The data associated with this paper are available for review via Dryad. The following is a temporary direct download link. Please copy and paste it directly into a web browser to download the data files to your computer (unfortunately this may not work as a link to click on)

https://datadryad.org/stash/share/lKrKZ_il19K-mIEARmhqKvIrBTDsMi3-2Rppck5NtW4

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[2,5], Lukardis CM Wencker[3], Massimo Delfino[3,6], Paul Upchurch[7]

1- 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

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

7- 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 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: SPECIES COMPARABILITY IN BIOLOGY AND PALAEONTOLOGY

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

Species are the fundamental biological 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 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, most famously Mayr's (1942) Biological Species Concept (Zachos 2016, 2018). However, most of these 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 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 different species concepts led researchers to develop various approaches to delimit species, which 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; Queiroz 2007; Hausdorf and Hennig 2010; Hausdorf 2011; Carstens et al. 2013). Any such recognized "species" unit (taxonomic species, sensu Simpson 1940) is an estimate of the naturally occurring species (real species, sensu Simpson 1940) and will approach the real species

to differing degrees. Hence, although these units are all called "species", they are not necessarily comparable (Cracraft 1987), and should not be used in a comparative context. 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). 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. 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 et al. 2013). Given that species are generally used as fundamental units in a variety of biological studies, these issues have wide-reaching implications (Sites and Marshall 2003, 2004; Balakrishnan 2005). We herein call this issue the "species comparability problem".

The species comparability problem is 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 restricted amount of data results from having few specimens per species, as well as having a limited range of data preserved in each specimen. Consequently, even if palaeontologists 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, 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). The species comparability problem can be added as a fourth issue, resulting from the "species nature problem" and the "species recognition problem" – as long as we delimit extant species using methodologies that 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 direct comparison to the holotype specimen (a typological interpretation of the species), or in comparison to observed intraspecific variability of a "type population", where the holotype may not represent the arithmetic mean (a polytypic or population interpretation of the species; Simpson 1940; Mayr 1942; Dzik 1985).

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 are often incomplete, hampering comparison 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 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 (Wiens 2007). 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, which would erroneously double the species count for sexually reproducing species (Wiley 1978).

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 any study using “species” as its basic unit, including analyses of

biodiversity through Deep Time (e.g., Carrasco 2013), extinction rates, 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 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). 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. Our study is the first to quantify intraspecific variability across a number of extant and extinct lacertid species, based on a large sample of osteological characters.

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 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 by Villa et al. (2017) and extended and modified by Tschopp et al. (2018b). These datasets 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.

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 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).

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. We included 99 extant specimens in the dataset for the dissimilarity analyses. Of the 24 sampled species, 16 were represented by two or more specimens (up to twelve; Supplementary Table 1); a total of 91 specimens were used for our calculations of intraspecific osteological variability. These include all eight sampled species of *Lacerta* (*L. agilis*, *L. 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 eight species solely contributed to the calculation of intrageneric and intergeneric variability.

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 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.

Specimen Identification.—Correct species identifications of the sampled specimens is paramount to studies 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 staff urged caution. All specimens with dubious identification were excluded from scoring. Many of the studied specimens were referred to a species and accessioned in collections 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 specimens, skeletal preparations included, catalogued as "*Lacerta ocellata*" and *Lacerta viridis* with associated locality information could still be attributed to their respective species.

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

302 *Phylogenetic Framework*

303 The phylogenetic framework we followed is based on earlier works (Carranza et al. 2004;
304 Arnold et al. 2007; Kapli et al. 2011; Pyron et al. 2013; Mendes et al. 2016; Čerňanský et al.
305 2016b, 2017; Cruzado-Caballero et al. 2019). Given that the use of our compiled morphological
306 matrix for phylogenetic inference may be limited (see Dataset – Character sampling), we refrain
307 from performing an independent analysis based on our own dataset. However, the main
308 importance for this study is that all included species belong to Lacertidae, so we can assess if
309 osteological intraspecific variability is consistent among the extant species in this particular
310 clade, and could reasonably be used as a guideline to delimit extinct lacertid species, as well.

311 Molecular, morphological, and total-evidence phylogenetic analyses all recover the
312 extant species in our dataset as members of Lacertidae. All three major lacertid clades are
313 represented in our dataset: *Gallotia* and *Psammodromus* are gallotiine lacertids (Carranza et al.
314 2004; Arnold et al. 2007; Pyron et al. 2013; Mendes et al. 2016; Čerňanský et al. 2016b, 2017;
315 Cruzado-Caballero et al. 2019), *Ophisops elegans* is an eremiadin lacertine (Kapli et al. 2011;
316 Pyron et al. 2013), and the remaining species belong to Lacertini (Carranza et al. 2004; Arnold et
317 al. 2007; Kapli et al. 2011; Pyron et al. 2013; Mendes et al. 2016).

318 The extinct species were identified as lacertids based on particular diagnostic characters
319 (mostly in the jaw; Supplementary Data 3). Some were later confirmed to be lacertids in
320 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. 2021). Based on these works, the species can be tentatively referred to the subclades 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 one 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. Pairwise dissimilarity based on a discrete character matrix was applied to delimit taxonomic units by Benson et al. (2012; in plesiosaurs, using mean values of species-level OTUs to delimit genera) and by Tschopp et al. (2015; in sauropod dinosaurs, where both species and genus 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 (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 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). 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

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 in our dataset, missing data is widespread in our sample. Given the highly divergent completeness 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 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 untouched. The real fossil specimens 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.

The intermediate simulation was done using two custom Python scripts. 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. If missing values 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.

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 Coordinate Analysis.—First, we performed a principal 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 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 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, sample size, and skeletal modularity in our dataset. We used a third custom Python script (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 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) (Tschopp et al. 2015; 2018a). 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 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) for the calculation of the overlap indices. The AOI and COI were calculated for every ingroup species assessed for intraspecific variability (Supplementary Data 9-12). 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 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), *L. bilineata* (N = 12), *L. viridis* (N = 11), and *Timon lepidus* (N = 10). For each taxon, resamples were done

459 with numbers of specimens ranging from two to the maximum sampled, with each sample size
460 replicated 100 times, and the maximum, minimum, and mean pairwise dissimilarity recorded.

461 RESULTS

462 *Pairwise Dissimilarity*

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

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

Pairwise comparisons of all extant taxa recovered most taxa as displaying statistically indistinguishable intraspecific dissimilarity (Fig. 2a) – thus, 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 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* were significantly less dissimilar than *L. agilis*, *L. bilineata*, *L. media*, *L. viridis*, and *T. lepidus*, *O. elegans* was also significantly less dissimilar than *L. schreiberi* and *L. trilineata*. Aside from these three outlier taxa, *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 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 qualitative dental characters, were all significantly more dissimilar

503 intraspecifically than the pooled variation. Quantitative dental characters were significantly less
504 intraspecifically dissimilar than the pooled variation.

