences per character compared. By doing so, we normalised the comparisons to the amount of data available for the analysis, reducing the impact of lacking anatomical overlap (following Tschopp et al. 2015). ~~This script was first executed to calculate dissimilarity in extant species followed by the extinct species (both represented in the complete dataset provided as Supplementary Material 2). Then, it was executed in two datasets where we simulated different amounts and distributions of missing data in extant species (Supplementary Material 3).~~ Statistical significance was assessed via ANOVA, using an *a priori* significance threshold of 0.05 and Tukey HSD post-hoc tests for all statistical comparisons.

Fossil ~~simulation~~Simulation

Missing data can be a serious issue in analyses of morphological disparity (Cope and Lacy 1992; Smith et al. 2014; Gerber 2019). Due to sampling and preservation biases in the fossils ~~included~~ in our dataset, missing data is widespread in our sample. Given the highly divergent completeness ~~scores~~ of the specimens of extant versus extinct species, we created two additional datasets to simulate the loss of data through fossilization observed in the extinct species using ~~six~~ extant partner species. The datasets simulate a loss of data equivalent to that of our real fossil sample and a loss intermediate between the extant and extinct values (see below).

We deleted entries from the extant species *Lacerta agilis*, *L. bilineata*, *L. trilineata*, *Podarcis muralis*, *Psammodromus algirus*, and *Timon lepidus* guided by the distribution of missing data in extinct species in our dataset. The specimens of the other extant species were left

For Peer Review Only

~~in their original state~~untouched~~, whereas~~. ~~the~~ The real fossil specimens ~~of the extinct species~~ were deleted from the dataset, so they could not impact the simulation. Each of the simulated extant species was assigned to an extinct partner species with an equal or lower number of scored specimens (*L. agilis* – *Plesiolacerta lydekkeri*; *L. bilineata* – *Dracaenosaurus croizeti*; *L. trilineata* – *Pseudeumeces cadurcensis*; *Podarcis muralis* – *Mediolacerta roceki*; *Psammodromus algirus* – *"L." filholi*; *Timon lepidus* – "*L.*" *siculimelitensis*). The distribution patterns of missing values in the extinct partner species were used as a model for the extant species~~.~~ .

~~We created~~The intermediate simulation was done using two custom Python scripts ~~for the simulated fossil dataset with intermediate loss of data (available upon request)~~. One script divides the character matrix into 25 sections with ten characters each (13 characters in the last section). It then calculates the percentage of missing values per character section in a predefined set of OTUs (Supplementary Data 5). The second script randomly deletes a predefined percentage of scored character states within any particular character section of a dataset (Supplementary Data 6). Using these scripts, we could adopt the distribution pattern and percentage of missing values found in an extinct species to simulate loss of data in the extant partner species; for the simulation with intermediate loss of data we used a percentage of missing values that was 20% lower for each section compared to the percentage observed in the extinct partner species, calculated over the entire set of specimens per species.

For the simulated dataset with extreme loss of data (equivalent to the amount of missing data observed in our extinct species), we matched single specimens within the extant and extinct partner species (Supplementary Table 3) and exactly adopted the distribution and number of missing values from the fossil to the extant partner specimens. ~~In cases when~~If missing values

For Peer Review Only

occurred in an extant specimen, but not in the fossil partner specimen, a character substitute from the same skeletal region was kept instead to obtain the exact same amount of missing data in the simulated extant specimen.

~~These~~ Our simulations excluded 48 to 68% (intermediate) and 69 to 94% (extreme) of the data scored for these extant taxa, amounting to total values of missing data of 61 to 76% (intermediate) and 77 to 95% (extreme). The resultant datasets (Supplementary Data 7) were analysed according to the procedure detailed above for the complete dataset.

Data Exploration and Sensitivity Analyses

Principal~~e~~ Coordinate Analysis.—~~In a f~~First ~~step~~, we performed a principal~~e~~ coordinates analysis (PCoA) implemented via the R package 'ape' (Paradis and Schliep 2019) using the complete dataset to explore morphospace occupation of extant species and genera based on a pairwise Euclidean distance matrix computed from our character scores. Principal coordinates analysis was selected as a data ordination method over other techniques such as ~~PCA~~ Principal component analysis because of its ability to accommodate missing data values and discrete, rather than continuous, data. However, PCoA is ineffective if specimens lack anatomical overlap, as no dissimilarity can be computed. Therefore, our PCoA only incorporated specimens of extant taxa, for which more complete scorings were available. We ~~further~~ used hierarchical clustering analysis implemented through the R package 'pvclust' (https://github.com/shimo-lab/pvclust) to determine whether PCoA clusters were able to discriminate between extant genera, and between species within the well-represented genera *Lacerta*, *Podarcis*, and *Timon*. We used a modification of Ward's clustering method, with a significance threshold of 0.05.

Missing Data.—Given the potential negative impact of missing data on disparity analyses (Gerber 2019), we conducted sensitivity analyses to further assess the effects of missing data,

For Peer Review Only

sample size, and skeletal modularity in our dataset. ~~First, w~~We used a third custom Python script (~~available upon request~~Supplementary Data 8) to calculate the percentage of missing data of all ingroup species for the complete dataset as well as for nine partitions: the cranial, dental, or postcranial character partitions, plus each of these ~~three~~ partitions divided into subsets of qualitative or quantitative characters (Supplementary Table 4).

In addition to quantifying missing data per se, we explored the dataset using the All Characters Overlap Index (AOI) and the Comparable Characters Overlap Index (COI) ~~(introduced by~~ Tschopp et al. ~~(~~2015~~) and formalised by~~ ; ~~Tschopp et al. (~~2018a). ~~Our case illustrates a use for data exploration by means of these overlap indices that has not previously been recognised. The AOI and COI help to determine whether some of the results from disparity analyses using discrete character data are impacted by restricted anatomical overlap, and how that relates to the impact of missing data per se.~~ When analysing pairwise dissimilarity scores, the AOI in particular is more meaningful than just calculating missing data, because only the characters with anatomical overlap provide information concerning pairwise dissimilarity within a certain group of OTUs. The AOI quantifies this amount of anatomical overlap within a group in relation to the possible total amount of anatomical overlap (Tschopp et al. 2015, 2018a). In a hypothetical case, two specimens could be scored for half the characters each, but could have no anatomical overlap whatsoever, resulting in a relatively high completeness score of 50% but AOIs and COIs of 0%, and no available data for pairwise dissimilarity analyses. By comparing the overall overlap indices with the indices restricted to particular anatomical partitions, such as the modules defined above for the completeness scores and sensitivity analysis, we can check if anatomical overlap is localised in a certain module or spread over the entire dataset. For this particular exploration, there is no point in dividing the characters into qualitative and quantitative

For Peer Review Only

sets, because we are solely interested in the impact of missing data and reduced anatomical overlap among skeletal regions. Different conceptual types of characters can only rarely contribute to an increase in missing data (e.g., when certain measurements are not available as a result of preservation; Mannion et al. 2013; Tschopp and Upchurch 2019).

We used the template file provided by Tschopp et al. (2018a~~: supplementary materials~~) for the calculation of the overlap indices. ~~Because the file locks itself once the first calculations are made, we created four different files with different matrices, one including the entire character set, and the other three restricted to cranial, dental, and postcranial characters.~~ The AOI and COI were ~~then~~ calculated for every ingroup species assessed for intraspecific variability (Supplementary Data 9-12) ~~(i.e., with more than three scored specimens)~~. These values allowed us to identify subsets of characters that are considerably more completely scored than other subsets, and hence less impacted by reduced anatomical overlap (Gerber 2019).

Furthermore, ~~To~~ to assess the impact of lacking anatomical overlap in our dataset directly, we computed every possible intraspecific pairwise comparison, recording the number of characters scored in both specimens. We then computed the percentage of the maximum intraspecific dissimilarity observed for each species that was achieved in each comparison.

Sample Size.—To assess the impact of sample size in our dataset, we conducted resampling with the four best-sampled taxa in our sample~~—~~: *Lacerta agilis* (N = 12), ~~*Lacerta*~~ *L. bilineata* (N = 12), ~~*Lacerta*~~ *L. viridis* (N = 11), and *Timon lepidus* (N = 10). For each taxon, resamples were done with numbers of specimens ranging from two to the maximum sampled, with each sample size replicated 100 times, and the maximum, minimum, and mean pairwise dissimilarity recorded.

For Peer Review Only

RESULTS

Pairwise Dissimilarity

Among extant taxa, intergeneric dissimilarity was consistently significantly greater than intrageneric/interspecific dissimilarity, which in turn was consistently significantly greater than intraspecific dissimilarity (Fig. 1). ~~Poorly sampled t~~Taxa sampled by two or fewer specimens, such as the three *Gallotia* species and *Iberolacerta monticola* showed insignificant differences between intergeneric and intrageneric/interspecific dissimilarity and intrageneric/interspecific and intraspecific dissimilarity, respectively. *Lacerta media* (~~sampled by four specimens~~N = 4) and *Podarcis siculus* (~~three specimens~~N =3) also showed insignificant differences between intrageneric/interspecific and intraspecific dissimilarity. Most, but not all, extinct taxa also had dissimilarity values that were significantly lower intraspecifically compared to intrageneric/interspecific variability, which was in turn significantly lower than intergeneric dissimilarity (Supplementary Table ~~5~~4).

"*Lacerta*" *filholi* showed equivalent intrageneric/interspecific and intergeneric dissimilarity, indicating that it was as dissimilar from other *Lacerta* species as it was to species placed in other genera. *Mediolacerta roceki* and *Plesiolacerta lydekkeri* showed equivalent intergeneric and intraspecific dissimilarity. By contrast, *Dracaenosaurus croizeti* and *Pseudeumeces cadurcensis* both showed significantly lower intraspecific dissimilarity than intergeneric dissimilarity, as did "*L.*" *siculimelitensis*, which showed an extant-like pattern with intrageneric/interspecific dissimilarity as intermediate between intraspecific and intergeneric dissimilarity.

For Peer Review Only

~~Figure 1. Intergeneric (between two specimens of two different genera), intrageneric (between two specimens of the same genus but different species), and interspecific dissimilarity for extant lacertid taxa. The horizontal black line in the boxplots represents the median. NS indicates statistically non-significant differences. Increasing number of stars refers to decreasing significance cutoff ("***"=0.001, "**"=0.01, "*"=0.05). Generally, intraspecific dissimilarity is significantly lower than intrageneric dissimilarity, which is in turn significantly lower than intergeneric dissimilarity. The exceptions are species with low sample size (1 specimen in the species of *Gallotia*; 2 specimens of *Iberolacerta monticola*; 4 specimens of *Lacerta media*; 3 specimens of *Podarcis siculus*).~~

Pairwise comparisons of all extant taxa recovered most taxa as displaying statistically indistinguishable intraspecific dissimilarity (Fig. 2a) – ~~that is~~thus, ~~that the~~ extant taxa generally showed similar degrees of intraspecific morphological variability. Five out of six (100/120 comparisons, exactly) of the pairwise comparisons were statistically insignificant. The significant differences in intraspecific dissimilarity mostly included ~~a few~~three outlier taxa. *Lacerta media* was significantly more dissimilar than *L. pamphylica*, *L. schreiberi*, *L. trilineata*, *Ophisops elegans*, *Podarcis muralis*, *Podarcis tiliguerta*, *Psammodromus algirus*, *Timon lepidus*, and *T. pater*. *Lacerta pamphylica* and *O. elegans* ~~was~~ were significantly less dissimilar than *L. agilis*, *L. bilineata*, *L. media*, *L. viridis*, and *T. lepidus*~~.~~, ~~*Ophisops*~~ *O.* *elegans* was ~~was~~ also significantly less dissimilar than ~~*L. agilis*, *L. bilineata*, *L. media*,~~ *L. schreiberi* and~~,~~ *L. trilineata*~~, *L. viridis*, and *T. lepidus*~~. Aside from these three outlier taxa ~~(*L. media*, *L. pamphylica*, and *O. elegans*)~~, *L. bilineata* was significantly more dissimilar than *L. trilineata* and *Podarcis muralis*. However, this signal appears to be an "edge effect" wherein the most and least intraspecifically dissimilar taxa are significantly different from one another, but not to the majority of taxa (Fig. 2a). Taken together, extant species showed a mean weighted pairwise intraspecific dissimilarity of 0.2076 ± 0.0579 character state differences per character scored. The non-outlier taxa, combined, showed a mean weighted pairwise intraspecific dissimilarity of 0.2089 ± 0.0557 character~~s~~ state differences per character scored. *Lacerta media* had a mean weighted pairwise intraspecific variation of 0.2631 ± 0.0786, *L. pamphylica* one of 0.1226 ± 0.0477, and *O. elegans* one of 0.1286 ± 0.0353 (all in units of character state differences per character scored). Within extant taxa, dissimilarity was dominated by qualitative cranial and postcranial characters, which did not differ significantly from the pooled intraspecific dissimilarity derived from all characters. Quantitative cranial and postcranial characters, and

For Peer Review Only

qualitative dental characters, were all significantly more dissimilar intraspecifically than the pooled variation. Quantitative dental characters were significantly less intraspecifically dissimilar than the pooled variation.

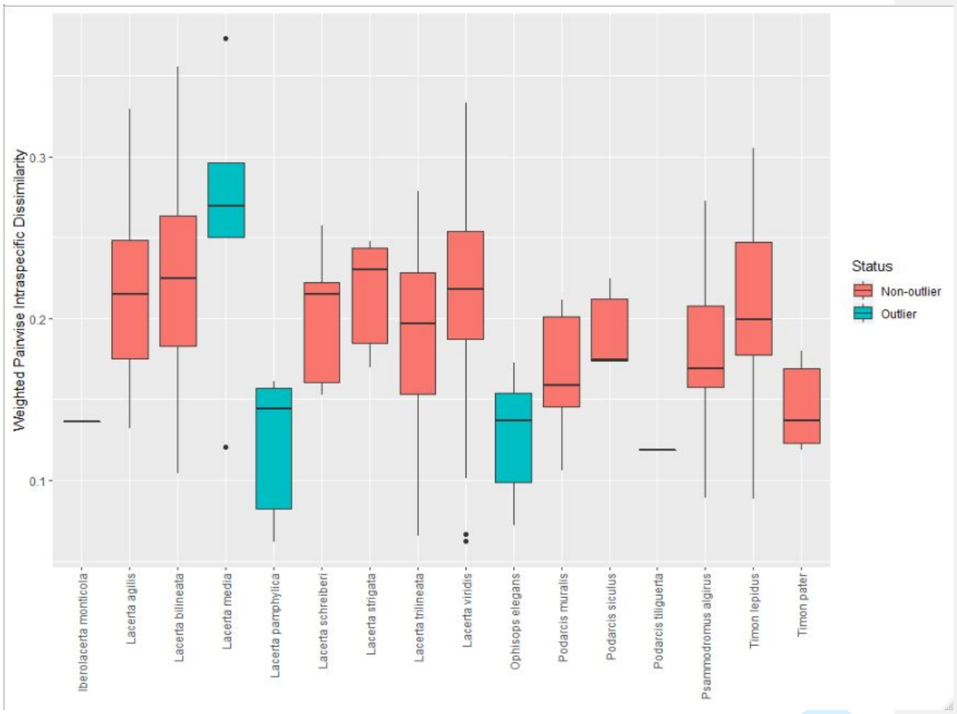
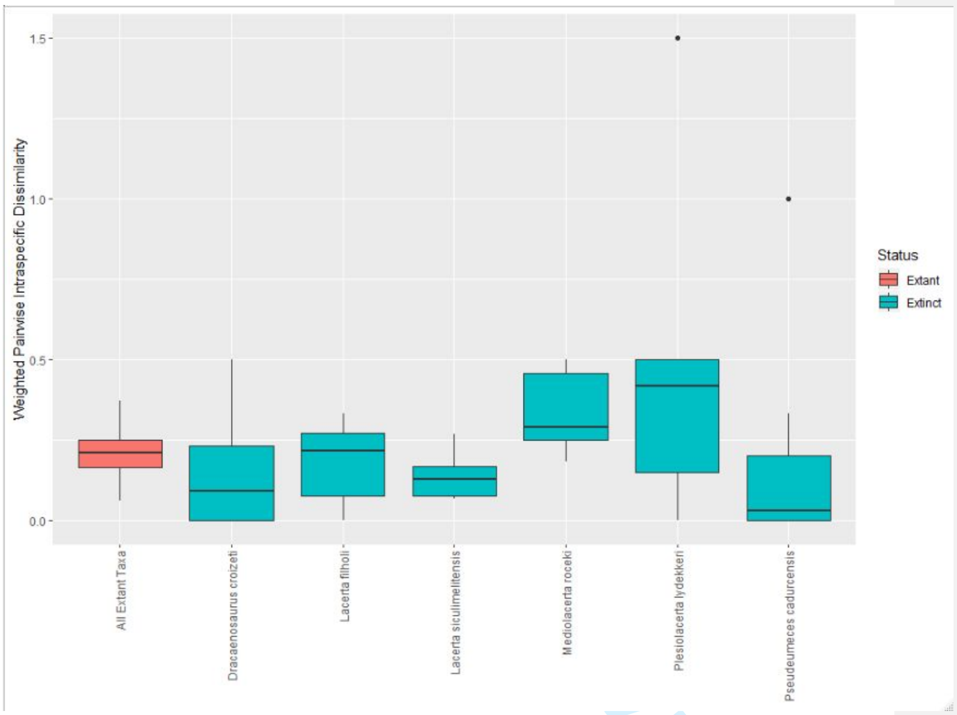


Figure 2: Intraspecific dissimilarities for all extant lacertid species in our dataset. Boxes in blue represent "outlier taxa" that were statistically distinguished from more than two other taxa in the dataset. Horizontal black lines in the box plots represents the median. Overall mean weighted pairwise dissimilarity is 0.2076 ± 0.0579 character state differences per character scored.

The weighted intraspecific pairwise dissimilarities of *Dracaenosaurus croizeti* and *Pseudeumeces cadurcensis* were significantly lower than the pooled intraspecific dissimilarities of the extant taxa, while *Mediolacerta roceki* and *Plesiolacerta lydekkeri* were significantly more dissimilar than the extant taxa (Fig. 2b3). "*Lacerta*" *filholi* and "*L.*" *siculimelitensis* did not differ from extant taxa. A total of 8 of 15 pairwise comparisons among the extinct taxa were statistically significant, indicating that the extinct taxa do not group with each other in terms of intraspecific dissimilarity, as the extant taxa do. *Dracaenosaurus croizeti*, "*LacertaL.*" *filholi*, "*L.*" *siculimelitensis*, and *Pseudeumeces cadurcensis* are all significantly less intraspecifically dissimilar than *Mediolacerta M. roceki* and *Plesiolacerta lydekkeri*. These results are unchanged if the outlier extant taxa *Lacerta L. media*, *L. pamphylica*, and *Ophisops elegans* are excluded from the dataset.

For Peer Review Only



~~Figure 3: Intraspecific dissimilarity of extinct lacertid lizards in our dataset (in blue), compared to bulk intraspecific dissimilarity of the sampled extant species (in red). The horizontal black line in the boxplots represents the median. Extinct species have much more variable intraspecific dissimilarity than extant species.~~

Under an intermediate fossilization simulation, a~~ll but one~~ single extant taxa ~~did not~~ approximated the patterns seen in ~~their~~ its extinct partner~~s~~ (~~the exception being~~ *Psammodromus algirus*, which was already similar to its extinct partner species "*Lacerta*" *filholi* when scored completely). In this simulation, only *Podarcis muralis* differed significantly in intraspecific dissimilarity from the average of the remaining "extant" species used in the simulations, being significantly less variable than extant taxa. On the contrary, its extinct partner species, *Mediolacerta roceki*, was significantly more variable than extant taxa, suggesting that the high variability observed in this species is not solely due to low anatomical overlap and/or sample size.

In the extreme ~~simulated~~ fossilization simulation, no simulated ~~fossil~~ "fossilized" taxon differed significantly from the extant taxa in terms of pairwise intraspecific dissimilarity, in contrast to the extinct taxa. When compared to their extinct partner species, four species approximated the intraspecific dissimilarity of their extinct partner (*Lacerta agilis* – *Plesiolacerta lydekkeri*; *L. bilineata* – *Dracaenosaurus croizeti*; *Psammodromus algirus* – "*L.*" *filholi*; *Timon lepidus* – "*L.*" *siculimelitensis*). For ~~*Lacerta*~~ *L.* *agilis* and *L. bilineata*, this statistical indistinctness appears to be an artifact of increasing variance, as mean and median values of intraspecific dissimilarity remain distinct and more similar to their original dataset than that of their extinct partner species. *Timon lepidus* shows a true approximation of the intraspecific dissimilarity of *"~~Lacerta~~L." siculimelitensis*, while *Psammodromus algirus* continues to resemble "~~*Lacerta*~~ *L.*" *filholi*, as it did in its original scoring and in the intermediate ~~simulated~~ fossilization simulation. *Lacerta trilineata* and *Podarcis muralis* remained significantly different ~~form~~ from their extinct partner species *Pseudeumeces cadurcensis* and *Mediolacerta roceki*, respectively, suggesting that the low observed intraspecific variability in

For Peer Review Only

Pseudeumeces cadurcensis and the high variability in ~~*Mediolacerta*~~ *M.* *roceki* are true signals. Although *L. agilis* and *L. bilineata* are not statistically distinguishable from their extinct partners under an extreme ~~simulated~~ fossilization simulation, the persistent differences in median values shown in Fig.~~.~~ure ~~4~~3 suggest that the patterns seen in *Plesiolacerta lydekkeri* and ~~*Dracaenosaurus*~~ *D.* *croizeti* may be true signals as well, ~~that~~ which are not distinguishable in our dataset due to sample size.

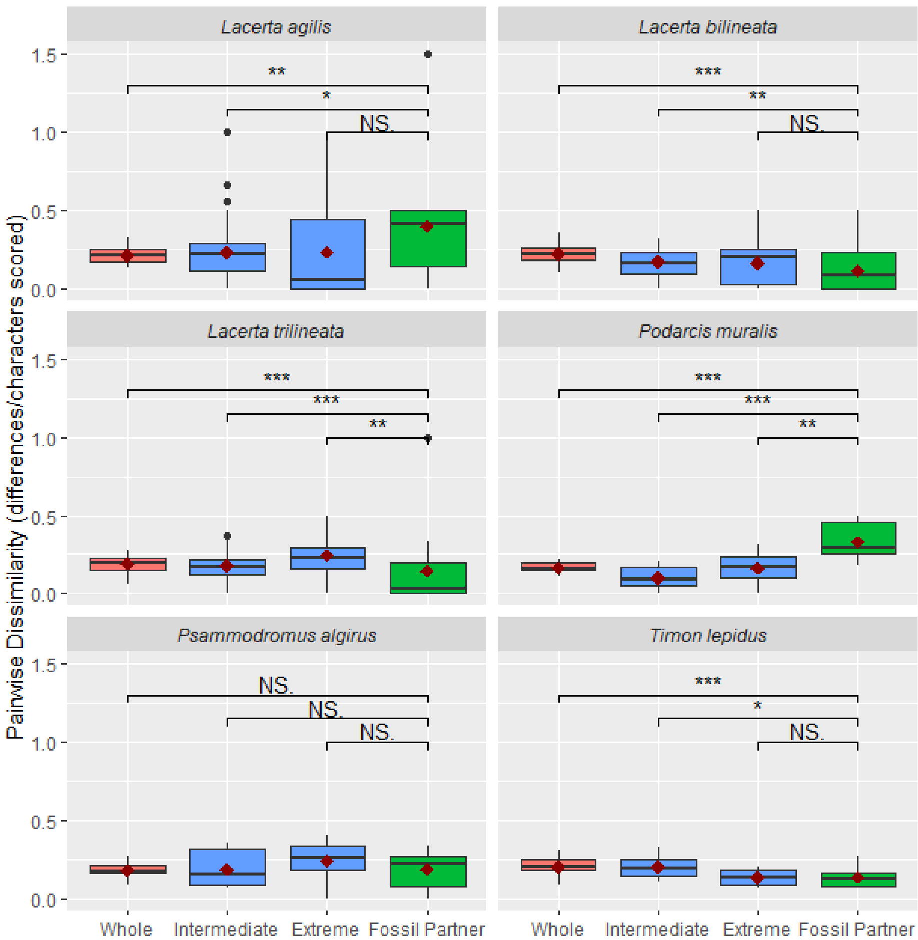


Figure 4: Simulation of missing data in extant species, following patterns observed in extinct species. Intraspecific, weighted pairwise dissimilarity scores (y-axis) are given for the whole dataset, the simulated dataset with intermediate values of missing data, the simulated dataset with the same characters missing from the comparison as in the extinct partner species, and the extinct partner species. The extinct partner species are *Plesiolacerta lydekkeri* (for *Lacerta agilis*), *Dracaenosaurus croizeti* (for *L. bilineata*), *Pseudeumeces cadurcensis* (for *L. trilineata*), *Mediolacerta roceki* (for *Podarcis muralis*), "*L.*" *filholi* (for *Psammodromus algirus*), "*L.*" *siculimelitensis* (for *Timon lepidus*). NS indicates non-significant differences. Increasing number of stars refers to decreasing significance cutoff ("***"=0.001, "**"=0.01, "*"=0.05). The black line in the box plots represents the median, red diamonds represent the mean.

Data Exploration and Sensitivity Analyses

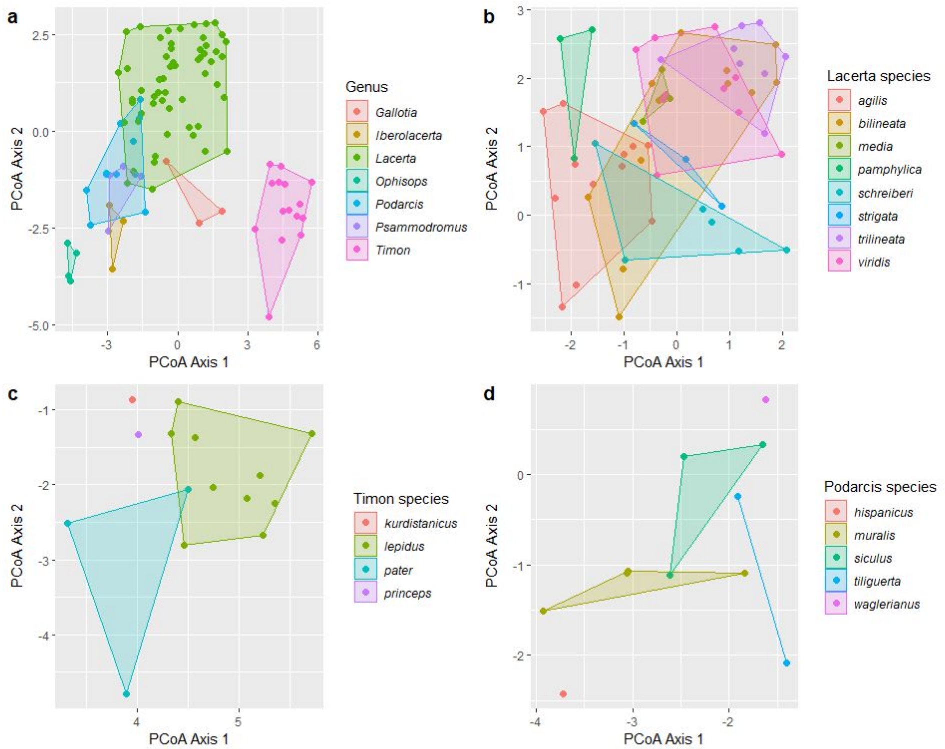
Principale Coordinates Analysis.—Principal coordinates analysis recovers a strong separation between *Timon* and all other genera, with *Gallotia*, *Podarcis*, and *Psammodromus* overlapping the *Lacerta* morphospace, and *Iberolacerta* and *Ophisops* forming separate clusters nearby (Fig. 45a). Hierarchical clustering analysis finds *Iberolacerta*, *Ophisops*, *Psammodromus*, and *Timon* to be the only genera to form statistically significant clusters with the exclusion of other genera. *Gallotia stehlini* and *G. simonyi* form a statistically significant cluster, but do not significantly group with *G. caesaris*. Although the individual species of *Lacerta* tend to cluster together, the only *Lacerta* species to cluster together significantly was *Lacerta pamphylica*, and several *Lacerta* specimens cluster with specimens of *Podarcis* rather than congeners.

Even with only specimens of *Lacerta* included, there is significant overlap between species, and there is no significant tendency for hierarchical clustering analysis to group

For Peer Review Only

specimens of a single species to the exclusion of those referred to others (Fig. 4~~5~~b). An exception is *Lacerta pamphylica*, of which all three specimens cluster together when all taxa are analysed, but this cluster does not include one of the specimens when only *Lacerta* is included in the analysis. This is probably a consequence of lacking more disparate species, which make *L. pamphylica* appear more distinct when they are included for comparison. Other species often cluster partially. For instance, *L. agilis* is split into one significant cluster of five specimens, with the other ~~7~~ seven tending to cluster insignificantly with two *L. viridis* specimens. Within *Podarcis* (Fig. 4~~5~~c), only *P. tiliguerta* forms a statistically significant cluster, with specimens of *P. muralis*, *P. siculus*, and *P. waglerianus* mixed together into a statistically insignificant cluster. Within *Timon* (Fig. ~~5~~4d), *T. lepidus* mostly forms one statistically insignificant cluster, but one specimen is recovered in a near significant cluster with *T. pater*. The sole specimens of *T. princeps* and *T. kurdistanicus* cluster together.

For Peer Review Only



~~Figure 5. Principle Coordinate Analysis based on dissimilarity scores highlighting the different genera (a), and species within *Lacerta* (b), *Timon* (c), and *Podarcis* (c). Genera can be more easily distinguished in this way than species.~~

Missing Data.—Within the matrix, missing data is distributed unequally (Table 1~~2~~), indicating that the absence of character scores in both extant and extinct species is non-random. Such a non-random distribution of missing entries ~~seems to be~~is fairly typical for morphological datasets, especially when they include extinct taxa (e.g., Smith et al. 2014; Gerber 2019). As expected, extant species have much higher completeness scores than extinct species that are nearly always represented by fragmentary specimens. Throughout the entire dataset, dental characters (quantitative and qualitative) have fewer missing entries than cranial characters (except for *Lacerta pamphylica*, which could only be scored from published figures), and cranial characters are more completely scored than postcranial characters (except for *L. viridis*, which is the sole extant species in our dataset that includes a specimen that only preserves postcranial material). In all extant species, quantitative postcranial characters are the ones with most missing entries per species, and quantitative dental characters make up the most completely scored subset (except for *L. pamphylica*). In extinct species, quantitative dental characters are not consistently the most completely scored subset: in *Dracaenosaurus croizeti* and *Pseudeumeces cadurcensis*, qualitative dental characters have the least missing entries. In the extant species, the amount of missing data ~~does seem to have a~~is slightly ~~correlation~~ correlated with numbers of specimens per species (Fig. 5~~6~~), but there is no correlation with number of characters per subset (Table 1~~2~~). However, as mentioned above, these absolute values of missing data are not necessarily correlated with the utility of the data~~-~~set for analyses of pairwise dissimilarity, which requires ~~anatomical overlap, i.e.,~~ at least two individuals scored for the same character.

~~Table 2: Completeness scores per species and partition.~~

	~~Species~~	~~OTUs~~	~~Complete~~	~~Cranial~~			~~Dental~~			~~Postcranial~~		
				~~Quant.~~	~~Qual.~~	~~Total~~	~~Quant.~~	~~Qual.~~	~~Total~~	~~Quant.~~	~~Qual.~~	~~Total~~
-	-	-	~~253~~	~~14~~	~~153~~	~~167~~	4	~~13~~	~~17~~	~~16~~	53	~~69~~
~~extant~~	~~*Lacerta agilis*~~	~~12~~	~~74%~~	~~66%~~	~~76%~~	~~75%~~	~~98%~~	~~85%~~	~~88%~~	~~57%~~	~~71%~~	~~68%~~
	~~*Lacerta bilineata*~~	~~12~~	~~68%~~	~~58%~~	~~76%~~	~~75%~~	~~98%~~	~~84%~~	~~87%~~	~~44%~~	~~50%~~	~~49%~~
	~~*Lacerta media*~~	4	~~46%~~	~~57%~~	~~53%~~	~~53%~~	~~88%~~	~~62%~~	~~68%~~	~~6%~~	~~30%~~	~~24%~~

For Peer Review Only

	Lacerta schreiberi	6	55%	51%	52%	52%	88%	63%	69%	31%	69%	60%
	Lacerta strigata	3	70%	83%	81%	81%	92%	92%	92%	31%	40%	38%
	Lacerta trilineata	8	78%	72%	81%	80%	100%	80%	85%	64%	74%	71%
	Lacerta viridis	11	66%	53%	66%	64%	91%	76%	80%	45%	70%	65%
	Lacerta pamphylica	3	52%	79%	71%	71%	67%	67%	67%	0%	0%	0%
	Ophisops elegans	4	67%	39%	81%	78%	94%	87%	88%	31%	39%	37%
	Podarcis muralis	4	68%	54%	73%	72%	94%	87%	88%	42%	59%	55%
	Podarcis siculus	3	80%	74%	86%	85%	100%	95%	96%	52%	69%	65%
	Psammodromus algirus	4	70%	64%	76%	75%	100%	92%	94%	41%	53%	50%
	Timon lepidus	10	62%	58%	61%	61%	93%	75%	79%	46%	62%	59%
	Timon pater	3	58%	43%	59%	58%	100%	87%	90%	35%	54%	50%
extinct	*Dracaenosaurus croizeti*	7	16%	23%	19%	19%	39%	43%	42%	3%	1%	1%
extinct	*"Lacerta" filholi*	4	5%	0%	4%	3%	44%	38%	40%	0%	0%	0%
extinct	*"Lacerta" siculimelitensis*	5	23%	17%	25%	24%	70%	35%	44%	10%	18%	16%
extinct	*Mediolacerta roceki*	4	7%	0%	6%	6%	44%	44%	44%	0%	0%	0%
extinct	*Plesiolacerta lydekkeri*	12	5%	5%	4%	4%	19%	15%	16%	2%	3%	3%
extinct	*Pseudeumeces cadurcensis*	8	8%	8%	8%	8%	34%	35%	35%	0%	0%	0%

Quantification of anatomical overlap shows that extant lacertids have AOIs ranging from 34% (*Lacerta media*) to 75% (*L. trilineata*) when analysing the entire dataset, whereas extinct species have values between 1% (*Plesiolacerta lydekkeri*) and 12% ("*L.*" *siculimelitensis*). COIs covering the entire dataset range from 56% (*L. media*) to 83% (*L. pamphylica*) in extant species, and from 14% (~~*Plesiolacerta*~~ <u>*P.*</u> *lydekkeri*) to 90% (*Mediolacerta roceki*) in extinct taxa (Table <u>1</u>~~3~~). The AOI and COI are slightly correlated with completeness values in extant taxa (the AOI slightly more so than the COI). Whereas the COI of extant species does not seem to correlate with the number of OTUs in a particular species, the AOI does so, even more than regular completeness values (Fig. <u>5</u>~~6~~). Overlap indices in the extinct taxa, however, show the opposite trends, with the COI being most correlated with number of specimens, AOI being stable, and completeness decreasing with higher numbers of specimens (Fig. <u>5</u>~~6~~). As ~~was to be~~ expected, extinct species generally have much lower absolute numbers of comparable characters (characters with anatomical overlap), total number of overlaps and AOI within the species compared to extant species. Total number of characters with anatomical overlap among OTUs of a particular extinct species range from 13 comparable characters (with 28 overlaps; "~~*Lacerta*~~<u>*L.*</u>" *filholi*) to 73 (with 127 overlaps; *Dracaenosaurus croizeti*). The lowest numbers in extant species are present in *L. pamphylica* with 142 comparable characters and 235 overlaps (<u>Supplementary</u> Table <u>6</u>~~3~~).

For Peer Review Only

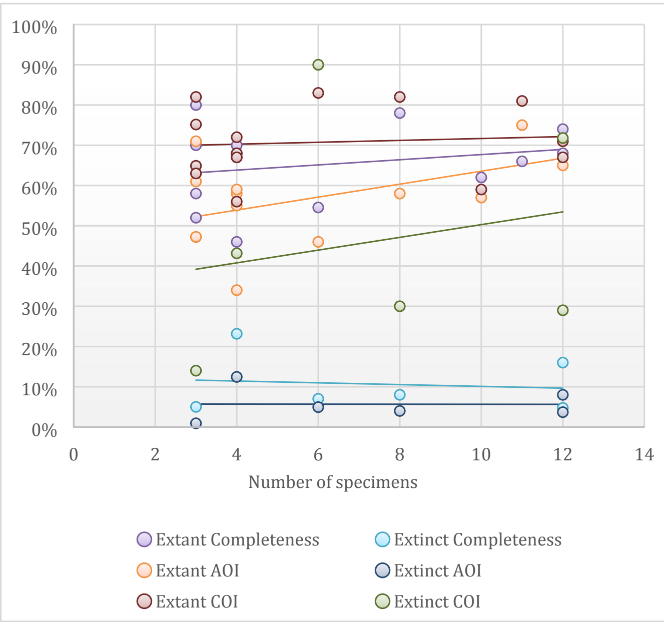


Figure 6: Correlation of average completeness score, AOI, and COI within a species (y-axis) and number of specimens per extant and extinct lacertid species (x-axis). AOI seems most correlated with sample size in extant species, but extinct species show different patterns.

For Review Only

Table 3: All Characters Overlap Index (AOI) and Comparable Characters Overlap Index (COI) within extant and extinct lacertid species in the complete dataset and partitions. CC, comparable characters; O, overlaps; OTUs, operational taxonomic units.

			Complete (253)			Cranial (167)			Dental (17)			Postcranial (69)		
-	Species	OTUs	CC	AOI	COI	CC	AOI	COI	CC	AOI	COI	CC	AOI	COI
extant	*Lacerta agilis*	12	74%251	71%	71%	75%165	72%	73%	88%17	84%	84%	68%69	64%	64%
	Lacerta bilineata	12	68%245	65%	67%	75%163	72%	73%	87%17	84%	84%	49%65	44%	47%
	Lacerta media	4	46%156	34%	56%	53%117	40%	57%	68%15	57%	64%	24%24	15%	44%
	Lacerta pamphylica	3	55%142	46%	83%	52%128	64%	84%	69%14	59%	71%	60%-	0%	0%
	Lacerta schreiberi	6	70%184	47%	65%	81%118	44%	62%	92%13	64%	83%	38%53	51%	67%
	Lacerta strigata	3	78%177	58%	82%	80%142	71%	83%	85%17	88%	88%	71%18	17%	67%
	Lacerta trilineata	8	66%233	75%	81%	64%155	77%	83%	80%17	82%	82%	65%61	67%	76%
	Lacerta viridis	11	52%245	61%	63%	71%159	60%	64%	67%17	76%	76%	0%69	60%	60%
	Ophisops elegans	4	67%205	55%	68%	78%146	67%	77%	88%15	76%	87%	37%44	21%	33%
	Podarcis muralis	4	68%216	58%	67%	72%147	63%	71%	88%16	78%	83%	55%53	40%	52%
	Podarcis siculus	3	80%219	71%	82%	85%151	77%	85%	96%16	88%	94%	65%52	52%	69%
	Psammodromus algirus	4	70%207	59%	72%	75%142	66%	78%	94%16	84%	90%	50%49	34%	48%
	Timon lepidus	10	62%242	57%	59%	61%157	56%	60%	79%17	75%	75%	59%68	53%	54%
	Timon pater	3	58%159	47%	75%	58%97	49%	84%	90%16	88%	94%	50%46	33%	50%

45

extinct	*Dracaenosaurus croizeti*	7	16%73	8%	29%	19%62	10%	26%	42%11	29%	45%	1% -	0%	0%
	"*Lacerta*" *filholi*	4	5%13	4%	72%	3%6	2%	61%	40%7	33%	81%	0% -	0%	0%
	"*Lacerta*" *siculimelitensis*	5	23%73	12%	43%	24%49	13%	45%	44%8	32%	69%	16%16	6%	25%
	Mediolacerta roceki	4	7%14	5%	90%	6%6	4%	100%	44%8	39%	83%	0% -	0%	0%
	Plesiolacerta lydekkeri	12	5%23	1%	14%	4%7	0%	12%	16%9	9%	17%	3%7	1%	12%
	Pseudeumeces cadurcensis	8	8%34	4%	30%	8%23	3%	22%	35%11	29%	45%	0% -	0%	0%

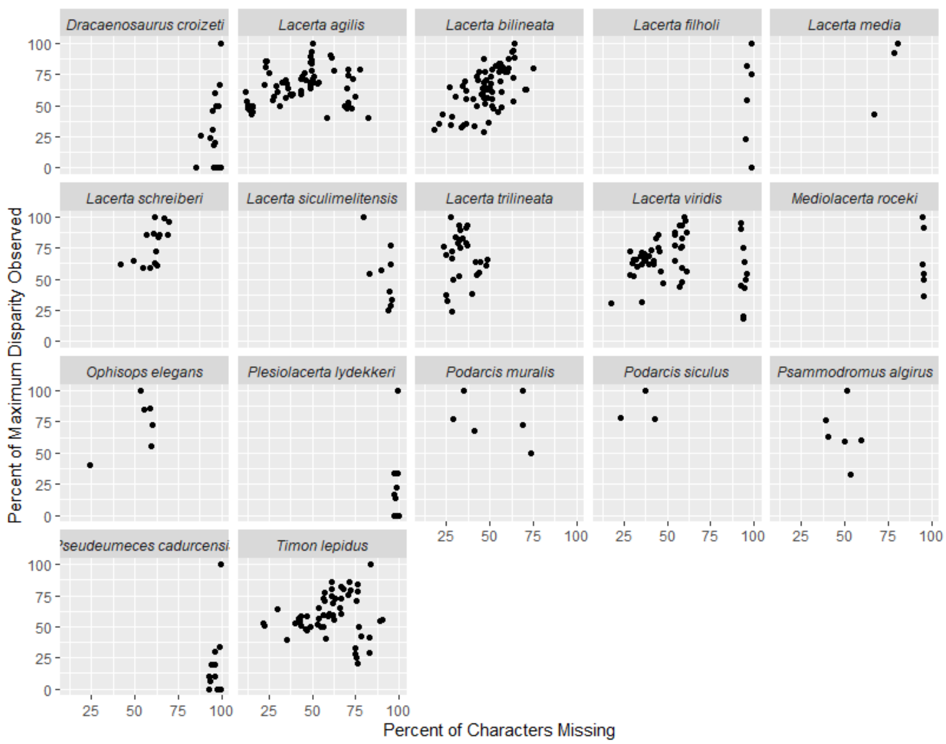
eer Review Only

Among the partitioned character sets, AOI and COI ~~values~~ are generally highest in the dental characters, and lowest in the postcranial characters, both in extant and extinct taxa. This is in part because all but two of the sampled extinct species entirely lack anatomical overlap in the postcranial partition (the exceptions are "*Lacerta*" *siculimelitensis* and *Plesiolacerta lydekkeri*). However, a lack of anatomical overlap in a particular partition does not necessarily mean that there are no characters scored in these taxa. For instance, ~~*Plesiolacerta*~~ *P.* *lydekkeri* was sampled by the most specimens of all ~~included~~ extinct species~~, ; and still~~ its ~~has the~~ lowest overall ~~values of~~ AOI and COI~~; lacking overlap here~~ result~~s~~ from different specimens having different bones preserved that cannot be directly compared. ~~In fact, comparison of the distribution of missing data and the overlap indices~~These results further highlight~~s~~ that the negative impact on dissimilarity analyses does not derive from the missing data per se, but from reduced anatomical overlap.