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

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

523 In the extreme fossilization simulation, no simulated "fossilized" taxon differed
524 significantly from the extant taxa in terms of pairwise intraspecific dissimilarity, in contrast to
525 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 *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 *"L." siculimelitensis*, while *Psammodromus algirus* continues to resemble "*L.*" *filholi*, as it did in its original scoring and in the intermediate fossilization simulation. *Lacerta trilineata* and *Podarcis muralis* remained significantly different from their extinct partner species *Pseudeumeces cadurcensis* and *Mediolacerta roceki*, respectively, suggesting that the low observed intraspecific variability in *Pseudeumeces cadurcensis* and the high variability in *M. roceki* are true signals. Although *L. agilis* and *L. bilineata* are not statistically distinguishable from their extinct partners under an extreme fossilization simulation, the persistent differences in median values shown in Figure 3 suggest that the patterns seen in *Plesiolacerta lydekkeri* and *D. croizeti* may be true signals as well, which are not distinguishable in our dataset due to sample size.

Data Exploration and Sensitivity Analyses

Principal 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. 4a). 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

549 together, the only *Lacerta* species to cluster together significantly was *Lacerta pamphylica*, and
550 several *Lacerta* specimens cluster with specimens of *Podarcis* rather than congeners.
551 Even with only specimens of *Lacerta* included, there is significant overlap between
552 species, and there is no significant tendency for hierarchical clustering analysis to group
553 specimens of a single species to the exclusion of those referred to others (Fig. 4b). An exception
554 is *Lacerta pamphylica*, of which all three specimens cluster together when all taxa are analysed,
555 but this cluster does not include one of the specimens when only *Lacerta* is included in the
556 analysis. This is probably a consequence of lacking more disparate species, which make *L.*
557 *pamphylica* appear more distinct when they are included for comparison. Other species often
558 cluster partially. For instance, *L. agilis* is split into one significant cluster of five specimens, with
559 the other seven tending to cluster insignificantly with two *L. viridis* specimens. Within *Podarcis*
560 (Fig. 4c), only *P. tiliguerta* forms a statistically significant cluster, with specimens of *P. muralis*,
561 *P. siculus*, and *P. waglerianus* mixed together into a statistically insignificant cluster. Within
562 *Timon* (Fig. 4d), *T. lepidus* mostly forms one statistically insignificant cluster, but one specimen
563 is recovered in a near significant cluster with *T. pater*. The sole specimens of *T. princeps* and *T.*
564 *kurdistanicus* cluster together.
565 *Missing Data*.—Within the matrix, missing data is distributed unequally (Table 1),
566 indicating that the absence of character scores in both extant and extinct species is non-random.
567 Such a non-random distribution of missing entries is fairly typical for morphological datasets,
568 especially when they include extinct taxa (e.g., Smith et al. 2014; Gerber 2019). As expected,
569 extant species have much higher completeness scores than extinct species that are nearly always
570 represented by fragmentary specimens. Throughout the entire dataset, dental characters
571 (quantitative and qualitative) have fewer missing entries than cranial characters (except for

572 *Lacerta pamphylica*, which could only be scored from published figures), and cranial characters
573 are more completely scored than postcranial characters (except for *L. viridis*, which is the sole
574 extant species in our dataset that includes a specimen that only preserves postcranial material). In
575 all extant species, quantitative postcranial characters are the ones with most missing entries per
576 species, and quantitative dental characters make up the most completely scored subset (except
577 for *L. pamphylica*). In extinct species, quantitative dental characters are not consistently the most
578 completely scored subset: in *Dracaenosaurus croizeti* and *Pseudeumeces cadurcensis*,
579 qualitative dental characters have the least missing entries. In the extant species, the amount of
580 missing data is slightly correlated with numbers of specimens per species (Fig. 5), but there is no
581 correlation with number of characters per subset (Table 1). However, as mentioned above, these
582 absolute values of missing data are not necessarily correlated with the utility of the dataset for
583 analyses of pairwise dissimilarity, which requires at least two individuals scored for the same
584 character.

585 Quantification of anatomical overlap shows that extant lacertids have AOIs ranging from
586 34% (*Lacerta media*) to 75% (*L. trilineata*) when analysing the entire dataset, whereas extinct
587 species have values between 1% (*Plesiolacerta lydekkeri*) and 12% ("*L.*" *siculimelitensis*). COIs
588 covering the entire dataset range from 56% (*L. media*) to 83% (*L. pamphylica*) in extant species,
589 and from 14% (*P. lydekkeri*) to 90% (*Mediolacerta roceki*) in extinct taxa (Table 1). The AOI
590 and COI are slightly correlated with completeness values in extant taxa (the AOI slightly more
591 so than the COI). Whereas the COI of extant species does not seem to correlate with the number
592 of OTUs in a particular species, the AOI does so, even more than regular completeness values
593 (Fig. 5). Overlap indices in the extinct taxa, however, show the opposite trends, with the COI
594 being most correlated with number of specimens, AOI being stable, and completeness decreasing

595 with higher numbers of specimens (Fig. 5). As expected, extinct species generally have much
596 lower absolute numbers of comparable characters (characters with anatomical overlap), total
597 number of overlaps and AOI within the species compared to extant species. Total number of
598 characters with anatomical overlap among OTUs of a particular extinct species range from 13
599 comparable characters (with 28 overlaps; "*L.*" *filholi*) to 73 (with 127 overlaps; *Dracaenosaurus*
600 *croizeti*). The lowest numbers in extant species are present in *L. pamphylica* with 142
601 comparable characters and 235 overlaps (Supplementary Table 6).Among the partitioned
602 character sets, AOI and COI are generally highest in the dental characters, and lowest in the
603 postcranial characters, both in extant and extinct taxa. This is in part because all but two of the
604 sampled extinct species entirely lack anatomical overlap in the postcranial partition (the
605 exceptions are "*Lacerta*" *siculimelitensis* and *Plesiolacerta lydekkeri*). However, a lack of
606 anatomical overlap in a particular partition does not necessarily mean that there are no characters
607 scored in these taxa. For instance, *P. lydekkeri* was sampled by the most specimens of all extinct
608 species; its lowest overall AOI and COI result from different specimens having different bones
609 preserved that cannot be directly compared. These results further highlight that the negative
610 impact on dissimilarity analyses does not derive from the missing data per se, but from reduced
611 anatomical overlap.

612 The number of characters missing from a particular pairwise comparison has no
613 consistent relationship with the recovered dissimilarity (Fig. 6). In most taxa, dissimilarity does
614 not seem to be correlated with the amount of lacking anatomical overlap, though some (most
615 notably *Lacerta bilineata*) appear to show a trend of increased dissimilarity with decreasing
616 overlap, and several show the highest dissimilarity between 150 and 200 missing characters (of a
617 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 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.

Sample Size.—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. 7). By a sample size of four or more individuals, the distribution 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.