~~When analysing the impact of lacking anatomical overlap (and thus, pairwise comparisons) on the calculation of pairwise dissimilarity, we see that t~~The number of characters missing from a particular pairwise comparison has no consistent relationship with the recovered dissimilarity (Fig. 6~~7~~). In most taxa, dissimilarity ~~is relatively consistent for a wide range~~does not seem to be correlated with the amount of lacking anatomical overlap, though some (most notably *Lacerta bilineata*) appear to show a trend of increased dissimilarity with decreasing overlap, and several show the highest dissimilarity between 150 and 200 missing characters (of a total of 253) whereas comparisons with even fewer characters show less overall dissimilarity. Minimum observed pairwise dissimilarity, on the other hand, did not show any correlation with lacking anatomical overlap. Thus, lacking anatomical overlap does not seem to have a great impact on average and minimum pairwise dissimilarity scores, whereas the highest dissimilarity

For Peer Review Only

seems to occur at a level of about 60- to 80% of absent anatomical overlap. However, smaller amounts of anatomical overlap (as observed in our extinct species) does not seem to artificially inflate dissimilarity.



~~Figure 7: Distribution of missing characters from the pairwise comparisons relative to percent of maximum dissimilarity observed in extant and extinct lacertid species. There does not seem to be a general trend of higher dissimilarity or ranges of dissimilarity with more missing characters.~~

Sample ~~s~~S*ize*.—At a sample size of two individuals, maximum, mean, and minimum dissimilarity are equal for each replicate as only one comparison is performed. As sample size increases, maximum and minimum dissimilarity diverge, with increasing numbers of replicates finding the observed maximum or minimum dissimilarity, and mean dissimilarity stabilizes (Fig. 78). By a sample size of four or more individuals, the distributions of maximum and minimum dissimilarity do not overlap each other and are almost distinct from the range of mean dissimilarities, although the variance remains high. With seven or eight individual specimens sampled, maximum and minimum dissimilarity do not overlap mean dissimilarity anymore, and variance in mean, minimum, and maximum values decreases considerably.

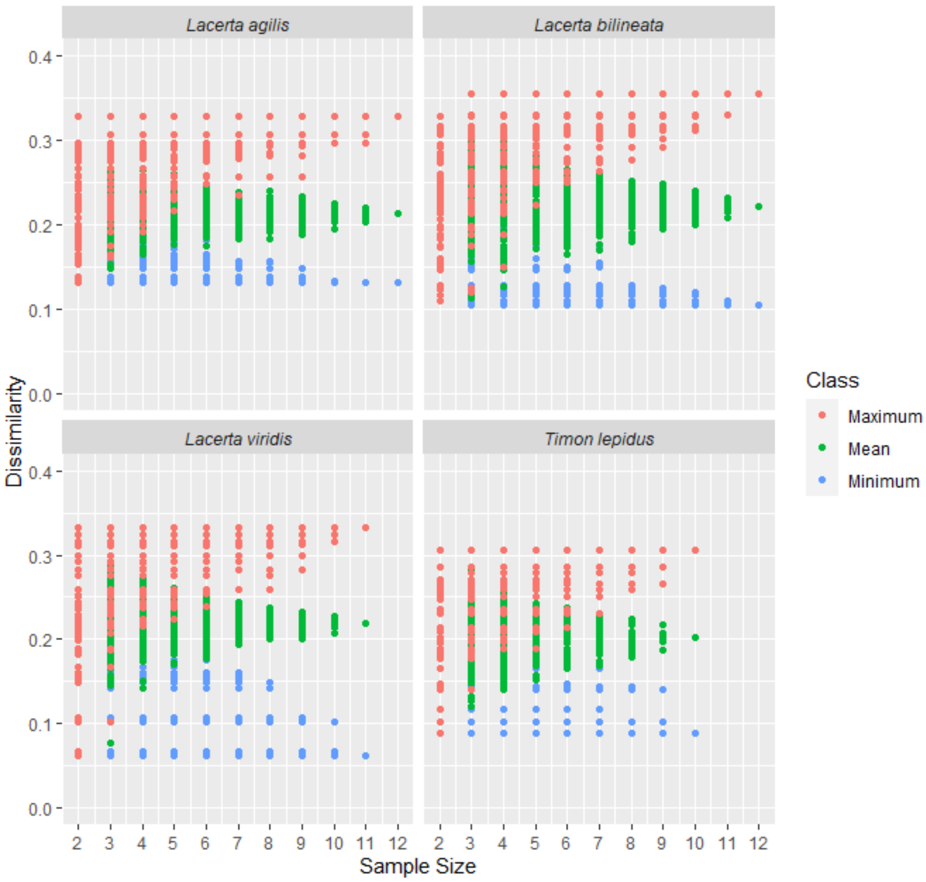


Figure 8: Observed dissimilarity values relative to sample size subsampled in the four best-represented lacertid species in our dataset. Maximum and minimum values do not overlap significantly with four or more specimens, whereas maximum and minimum values also do not overlap with average values when sampling seven or more specimens.

Discussion

Extant Lacertid Species are Comparable Units

All 14 extant lacertid lizard species we analysed for intraspecific variability display comparable degrees of pairwise dissimilarity, with only three outlier taxa being significantly more, or less, dissimilar than some (but not all) other species. Assuming the identification of the specimens referred to these species was mostly based on external morphology and provenance, and assuming it is correct, it is reassuring to see that all these species comprise a comparable degree of skeletal dissimilarity. Moreover, the included species vary in body size, ecological niche, and phylogenetic history, the species were represented by divergent sample sizes, and specimens showed different degrees of anatomical overlap. Notwithstanding these differences in their biology, sampling procedure, and available data, pairwise dissimilarity remains consistent. The three partial exceptions are *Lacerta media*, *L. pamphylica*, and *Ophisops elegans*.

Lacerta media was found to have a significantly higher intraspecific variability than eight other species within our dataset, whereas no significant difference was found with five other species. It was sampled by four specimens in our dataset and also has the lowest anatomical overlap scores over all characters as well as within the cranial, dental, and postcranial subsets, and it was sampled by four specimens in our dataset (Tables 1, 31). The high overall dissimilarity is driven by high variability in qualitative cranial and postcranial characters

For Peer Review Only

(Supplementary Table 54). *Lacerta media* is less, or similarly, variable than many other species in character subsets that generally show high variability (e.g., quantitative cranial and qualitative dental characters). Our findings could be a result of sampling of specimens from distinct lineages within *L. media* currently recognized as subspecies (probably *L. m. media* in Turkey and *L. m. wolterstorffi* in Israel; Ahmadzadeh et al. 2013), suggesting that their morphological dissimilarity would support distinction at species level of at least the northern and southern clades recognized by Ahmadzadeh et al. (2013). However, aAdditional sampling of the various subspecies (which all occur relatively close to each other around the eastern coast of the Mediterranean) as well as a more complete sampling of the entire geographical range of *L. m. media* (which reaches as far east as northern Iran and the Caspian Sea) would provide an interesting case study to understand if morphological variability corresponds to genetic or geographic distance in this species. If geographically widespread species within a certain clade would also be morphologically more variable than to address this particular question will be necessary to understand if this distinct variability pattern is real or an artifact of low sample size and limited anatomical overlap.other species within that clade, any method for species delimitation based on our results would have to normalize disparity values based on geographical distance among specimens.

Lacerta pamphylica has a significantly lower variability compared to four other species (no significant difference is found with nine other species), although the three specimens sampled cover a juvenile and an adult male and female. These were scored based on the figures provided by Čerňanský and Syromyatnikova (2019; see Supplementary Table 1), which only figured part of the skull, so no postcranial material could be compared. *Lacerta pamphylica* has a relatively small geographic distribution and no distinct lineages are known below the species level (Ahmadzadeh et al. 2013; Kornilios et al. 2020), which may be a reason for low overall

For Peer Review Only

osteological variability. This low overall variability is mostly driven by a low variability in the qualitative cranial characters (0.0991 ± 0.0482; Supplementary Table 54), which constitute the majority of the included characters. This is the lowest value of intraspecific variability among qualitative cranial characters for all extant species; it is significantly lower than the binned qualitative cranial dissimilarity of the other extant species. The absence of scores for postcranial characters may have artificially increased the impact of this character subset on the entire values, but given the comparatively low dissimilarity, it remains possible that there is a genuine signal that should be further explored with more extensive sampling.

Ophisops elegans has a comparably low intraspecific variability~~, as~~ to that in *Lacerta pamphylica*, being significantly different from six, but similar to seven other species. It is sampled by four specimens and has intermediate levels of completeness and anatomical overlap. Three specimens are from the Greek island of Samos, and one is from Armenia, so they probably represent specimens of the subspecies *O. e. macrodactylus* and *O. e. persicus*, respectively (Montgelard et al. 2020). No information is available on their sex and maturity. As in the other outlier species, the pattern of variability is intriguing, especially because Montgelard et al. (2020) proposed to elevate *O. e. persicus* to species level (so we would have two species represented in our sample), but the low sample size, with three of four specimens coming from Samos Island, casts doubt on this pattern being a genuine representation of intraspecific variability across the entire species. Low overall variability of *O. elegans* is driven by a low dissimilarity among qualitative cranial characters, as in *L. pamphylica*, but it also has very low to non-existent variability in quantitative cranial and dental characters and qualitative dental and postcranial characters (Supplementary Table 54). Quantitative cranial and qualitative dental characters are otherwise more variable than average, so their low values in *O. elegans* is peculiar. Given that

For Peer Review Only

this was the only representative of the lacertid subclade Eremiadini (and that the sampled specimens may represent two distinct species), one might take this as an indication that patterns within Eremiadini are different from other lacertids, but additional species, subspecies, and specimens will have to be sampled in this clade to confirm this. At present, we cannot confidently exclude that the significant differences in intraspecific variability between these three outlier species and some (though not all) other species are artefacts of low sample size and restricted anatomical overlap.

The results from our studies corroborate that current species delimitation is generally robust in the extant species we analysed, and that these taxa do not suffer considerably from the species comparability problem. This stability~~, in turn,~~ suggests that osteological intraspecific variability can be used as a proxy for other secondary defining properties and may be suitable for species delimitation even in the absence of autapomorphic osteological features in a particular species (as is the case in some of the analysed lacertids; Villa et al. 2017). Hence, these values may also be of use to delimit extinct lacertid species. However, our results in the analysis of mean pairwise dissimilarity in extinct lacertid species shows that some extinct species we examined had divergent dissimilarity values compared to extant species.

Reasons for Incongruence in Dissimilarity Between Extant and Extinct Lacertids

The reasons for the diverging results in intraspecific osteological variability in the sampled fossil taxa ~~are likely to be manifold. They~~ could include matrix and OTU construction, missing data, the inclusion or exclusion of sexual dimorphisms and/or ontogenetic differences, and differing interpretations by researchers of intraspecific variability. ~~All these factors may effectively lead to incongruent species delimitation.~~ Moreover, palaeontology provides a unique opportunity to study species through time, which while generally beneficial, could lead to time-

For Peer Review Only

averaging – i.e., fossils of a species lineage sampled across a few thousand or tens of thousands of years might include more 'evolution' and thus be more dissimilar than an extant specimen set derived from a single time plane.

Matrix Construction.—The effect of matrix construction on disparity analyses has been discussed in detail by Lloyd (2016) and Gerber (2019). Herein, wWe followed their recommendations that the matrix should include as many characters as possible, irrespective of their homoplasy rate (see Dataset – Character sampling). Additionally, we tested the impact of OTU construction on our dataset by including two conceptual types of OTUs. Generally, the species we compared in this analysis were scored at specimen-level, with one exception ("*Lacerta*" *siculimelitensis*), which comprises locality-level OTUs, so it is possible that some of the observed variability among single specimens is obscured. Given that polymorphic characters in these locality-level OTUs were scored based on a frequency scoring approach (see Dataset – Extinct taxon and specimen sampling),). it is possible that some of the observed variability among single specimens is obscured. OTU construction may thus have artificially lowered intraspecific osteological variability in "*L.*" *siculimelitensis* (see below for a detailed assessment of this species). Given that observed dissimilarity with these locality-level OTUs of "*L.*" *siculimelitensis* (0.1364 differences per character scored) is not significantly different from that of extant species, it is possible that real intraspecific variability in "*L.*" *siculimelitensis* would be higher.

Missing Data and Low low Sample Size.—The fossil record is notoriously incomplete, resulting both in OTUs with an increased amount of missing entries due to incomplete preservation of the fossils and a generally low sample size within species. Missing entries in our dataset indeed result in much lower numbers of anatomical overlaps in extinct versus extant

For Peer Review Only

species (Fig. 6~~7~~; Table 1~~3~~), which can have a substantial impact on pairwise dissimilarity analyses (Smith et al. 2014; Gerber 2019).

~~A~~However, non-randomly distributed missing entries, as present in our dataset, seem to have a less significant impact on disparity analyses than randomly distributed ones ~~shown by simulations, which removed data mimicking observed patterns during degradation and fossilisation~~ (Smith et al. 2014)~~, non-randomly distributed missing entries seem to have a less significant impact on disparity analyses than randomly distributed ones~~. This pattern was partially confirmed by our ~~own~~ simulations. ~~Visualizing the dissimilarity of all pairwise comparisons for all taxa plotted against the number of missing characters~~ Given~~shows~~ that most taxa display either a flat relationship between low anatomical overlap and dissimilarity, or show the highest dissimilarity with intermediate amounts of anatomical overlap, ~~making~~ it is unlikely that the high number of missing scores for extinct taxa is the only ~~reason~~ factor generating an artificially high or low dissimilarity. Furthermore, our “simulated fossil” datasets only found two extant species with artificially removed character scores to approximate intraspecific variation patterns seen in those four extinct partner species that diverged from the general average ~~observed~~ (Fig. 3~~4~~). *Mediolacerta roceki* and *Pseudeumeces cadurcensis* remained significantly different compared to their extant partner species with an equivalent number of removed character scores.

The significantly higher variability of *Mediolacerta roceki* and the significantly lower intraspecific variability of *Pseudeumeces cadurcensis* are not solely artifacts of missing data and low anatomical overlap, but include a true signal of the osteological variability that our dataset captures despite the incompleteness of the fossil record. These two species have intermediate values of completeness and AOI compared to the other extinct species~~. This~~, ~~suggests~~ suggesting

For Peer Review Only

http://mc.manuscriptcentral.com/systbiol

that analysis of weighted mean pairwise dissimilarity can yield meaningful results even at high levels of missing data, but that there is no clear correlation between completeness and significance of the result. Simulations as proposed in our study will be paramount in future assessments to evaluate if the recovered signal is in fact true or if it is ~~likely~~ impacted by specimen and species incompleteness.

The low sample size in our dataset for both extant and extinct species (up to a maximum of twelve specimens per species) may seem problematic at first, but does not appear to impact our results considerably, corroborating earlier studies that showed little effect of low sample size on mean pairwise dissimilarity analyses (Foote 1992b, 1993; Ciampaglio et al. 2001). In molecular specimen-level phylogenetic analyses, genetic variation is thought to be covered sufficiently to yield accurate trees, if ten specimens per species are included in an analysis (Saunders et al. 1984; Carstens et al. 2013). For the study of morphological variation, ~~workers~~ it has been~~have~~ suggested that at least eight (Roth 1992), ten (Ciampaglio et al. 2001), or 20 (Cope and Lacy 1992) specimens need to be sampled to cover a significant portion of the actual variability present in a species.

Our sensitivity analyses ~~with *Lacerta agilis*, *L. bilineata*, *L. viridis*, and *Timon lepidus* (Fig. 8)~~ suggest that mean dissimilarity values do not change significantly when analysing four or more specimens, and that minimum and maximum values do not overlap recovered mean values when sampling at least seven or eight specimens (Fig. 7~~8~~). Thus, taxa represented by seven or more individuals in our dataset ~~while poorly sampled species in our dataset may have overestimated minimum dissimilarity and underestimated maximum dissimilarity, their mean dissimilarities are~~ probably show representative mean dissimilarity and ~~and~~ variance that are comparable ~~with those of well-sampled species~~among each other~~. This is especially true for taxa~~

For Peer Review Only

~~represented by seven or more individuals in our dataset~~, whereas some doubts remain for those species sampled by ~~four or~~ fewer specimens – especially those with divergent results (~~see the discussion of~~as is the case in the outlier species discussed above). This result ~~is promising in many ways. It~~ shows that low sample sizes should not be regarded as impeding research on morphological dissimilarity, and that the low number of available osteological specimens in museum collections (Bell and Mead 2014) is not necessarily a barrier to applying the approaches advocated here. However, it will be interesting to see studies with tens to hundreds of specimens of a single species in future; ~~The possibility of obtaining meaningful results from osteology may also reduce the necessity~~with the ever-increasing availability of ~~producing~~ CT scans of wet-specimens, ~~although inclusion of osteological data from CT scans (and of course soft tissue) would still be beneficial as they~~ provid~~e~~ing a wealth of additional information that is not visible in skeletal preparations, this should only be a matter of time. ~~Last, but not least, it shows that concerted efforts to elucidate the osteology of modern taxa in a comparative way, as is being done for European lizards (Barahona and Barbadillo 1997, 1998; Barahona et al. 2000; Klembara et al. 2014; Villa et al. 2017, 2018; Camaiti et al. 2019; Čerňanský and Syromyatnikova 2019; Čerňanský et al. 2019; Villa and Delfino 2019), yield meaningful data that can be used to study both diversity and disparity in extant and extinct taxa. Although~~ In sum, an inclusion of seven or more specimens per species is advisable, but dissimilarity analyses ~~can yield significant results even at~~ with lower sample sizes~~, but these~~ may yield meaningful results ~~will have to be~~ if they are carefully assessed for potential shortcomings ~~when studying ranges of variability~~due to low sample size.

Uneven Sampling of Ontogeny and Sexes.—Intraspecific variability is greatly affected by sexually dimorphic features and ontogenetic changes. ~~However, i~~In a complete sample including

For Peer Review Only

members of both sexes and from various ontogenetic stages, these two factors should probably not have a large impact on the mean dissimilarity value, although they may increase the observed ranges in dissimilarity considerably. Given that sex and ontogenetic stage are not known for many fossil specimens, especially if they are only partially preserved, it remains difficult to quantify the amount of variability that is absent in the extinct dataset. Hence, the expected impact on studies of extinct species, if any, would be a lower range in variability compared to more completely extensively sampled extant species, which is similar to the effect of low anatomical overlap and sample sizes in general. This would be especially the case if sexual morphs, instead of being recognized as different sexes of a single species, are erroneously treated as distinct species given their diverging morphology, something that is very difficult to assess in palaeontological samples (Wiley 1978; Tschopp and Upchurch 2019). In any case, sexual dimorphisms and ontogenetically variable characters often affect certain character complexes. In the case of sexual dimorphism, these are often restricted to soft tissue morphology associated with the reproductive tract, which is generally not preserved in fossils, or to features bearing a display function that may or may not have osteological correlates (and if they have, they may not be recognised as such in fossils; Mallon 2017). Restricting our dataset to osteological characters, and analysing mean pairwise dissimilarity over a complete set of cranial, dental, and postcranial characters can probably be expected to reduce the confounding impact of these types of intraspecific variability – even though we also deliberately included characters that are known to be variable between sexes and through ontogeny.

Time-averaging in Fossils.—Time-averaging can result from a sampling of fossil specimens from different geological ages (even if only thousands of years, a time span too short to be recognisable in many geological contexts). Hence, fossil samples may combine variability

For Peer Review Only

that had accumulated over time while the species was adapting to changing environmental ~~circumstances~~ conditions through natural selection (Simpson 1937, 1951). The inclusion of specimens from potentially different evolutionary stages within the same species would be expected to ~~particularly~~ increase ~~the~~ mean dissimilarity as well as range of variability because such variability cannot be observed in samples of extant species (Kelley 1986). The resulting higher observed variability in time-averaged fossil samples could counteract or even overwhelm the impact of missing data and uneven sampling of sexes and ontogenetic stages. ~~Whereas we can imagine the impact of time-averaging on the range of variability, its impact on the mean dissimilarity within a species is difficult to assess, and probably depends on the presence of selection for particular traits, which might reduce overall mean dissimilarity over time. However, we would assume that these changes only affect a reduced number of characters, so that the impact on dissimilarity values calculated from a large set of characters from all body regions remains minimal.~~

The Species Comparability Problem in Extinct Lacertids

The extinct species examined have more variable dissimilarity scores compared to extant species, suggesting that a species comparability problem occurs both between extant and extinct species, as well as among extinct species only. In two extinct species (*Mediolacerta roceki* and *Plesiolacerta lydekkeri*), these intraspecific differences are as pronounced or larger than intrageneric dissimilarity in extant genera, supporting earlier claims that these two extinct taxonomic species units are more inclusive than extant taxonomic species units and more closely compare to genera. At the same time, the other four extinct lacertid species are equally, or less, variable than extant species (most importantly *Dracaenosaurus croizeti*). ~~Hence, overall~~

For Peer Review Only

osteological variability in the six extinct species we examined yielded divergent results, which probably have a number of potential underlying causes.

All aspects discussed above probably impacted our mean pairwise dissimilarity values obtained from the sampled extinct species, but it remains difficult to estimate the contribution of each of those factors, especially in the two species that remain significantly different even from their extant partner species in the simulations. The low mean dissimilarity in *Pseudeumeces cadurcensis* and the large variability in *Mediolacerta roceki*, in particular, indicate that these results are a consequence of taxonomists holding diverging views on the "acceptable" or "typical" amounts of intraspecific osteological variability within a species. What this means for the extinct species analysed here is discussed below.

Dracaenosaurus croizeti.—The observed variability in *D. croizeti* is significantly lower than any other species we analysed, be it extant or extinct. This is true for the whole dataset as well as the modules of qualitative cranial characters (the majority of characters in the dataset) and qualitative and quantitative dental characters (Supplementary Table 45). However, vVariability is not significantly lower compared to the values obtained in its extant partner species (*L. acerta bilineata*) with artificially decreased anatomical overlap (although the mean value remains much lower; 0.1182±0.1425 in *D. croizeti*; 0.1632 ±0.1297 in *L. bilineata*). Thus, we cannot completely rule out that low anatomical overlap is driving these discordant values. However, additional factors may have played a role, especially when considering that time-averaging would have worked against the low values (the sampled specimens cover 900'000 years, whereas the species (as currently understood) may have been present for up to 16 My, depending on the actual strata where it was found in Quercy (France; Böhme and Ilg 2003). Another reason for the low values might be the fact that This is a very considerable amount of

~~time in terms of generations (i.e., we would expect variability to be higher than in extant species because of time-averaging) although it is less than what is covered in the other included species (see below).~~ D~~racaenosaurus~~. *croizeti* is highly specialized, with its strongly enlarged posterior teeth adapted for crushing and the generally stout skull and jaws for the attachment of strong musculature (Hoffstetter 1944; Müller 2004; Čerňanský et al. 2017). ~~It is possible that s~~Such an advanced specialization was possibly favoured by strong natural selection~~, which~~ that ultimately constrained aspects of morphology and so reduced intraspecific variability, especially in cranial and dental characters. Postcranial material of the sampled *D. croizeti* specimens was excluded from contributing to the calculation of intraspecific variability because of the absence of anatomical overlap in this module (Table 1~~3~~), which was probably less ~~impacted~~ constrained in morphology by this feeding specialization (and may thus have increased the dissimilarity) ~~than cranial and dental features~~. ~~Thus, natural selection favouring strong adaptation to durophagy may be partly responsible for constraining the morphology and therefore the low variability observed in this species.~~ Additionally, researchers may have been overly cautious in referring specimens with slightly diverging morphologies to this species, ~~i.e., researchers~~ thereby applying a more strictly typological species concept when identifying fossils. This would suggest that additional material now referred to "*Dracaenosaurus* sp." should indeed be assigned to *D. croizeti* as well. ~~These s~~Specimens that were not identified to species level are all from the same localities in France and Germany that also produced specimens referred to *D. croizeti* (Böhme and Ilg 2003; Čerňanský et al. 2016a), further supporting our suggestion. ~~However, inclusion of all this material in our dataset would be required to understand how this impacts variability, completeness, and statistical significance.~~

For Peer Review Only

"*Lacerta*" *filholi*.—No significant difference was found in recovered intraspecific variability of "*L.*" *filholi* compared to extant species, as well as compared to its extant partner species (*Psammodromus algirus*) in both the original dataset and the simulated dataset with artificially reduced anatomical overlap. This indicates that the species "*L.*" *filholi* may represent a unit comparable to extant species, although the included specimens cover a time span of approximately 1.2 Myr~~, and all referred specimens would cover 11.9 My~~ (Böhme and Ilg 2003). We interpret our results with caution because the species is represented in our dataset by very few, disarticulated specimens, and there may be much more variability occurring in the entire duration of the species as currently understood. Only 13 characters could be compared in this species, but ~~the high COI of 72% shows that~~ the few comparable characters were shared in several ~~of the~~ specimens. These are almost entirely restricted to dental and mandibular features (the specimens referred to this species by Augé (2005) only include dentaries, maxillae, a few premaxillae, and a coronoid), and it remains to be seen if other cranial and postcranial material would alter the observed variability. In any case, the calculated intrageneric variability with other *Lacerta* specimens was found to be significantly larger than normal intrageneric variability within extant taxa, and even exceeded most of the recovered dissimilarity scores calculated between extant genera (Supplementary Table ~~4~~5). This finding ~~suggests~~ supports earlier studies suggesting that the referral of this species to the genus *Lacerta* is questionable ~~, as was also recognized by~~ (Augé ~~(~~2005~~) and~~; Augé and Hervet ~~(~~2009; Wencker et al. 2021)~~, who consistently referred to this species as *Lacerta* s.l. *filholi*~~.

"*Lacerta*" *siculimelitensis*.—As for "*L.*" *filholi*, also "*L.*" *siculimelitensis* is comparable to extant taxa in its intraspecific osteological variability, although it has a relatively low dissimilarity score (0.1364 ± 0.0641; Supplementary Table ~~4~~5). The five OTUs included in the

For Peer Review Only

present analysis span a time range of 1.72 Myr (Delfino and Bailon 2000; Böhme and Ilg 2003; Tschopp et al. 2018b~~), and~~) and occur in southern continental Italy and on the islands Sardinia, Sicily, and Malta. The relatively low variability probably underestimates true dissimilarity due to the construction of the locality-level OTUs used in our analysis, so we may expect higher values (i.e., values more closely matching the extant ~~average~~mean) being present if individual specimens were scored separately. Thus, we expect this species to be comparable to extant species. ~~Attribution of this species to the genus *Lacerta* is somewhat better supported than for "*L.*" *filholi*, but i~~Intrageneric variability observed in "*L.*" *siculimelitensis* was found to be higher than in most extant species (although less so than in "*L.*" *filholi*), suggesting that an attribution to a distinct genus may be better supported by morphology~~, although this will have to be confirmed by phylogenetic analysis~~.

Mediolacerta roceki.—This species is significantly more variable than extant species. It exceeds variability of extant species in almost all character modules that could be analysed~~, though~~ (several of them were scored too incompletely to yield any data~~;~~ ~~(~~Supplementary Table ~~5~~4). *Mediolacerta roceki* is also significantly different from its extant partner species *Podarcis muralis*~~, which did not end up with significantly higher variability~~ when deleting the same characters as are missing in *M. roceki* (Fig. ~~3~~4). Thus, reduced anatomical overlap alone cannot explain the difference in dissimilarity. The time covered by the included specimens is probably around 1.2 Myr (Böhme and Ilg 2003), and thus comparable to, or less than, "*L.*" *filholi* and "*L.*" *siculimelitensis*, which have a significantly lower intraspecific variability. *Mediolacerta roceki* ~~This species~~ was initially defined based on a dentition that is intermediate between the conditions in "*L.*" *filholi* and the more clearly amblyodont *Amblyolacerta dolnicensis* (Augé 2005). It is possible that this differential diagnosis is too vague to unambiguously identify lacertid tooth-

For Peer Review Only

bearing bones, so some of the referred specimens may actually belong to the less or more amblyodont species, rather than to *M. roceki*.

Plesiolacerta lydekkeri.—As with *Mediolacerta roceki*, intraspecific osteological variability was found to be significantly higher in *Plesiolacerta lydekkeri* compared to extant species. The high variability seems to be mostly driven by it having by far the highest dissimilarity in qualitative dental characters, which are already among the most variable characters in our dataset – other character modules that could be analysed show comparable values to extant species (Supplementary Table ~~4~~5). ~~However, i~~Its extant partner species in the simulated dataset (*Lacerta agilis*) did approximate the pattern observed in *P. lydekkeri* when deleting nearly the exact same characters as those missing in *P. lydekkeri* (Fig. 3~~4~~). Thus, the high variability in the teeth of the sampled *P. lydekkeri* specimens may not be a true signal. Indeed, although we scored 12 specimens, its completeness score is the lowest among the extinct species (together with "*L.*" *filholi*; Table 1~~2~~), and only 35 overlaps (in 23 characters) and thus pairwise comparisons occur between these 12 specimens (resulting in an AOI of 1%; Table 1~~3~~). ~~Six out of the 12 specimens preserve dental material, but only 15% of the qualitative dental characters could be scored, with 11 overlaps in total.~~ Additionally, *P. lydekkeri* could only be scored for 4% of qualitative cranial characters, which generally drive average mean dissimilarity scores within other species. The paucity of available data in general, and of data from the apparently most relevant skeletal module, is probably the reason why intraspecific variability was found to be higher than interspecific variability (which compares *P. lydekkeri* specimens with specimens from other species, so that the number of comparable characters is much higher; Fig. 3~~4~~). This surprising pattern also holds true among qualitative dental characters, the module that is mostly responsible for driving the values observed in *P. lydekkeri* (Supplementary Table

For Peer Review Only

4~~4~~5)~~. This unexpected result~~, which is another indication that the value obtained within *P. lydekkeri* represents an outlier far from the true mean dissimilarity value of the species~~; all other species follow the expectation that intraspecific variability is lower than interspecific variability~~ (Fig. 2b~~1~~). ~~Nevertheless, the fragmentary preservation of the referred specimens does not necessarily explain the recovered high variability in qualitative dental characters. T~~Additionally, the high variability in the dentition of *P. lydekkeri* could be a result of time-averaging; the included specimens cover a period of 4.2 Myr (~~the entire species seems to have a temporal range of 10 My;~~ Böhme and Ilg 2003), the highest of all extinct species represented in our dataset. ~~Alternatively, or additionally~~Moreover, this result could reflect the fact that the holotype – in contrast to almost all other extinct lacertid species – does not include any cranial material but consists of a relatively large dorsal vertebra (Hoffstetter 1942; Čerňanský and Augé 2013). Consequently, and because no articulated specimen is currently known, most of the referred material was probably assigned to the species based on size instead of shared apomorphic features. The high dental variability then would suggest that more than one large-sized lacertid was present in the Oligocene of Europe, but additional sampling (and probably the find of an at least partially articulated skeleton) would be required to test this in detail. ~~Combining single bones into locality-level OTUs as we have done in "*L.*" *siculimelitensis* could possibly help to understand this variability but would also run the risk of combining elements of different species and artificially masking interspecies variability as intraspecific.~~

Pseudeumeces cadurcensis.—This species is the second in our dataset with a significantly lower intraspecific variability compared to extant taxa. It is also significantly different from its extant partner species ~~(~~*Lacerta trilineata*~~)~~ when simulating missing data. The sampled specimens of *P. cadurcensis* cover a time span of 1.2 to 5 Myr (depending on the ~~strata~~

For Peer Review Only

stratum that yielded the historical material ~~was found,~~ which was not reported), which is thus comparable to the other extinct species. Like *Dracaenosaurus croizeti*, *P. cadurcensis* is a strongly amblyodont taxon, although slightly less so than the former. Thus, the same considerations regarding specialisation and strong stabilising selection leading to lower variability also ~~applies~~ apply to this species; the fact that *D. croizeti*, with stronger amblyodonty, is less variable than *P. cadurcensis* may add further support to this hypothesis. However, it is also likely that material identified as *Pseudeumeces* cf. *cadurcensis* from Herrlingen in Germany (Čerňanský et al. 2016a) can be assigned to the species, and possibly even material currently referred to *Pseudeumeces* sp.

Inconsistent Morphological Species Delimitation~~,~~ and its Effects~~, and how to Overcome them~~

The differences in intraspecific dissimilarity seen in the extinct lacertids ~~lizards in our dataset~~ indicate that species delimitation approaches are not always consistent between neontology and palaeontology, even though most specimen identifications were probably based on morphology. In at least two of the six extinct species we sampled, low anatomical overlap did not significantly skew the recovered dissimilarity values. *Pseudeumeces cadurcensis* is significantly less disparate than any sampled extant taxon, indicating that palaeontologists have been overly strict when referring specimens to this species, whereas *Mediolacerta roceki* is significantly more variable, suggesting that some specimens referred to this species should be assigned to other taxa.

Our results indicate that ~~different researchers appear to have divergent assumptions regarding the expected skeletal dissimilarity within extinct species. This suggests that~~ the assessment of Wiens and Servedio (2000) and Wiens (2007) that there has been little progress in the methodology of species delimitation based on morphology, still holds true today. This could

For Peer Review Only

partially result from the fact that the taxonomy of extant species continues to change with the identification of cryptic lineages based on phylogenomic approaches (e.g., Ahmadzadeh et al. 2013; Kornilios et al. 2020; Montgelard et al. 2020), as is also the case in many other vertebrate clades (Brochu and Sumrall 2020). It is difficult to keep up with the pace of these phylogenomic taxonomic revisions when analysing morphological disparity and intraspecific variability, because acquisition of significant amounts of data takes time and often requires specimen loans or collection visits (Brochu and Sumrall 2020). However, if intraspecific skeletal dissimilarity values among modern species are consistent, this has great potential to help systematists to develop and apply morphological species delimitation in the future, and thereby overcome the species comparability problem.

It is important to avoid divergences between extinct and extant species: ideally, we need to render the "species" a comparable taxonomic unit in both fossil and recent datasets (Barnosky et al. 2011; Brochu and Sumrall 2020). ~~While Barnosky et al. (2011) highlighted the important issues relating to the use of palaeodiversity data to inform our understanding of current biodiversity trajectories, the species comparability problem has deleterious effects on an even wider array of biological studies.~~ Attempts to reconstruct the diversity of taxa through deep time are fundamental to palaeobiology, being used to identify radiations and extinctions that can then be correlated with intrinsic and extrinsic factors (e.g., Mannion et al. 2015; Tennant et al. 2016). Yet inconsistent taxonomic practices might inflate or deflate species counts in particular time bins, geographic regions, or clades, in ways that create noise or even artefactual patterns, such as the so-called Pull of the Recent, which summarises potential biases leading to higher diversity in extant compared to extinct taxa (see e.g., Raup 1972; Sahney and Benton 2017). ~~Essentially, i~~If an extinct species, as a taxonomic unit, would be comparable to an extant genus~~, which includes~~

For Peer Review Only

including several species, observed patterns of species diversity, speciation rate, species longevity, and others, could simply reflect inconsistency of what we mean by species today and in the fossil record. Fortunately, this does not appear to be the case in all lacertids, but it remains to be seen if this also applies to other vertebrates. Nevertheless, the availability of large amounts of comparative data of various types (e.g., DNA, soft tissue, ecology, etc.) to establish and delimit species living today, potentially leads to the recognition of many extant species that cannot be diagnosed using fixed, apomorphic skeletal features and thus cannot be recognised in the fossil record, resulting in lower numbers of extinct compared to extant taxa (Brochu and Sumrall 2020). In fact, ~~it has been shown in~~ small North American mammals ~~that~~ show an ~~observed~~ apparent increase in diversity from the Holocene to Modern times, but this results from the presence of several extant species recognised based on molecular or soft tissue characters only, so the apparent diversity increase solely reflects such a taxonomic bias (Carrasco 2013). Another problem stems from the need to adjust for the uneven sampling of the fossil record when assessing changes in palaeodiversity. Methods aimed at ameliorating the effects of uneven sampling of the fossil record and other biases depend on our ability to accurately identify distinct species, assign specimens to species, and count species, since such data affect parameters such as Goods U in SQS (Alroy 2010) or the number of samples per time bin in TRIPS (Starrfelt and Liow 2016). A similar case can be made for historical biogeographic studies: the spatiotemporal ranges of species are required for such analyses (e.g., Matzke 2013, 2014; Poropat et al. 2016; O'Donovan et al. 2018; Xu et al. 2018) and many less quantitative (i.e., narrative) approaches to palaeobiogeography base their inferences on the ranges of notional species. It is common practice, for example, to infer that two geographic regions are likely to have been in contact (or at least linked by a viable dispersal route) if they share species in common – this implies gene

For Peer Review Only

flow and therefore continuity of areas and populations. Clearly, such palaeobiogeographic analyses are likely to produce incorrect or distorted results if the paleontological species units they use have been recognized in an inconsistent manner with respect to geographic and/or temporal ranges. Thus, the development of data sets in which the equivalence or comparability of its species units has been assessed and standardised as much as possible, is vital if we are to ensure that they do not obscure true macroevolutionary or sampling bias patterns ~~(see also Carrasco 2013)~~.

~~Our finding that skeletal variability is consistent within extant lacertid species that were confirmed to be "species" under a variety of species delimitation approaches means that osteological variability among fossil skeletons of extinct species can be equally assessed. Averages of mean pairwise dissimilarity and their standard deviations in extant species can be applied to delimit extinct species of the same clade.~~

Can we ~~develop~~ Develop Species Delimitation ~~methods~~ Methods based on ~~morphological~~ Morphological ~~clusters~~Clusters?

Irrespective of what species concept is preferred, speciation will eventually lead to accumulation of unique genetic and most likely also phenotypic traits, justifying the use of genetic or phenotypic clustering methods for species delimitation (Hausdorf 2011). However, different evolutionary processes can act on different species. These processes can affect distinct morphological characters or character complexes, which may in turn result in varying variability patterns across skeletal regions (e.g., feeding adaptations versus locomotion). Thus, it is important to study overall skeletal variability instead of single traits or trait complexes.