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 has the lowest anatomical overlap scores over all characters as well as within the cranial, dental, and postcranial subsets (Table 1). The high overall dissimilarity is driven by high variability in qualitative cranial and postcranial characters (Supplementary Table 5). *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). Additional 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 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 figures provided by Čerňanský and Syromyatnikova (2019), 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 osteological variability. This low overall variability is mostly driven by a low variability in the qualitative cranial characters (0.0991 ± 0.0482; Supplementary Table 5), 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 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

686 outlier species, the pattern of variability is intriguing, especially because Montgelard et al. (2020)
687 proposed to elevate *O. e. persicus* to species level (so we would have two species represented in
688 our sample), but the low sample size, with three of four specimens coming from Samos Island,
689 casts doubt on this pattern being a genuine representation of intraspecific variability across the
690 entire species. Low overall variability of *O. elegans* is driven by a low dissimilarity among
691 qualitative cranial characters, as in *L. pamphylica*, but it also has very low to non-existent
692 variability in quantitative cranial and dental characters and qualitative dental and postcranial
693 characters (Supplementary Table 5). Quantitative cranial and qualitative dental characters are
694 otherwise more variable than average, so their low values in *O. elegans* is peculiar. Given that
695 this was the only representative of the lacertid subclade Eremiadini (and that the sampled
696 specimens may represent two distinct species), one might take this as an indication that patterns
697 within Eremiadini are different from other lacertids, but additional species, subspecies, and
698 specimens will have to be sampled in this clade to confirm this. At present, we cannot
699 confidently exclude that the significant differences in intraspecific variability between these
700 three outlier species and some (though not all) other species are artefacts of low sample size and
701 restricted anatomical overlap.
702 The results from our studies corroborate that current species delimitation is generally
703 robust in the extant species we analysed, and that these taxa do not suffer considerably from the
704 species comparability problem. This stability suggests that osteological intraspecific variability
705 can be used as a proxy for other secondary defining properties and may be suitable for species
706 delimitation even in the absence of autapomorphic osteological features in a particular species
707 (as is the case in some of the analysed lacertids; Villa et al. 2017). Hence, these values may also
708 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 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. Moreover, palaeontology provides a unique opportunity to study species through time, which while generally beneficial, could lead to time-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). We 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 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 (see Dataset – Extinct taxon and specimen sampling). OTU construction may thus have artificially lowered intraspecific osteological variability in "*L*." *siculimelitensis* (see below for a detailed assessment of this species).

Missing Data and low Sample Size.—Missing entries in our dataset result in much lower numbers of anatomical overlaps in extinct versus extant species (Fig. 6; Table 1), which can have

a substantial impact on pairwise dissimilarity analyses (Smith et al. 2014; Gerber 2019). 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 (Smith et al. 2014). This pattern was partially confirmed by our simulations. Given 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, it is unlikely that the high number of missing scores for extinct taxa is the only 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 (Fig. 3). *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, suggesting 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 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, it has been 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 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). Thus, taxa represented by seven or more individuals in our dataset probably show representative mean dissimilarity and variance that are comparable among each other, whereas some doubts remain for those species sampled by fewer specimens – especially those with divergent results (as is the case in the outlier species discussed above). This result 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; with the ever-increasing availability of CT scans of wet-specimens, providing a wealth of additional information that is not visible in skeletal preparations, this should only be a matter of time. In sum, an inclusion of seven or more specimens per species is advisable, but dissimilarity analyses with lower sample sizes may yield meaningful results if they are carefully assessed for potential shortcomings due to low sample size.

778 *Uneven Sampling of Ontogeny and Sexes*.—Intraspecific variability is greatly affected by
779 sexually dimorphic features and ontogenetic changes. In a complete sample including members
780 of both sexes and from various ontogenetic stages, these two factors should probably not have a
781 large impact on the mean dissimilarity value, although they may increase the observed ranges in
782 dissimilarity considerably. Given that sex and ontogenetic stage are not known for many fossil
783 specimens, especially if they are only partially preserved, it remains difficult to quantify the
784 amount of variability that is absent in the extinct dataset. Hence, the expected impact on studies
785 of extinct species would be a lower range in variability compared to more extensively sampled
786 extant species, similar to the effect of low anatomical overlap and sample sizes in general. This
787 would be especially the case if sexual morphs, instead of being recognized as different sexes of a
788 single species, are erroneously treated as distinct species given their diverging morphology,
789 something that is very difficult to assess in palaeontological samples (Wiley 1978; Tschopp and
790 Upchurch 2019). In any case, sexual dimorphisms and ontogenetically variable characters often
791 affect certain character complexes. In the case of sexual dimorphism, these are often restricted to
792 soft tissue morphology associated with the reproductive tract, which is generally not preserved in
793 fossils, or to features bearing a display function that may or may not have osteological correlates
794 (and if they have, they may not be recognised as such in fossils; Mallon 2017). Restricting our
795 dataset to osteological characters, and analysing mean pairwise dissimilarity over a complete set
796 of cranial, dental, and postcranial characters can probably be expected to reduce the confounding
797 impact of these types of intraspecific variability – even though we also deliberately included
798 characters that are known to be variable between sexes and through ontogeny.
799 *Time-averaging in Fossils*.—Time-averaging can result from a sampling of fossil
800 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 that had accumulated over time while the species was adapting to changing environmental conditions through natural selection (Simpson 1937, 1951). The inclusion of specimens from potentially different evolutionary stages within the same species would be expected to increase 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.

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*).

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

824 "typical" amounts of intraspecific osteological variability within a species. What this means for
825 the extinct species analysed here is discussed below.
826 *Dracaenosaurus croizeti*.—The observed variability in *D. croizeti* is significantly lower
827 than any other species we analysed, be it extant or extinct. This is true for the whole dataset as
828 well as the modules of qualitative cranial characters (the majority of characters in the dataset)
829 and qualitative and quantitative dental characters (Supplementary Table 5). Variability is not
830 significantly lower compared to the values obtained in its extant partner species *Lacerta*
831 *bilineata* with artificially decreased anatomical overlap (although the mean value remains much
832 lower; 0.1182±0.1425 in *D. croizeti*; 0.1632 ±0.1297 in *L. bilineata*). Thus, we cannot
833 completely rule out that low anatomical overlap is driving these discordant values. However,
834 time-averaging would have worked against the low values (the sampled specimens cover
835 900'000 years; Böhme and Ilg 2003). Another reason for the low values might be the fact that *D.*
836 *croizeti* is highly specialized, with its strongly enlarged posterior teeth adapted for crushing and
837 the generally stout skull and jaws for the attachment of strong musculature (Hoffstetter 1944;
838 Müller 2004; Čerňanský et al. 2017). Such an advanced specialization was possibly favoured by
839 strong natural selection that ultimately constrained aspects of morphology and so reduced
840 intraspecific variability, especially in cranial and dental characters. Postcranial material of the
841 sampled *D. croizeti* specimens was excluded from contributing to the calculation of intraspecific
842 variability because of the absence of anatomical overlap in this module (Table 1), which was
843 probably less constrained in morphology by this feeding specialization (and may thus have
844 increased the dissimilarity). Additionally, researchers may have been overly cautious in referring
845 specimens with slightly diverging morphologies to this species, thereby applying a more strictly
846 typological species concept when identifying fossils. This would suggest that additional material

847 now referred to "*Dracaenosaurus* sp." should indeed be assigned to *D. croizeti* as well.
848 Specimens that were not identified to species level are all from the same localities in France and
849 Germany that also produced specimens referred to *D. croizeti* (Böhme and Ilg 2003; Čerňanský
850 et al. 2016a), further supporting our suggestion.
851 "*Lacerta*" *filholi*.—No significant difference was found in recovered intraspecific
852 variability of "*L.*" *filholi* compared to extant species, as well as compared to its extant partner
853 species (*Psammodromus algirus*) in both the original dataset and the simulated dataset with
854 artificially reduced anatomical overlap. This indicates that the species "*L.*" *filholi* may represent
855 a unit comparable to extant species, although the included specimens cover a time span of
856 approximately 1.2 Myr (Böhme and Ilg 2003). We interpret our results with caution because the
857 species is represented in our dataset by very few, disarticulated specimens, and there may be
858 much more variability occurring in the entire duration of the species as currently understood.
859 Only 13 characters could be compared in this species, but the few comparable characters were
860 shared in several specimens. These are almost entirely restricted to dental and mandibular
861 features (the specimens referred to this species by Augé (2005) only include dentaries, maxillae,
862 a few premaxillae, and a coronoid), and it remains to be seen if other cranial and postcranial
863 material would alter the observed variability. In any case, the calculated intrageneric variability
864 with other *Lacerta* specimens was found to be significantly larger than normal intrageneric
865 variability within extant taxa, and even exceeded most of the recovered dissimilarity scores
866 calculated between extant genera (Supplementary Table 5). This finding supports earlier studies
867 suggesting that the referral of this species to the genus *Lacerta* is questionable (Augé 2005; Augé
868 and Hervet 2009; Wencker et al. 2021).

"*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 5). The five OTUs included in the present analysis span a time range of 1.72 Myr (Delfino and Bailon 2000; Böhme and Ilg 2003; Tschopp et al. 2018b) 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 mean) being present if individual specimens were scored separately. Thus, we expect this species to be comparable to extant species. 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.

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 (several of them were scored too incompletely to yield any data; Supplementary Table 5). *Mediolacerta roceki* is also significantly different from its extant partner species *Podarcis muralis* when deleting the same characters as are missing in *M. roceki* (Fig. 3). 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* 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-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 5). 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). 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), and only 35 overlaps (in 23 characters) and thus pairwise comparisons occur between these 12 specimens (resulting in an AOI of 1%; Table 1). 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). 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 5), which is another indication that the value obtained within *P. lydekkeri* represents an outlier far from the true mean dissimilarity value of

the species (Fig. 2b). 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 (Böhme and Ilg 2003), the highest of all extinct species represented in our dataset. 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.

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 stratum that yielded the historical material - 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 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

The differences in intraspecific dissimilarity seen in the extinct lacertids 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 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 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). 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). If an extinct species, as a taxonomic unit, would be comparable to an extant genus 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, small North American mammals show an 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 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.

Can we Develop Species Delimitation Methods based on Morphological 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 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 Data 2); our findings that these three species have comparable dissimilarity is probably because our dataset covers the entire skeleton. Focusing on one or few traits is useful to analyse function of convergently acquired features, but it does not contribute much to species delimitation.

Numerous operational criteria have been proposed to 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

1028 divergence (Van Valen 1976) can have significant effects on morphology but does not
1029 necessarily impact the same traits or sets of traits – nor are these changes associated with single
1030 genes (Highton 1990). With time, these processes may lead to evolution and fixation of new
1031 morphological traits that can be considered diagnostic for a particular species (Kimura et al.
1032 2016). However, asserting that a trait is truly fixed is statistically nearly impossible, and even
1033 allowing a 95% fixation rate within a population, sample sizes have to be very large to confirm
1034 that a particular trait can be considered diagnostic (Wiens and Servedio 2000). After prolonged
1035 diverging evolution, diagnostic morphological features may also occur across the entire
1036 organism, but recently diverged species may not have evolved widespread diagnosability, and if
1037 single diagnostic traits occur, they might be in discordance with other features and thus difficult
1038 to interpret (Hausdorf 2011; Harrison and Larson 2014). Using overall variability scores derived
1039 from a set of diverse morphological characters circumvents these problems and may cover all
1040 aspects resulting from evolutionary mechanisms culminating in speciation, even if these incipient
1041 species are not yet diagnosable by particular, fixed, apomorphic traits.
1042 The accuracy of species delimitation based on genetic data also depends on the number of
1043 sampled loci. Single-locus analyses are prone to failure in detecting species status of recently
1044 diverged lineages, whereas combining information from multiple loci resulted in a decrease of
1045 such false-negatives (Knowles and Carstens 2007; Hausdorf and Hennig 2010). By reducing the
1046 number of loci (or morphological characters, for that matter) in an analysis of species
1047 boundaries, one is more likely to be misled by a mismatch of the evolutionary assumptions
1048 underlying the delimitation methodology and the actual evolutionary processes leading to lineage
1049 splitting. If speciation is driven by strong selection resulting in fast evolutionary rates in the
1050 fixation of a genotypic or phenotypic trait, focusing species delimitation on slow-evolving traits

1051 will not be capable of recognizing this recent and rapid lineage-splitting event (Knowles and
1052 Carstens 2007).

1053 Focusing species delimitation on a variety of slow- and fast-evolving traits reflects a
1054 polytypic understanding of species as morphologically (and genetically) variable populations
1055 (following Mayr 1942). It is equivalent to defining the "taxonomic space" as intended by Hull
1056 (1965), using morphological variability as an indicator for the presence of one or more secondary
1057 defining properties (sensu Queiroz 2007) that renders a metapopulation a distinct species. Such
1058 an approach is supported by evidence from *Drosophila*, where morphological differences in male
1059 genitalia are associated with mutations in genes from all chromosomes (Coyne and Kreitman
1060 1986), which led Highton (1990) to propose species delimitation based on genetic distance
1061 calculated across many loci scattered throughout the genome. This overall divergence was
1062 observed to be fairly generalized across non-avian vertebrates (Thorpe 1982), supporting its use
1063 as a proxy for a speciation event (Sites and Marshall 2003), just as we propose to use overall
1064 morphological distance as an indicator for species boundaries.