Analysing variability in single traits or skeletal regions cannot capture overall morphological variability. These approaches may be useful to assess if certain characters

For Peer Review Only

proposed to be diagnostic for particular extinct species are valid or if they fall within the range of variability observed in extant species (e.g., Barahona et al. 2000). However, they do not permit the development of a more general approach to species delimitation applicable both to extant and extinct species. In fact, they may reveal conflicting results. For example, variability in scale patterns on the skull roof in three species of lacertids also included in the present analysis (*Lacerta bilineata*, *Podarcis muralis*, *P. siculus*) revealed that *P. muralis* was about 1.4 times more variable in this trait complex than *P. siculus* and nearly twice as variable compared to *L. bilineata* (Bruner and Costantini 2009). Our dataset includes discrete characters describing the skull roof patterns quantified and analysed by Bruner and Costantini (2009; see Supplementary ~~Material 4~~Data 2)~~, so this apparent contradiction with~~; our ~~own results~~findings that these three species have comparable dissimilarity is probably because our dataset covers the entire skeleton. Focusing on one or few traits is ~~extremely promising~~useful to analyse function of convergently acquired features, but it does not contribute much to ~~help delimiting~~ species delimitation.

Numerous operational criteria have been proposed to ~~define or~~ delimit species, based on varying interpretations of which defining property marks the completion of the speciation process (see reviews in Sites and Marshall 2003, 2004; Queiroz 2005, 2007). If we accept that any single one of these criteria may suffice to result in speciation, different species can have distinct defining properties (Queiroz 2005, 2007). These properties may affect morphology in disparate ways, as well, and they also will affect different character complexes. Just as in the example above of feeding versus locomotion, the evolution of reproductive isolation (Mayr 1942) or ecological divergence (Van Valen 1976) can have significant effects on morphology but does not necessarily impact the same traits or sets of traits – nor are these changes associated with single genes (Highton 1990). With time, these processes may lead to evolution and fixation

For Peer Review Only

of new morphological traits that can be considered diagnostic for a particular species (Kimura et al. 2016). However, asserting that a trait is truly fixed is statistically nearly impossible, and even allowing a 95% fixation rate within a population, sample sizes have to be very large to confirm ~~it~~ that a particular trait can be considered diagnostic (Wiens and Servedio 2000). After prolonged diverging evolution, diagnostic morphological features may also occur across the entire organism, but recently diverged species may not have evolved widespread diagnosability, and if single diagnostic traits occur, they might be in discordance with other features and thus difficult to interpret (Hausdorf 2011; Harrison and Larson 2014). Using overall variability scores derived from a set of diverse morphological characters circumvents these problems and may cover all aspects resulting from evolutionary mechanisms culminating in speciation, even if these incipient species are not yet diagnosable by particular, fixed, apomorphic traits~~ (as is the case in different clades of lizards; Wiens and Penkrot 2002; Villa et al. 2017)~~.~~ This approach is similar to the use of multiple loci in estimating species boundaries.~~

The accuracy of species delimitation based on genetic data also depends on the number of sampled loci. Single-locus analyses are prone to failure in detecting species status of recently diverged lineages, whereas combining information from multiple loci resulted in a decrease of such false-negatives (Knowles and Carstens 2007; Hausdorf and Hennig 2010). By reducing the number of loci (or morphological characters, for that matter) in an analysis of species boundaries, one is more likely to be misled by a mismatch of the evolutionary assumptions underlying the delimitation methodology and the actual evolutionary processes leading to lineage splitting. ~~For instance, i~~If speciation is driven by strong selection resulting in fast evolutionary rates in the fixation of a genotypic or phenotypic trait, focusing species delimitation on slow-

For Peer Review Only

evolving traits will not be capable of recognizing this recent and rapid lineage-splitting event (Knowles and Carstens 2007).

Focusing species delimitation on a variety of slow- and fast-evolving traits also embraces reflects a polytypic understanding of species as morphologically (and genetically) variable populations (following Mayr 1942). It is equivalent to defining the "taxonomic space" as intended by Hull (1965), using morphological variability as an indicator for the presence of one or more secondary defining properties (sensu Queiroz 2007) that renders a metapopulation a distinct species. Such an approach is supported by evidence from *Drosophila*, where morphological differences in male genitalia are associated with mutations in genes from all chromosomes (Coyne and Kreitman 1986), which led Highton (1990) to propose species delimitation based on genetic distance calculated across many loci scattered throughout the genome. This overall divergence was observed to be fairly generalized across non-avian vertebrates (Thorpe 1982), supporting its use as a proxy for a speciation event (Sites and Marshall 2003), just as we propose to use overall morphological distance as an indicator for species boundaries.

Avoiding Extreme extreme Values that May Derive from Observation Errors.—Observation errors are difficult to avoid and may stem from diverse reasons. Using average values of mean pairwise dissimilarity and their standard deviations for species delimitation rather than ranges based on maximum and minimum values alleviates this issue. First, specimen identification in scientific collections can be wrong and are often not documented so testing and/or confirming these identifications becomes difficult. Specimens can also be scored incorrectly, or errors can occur when transcribing scores from one version of a character matrix to another. Finally, different researchers may interpret certain characters in a slightly different

For Peer Review Only

~~way. Maximum disparity in particular may result from such erroneous inclusion of specimens within a given species, and errors in scoring.~~

~~Furthermore, o~~Our sensitivity analyses indicate that the range of maximum to minimum pairwise dissimilarity correlates with sample size (Fig. 7~~8~~), and that maximum dissimilarity is also affected by lacking anatomical overlap (Fig. 6~~7~~). Additionally, dissimilarity values can be impacted by observation errors, which may stem from diverse reasons (e.g., wrong specimen identifications in collections, errors in scoring, divergent interpretations of morphological characters by researchers). On the other hand, average mean pairwise dissimilarity scores are not likely to change significantly when adding more specimens per species, though the standard deviation of the mean dissimilarity does decrease with increasing sample size. ~~Thus, maximum and minimum values are more variable, and thus less reliable, particularly at low sample sizes.~~ Restricting species delimitation to the use of average and standard deviation, lowers the impact of such potential sampling error producing extreme values (Cope and Lacy 1992), especially as more specimens are sampled per species. ~~However,~~

Species Clusters versus Genus Clusters.— ~~further tests are necessary to understand how species clusters can be distinguished from genus clusters. The~~Our ~~Principle~~P Co~~ordinate~~ A~~nalyses we conducted~~ showed that while specimens of a single genus form distinct clusters in the morphospace, species clusters are not always recognizable (Fig. 4~~5~~). ~~Nevertheless~~Thus, ~~intrageneric variability is nearly always significantly higher than intraspecific variability in our lacertid dataset. Consequently,~~ species delimitation approaches have to be based on variability scores, whereas the morphospace analysis could be used to distinguish higher-level clusters that could be used to delimit genera. If such numerical boundaries are stable in established and well-accepted~~,~~ closely related species, they could be applied to other species complexes of the same

For Peer Review Only

clade, be they extant or extinct, where species boundaries are unclear. Combined with a meaningful specimen-level phylogenetic analysis, variability scores can be calculated between closely related specimens and added up in a stepwise manner until the species threshold is reached (a similar approach was used by Tschopp et al. (2015). Thereby, morphological species delimitation would combine historical consensus on species boundaries in well-known species and phenetic clustering based on a sound phylogenetic framework. Whether this is best done with discrete character matrices, geometric morphometrics, or a combination of the two, and whether this would be applicable to a wide variety of vertebrate clades, forms a rich field for further investigation.

Intraspecific ~~variability~~ Variability in ~~vertebrates~~Vertebrates

Our analysis of lacertid intraspecific variability adds to earlier studies on other vertebrate clades (e.g., Roth 1992; Wiens and Penkrot 2002; Bever 2009; Foth et al. 2015). Exact values of variability depend on character and taxon sampling, and can also vary between clades (e.g., tooth variability in different clades of mammals; Roth 1992). In fact, evidence from other lizards indicates that intraspecific variability may be higher than interspecific variability in certain genera (e.g., *Sceloporus*; Wiens and Penkrot 2002)~~, and~~. ~~cranial~~ Cranial intraspecific variability in the turtle *Pseudemys texana* was found to be at least 27% (Bever 2009), and thus also considerably higher than the observed 21% in our lacertid sample ~~(21%, -~~ although this may be partially the result of slightly different methodologies, and the restriction to cranial osteology in Bever (2009). However, in clades where all extant members have comparable intraspecific variability, and where these are consistently and significantly lower than interspecific variability, those values can be used to delimit extinct species of the same clade. Where data on skeletal variability in extant members of a clade are available, we can relatively easily quantify

For Peer Review Only

dissimilarity based on both discrete characters (e.g., Hetherington et al. 2015; this study), and geometric morphometrics (e.g., Bruner and Costantini 2009; Foth et al. 2015; Hetherington et al. 2015; Tayhan et al. 2016; Cooney et al. 2017; Cerio and Witmer 2019; Gray et al. 2019; Watanabe et al. 2019). Applying such values to delimit extinct and extant species consistently throughout a clade would be a straightforward approach to overcome the species comparability problem between neontological and palaeontological species, at least in clades that have extant members (Brochu and Sumrall 2020). Even in a clade that lacks extant taxa, we could still apply these approaches in order to investigate the consistency with which different workers have recognized species (as has been done by Benson et al. 2012 and Tschopp et al. 2015). Such an approach should lead to insights into data set quality, and highlight areas where disagreement is most extreme, and thus indicate where taxonomic revision is best focused in order to achieve greater consistency. As a result, we would have much more robustly delineated extinct species and more consistent ways to compare extinct and extant species numbers for any kind of macroevolutionary study through Deep Time.

CONCLUSION

Skeletal pairwise dissimilarity was found to be consistent within extant species of lacertid lizards, which were originally identified based on non-osteological features and partially delimited based on non-morphological species criteria. Extinct lacertid species delimited based on osteological grounds have more widely diverging ranges and averages of mean pairwise dissimilarity. This incongruence highlights that the species comparability problem, the fact that species delimited based on different species criteria are not comparable biological units, is still an issue, in particular in studies comparing species numbers through Deep Time and including extant taxa. However, given that intraspecific osteological variability is consistent and stable

For Peer Review Only

among, and within, extant lacertid species, we propose that dissimilarity values can, and should, be used to delimit extinct species as well. Quantifying osteological intraspecific variability in extant members of a clade and applying them to extinct members of the same clade, is a way to overcome the species comparability problem in a particular clade. Similar approaches should be applied to other vertebrate clades in order to assess if our results can be generalised, and to ensure the comparability of extinct and extant species from different time periods or geographic regions, before attempting to study biodiversity changes and other macroevolutionary patterns through Deep Time.

Supplementary Material

Supplementary material can be found in the Dryad digital data repository (http://dx.doi.org/10.5061/dryad.[NNNN]).

Funding

Funding for the project was provided through a postdoctoral fellowship to E. Tschopp by the European Union's Seventh Framework programme for research and innovation under the Marie Skłodowska–Curie grant agreement No. 609402-2020 researchers: Train to Move (T2M). Collection visits by E. Tschopp were possible thanks to several SYNTHESYS Projects (http://www.synthesys.info/), financed by the European Community Research Infrastructure Action under the FP7 'Capacities' Programme (MNHN: FR-TAF-5839; NHMW: AT-TAF-5725) and two Erasmus+ Traineeships to L.C.M. Wencker (University College London, UK; NHMW, Austria). J. Napoli is supported by a Richard Gilder Graduate School fellowship and by the Newt and Callista Gingrich endowment of the American Museum of Natural History, New York. L.C.M. Wencker is funded through a PhD fellowship provided by the Università degli Studi di Torino. M. Delfino is supported by Fondi di Ricerca Locale UNITO 2017-2019.

For Peer Review Only

ACKNOWLEDGEMENTS

We thank all the institutions and curators and collection managers that allowed us to study specimens under their care, including: the Muséum national d'Histoire naturelle, Paris, France (late Jean-Claude Rage, Salvador Bailon, Virginie Bouetel); Naturhistorisches Museum Wien, Vienna, Austria (Heinz Grillitsch, Georg Gassner, Silke Schweiger); Osteoteca, Laboratorio Arqueociencias, Lisbon, Portugal (Sónia Gabriel, Simon Davis); Musée Royal de l'Afrique Centrale, Tervuren, Belgium (Annelise Folie); Natural History Museum, London, UK (Patrick Campbell, Jeff Streicher); Museo Nacional de Ciencias Naturales, Madrid, Spain (Marta Calvo, Alberto Sanchez); Institute of Systematics and Evolution of Animals, Polish Academy of Sciences, Krakow, Poland (Zbigniew Szyndlar); Paläontologisches Institut und Museum der Universität Zürich, Switzerland (Christian Klug, Winand Brinkmann); Grant Museum of Zoology, London, UK (Tannis Davidson, Hannah Cornish); Osteological Collections, Hebrew University of Jerusalem, Israel (Rebecca Biton); Universidad Autónoma de Madrid (Francisco Ortega); Università Roma 3, Rome, Italy (Tassos Kotsakis); Soprintendenza Archeologia, Belle Arti e Paesaggio per le province di Sassari e Nuoro, Nuoro, Italy (Caterinella Tuveri); Zoologisches Museum Hamburg, Germany (Jakob Hallermann, Alexander Haas). Frank Glaw (Zoologische Staatssammlung München, Germany) provided additional data on specimens included based on literature. We thank the Mesquite Project Team, the Willi Hennig Society, and the R core team for providing their software MESQUITE, TNT, and R, respectively, for free online. Moreover, we thank the executive committee of MorphoBank for maintaining the site, and the National Science Foundation's Advances in Biological Informatics program, which is funding the current version and makes it available online for free. We further thank Tim Breu for his support during the construction of the Python scripts. ~~An e~~Earlier submissions of the MS ~~has~~

For Peer Review Only

have greatly benefited by reviews from Andrej Čerňanský (who also shared CT scan data for the revision of the dataset), ~~and two~~ three anonymous reviewers, and the managing editor Lisa Barrow.

LITERATURE CITED

Agapow P., Bininda-Emonds O.R.P., Crandall K.A., Gittleman J.L., Mace G.M., Marshall J.C., Purvis A. 2004. The impact of species concept on biodiversity studies. Q. Rev. Biol. 79:161–179.

Ahmadzadeh F., Flecks M., Rödder D., Böhme W., Ilgaz Ç., Harris D.J., Engler J.O., Üzüm N., Carretero M.A. 2013. Multiple dispersal out of Anatolia: biogeography and evolution of oriental green lizards. Biol. J. Linn. Soc. 110:398–408.

Allmon W.D. 2013. Species, speciation and palaeontology up to the Modern Synthesis: persistent themes and unanswered questions. Palaeontology. 56:1199–1223.

Allmon W.D. 2016. Studying species in the fossil record: a review and recommendations for a more unified approach. In: Allmon W.D., Yacobucci M.M., editors. Species and speciation in the fossil record. Chicago, USA: University of Chicago Press. p. 59–120.

Allmon W.D., Smith U.E. 2011. What, if anything, can we learn from the fossil record about speciation in marine gastropods? Biological and geological considerations. Am. Malacol. Bull. 29:247–276.

Alroy J. 2010. Fair sampling of taxonomic richness and unbiased estimation of origination and extinction rates. Paleontol. Soc. Pap. 16:55–80.

For Peer Review Only

Anderson E., Abbe E.C. 1934. A quantitative comparison of specific and generic differences in the Betulaceae. J. Arnold Arbor. 15:43–49.

Anderson E., Whitaker T.W. 1934. Speciation in Uvularia. J. Arnold Arbor. 15:28–42.

~~Arnold E.N. 1973. Relationships of the palaeartic lizards assigned to the genera *Lacerta*, *Algyroides* and *Psammodromus* (Reptilia: Lacertidae). Bull. Br. Mus. Nat. Hist. Zool. 25:291–366.~~

Arnold E.N., Arribas O., Carranza S. 2007. Systematics of the Palaearctic and Oriental lizard tribe Lacertini (Squamata: Lacertidae: Lacertinae), with descriptions of eight new genera. Zootaxa. 1430:3–86.

Arsdale A.P.V., Wolpoff M.H. 2013. A single lineage in Early Pleistocene *Homo*: Size variation continuity in Early Pleistocene *Homo* crania from East Africa and Georgia. Evolution. 67:841–850.

~~Augé M. 1988. Une nouvelle espèce de Lacertidae (Sauria, Lacertilia) de l'Oligocène français: *Lacerta filholi*. Place de cette espèce dans l'histoire des Lacertidae de l'Eocène supérieur au Miocène inférieur. Neues Jahrb. Für Geol. Paläontol. Monatshefte. 8:464–478.~~

~~Augé M., Smith R. 2009. An assemblage of early Oligocene lizards (Squamata) from the locality of Boutersem (Belgium), with comments on the Eocene–Oligocene transition. Zool. J. Linn. Soc. 155:148–170.~~

Augé M.L. 2005. Evolution des lézards du Paléogène en Europe. Mém. Muséum Natl. Hist. Nat. 1993. 192.

For Peer Review Only

Augé M.L., Hervet S. 2009. Fossil lizards from the locality of Gannat (late Oligocene–early Miocene, France) and a revision of the genus *Pseudeumeces* (Squamata, Lacertidae). Palaeobiodiversity Palaeoenvironments. 89:191.

Augé M.L., Smith R. 2009. An assemblage of early Oligocene lizards (Squamata) from the locality of Boutersem (Belgium), with comments on the Eocene–Oligocene transition. Zool. J. Linn. Soc. 155:148–170.

Bailon S., Boistel R., Bover P., Alcover J.A. 2014. *Maioricalacerta rafelinensis*, gen. et sp. nov. (Squamata, Lacertidae), from the Early Pliocene of Mallorca (Balearic Islands, Western Mediterranean Sea). J. Vertebr. Paleontol. 34:318–326.

Balakrishnan R. 2005. Species concepts, species boundaries and species identification: A view from the tropics. Syst. Biol. 54:689–693.

~~Barahona F., Barbadillo L.J. 1997. Identification of some Iberian lacertids using skull characters. Rev. Esp. Herpetol. 11:47–62.~~

Barahona F., Barbadillo L.J. 1998. Inter-and intraspecific variation in the post-natal skull of some lacertid lizards. J. Zool. 245:393–405.

Barahona F., Evans S.E., Mateo J.A., García-Márquez M., López-Jurado L.F. 2000. Endemism, gigantism and extinction in island lizards: the genus *Gallotia* on the Canary Islands. J. Zool. 250:373–388.

For Peer Review Only

Barnosky A.D., Matzke N., Tomiya S., Wogan G.O.U., Swartz B., Quental T.B., Marshall C., McGuire J.L., Lindsey E.L., Maguire K.C., Mersey B., Ferrer E.A. 2011. Has the Earth's sixth mass extinction already arrived? Nature. 471:51–57.

Baum D.A. 1998. Individuality and the existence of species through time. Syst. Biol. 47:641–653.

Bell C.J., Mead J.I. 2014. Not enough skeletons in the closet: collections-based anatomical research in an age of conservation conscience. Anat. Rec. 297:344–348.

Bell P.R., Hendrickx C. 2021. Epidermal complexity in the theropod dinosaur Juravenator from the Upper Jurassic of Germany. Palaeontology. 64:203–223.

Benson R.B.J., Evans M., Druckenmiller P.S. 2012. High diversity, low disparity and small body size in plesiosaurs (Reptilia, Sauropterygia) from the Triassic–Jurassic boundary. PLoS ONE. 7:e31838.

Benton M.J., Pearson P.N. 2001. Speciation in the fossil record. Trends Ecol. Evol. 16:405–411.

Bernardi M., Minelli A. 2011. Il concetto di specie e la paleontologia: una rassegna introduttiva. Rendiconti Online Della Soc. Geol. Ital. 13:2–26.

Bever G.S. 2009. The postnatal skull of the extant North American turtle *Pseudemys texana* (Cryptodira: Emydidae), with comments on the study of discrete intraspecific variation. J. Morphol. 270:97–128.

Böhme M., Ilg A. 2003. fosFARbase. Www.wahre-Staerke.com. (accessioned March 2017).

For Peer Review Only

Böhme W., Zammit-Maempel G. 1982. *Lacerta siculimelitensis* sp. n. (Sauria: Lacertidae), a giant lizard from the Late Pleistocene of Malta. Amphib.-Reptil. 3:257–268.

Borczyk B., Kusznierz J., Paśko \Lukasz, Turniak E. 2014. Scaling of the sexual size and shape skull dimorphism in the sand lizard (Lacerta agilis L.). .

Brazeau M.D. 2011. Problematic character coding methods in morphology and their effects. Biol. J. Linn. Soc. 104:489–498.