1065 *Avoiding extreme Values*.—Our sensitivity analyses indicate that the range of maximum
1066 to minimum pairwise dissimilarity correlates with sample size (Fig. 7), and that maximum
1067 dissimilarity is also affected by lacking anatomical overlap (Fig. 6). Additionally, dissimilarity
1068 values can be impacted by observation errors, which may stem from diverse reasons (e.g., wrong
1069 specimen identifications in collections, errors in scoring, divergent interpretations of
1070 morphological characters by researchers). On the other hand, average mean pairwise
1071 dissimilarity scores are not likely to change significantly when adding more specimens per
1072 species, though the standard deviation of the mean dissimilarity does decrease with increasing
1073 sample size. Restricting species delimitation to the use of average and standard deviation, lowers

1074 the impact of such potential sampling error producing extreme values (Cope and Lacy 1992),
1075 especially as more specimens are sampled per species.

1076 *Species Clusters versus Genus Clusters*.—Our PCoA showed that while specimens of a
1077 single genus form distinct clusters in the morphospace, species clusters are not always
1078 recognizable (Fig. 4). Thus, species delimitation approaches have to be based on variability
1079 scores, whereas the morphospace analysis could be used to distinguish higher-level clusters that
1080 could be used to delimit genera. If such numerical boundaries are stable in established and well-
1081 accepted closely related species, they could be applied to other species complexes of the same
1082 clade, be they extant or extinct, where species boundaries are unclear. Combined with a
1083 meaningful specimen-level phylogenetic analysis, variability scores can be calculated between
1084 closely related specimens and added up in a stepwise manner until the species threshold is
1085 reached (a similar approach was used by Tschopp et al. (2015). Thereby, morphological species
1086 delimitation would combine historical consensus on species boundaries in well-known species
1087 and phenetic clustering based on a sound phylogenetic framework. Whether this is best done
1088 with discrete character matrices, geometric morphometrics, or a combination of the two, and
1089 whether this would be applicable to a wide variety of vertebrate clades, forms a rich field for
1090 further investigation.

1091 *Intraspecific Variability in Vertebrates*

1092 Our analysis of lacertid intraspecific variability adds to earlier studies on other vertebrate
1093 clades (e.g., Roth 1992; Wiens and Penkrot 2002; Bever 2009; Foth et al. 2015). Exact values of
1094 variability depend on character and taxon sampling, and can also vary between clades (e.g., tooth
1095 variability in different clades of mammals; Roth 1992). In fact, evidence from other lizards
1096 indicates that intraspecific variability may be higher than interspecific variability in certain

genera (e.g., *Sceloporus*; Wiens and Penkrot 2002). 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 - 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 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 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.

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. Earlier submissions of the MS have greatly benefited by reviews from Andrej Čerňanský (who also shared CT scan data for the revision of the dataset), 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.

1184 Allmon W.D. 2016. Studying species in the fossil record: a review and recommendations for a
1185 more unified approach. In: Allmon W.D., Yacobucci M.M., editors. Species and
1186 speciation in the fossil record. Chicago, USA: University of Chicago Press. p. 59–120.

1187 Allmon W.D., Smith U.E. 2011. What, if anything, can we learn from the fossil record about
1188 speciation in marine gastropods? Biological and geological considerations. Am. Malacol.
1189 Bull. 29:247–276.

1190 Alroy J. 2010. Fair sampling of taxonomic richness and unbiased estimation of origination and
1191 extinction rates. Paleontol. Soc. Pap. 16:55–80.

1192 Anderson E., Abbe E.C. 1934. A quantitative comparison of specific and generic differences in
1193 the Betulaceae. J. Arnold Arbor. 15:43–49.

1194 Anderson E., Whitaker T.W. 1934. Speciation in Uvularia. J. Arnold Arbor. 15:28–42.

1195 Arnold E.N., Arribas O., Carranza S. 2007. Systematics of the Palaearctic and Oriental lizard
1196 tribe Lacertini (Squamata: Lacertidae: Lacertinae), with descriptions of eight new genera.
1197 Zootaxa. 1430:3–86.

1198 Arsdale A.P.V., Wolpoff M.H. 2013. A single lineage in Early Pleistocene *Homo*: Size variation
1199 continuity in Early Pleistocene *Homo* crania from East Africa and Georgia. Evolution.
1200 67:841–850.

1201 Augé M.L. 2005. Evolution des lézards du Paléogène en Europe. Mém. Muséum Natl. Hist. Nat.
1202 1993. 192.

1203 Augé M.L., Hervet S. 2009. Fossil lizards from the locality of Gannat (late Oligocene–early
1204 Miocene, France) and a revision of the genus *Pseudeumeces* (Squamata, Lacertidae).
1205 Palaeobiodiversity Palaeoenvironments. 89:191.

1206 Augé M.L., Smith R. 2009. An assemblage of early Oligocene lizards (Squamata) from the
1207 locality of Boutersem (Belgium), with comments on the Eocene–Oligocene transition.
1208 Zool. J. Linn. Soc. 155:148–170.

1209 Bailon S., Boistel R., Bover P., Alcover J.A. 2014. *Maioricalacerta rafelinensis*, gen. et sp. nov.
1210 (Squamata, Lacertidae), from the Early Pliocene of Mallorca (Balearic Islands, Western
1211 Mediterranean Sea). J. Vertebr. Paleontol. 34:318–326.

1212 Balakrishnan R. 2005. Species concepts, species boundaries and species identification: A view
1213 from the tropics. Syst. Biol. 54:689–693.

1214 Barahona F., Barbadillo L.J. 1998. Inter-and intraspecific variation in the post-natal skull of
1215 some lacertid lizards. J. Zool. 245:393–405.

1216 Barahona F., Evans S.E., Mateo J.A., García-Márquez M., López-Jurado L.F. 2000. Endemism,
1217 gigantism and extinction in island lizards: the genus *Gallotia* on the Canary Islands. J.
1218 Zool. 250:373–388.

1219 Barnosky A.D., Matzke N., Tomiya S., Wogan G.O.U., Swartz B., Quental T.B., Marshall C.,
1220 McGuire J.L., Lindsey E.L., Maguire K.C., Mersey B., Ferrer E.A. 2011. Has the Earth's
1221 sixth mass extinction already arrived? Nature. 471:51–57.

1222 Baum D.A. 1998. Individuality and the existence of species through time. Syst. Biol. 47:641–
1223 653.

1224 Bell C.J., Mead J.I. 2014. Not enough skeletons in the closet: collections-based anatomical
1225 research in an age of conservation conscience. Anat. Rec. 297:344–348.

1226 Bell P.R., Hendrickx C. 2021. Epidermal complexity in the theropod dinosaur Juravenator from
1227 the Upper Jurassic of Germany. Palaeontology. 64:203–223.

1228 Benson R.B.J., Evans M., Druckenmiller P.S. 2012. High diversity, low disparity and small body
1229 size in plesiosaurs (Reptilia, Sauropterygia) from the Triassic–Jurassic boundary. PLoS
1230 ONE. 7:e31838.

1231 Benton M.J., Pearson P.N. 2001. Speciation in the fossil record. Trends Ecol. Evol. 16:405–411.

1232 Bernardi M., Minelli A. 2011. Il concetto di specie e la paleontologia: una rassegna introduttiva.
1233 Rendiconti Online Della Soc. Geol. Ital. 13:2–26.

1234 Bever G.S. 2009. The postnatal skull of the extant North American turtle *Pseudemys texana*
1235 (Cryptodira: Emydidae), with comments on the study of discrete intraspecific variation. J.
1236 Morphol. 270:97–128.

1237 Böhme M., Ilg A. 2003. fosFARbase. Www.wahre-Staerke.com. (accessioned March 2017).

1238 Borczyk B., Kusznierz J., Paśko \Lukasz, Turniak E. 2014. Scaling of the sexual size and shape
1239 skull dimorphism in the sand lizard (Lacerta agilis L.). .

1240 Brazeau M.D. 2011. Problematic character coding methods in morphology and their effects.
1241 Biol. J. Linn. Soc. 104:489–498.

1242 Brecko J., Huyghe K., Vanhooydonck B., Herrel A., Grbac I., Van Damme R. 2008. Functional
1243 and ecological relevance of intraspecific variation in body size and shape in the lizard
1244 *Podarcis melisellensis* (Lacertidae). Biol. J. Linn. Soc. 94:251–264.

1245 Briggs D.E.G., Fortey R.A., Wills M.A. 1992. Morphological disparity in the Cambrian. Science.
1246 256:1670–1673.

1247 Brizuela S. 2010. Los lagartos continentales fósiles de la Argentina (excepto Iguania). .

1248 Brochu C.A., Sumrall C.D. 2020. Modern cryptic species and crocodylian diversity in the fossil
1249 record. Zool. J. Linn. Soc. 189:700–711.

1250 Brown E.E., Cashmore D.D., Simmons N.B., Butler R.J. 2019. Quantifying the completeness of
1251 the bat fossil record. Palaeontology. 62:757–776.

1252 Bruner E. 2004. Evolution, actuality and species concept: a need for a palaeontological tool.
1253 Hum. Evol. 19:93–112.

1254 Bruner E., Costantini D. 2009. Morfología de la cabeza y el grado de variación en *Lacerta*
1255 *bilineata*, *Podarcis muralis* y *Podarcis sicula*. Int. J. Morphol. 27:667–676.

1256 Bruner E., Costantini D., Fanfani A., Dell'Omo G. 2005. Morphological variation and sexual
1257 dimorphism of the cephalic scales in Lacerta bilineata. Acta Zool. 86:245–254.

1258 Brusatte S.L. 2010. Representing supraspecific taxa in higher-level phylogenetic analyses:
1259 guidelines for palaeontologists. Palaeontology. 53:1–9.

1260 Cain A.J., Harrison G.A. 1958. An analysis of the taxonomist's judgment of affinity. Proc. Zool.
1261 Soc. Lond. 131:85–98.

1262 Carranza S., Arnold E.N., Amat F. 2004. DNA phylogeny of *Lacerta* (*Iberolacerta*) and other
1263 lacertine lizards (Reptilia: Lacertidae): did competition cause long-term mountain
1264 restriction? Syst. Biodivers. 2:57–77.

1265 Carrasco M.A. 2013. The impact of taxonomic bias when comparing past and present species
1266 diversity. Palaeogeogr. Palaeoclimatol. Palaeoecol. 372:130–137.

1267 Carstens B.C., Pelletier T.A., Reid N.M., Satler J.D. 2013. How to fail at species delimitation.
1268 Mol. Ecol. 22:4369–4383.

1269 Cerio D.G., Witmer L.M. 2019. Intraspecific variation and symmetry of the inner-ear labyrinth
1270 in a population of wild turkeys: implications for paleontological reconstructions. PeerJ.
1271 7:e7355.

1272 Čerňanský A., Augé M.L. 2013. New species of the genus *Plesiolacerta* (Squamata: Lacertidae)
1273 from the upper Oligocene (MP28) of Southern Germany and a revision of the type
1274 species *Plesiolacerta lydekkeri*. Palaeontology. 56:79–94.

1275 Čerňanský A., Bolet A., Müller J., Rage J.-C., Augé M., Herrel A. 2017. A new exceptionally
1276 preserved specimen of *Dracaenosaurus* (Squamata, Lacertidae) from the Oligocene of
1277 France as revealed by micro-computed tomography. J. Vertebr. Paleontol. 0:e1384738.

1278 Čerňanský A., Klembara J., Műller J. 2016a. The new rare record of the late Oligocene lizards
1279 and amphisbaenians from Germany and its impact on our knowledge of the European
1280 terminal Palaeogene. Palaeobiodiversity Palaeoenvironments. 96:559–587.

1281 Čerňanský A., Klembara J., Smith K.T. 2016b. Fossil lizard from central Europe resolves the
1282 origin of large body size and herbivory in giant Canary Island lacertids. Zool. J. Linn.
1283 Soc. 176:861–877.

1284 Čerňanský A., Syromyatnikova E.V. 2019. The first Miocene fossils of *Lacerta* cf. *trilineata*
1285 (Squamata, Lacertidae) with a comparative study of the main cranial osteological
1286 differences in green lizards and their relatives. PLOS ONE. 14:e0216191.

1287 Christiansen N.A., Tschopp E. 2010. Exceptional stegosaur integument impressions from the
1288 Upper Jurassic Morrison Formation of Wyoming. Swiss J. Geosci. 103:163–171.

1289 Ciampaglio C.N., Kemp M., McShea D.W. 2001. Detecting changes in morphospace occupation
1290 patterns in the fossil record: characterization and analysis of measures of disparity.
1291 Paleobiology. 27:695–715.

1292 Cleary T.J., Moon B.C., Dunhill A.M., Benton M.J. 2015. The fossil record of ichthyosaurs,
1293 completeness metrics and sampling biases. Palaeontology. 58:521–536.

1294 Conrad J.L. 2008. Phylogeny and systematics of Squamata (Reptilia) based on morphology.
1295 Bull. Am. Mus. Nat. Hist. 2008:1–182.

1296 Cooney C.R., Bright J.A., Capp E.J.R., Chira A.M., Hughes E.C., Moody C.J.A., Nouri L.O.,
1297 Varley Z.K., Thomas G.H. 2017. Mega-evolutionary dynamics of the adaptive radiation
1298 of birds. Nature. 542:344–347.

1299 Cope D.A., Lacy M.G. 1992. Falsification of a single species hypothesis using the coefficient of
1300 variation: A simulation approach. Am. J. Phys. Anthropol. 89:359–378.

1301 Coyne J.A., Kreitman M. 1986. Evolutionary genetics of two sibling species, *Drosophila*
1302 *simulans* and *D. sechellia*. Evolution. 40:673–691.