Brecko J., Huyghe K., Vanhooydonck B., Herrel A., Grbac I., Van Damme R. 2008. Functional and ecological relevance of intraspecific variation in body size and shape in the lizard *Podarcis melisellensis* (Lacertidae). Biol. J. Linn. Soc. 94:251–264.

Briggs D.E.G., Fortey R.A., Wills M.A. 1992. Morphological disparity in the Cambrian. Science. 256:1670–1673.

Brizuela S. 2010. Los lagartos continentales fósiles de la Argentina (excepto Iguania). .

Brochu C.A., Sumrall C.D. 2020. Modern cryptic species and crocodylian diversity in the fossil record. Zool. J. Linn. Soc. 189:700–711.

Brown E.E., Cashmore D.D., Simmons N.B., Butler R.J. 2019. Quantifying the completeness of the bat fossil record. Palaeontology. 62:757–776.

Bruner E. 2004. Evolution, actuality and species concept: a need for a palaeontological tool. Hum. Evol. 19:93–112.

Bruner E., Costantini D. 2009. Morfología de la cabeza y el grado de variación en *Lacerta bilineata*, *Podarcis muralis* y *Podarcis sicula*. Int. J. Morphol. 27:667–676.

For Peer Review Only

Bruner E., Costantini D., Fanfani A., Dell'Omo G. 2005. Morphological variation and sexual dimorphism of the cephalic scales in Lacerta bilineata. Acta Zool. 86:245–254.

Brusatte S.L. 2010. Representing supraspecific taxa in higher-level phylogenetic analyses: guidelines for palaeontologists. Palaeontology. 53:1–9.

Cain A.J., Harrison G.A. 1958. An analysis of the taxonomist's judgment of affinity. Proc. Zool. Soc. Lond. 131:85–98.

~~Caloi L., Kotsakis T., Palombo M.R. 1986. La fauna a vertebrati terrestri del Pleistocene delle isole del Mediterraneo. Geol. Romana. 25:235–256.~~

~~Camaiti M., Villa A., Wencker L.C.M., Bauer A.M., Stanley E.L., Delfino M. 2019. Descriptive osteology and patterns of limb loss of the European limbless skink *Ophiomorus punctatissimus* (Squamata, Scincidae). J. Anat. 235:313–345.~~

Carranza S., Arnold E.N., Amat F. 2004. DNA phylogeny of *Lacerta* (*Iberolacerta*) and other lacertine lizards (Reptilia: Lacertidae): did competition cause long-term mountain restriction? Syst. Biodivers. 2:57–77.

Carrasco M.A. 2013. The impact of taxonomic bias when comparing past and present species diversity. Palaeogeogr. Palaeoclimatol. Palaeoecol. 372:130–137.

Carstens B.C., Pelletier T.A., Reid N.M., Satler J.D. 2013. How to fail at species delimitation. Mol. Ecol. 22:4369–4383.

For Peer Review Only

Cerio D.G., Witmer L.M. 2019. Intraspecific variation and symmetry of the inner-ear labyrinth in a population of wild turkeys: implications for paleontological reconstructions. PeerJ. 7:e7355.

Čerňanský A., Augé M.L. 2013. New species of the genus *Plesiolacerta* (Squamata: Lacertidae) from the upper Oligocene (MP28) of Southern Germany and a revision of the type species *Plesiolacerta lydekkeri*. Palaeontology. 56:79–94.

Čerňanský A., Bolet A., Müller J., Rage J.-C., Augé M., Herrel A. 2017. A new exceptionally preserved specimen of *Dracaenosaurus* (Squamata, Lacertidae) from the Oligocene of France as revealed by micro-computed tomography. J. Vertebr. Paleontol. 0:e1384738.

Čerňanský A., Klembara J., Műller J. 2016a. The new rare record of the late Oligocene lizards and amphisbaenians from Germany and its impact on our knowledge of the European terminal Palaeogene. Palaeobiodiversity Palaeoenvironments. 96:559–587.

Čerňanský A., Klembara J., Smith K.T. 2016b. Fossil lizard from central Europe resolves the origin of large body size and herbivory in giant Canary Island lacertids. Zool. J. Linn. Soc. 176:861–877.

Čerňanský A., Syromyatnikova E.V. 2019. The first Miocene fossils of *Lacerta* cf. *trilineata* (Squamata, Lacertidae) with a comparative study of the main cranial osteological differences in green lizards and their relatives. PLOS ONE. 14:e0216191.

Čerňanský A., Yaryhin O., Ciceková J., Werneburg I., Hain M., Klembara J. 2019. Vertebral comparative anatomy and morphological differences in anguine lizards with a special reference to *Pseudopus apodus*. Anat. Rec. 302:232–257.

For Peer Review Only

Christiansen N.A., Tschopp E. 2010. Exceptional stegosaur integument impressions from the Upper Jurassic Morrison Formation of Wyoming. Swiss J. Geosci. 103:163–171.

Ciampaglio C.N., Kemp M., McShea D.W. 2001. Detecting changes in morphospace occupation patterns in the fossil record: characterization and analysis of measures of disparity. Paleobiology. 27:695–715.

Cleary T.J., Moon B.C., Dunhill A.M., Benton M.J. 2015. The fossil record of ichthyosaurs, completeness metrics and sampling biases. Palaeontology. 58:521–536.

Conrad J.L. 2008. Phylogeny and systematics of Squamata (Reptilia) based on morphology. Bull. Am. Mus. Nat. Hist. 2008:1–182.

Cooney C.R., Bright J.A., Capp E.J.R., Chira A.M., Hughes E.C., Moody C.J.A., Nouri L.O., Varley Z.K., Thomas G.H. 2017. Mega-evolutionary dynamics of the adaptive radiation of birds. Nature. 542:344–347.

Cope D.A., Lacy M.G. 1992. Falsification of a single species hypothesis using the coefficient of variation: A simulation approach. Am. J. Phys. Anthropol. 89:359–378.

Coyne J.A., Kreitman M. 1986. Evolutionary genetics of two sibling species, *Drosophila simulans* and *D. sechellia*. Evolution. 40:673–691.

Cracraft J. 1987. Species concepts and the ontology of evolution. Biol. Philos. 2:329–346.

Cruzado-Caballero P., Castillo Ruiz C., Bolet A., Colmenero J.R., De la Nuez J., Casillas R., Llacer S., Bernardini F., Fortuny J. 2019. First nearly complete skull of *Gallotia auaritae*

For Peer Review Only

(lower-middle Pleistocene, Squamata, Gallotiinae) and a morphological phylogenetic analysis of the genus *Gallotia*. Sci. Rep. 9:16629.

Dayan T., Wool D., Simberloff D. 2002. Variation and covariation of skulls and teeth: modern carnivores and the interpretation of fossil mammals. Paleobiology. 28:508–526.

~~De Queiroz K. 1987. Phylogenetic systematics of iguanine lizards. Univ. Calif. Publ. Zool. 118:1–223.~~

~~Delfino M. 2001. Early Pleistocene *Lacerta* remains from Southern Italy (Apricena, Foggia). Is it *Lacerta siculimelitensis*? Mediterrean Basin Lacertid Lizards - Biol. Approach.:21–26.~~

~~Delfino M. 2002. Erpetofaune italiane del Neogene e del Quaternario. .~~

Delfino M., Bailon S. 2000. Early Pleistocene herpetofauna from Cava Dell'Erba and Cava Pirro (Apulia, Southern Italy). Herpetol. J. 10:95–110.

Denton R.K.Jr., O'Neill R.C. 1995. *Prototeius stageri*, gen. et sp. nov., a new teiid lizard from the Upper Cretaceous Marshalltown Formation of New Jersey, with a preliminary phylogenetic revision of the Teiidae. J. Vertebr. Paleontol. 15:235–253.

Doan T.M., Castoe T.A. 2003. Using morphological and molecular evidence to infer species boundaries within *Proctoporus bolivianus* Werner (Squamata: Gymnophthalmidae). Herpetologica. 59:432–449.

Dunbar C.O. 1950. The species concept: Further discussion. Evolution. 4:175–176.

Dzik J. 1985. Typologic versus population concepts of chronospecies: implications for ammonite biostratigraphy. Acta Palaeontol. Pol. 30.

For Peer Review Only

~~Estes R. 1983. Handbuch der Paläoherpetologie. Encyclopedia of Paleoherpetology. T. 10A. Sauria terrestria, Amphisbaenia. Stuttgart, Portland-USA, G. Fischer.~~

Estes R., ~~De~~ Queiroz K. de, Gauthier J. 1988. Phylogenetic relationships within Squamata. In: Estes R., Pregill G., editors. Phylogenetic Relationships of the Lizard Families. Palo Alto, USA: Stanford University Press. p. 119–281.

~~Esu D., Kotsakis T., Burgio E. 1986. I vertebrati e i molluschi continentali pleistocenici di Poggio Schinaldo (Palermo, Sicilia). Ital. J. Geosci. 105:233–241.~~

Fabbri M., Wiemann J., Manucci F., Briggs D.E.G. 2020. Three-dimensional soft tissue preservation revealed in the skin of a non-avian dinosaur. Palaeontology. 63:185–193.

~~Filhol H. 1877. Recherches sur les phosphorites du Quercy: étude des fossiles qu'on y rencontre et spécialement des mammifères. Ann. Sci. Géologiques. 8:1–340.~~

Foote M. 1990. Nearest-neighbor analysis of trilobite morphospace. Syst. Biol. 39:371–382.

Foote M. 1992a. Paleozoic record of morphological diversity in blastozoan echinoderms. Proc. Natl. Acad. Sci. 89:7325–7329.

Foote M. 1992b. Rarefaction analysis of morphological and taxonomic diversity. Paleobiology. 18:1–16.

Foote M. 1993. Discordance and concordance between morphological and taxonomic diversity. Paleobiology. 19:185–204.

Foth C., Bona P., Desojo J.B. 2015. Intraspecific variation in the skull morphology of the black caiman *Melanosuchus niger* (Alligatoridae, Caimaninae). Acta Zool. 96:1–13.

For Peer Review Only

Gauthier J.A., Kearney M., Maisano J.A., Rieppel O., Behlke A.D.B. 2012. Assembling the squamate tree of life: perspectives from the phenotype and the fossil record. Bull. Peabody Mus. Nat. Hist. 53:3–308.

Gerber S. 2019. Use and misuse of discrete character data for morphospace and disparity analyses. Palaeontology. 62:305–319.

~~Gervais P. 1848. Zoologie et paléontologie françaises (animaux vertébrés): ou nouvelles recherches sur les animaux vivants et fossiles de la France. Paris, France: Arthus Bertrand.~~

Gingerich P.D. 1981. Variation, sexual dimorphism, and social structure in the Early Eocene horse *Hyracotherium* (Mammalia, Perissodactyla). Paleobiology. 7:443–455.

Gingerich P.D. 1985. Species in the fossil record: concepts, trends, and transitions. Paleobiology. 11:27–41.

Gray J.A., Sherratt E., Hutchinson M.N., Jones M.E.H. 2019. Evolution of cranial shape in a continental-scale evolutionary radiation of Australian lizards. Evolution. 73:2216–2229.

Haffer J. 1986. Superspecies and species limits in vertebrates. J. Zool. Syst. Evol. Res. 24:169–190.

Harrison R.G. 1998. Linking evolutionary pattern and process. Endless Forms.:19–31.

Harrison R.G., Larson E.L. 2014. Hybridization, introgression, and the nature of species boundaries. J. Hered. 105:795–809.

Hausdorf B. 2011. Progress toward a general species concept. Evolution. 65:923–931.

For Peer Review Only

Hausdorf B., Hennig C. 2010. Species delimitation using dominant and codominant multilocus markers. Syst. Biol. 59:491–503.

Hetherington A.J., Sherratt E., Ruta M., Wilkinson M., Deline B., Donoghue P.C.J. 2015. Do cladistic and morphometric data capture common patterns of morphological disparity? Palaeontology. 58:393–399.

Hey J., Waples R.S., Arnold M.L., Butlin R.K., Harrison R.G. 2003. Understanding and confronting species uncertainty in biology and conservation. Trends Ecol. Evol. 18:597–603.

Highton R. 1990. Taxonomic treatment of genetically differentiated populations. Herpetologica. 46:114–121.

Hoffstetter R. 1942. Sur les restes de Sauria du nummulitique Européen rapportés à la famille des Iguanidae. Bull. Muséum Natl. Hist. Nat. 14:233–240.

Hoffstetter R. 1944. Sur les Scincidae fossiles. I. Formes européennes et nord-américaines. Bull. Muséum Natl. Hist. Nat. 16:547–553.

~~Holman J.A. 1998. Pleistocene Amphibians and Reptiles in Britain and Europe. Oxford, UK: Oxford University Press.~~

Hull D.L. 1965. The effect of essentialism on taxonomy--two thousand years of stasis (II). Br. J. Philos. Sci. 16:1–18.

Joyce W.G., Micklich N., Schaal S.F.K., Scheyer T.M. 2012. Caught in the act: the first record of copulating fossil vertebrates. Biol. Lett. 8:846–848.

For Peer Review Only

Kapli P., Poulakakis N., Lymberakis P., Mylonas M. 2011. A re-analysis of the molecular phylogeny of Lacertidae with currently available data. Basic Appl. Herpetol. 25:97–104.

Kay R.F. 1982. Sexual dimorphism in Ramapithecinae. Proc. Natl. Acad. Sci. 79:209–212.

Kelley J. 1986. Species recognition and sexual dimorphism in *Proconsul* and *Rangwapithecus*. J. Hum. Evol. 15:461–495.

Kimura Y., Flynn L.J., Jacobs L.L. 2016. A palaeontological case study for species delimitation in diverging fossil lineages. Hist. Biol. 28:189–198.

Kirchhof S., Jacobsen N., Richter K. 2012. Intraspecific variation of morphology, colouration, pholidosis, and tail loss rate in a relic lacertid of South Africa, *Australolacerta rupicola* (Sauria: Lacertidae). Salamandra. 48:125–132.

~~Klembara J., Hain M., Dobiašová K. 2014. Comparative anatomy of the lower jaw and dentition of *Pseudopus apodus* and the interrelationships of species of subfamily Anguinae (Anguimorpha, Anguidae). Anat. Rec. 297:516–544.~~

Knowles L.L., Carstens B.C. 2007. Delimiting species without monophyletic gene trees. Syst. Biol. 56:887–895.

Kornilios P., Thanou E., Lymberakis P., Ilgaz Ç., Kumlutaş Y., Leaché A. 2020. A phylogenomic resolution for the taxonomy of Aegean green lizards. Zool. Scr. 49:14–27.

~~Kotsakis T. 1996. Anfibi e rettili. In: Basile B., Chilardi S., editors. Le ossa dei Giganti. Lo scavo paleontologico di Contrada Fusca. Siracusa, Italy: Arnaldo Lombardi. p. 56–60.~~

For Peer Review Only

Lee M.S.Y. 1998. Convergent evolution and character correlation in burrowing reptiles: towards a resolution of squamate relationships. Biol. J. Linn. Soc. 65:369–453.

Lloyd G.T. 2016. Estimating morphological diversity and tempo with discrete character-taxon matrices: implementation, challenges, progress, and future directions. Biol. J. Linn. Soc. 118:131–151.

Lordkipanidze D., León M.S.P. de, Margvelashvili A., Rak Y., Rightmire G.P., Vekua A., Zollikofer C.P.E. 2013. A complete skull from Dmanisi, Georgia, and the evolutionary biology of early *Homo*. Science. 342:326–331.

Lupia R. 1999. Discordant morphological disparity and taxonomic diversity during the Cretaceous angiosperm radiation: North American pollen record. Paleobiology. 25:1–28.

Maddison W.P., Maddison D.R. 2017. Mesquite: a modular system for evolutionary analysis. .

Mallon J.C. 2017. Recognizing sexual dimorphism in the fossil record: lessons from nonavian dinosaurs. Paleobiology. 43:495–507.

Mannion P.D., Benson R.B.J., Carrano M.T., Tennant J.P., Judd J., Butler R.J. 2015. Climate constrains the evolutionary history and biodiversity of crocodylians. Nat. Commun. 6:ncomms9438.

Mannion P.D., Upchurch P. 2010. Completeness metrics and the quality of the sauropodomorph fossil record through geological and historical time. Paleobiology. 36:283–302.

For Peer Review Only

Mannion P.D., Upchurch P., Barnes R.N., Mateus O. 2013. Osteology of the Late Jurassic Portuguese sauropod dinosaur *Lusotitan atalaiensis* (Macronaria) and the evolutionary history of basal titanosauriforms. Zool. J. Linn. Soc. 168:98–206.

Marshall J.C., Arévalo E., Benavides E., Sites J.L., Sites J.W. 2006. Delimiting species: Comparing methods for Mendelian characters using lizards of the *Sceloporus grammicus* (Squamata: Phrynosomatidae) complex. Evolution. 60:1050–1065.

Mateo J.A. 1988. Estudio sistemático y zoogeográfico de los lagartos ocelados. PhD thesis.

Matthew W.D. 1930. Range and limitations of species as seen in fossil mammal faunas. GSA Bull. 41:271–274.

Matzke N.J. 2013. Probabilistic historical biogeography: new models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. Front. Biogeogr. 5.

Matzke N.J. 2014. Model Selection in Historical Biogeography Reveals that Founder-Event Speciation Is a Crucial Process in Island Clades. Syst. Biol. 63:951–970.

Mayr E. 1942. Systematics and the Origin of the Species. New York, USA: Columbia University Press.

Mendes J., Harris D.J., Carranza S., Salvi D. 2016. Evaluating the phylogenetic signal limit from mitogenomes, slow evolving nuclear genes, and the concatenation approach. New insights into the Lacertini radiation using fast evolving nuclear genes and species trees. Mol. Phylogenet. Evol. 100:254–267.

For Peer Review Only

Miller III W. 2016. The species problem: concepts, conflicts, and patterns preserved in the fossil record. In: Allmon W.D., Yacobucci M.M., editors. Species and speciation in the fossil record. Chicago, USA: University of Chicago Press. p. 28–58.

Montgelard C., Behrooz R., Arnal V., Asadi A., Geniez P., Kaboli M. 2020. Diversification and cryptic diversity of *Ophisops elegans* (Sauria, Lacertidae). J. Zool. Syst. Evol. Res. 58:1276–1289.

Müller J. 2004. Cranial osteology from *Dracaenosaurus croizeti*, a lacertid lizard from the Oligocene of France (Reptilia, Squamata). Neues Jahrb. Für Geol. Paläontol. - Abh.:253–266.

O'Donovan C., Meade A., Venditti C. 2018. Dinosaurs reveal the geographical signature of an evolutionary radiation. Nat. Ecol. Evol. 2:452–458.

O'Leary M.A., Kaufman S.G. 2012. MorphoBank 3.0: Web application for morphological phylogenetics and taxonomy. http://www.morphobank.org:.

Paradis E., Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. 35:526–528.

Poropat S.F., Mannion P.D., Upchurch P., Hocknull S.A., Kear B.P., Kundrát M., Tischler T.R., Sloan T., Sinapius G.H.K., Elliott J.A., Elliott D.A. 2016. New Australian sauropods shed light on Cretaceous dinosaur palaeobiogeography. Sci. Rep. 6:34467.

Prendini L. 2001. Species or supraspecific taxa as terminals in cladistic analysis? Groundplans versus exemplars revisited. Syst. Biol. 50:290–300.

For Peer Review Only

Pyron R.A., Burbrink F.T., Wiens J.J. 2013. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. BMC Evol. Biol. 13:93.

Quadros A.B., Chafrat P., Zaher H. 2018. A new teiid lizard of the genus *Callopistes* Gravenhorst, 1838 (Squamata, Teiidae), from the Lower Miocene of Argentina. J. Vertebr. Paleontol. 38:(1)-(18).

Queiroz K. de. 1987. Phylogenetic systematics of iguanine lizards. Univ. Calif. Publ. Zool. 118:1–223.

Queiroz K. de. 1998. The general lineage concept of species, species criteria, and the process of speciation. In: Howard D.J., Berlocher S.H., editors. Endless Forms: Species and Speciation. Oxford, UK: Oxford University Press. p. 57–75.

Queiroz K. de. 2005. Different species problems and their resolution. BioEssays. 27:1263–1269.

Queiroz K. de. 2007. Species concepts and species delimitation. Syst. Biol. 56:879–886.

Rauhut O.W.M., Foth C., Tischlinger H., Norell M.A. 2012. Exceptionally preserved juvenile megalosauroid theropod dinosaur with filamentous integument from the Late Jurassic of Germany. Proc. Natl. Acad. Sci. 109:11746–11751.

Raup D.M. 1972. Taxonomic diversity during the Phanerozoic. Science. 177:1065–1071.