1303 Cracraft J. 1987. Species concepts and the ontology of evolution. Biol. Philos. 2:329–346.

1304 Cruzado-Caballero P., Castillo Ruiz C., Bolet A., Colmenero J.R., De la Nuez J., Casillas R.,
1305 Llacer S., Bernardini F., Fortuny J. 2019. First nearly complete skull of *Gallotia auaritae*
1306 (lower-middle Pleistocene, Squamata, Gallotiinae) and a morphological phylogenetic
1307 analysis of the genus *Gallotia*. Sci. Rep. 9:16629.

1308 Dayan T., Wool D., Simberloff D. 2002. Variation and covariation of skulls and teeth: modern
1309 carnivores and the interpretation of fossil mammals. Paleobiology. 28:508–526.

1310 Delfino M., Bailon S. 2000. Early Pleistocene herpetofauna from Cava Dell'Erba and Cava Pirro
1311 (Apulia, Southern Italy). Herpetol. J. 10:95–110.

1312 Denton R.K.Jr., O'Neill R.C. 1995. *Prototeius stageri*, gen. et sp. nov., a new teiid lizard from
1313 the Upper Cretaceous Marshalltown Formation of New Jersey, with a preliminary
1314 phylogenetic revision of the Teiidae. J. Vertebr. Paleontol. 15:235–253.

1315 Doan T.M., Castoe T.A. 2003. Using morphological and molecular evidence to infer species
1316 boundaries within *Proctoporus bolivianus* Werner (Squamata: Gymnophthalmidae).
1317 Herpetologica. 59:432–449.

1318 Dunbar C.O. 1950. The species concept: Further discussion. Evolution. 4:175–176.

1319 Dzik J. 1985. Typologic versus population concepts of chronospecies: implications for ammonite
1320 biostratigraphy. Acta Palaeontol. Pol. 30.

1321 Estes R., Queiroz K. de, Gauthier J. 1988. Phylogenetic relationships within Squamata. In: Estes
1322 R., Pregill G., editors. Phylogenetic Relationships of the Lizard Families. Palo Alto,
1323 USA: Stanford University Press. p. 119–281.

1324 Fabbri M., Wiemann J., Manucci F., Briggs D.E.G. 2020. Three-dimensional soft tissue
1325 preservation revealed in the skin of a non-avian dinosaur. Palaeontology. 63:185–193.

1326 Foote M. 1990. Nearest-neighbor analysis of trilobite morphospace. Syst. Biol. 39:371–382.

1327 Foote M. 1992a. Paleozoic record of morphological diversity in blastozoan echinoderms. Proc.
1328 Natl. Acad. Sci. 89:7325–7329.

1329 Foote M. 1992b. Rarefaction analysis of morphological and taxonomic diversity. Paleobiology.
1330 18:1–16.

1331 Foote M. 1993. Discordance and concordance between morphological and taxonomic diversity.
1332 Paleobiology. 19:185–204.

1333 Foth C., Bona P., Desojo J.B. 2015. Intraspecific variation in the skull morphology of the black
1334 caiman *Melanosuchus niger* (Alligatoridae, Caimaninae). Acta Zool. 96:1–13.

1335 Gauthier J.A., Kearney M., Maisano J.A., Rieppel O., Behlke A.D.B. 2012. Assembling the
1336 squamate tree of life: perspectives from the phenotype and the fossil record. Bull.
1337 Peabody Mus. Nat. Hist. 53:3–308.

1338 Gerber S. 2019. Use and misuse of discrete character data for morphospace and disparity
1339 analyses. Palaeontology. 62:305–319.

1340 Gingerich P.D. 1981. Variation, sexual dimorphism, and social structure in the Early Eocene
1341 horse *Hyracotherium* (Mammalia, Perissodactyla). Paleobiology. 7:443–455.

1342 Gingerich P.D. 1985. Species in the fossil record: concepts, trends, and transitions. Paleobiology.
1343 11:27–41.

1344 Gray J.A., Sherratt E., Hutchinson M.N., Jones M.E.H. 2019. Evolution of cranial shape in a
1345 continental-scale evolutionary radiation of Australian lizards. Evolution. 73:2216–2229.

1346 Haffer J. 1986. Superspecies and species limits in vertebrates. J. Zool. Syst. Evol. Res. 24:169–
1347 190.

1348 Harrison R.G. 1998. Linking evolutionary pattern and process. Endless Forms.:19–31.

1349 Harrison R.G., Larson E.L. 2014. Hybridization, introgression, and the nature of species
1350 boundaries. J. Hered. 105:795–809.

1351 Hausdorf B. 2011. Progress toward a general species concept. Evolution. 65:923–931.

1352 Hausdorf B., Hennig C. 2010. Species delimitation using dominant and codominant multilocus
1353 markers. Syst. Biol. 59:491–503.

1354 Hetherington A.J., Sherratt E., Ruta M., Wilkinson M., Deline B., Donoghue P.C.J. 2015. Do
1355 cladistic and morphometric data capture common patterns of morphological disparity?
1356 Palaeontology. 58:393–399.

1357 Hey J., Waples R.S., Arnold M.L., Butlin R.K., Harrison R.G. 2003. Understanding and
1358 confronting species uncertainty in biology and conservation. Trends Ecol. Evol. 18:597–
1359 603.

1360 Highton R. 1990. Taxonomic treatment of genetically differentiated populations. Herpetologica.
1361 46:114–121.

1362 Hoffstetter R. 1942. Sur les restes de Sauria du nummulitique Européen rapportés à la famille
1363 des Iguanidae. Bull. Muséum Natl. Hist. Nat. 14:233–240.

1364 Hoffstetter R. 1944. Sur les Scincidae fossiles. I. Formes européennes et nord-américaines. Bull.
1365 Muséum Natl. Hist. Nat. 16:547–553.

1366 Hull D.L. 1965. The effect of essentialism on taxonomy--two thousand years of stasis (II). Br. J.
1367 Philos. Sci. 16:1–18.

1368 Joyce W.G., Micklich N., Schaal S.F.K., Scheyer T.M. 2012. Caught in the act: the first record
1369 of copulating fossil vertebrates. Biol. Lett. 8:846–848.

1370 Kapli P., Poulakakis N., Lymberakis P., Mylonas M. 2011. A re-analysis of the molecular
1371 phylogeny of Lacertidae with currently available data. Basic Appl. Herpetol. 25:97–104.

1372 Kay R.F. 1982. Sexual dimorphism in Ramapithecinae. Proc. Natl. Acad. Sci. 79:209–212.

1373 Kelley J. 1986. Species recognition and sexual dimorphism in *Proconsul* and *Rangwapithecus*. J.
1374 Hum. Evol. 15:461–495.

1375 Kimura Y., Flynn L.J., Jacobs L.L. 2016. A palaeontological case study for species delimitation
1376 in diverging fossil lineages. Hist. Biol. 28:189–198.