R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

For Peer Review Only

Reichenbacher B., Sienknecht U., Küchenhoff H., Fenske N. 2007. Combined otolith morphology and morphometry for assessing taxonomy and diversity in fossil and extant killifish (*Aphanius*, †*Prolebias*). J. Morphol. 268:898–915.

Rieppel O. 2009. Species as a process. Acta Biotheor. 57:33–49.

Rightmire G.P., Margvelashvili A., Lordkipanidze D. 2019. Variation among the Dmanisi hominins: Multiple taxa or one species? Am. J. Phys. Anthropol. 168:481–495.

Roth V.L. 1992. Quantitative variation in elephant dentitions: Implications for the delimitation of fossil species. Paleobiology. 18:184–202.

Sahney S., Benton M.J. 2017. The impact of the Pull of the Recent on the fossil record of tetrapods. Evol. Ecol. Res. 18:7–23.

Saunders I.W., Tavaré S., Watterson G.A. 1984. On the genealogy of nested subsamples from a haploid population. Adv. Appl. Probab. 16:471–491.

Scanlon J.D. 1996. Studies in the palaeontology and systematics of Australian snakes. .

Schwartz J.H., Tattersall I., Chi Z. 2014. Comment on "A complete skull from Dmanisi, Georgia, and the evolutionary biology of early *Homo*." Science. 344:360–360.

Simpson G.G. 1937. Patterns of phyletic evolution. GSA Bull. 48:303–314.

Simpson G.G. 1940. Types in modern taxonomy. Am. J. Sci. 238:413–431.

Simpson G.G. 1941. The species of *Hoplophoneus*. Am. Mus. Novit. 1123.

Simpson G.G. 1951. The species concept. Evolution. 5:285–298.

For Peer Review Only

Sites J.W., Marshall J.C. 2003. Delimiting species: a Renaissance issue in systematic biology. Trends Ecol. Evol. 18:462–470.

Sites J.W., Marshall J.C. 2004. Operational criteria for delimiting species. Annu. Rev. Ecol. Evol. Syst. 35:199–227.

Smith A.J., Rosario M.V., Eiting T.P., Dumont E.R. 2014. Joined at the hip: Linked characters and the problem of missing data in studies of disparity. Evolution. 68:2386–2400.

Sneath P.H., Sokal R.R. 1973. Numerical taxonomy. The principles and practice of numerical classification. San Francisco, USA: W. H. Freeman and Company.

Starrfelt J., Liow L.H. 2016. How many dinosaur species were there? Fossil bias and true richness estimated using a Poisson sampling model. Phil Trans R Soc B. 371:20150219.

Tayhan Y., Yakin B.Y., Tok C.V. 2016. Shape variability of the head of *Ophisops elegans* Ménétriés, 1832 (Reptilia: Lacertidae) from Konya, Turkey. Ital. J. Zool. 83:208–212.

Tennant J.P., Mannion P.D., Upchurch P. 2016. Sea level regulated tetrapod diversity dynamics through the Jurassic/Cretaceous interval. Nat. Commun. 7:12737.

Thorpe J.P. 1982. The molecular clock hypothesis: Biochemical evolution, genetic differentiation and systematics. Annu. Rev. Ecol. Syst. 13:139–168.

Trueman A.E. 1924. The species-concept in palaeontology. Geol. Mag. 61:355–360.

Tschopp E., Mateus O., Benson R.B.J. 2015. A specimen-level phylogenetic analysis and taxonomic revision of Diplodocidae (Dinosauria, Sauropoda). PeerJ. 3:e857.

For Peer Review Only

Tschopp E., Tschopp F.A., Mateus O. 2018a. Overlap Indices: Tools to quantify the amount of anatomical overlap among groups of incomplete terminal taxa in phylogenetic analyses. Acta Zool. 99:169–176.

Tschopp E., Upchurch P. 2019. The challenges and potential utility of phenotypic specimen-level phylogeny based on maximum parsimony. Earth Environ. Sci. Trans. R. Soc. Edinb. 109:301–323.

Tschopp E., Villa A., Camaiti M., Ferro L., Tuveri C., Rook L., Arca M., Delfino M. 2018b. The first fossils of *Timon* (Squamata: Lacertinae) from Sardinia (Italy) and potential causes for its local extinction in the Pleistocene. Zool. J. Linn. Soc. 184:825–856.

Van Valen L. 1976. Ecological Species, Multispecies, and Oaks. Taxon. 25:233–239.

Villa A., Daza J.D., Bauer A.M., Delfino M. 2018. Comparative cranial osteology of European gekkotans (Reptilia, Squamata). Zool. J. Linn. Soc. 184:857–895.

Villa A., Delfino M. 2019. A comparative atlas of the skull osteology of European lizards (Reptilia: Squamata). Zool. J. Linn. Soc. 187:829–928.

Villa A., Kirchner M., Alba D.M., Bernardini F., Bolet A., Luján À.H., Fortuny J., Hipsley C.A., Müller J., Sindaco R., Tuniz C., Delfino M. 2019. Comparative cranial osteology of *Blanus*(Squamata: Amphisbaenia). Zool. J. Linn. Soc. 185:693–716.

Villa A., Tschopp E., Georgalis G.L., Delfino M. 2017. Osteology, fossil record and palaeodiversity of the European lizards. Amphib.-Reptil. 38:79–88.

For Peer Review Only

Watanabe A. 2016. The impact of poor sampling of polymorphism on cladistic analysis. Cladistics. 32:317–334.

Watanabe A., Fabre A.-C., Felice R.N., Maisano J.A., Müller J., Herrel A., Goswami A. 2019. Ecomorphological diversification in squamates from conserved pattern of cranial integration. Proc. Natl. Acad. Sci. 116:14688–14697.

Wencker L.C.M, Tschopp E., Villa A., Augé M.L., Delfino M. ~~(in review)~~2021. Phylogenetic value of jaw elements of lacertid lizards (Squamata: Lacertoidea): A case study with Oligocene material from France. Cladistics. Early View.

Wiens J.J. 1995. Polymorphic characters in phylogenetic systematics. Syst. Biol. 44:482–500.

Wiens J.J. 1998. Testing phylogenetic methods with tree congruence: phylogenetic analysis of polymorphic morphological characters in phrynosomatid lizards. Syst. Biol. 47:427–444.

Wiens J.J. 2000. Coding morphological variation within species and higher taxa for phylogenetic analysis. In: Wiens J.J., editor. Phylogenetic analysis of morphological data. Washington, DC, USA: Smithsonian Institution Press. p. 115–145.

Wiens J.J. 2007. Species delimitation: New approaches for discovering diversity. Syst. Biol. 56:875–878.

Wiens J.J., Penkrot T.A. 2002. Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (*Sceloporus*). Syst. Biol. 51:69–91.

Wiens J.J., Servedio M.R. 2000. Species delimitation in systematics: inferring diagnostic differences between species. Proc. R. Soc. Lond. B Biol. Sci. 267:631–636.

For Peer Review Only

1930 Wiley E.O. 1978. The evolutionary species concept reconsidered. Syst. Biol. 27:17–26.

1931 Wilkinson M. 1997. Characters, congruence and quality: A study of neuroanatomical and traditional data in caecilian phylogeny. Biol. Rev. 72:423–470.

1933 Wood B.A., Li Y., Willoughby C. 1991. Intraspecific variation and sexual dimorphism in cranial and dental variables among higher primates and their bearing on the hominid fossil record. J. Anat. 174:185–205.

1936 Wood H.E. 2nd. 1931. Lower Oligocene rhinoceroses of the genus *Trigonias*. J. Mammal. 12:414–428.

1938 Xu X., Upchurch P., Mannion P.D., Barrett P.M., Regalado-Fernandez O.R., Mo J., Ma J., Liu H. 2018. A new Middle Jurassic diplodocoid suggests an earlier dispersal and diversification of sauropod dinosaurs. Nat. Commun. 9:2700.

1941 Zachos F.E. 2016. Tree thinking and species delimitation: Guidelines for taxonomy and phylogenetic terminology. Mamm. Biol. - Z. Für Säugetierkd. 81:185–188.

1943 Zachos F.E. 2018. (New) Species concepts, species delimitation and the inherent limitations of taxonomy. J. Genet. 97:811–815.

1945 Zarapkin S.R. 1939. Das Divergenzprinzip in der Bestimmung kleiner systematischer Kategorien. Verhandl VII Intern Kong Entomol. 1:494–518.

1947 Zheng X., O'Connor J.K., Wang X., Pan Y., Wang Y., Wang M., Zhou Z. 2017. Exceptional preservation of soft tissue in a new specimen of *Eoconfuciusornis* and its biological implications. Natl. Sci. Rev. 4:441–452.

For Peer Review Only

Zollikofer C.P.E., León M.S.P. de, Margvelashvili A., Rightmire G.P., Lordkipanidze D. 2014. Response to comment on "A complete skull from Dmanisi, Georgia, and the evolutionary biology of early *Homo*." Science. 344:360–360.

For Peer Review Only

For Peer Review Only

CAPTIONS

Figures and table

Figure 1. Intergeneric (between two specimens of two different genera), intrageneric (between two specimens of ~~the same~~a single genus but different species), and interspecific dissimilarity for extant lacertid taxa (left, middle, and right columns, respectively). The horizontal black line in the boxplots represents the median. NS indicates statistically non-significant differences. Increasing number of stars refers to decreasing significance cutoff ("***"=0.001, "**"=0.01, "*"=0.05). Generally, intraspecific dissimilarity is significantly lower than intrageneric dissimilarity, which is ~~in turn~~ significantly lower than intergeneric dissimilarity. The exceptions are species with low sample size (1 specimen ~~in the~~ per species of *Gallotia*; 2 specimens of *Iberolacerta monticola*; 4 ~~specimens~~ of *Lacerta media*; 3 ~~specimens~~ of *Podarcis siculus*).

Figure 2: Intraspecific dissimilarities for all extant (a) and extinct (b) lacertid species in our dataset. Horizontal black lines in the box plots represent the median. ~~Boxes~~ Dark boxes ~~in blue~~ represent “outlier taxa” that were statistically distinguished from more than two other taxa in the dataset. (b) Extinct lacertid species are compared to ~~Horizontal black lines in the box plots represents the median. O~~overall mean weighted pairwise dissimilarity of extant species, which is 0.2076 ± 0.0579 character state differences per character scored.

~~Figure 3: Intraspecific dissimilarity of extinct lacertid lizards in our dataset (in blue), compared to bulk intraspecific dissimilarity of the sampled extant species (in red). The horizontal black line in the boxplots represents the median.~~ Extinct species have much more variable intraspecific dissimilarity than extant species.

For Peer Review Only

Figure ~~4~~3: Simulation of missing data in extant species, following patterns observed in extinct species. Intraspecific, weighted pairwise dissimilarity scores (y-axis) are given for the whole dataset, the simulated dataset with intermediate values of missing data, the simulated dataset with the same characters missing from the comparison as in the extinct partner species, and the extinct partner species. The extinct partner species are *Plesiolacerta lydekkeri* (for *Lacerta agilis*), *Dracaenosaurus croizeti* (for *L. bilineata*), *Pseudeumeces cadurcensis* (for *L. trilineata*), *Mediolacerta roceki* (for *Podarcis muralis*), “*L.*” *filholi* (for *Psammodromus algirus*), “*L.*” *siculimelitensis* (for *Timon lepidus*). NS indicates non-significant differences. Increasing number of stars refers to decreasing significance cutoff (“***”=0.001, “**”=0.01, “*”=0.05). The black line in the box plots represents the median, ~~red~~ diamonds represent the mean.

Figure ~~5~~4. Principa~~le~~l Coordinate Analysis based on dissimilarity scores highlighting the different genera (a), and species within *Lacerta* (b), *Timon* (c), and *Podarcis* (~~e~~d). Genera can be more easily distinguished in this way than species.

Figure ~~6~~5: Correlation of average completeness score, AOI, and COI within a species (y-axis) and number of specimens per extant (squares) and extinct (circles) lacertid species (x-axis). Trendlines are indicated with solid lines for extant and dashed lines for extinct species (completeness, long dashes; AOI, intermediate length of dashes; COI, short dashes). AOI seems most correlated with sample size in extant species, but extinct species show different patterns.

Figure ~~7~~6: Distribution of missing characters from the pairwise comparisons relative to percent of maximum dissimilarity observed in extant and extinct lacertid species. There does not seem to be a general trend of higher dissimilarity or ranges of dissimilarity with more missing characters.

For Peer Review Only

Figure ~~8~~7: Observed dissimilarity values relative to sample size subsampled in the four best-represented lacertid species in our dataset. ~~Maximum and minimum values do not overlap significantly with four or more specimens, whereas~~Variability in the average values of mean pairwise dissimilarity (triangles) does not overlap with observed maximum (dots) and minimum values (squares) ~~also do not overlap with average values when~~once -sampling includes seven or more specimens.

Table 1: Completeness (C), All Characters Overlap Index (AOI), and Comparable Characters Overlap Index (COI) within extant and extinct lacertid species in the complete dataset and partitions.~~Table 1: Species sampling for disparity analyses of osteological variability in extant lacertids.~~

~~Table 2: Completeness scores per species and partition.~~

~~Table 3: Overlap Indices within extant and extinct lacertid species in the complete dataset and partitions.~~

~~Supplementary Material~~

Commented [ec1]: update

~~1- Locked excel files with calculations of AOIs and COIs for complete character set and partitions.~~

~~2- Matrix for disparity analysis. Excel file.~~

~~3- Matrices for the disparity analyses with the fossil simulation. Excel file.~~

~~4- Character list. Doc file.~~

For Peer Review Only

5- Supplementary Tab. 1: Specimens of extant lacertid species included in disparity analyses. Skeletal maturity, sex, locality, and snout-vent-length (SVL) provided where known, plus source for scores.

6- Supplementary Tab. 2: Specimens and locality-level OTUs of extinct lacertid species included in disparity analyses. Skeletal content, locality, and geological age provided where known, plus source for scores. Holotype specimens included are indicated in the “type” column.

7- Supplementary Tab. 3: List of partner specimens of extant and extinct partner species for the fossil simulation study.

For Peer Review Only

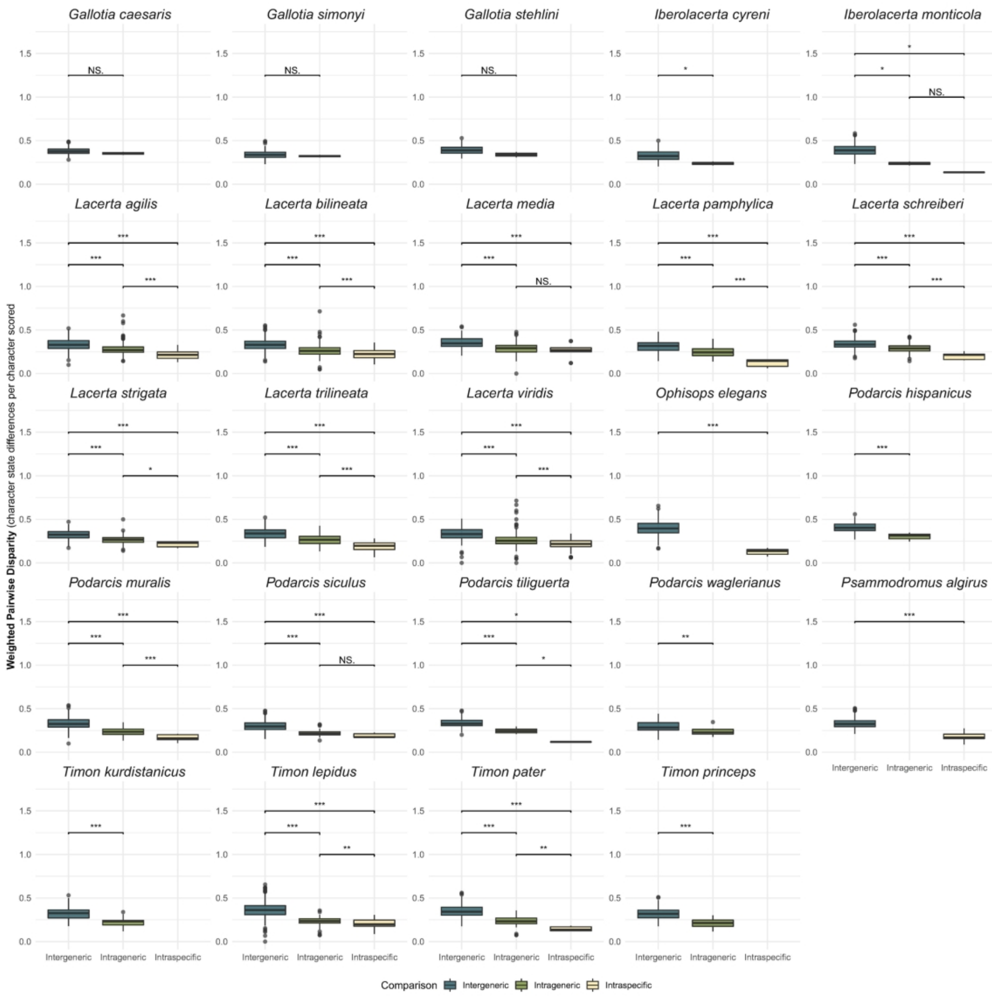


Figure 1. Intergeneric (between two specimens of two different genera), intrageneric (between two specimens of a single genus but different species), and interspecific dissimilarity for extant lacertid taxa (left, middle, and right columns, respectively). The horizontal black line in the boxplots represents the median. NS indicates statistically non-significant differences. Increasing number of stars refers to decreasing significance cutoff (***=0.001, **=0.01, *=0.05). Generally, intraspecific dissimilarity is significantly lower than intrageneric dissimilarity, which is significantly lower than intergeneric dissimilarity. The exceptions are species with low sample size (1 specimen per species of *Gallotia*; 2 specimens of *Iberolacerta monticola*; 4 of *Lacerta media*; 3 of *Podarcis siculus*).

202x202mm (300 x 300 DPI)

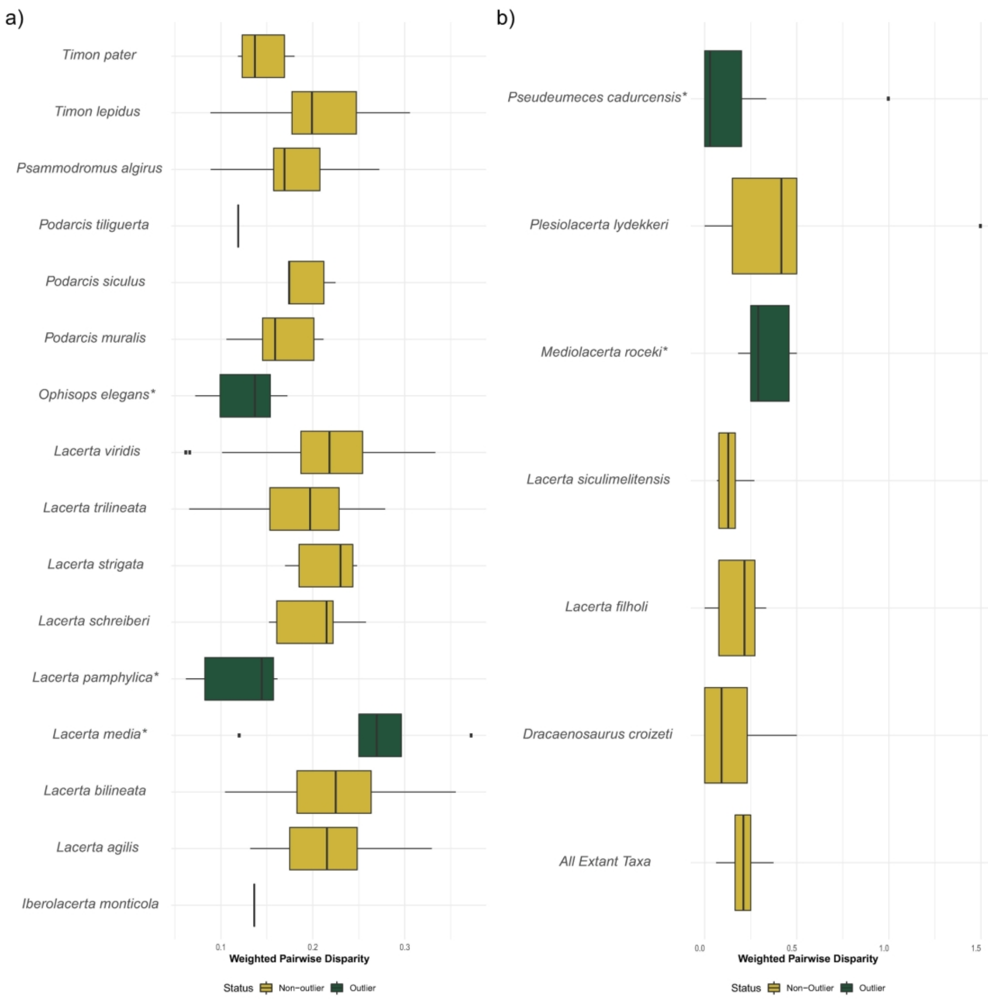


Figure 2: Intraspecific dissimilarities for all extant (a) and extinct (b) lacertid species in our dataset. Horizontal black lines in the box plots represent the median. Dark boxes represent “outlier taxa” that were statistically distinguished from more than two other taxa in the dataset. (b) Extinct lacertid species are compared to overall mean weighted pairwise dissimilarity of extant species, which is 0.2076 ± 0.0579 character state differences per character scored. Extinct species have much more variable intraspecific dissimilarity than extant species.

203x203mm (300 x 300 DPI)

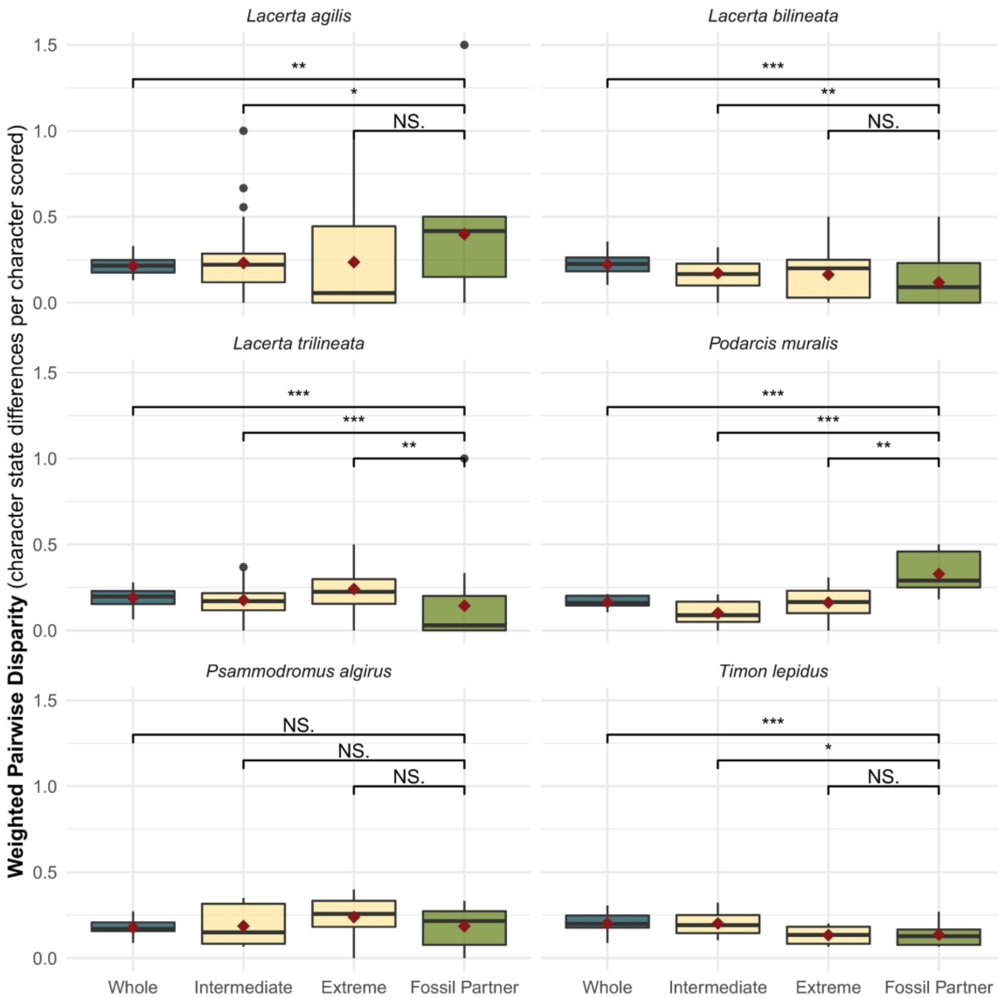


Figure 3: Simulation of missing data in extant species, following patterns observed in extinct species. Intraspecific, weighted pairwise dissimilarity scores (y-axis) are given for the whole dataset, the simulated dataset with intermediate values of missing data, the simulated dataset with the same characters missing from the comparison as in the extinct partner species, and the extinct partner species. The extinct partner species are *Plesiolacerta lydekkeri* (for *Lacerta agilis*), *Dracaenosaurus croizeti* (for *L. bilineata*), *Pseudeumeces cadurcensis* (for *L. trilineata*), *Mediolacerta roceki* (for *Podarcis muralis*), "*L.*" *filholi* (for *Psammodromus algirus*), "*L.*" *siculimelitensis* (for *Timon lepidus*). NS indicates non-significant differences. Increasing number of stars refers to decreasing significance cutoff (***=0.001, **=0.01, *=0.05). The black line in the box plots represents the median, diamonds represent the mean.

199x199mm (300 x 300 DPI)

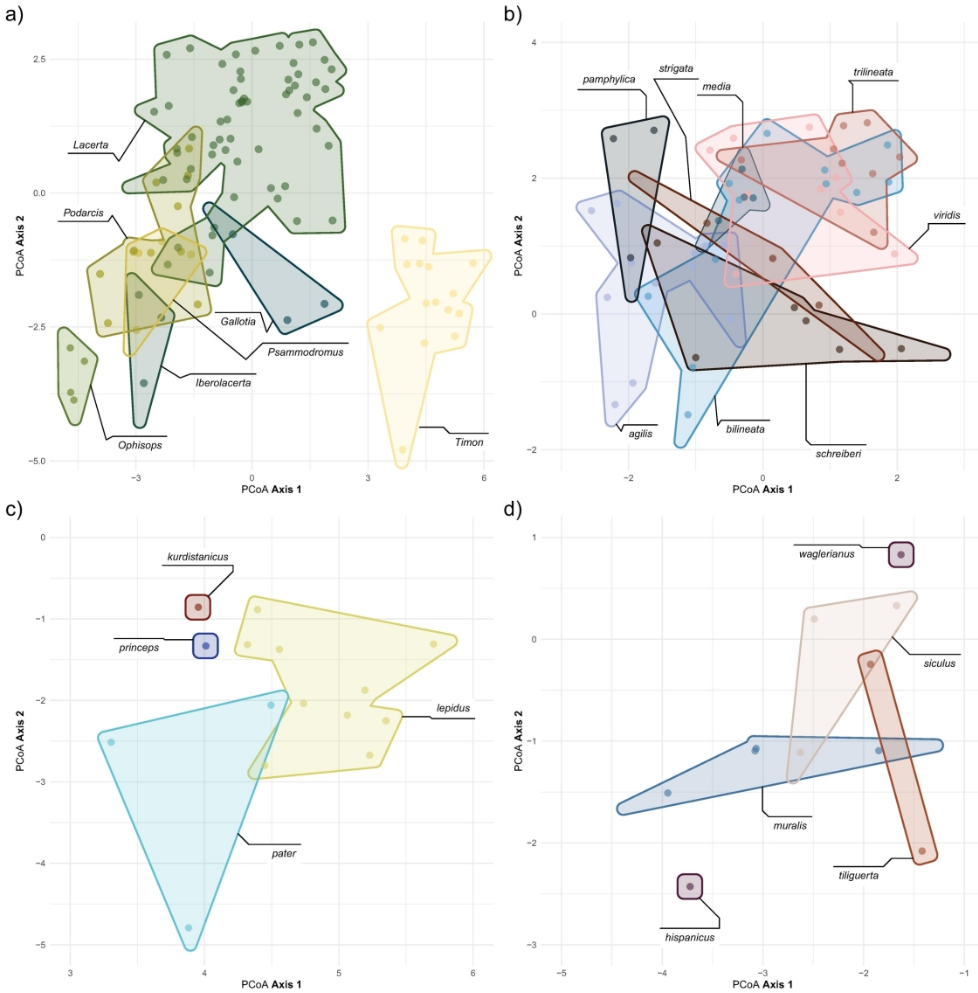


Figure 4. Principal Coordinate Analysis based on dissimilarity scores highlighting the different genera (a), and species within *Lacerta* (b), *Timon* (c), and *Podarcis* (d). Genera can be more easily distinguished in this way than species.

203x203mm (300 x 300 DPI)

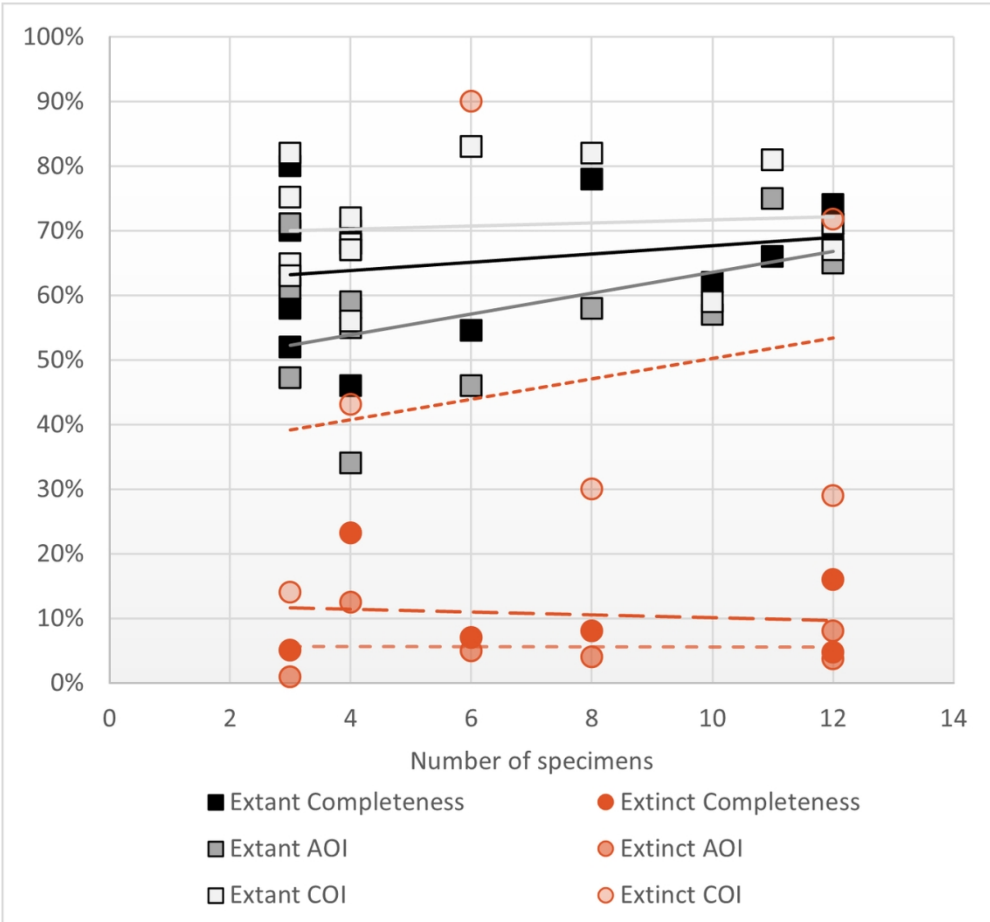


Figure 5: Correlation of average completeness score, AOI, and COI within a species (y-axis) and number of specimens per extant (squares) and extinct (circles) lacertid species (x-axis). Trendlines are indicated with solid lines for extant and dashed lines for extinct species (completeness, long dashes; AOI, intermediate length of dashes; COI, short dashes). AOI seems most correlated with sample size in extant species, but extinct species show different patterns.

115x107mm (300 x 300 DPI)

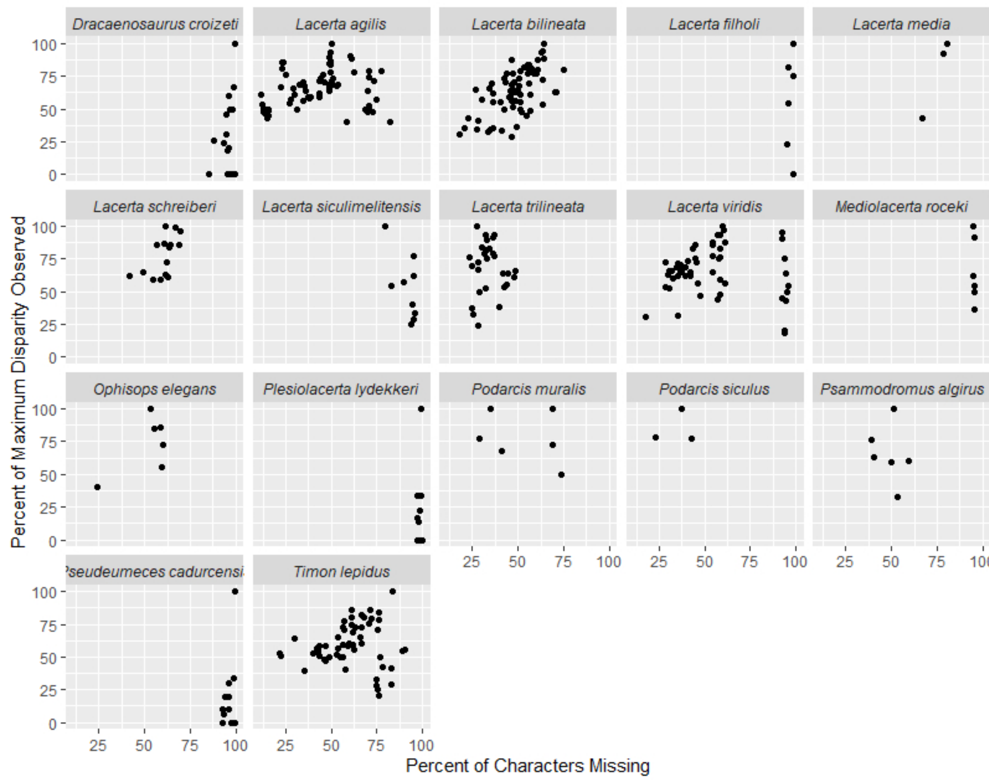


Figure 6: Distribution of missing characters from the pairwise comparisons relative to percent of maximum dissimilarity observed in extant and extinct lacertid species. There does not seem to be a general trend of higher dissimilarity or ranges of dissimilarity with more missing characters.

538x423mm (38 x 38 DPI)

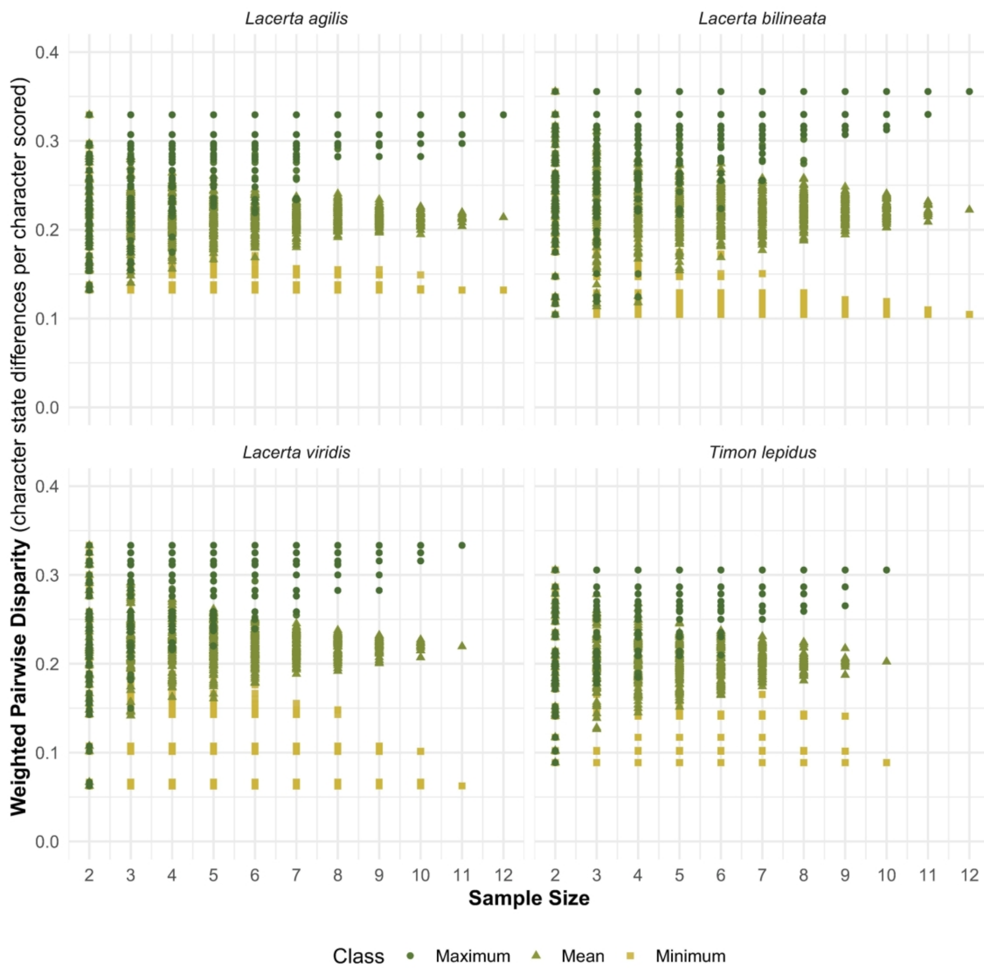


Figure 7: Observed dissimilarity values relative to sample size subsampled in the four best-represented lacertid species in our dataset. Variability in the average values of mean pairwise dissimilarity (triangles) does not overlap with observed maximum (dots) and minimum values (squares) once sampling includes seven or more specimens.

203x203mm (300 x 300 DPI)

Table 1: Completeness (C), All Characters Overlap Index (AOI), and Comparable Characters Overlap Index (COI) within extant and extinct lacertid species in the complete dataset and partitions.

			Complete (253)			Cranial (167)			Dental (17)			Postcranial (69)		
	Species	OTUs	C	AOI	COI	C	AOI	COI	C	AOI	COI	C	AOI	COI
extant	*Lacerta agilis*	12	74%	71%	71%	75%	72%	73%	88%	84%	84%	68%	64%	64%
	Lacerta bilineata	12	68%	65%	67%	75%	72%	73%	87%	84%	84%	49%	44%	47%
	Lacerta media	4	46%	34%	56%	53%	40%	57%	68%	57%	64%	24%	15%	44%
	Lacerta pamphylica	3	55%	46%	83%	52%	64%	84%	69%	59%	71%	60%	0%	0%
	Lacerta schreiberi	6	70%	47%	65%	81%	44%	62%	92%	64%	83%	38%	51%	67%
	Lacerta strigata	3	78%	58%	82%	80%	71%	83%	85%	88%	88%	71%	17%	67%
	Lacerta trilineata	8	66%	75%	81%	64%	77%	83%	80%	82%	82%	65%	67%	76%
	Lacerta viridis	11	52%	61%	63%	71%	60%	64%	67%	76%	76%	0%	60%	60%
	Ophisops elegans	4	67%	55%	68%	78%	67%	77%	88%	76%	87%	37%	21%	33%
	Podarcis muralis	4	68%	58%	67%	72%	63%	71%	88%	78%	83%	55%	40%	52%
	Podarcis siculus	3	80%	71%	82%	85%	77%	85%	96%	88%	94%	65%	52%	69%
	Psammodromus algirus	4	70%	59%	72%	75%	66%	78%	94%	84%	90%	50%	34%	48%
	Timon lepidus	10	62%	57%	59%	61%	56%	60%	79%	75%	75%	59%	53%	54%
	Timon pater	3	58%	47%	75%	58%	49%	84%	90%	88%	94%	50%	33%	50%
extinct	*Dracaenosaurus croizeti*	7	16%	8%	29%	19%	10%	26%	42%	29%	45%	1%	0%	0%
	"*Lacerta*" *filholi*	4	5%	4%	72%	3%	2%	61%	40%	33%	81%	0%	0%	0%
	"*Lacerta*" *siculimelitensis*	5	23%	12%	43%	24%	13%	45%	44%	32%	69%	16%	6%	25%
	Mediolacerta roceki	4	7%	5%	90%	6%	4%	100%	44%	39%	83%	0%	0%	0%
	Plesiolacerta lydekkeri	12	5%	1%	14%	4%	0%	12%	16%	9%	17%	3%	1%	12%
	Pseudeumeces cadurcensis	8	8%	4%	30%	8%	3%	22%	35%	29%	45%	0%	0%	0%

For Review Only