1377 Kirchhof S., Jacobsen N., Richter K. 2012. Intraspecific variation of morphology, colouration,
1378 pholidosis, and tail loss rate in a relic lacertid of South Africa, *Australolacerta rupicola*
1379 (Sauria: Lacertidae). Salamandra. 48:125–132.

1380 Knowles L.L., Carstens B.C. 2007. Delimiting species without monophyletic gene trees. Syst.
1381 Biol. 56:887–895.

1382 Kornilios P., Thanou E., Lymberakis P., Ilgaz Ç., Kumlutaş Y., Leaché A. 2020. A
1383 phylogenomic resolution for the taxonomy of Aegean green lizards. Zool. Scr. 49:14–27.

1384 Lee M.S.Y. 1998. Convergent evolution and character correlation in burrowing reptiles: towards
1385 a resolution of squamate relationships. Biol. J. Linn. Soc. 65:369–453.

1386 Lloyd G.T. 2016. Estimating morphological diversity and tempo with discrete character-taxon
1387 matrices: implementation, challenges, progress, and future directions. Biol. J. Linn. Soc.
1388 118:131–151.

1389 Lordkipanidze D., León M.S.P. de, Margvelashvili A., Rak Y., Rightmire G.P., Vekua A.,
1390 Zollikofer C.P.E. 2013. A complete skull from Dmanisi, Georgia, and the evolutionary
1391 biology of early *Homo*. Science. 342:326–331.

1392 Lupia R. 1999. Discordant morphological disparity and taxonomic diversity during the
1393 Cretaceous angiosperm radiation: North American pollen record. Paleobiology. 25:1–28.

1394 Maddison W.P., Maddison D.R. 2017. Mesquite: a modular system for evolutionary analysis. .

1395 Mallon J.C. 2017. Recognizing sexual dimorphism in the fossil record: lessons from nonavian
1396 dinosaurs. Paleobiology. 43:495–507.

1397 Mannion P.D., Benson R.B.J., Carrano M.T., Tennant J.P., Judd J., Butler R.J. 2015. Climate
1398 constrains the evolutionary history and biodiversity of crocodylians. Nat. Commun.
1399 6:ncomms9438.

1400 Mannion P.D., Upchurch P. 2010. Completeness metrics and the quality of the sauropodomorph
1401 fossil record through geological and historical time. Paleobiology. 36:283–302.

1402 Mannion P.D., Upchurch P., Barnes R.N., Mateus O. 2013. Osteology of the Late Jurassic
1403 Portuguese sauropod dinosaur *Lusotitan atalaiensis* (Macronaria) and the evolutionary
1404 history of basal titanosauriforms. Zool. J. Linn. Soc. 168:98–206.

1405 Marshall J.C., Arévalo E., Benavides E., Sites J.L., Sites J.W. 2006. Delimiting species:
1406 Comparing methods for Mendelian characters using lizards of the *Sceloporus grammicus*
1407 (Squamata: Phrynosomatidae) complex. Evolution. 60:1050–1065.

1408 Mateo J.A. 1988. Estudio sistemático y zoogeográfico de los lagartos ocelados. PhD thesis.

1409 Matthew W.D. 1930. Range and limitations of species as seen in fossil mammal faunas. GSA
1410 Bull. 41:271–274.

1411 Matzke N.J. 2013. Probabilistic historical biogeography: new models for founder-event
1412 speciation, imperfect detection, and fossils allow improved accuracy and model-testing.
1413 Front. Biogeogr. 5.

1414 Matzke N.J. 2014. Model Selection in Historical Biogeography Reveals that Founder-Event
1415 Speciation Is a Crucial Process in Island Clades. Syst. Biol. 63:951–970.

1416 Mayr E. 1942. Systematics and the Origin of the Species. New York, USA: Columbia University
1417 Press.

1418 Mendes J., Harris D.J., Carranza S., Salvi D. 2016. Evaluating the phylogenetic signal limit from
1419 mitogenomes, slow evolving nuclear genes, and the concatenation approach. New

1420 insights into the Lacertini radiation using fast evolving nuclear genes and species trees.
1421 Mol. Phylogenet. Evol. 100:254–267.

1422 Miller III W. 2016. The species problem: concepts, conflicts, and patterns preserved in the fossil
1423 record. In: Allmon W.D., Yacobucci M.M., editors. Species and speciation in the fossil
1424 record. Chicago, USA: University of Chicago Press. p. 28–58.

1425 Montgelard C., Behrooz R., Arnal V., Asadi A., Geniez P., Kaboli M. 2020. Diversification and
1426 cryptic diversity of *Ophisops elegans* (Sauria, Lacertidae). J. Zool. Syst. Evol. Res.
1427 58:1276–1289.

1428 Müller J. 2004. Cranial osteology from *Dracaenosaurus croizeti*, a lacertid lizard from the
1429 Oligocene of France (Reptilia, Squamata). Neues Jahrb. Für Geol. Paläontol. - Abh.:253–
1430 266.

1431 O'Donovan C., Meade A., Venditti C. 2018. Dinosaurs reveal the geographical signature of an
1432 evolutionary radiation. Nat. Ecol. Evol. 2:452–458.

1433 O'Leary M.A., Kaufman S.G. 2012. MorphoBank 3.0: Web application for morphological
1434 phylogenetics and taxonomy. http://www.morphobank.org.

1435 Paradis E., Schliep K. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary
1436 analyses in R. Bioinformatics. 35:526–528.

1437 Poropat S.F., Mannion P.D., Upchurch P., Hocknull S.A., Kear B.P., Kundrát M., Tischler T.R.,
1438 Sloan T., Sinapius G.H.K., Elliott J.A., Elliott D.A. 2016. New Australian sauropods shed
1439 light on Cretaceous dinosaur palaeobiogeography. Sci. Rep. 6:34467.

1440 Prendini L. 2001. Species or supraspecific taxa as terminals in cladistic analysis? Groundplans
1441 versus exemplars revisited. Syst. Biol. 50:290–300.

1442 Pyron R.A., Burbrink F.T., Wiens J.J. 2013. A phylogeny and revised classification of Squamata,
1443 including 4161 species of lizards and snakes. BMC Evol. Biol. 13:93.

1444 Quadros A.B., Chafrat P., Zaher H. 2018. A new teiid lizard of the genus *Callopistes*
1445 Gravenhorst, 1838 (Squamata, Teiidae), from the Lower Miocene of Argentina. J.
1446 Vertebr. Paleontol. 38:(1)-(18).

1447 Queiroz K. de. 1987. Phylogenetic systematics of iguanine lizards. Univ. Calif. Publ. Zool.
1448 118:1–223.

1449 Queiroz K. de. 1998. The general lineage concept of species, species criteria, and the process of
1450 speciation. In: Howard D.J., Berlocher S.H., editors. Endless Forms: Species and
1451 Speciation. Oxford, UK: Oxford University Press. p. 57–75.

1452 Queiroz K. de. 2005. Different species problems and their resolution. BioEssays. 27:1263–1269.

1453 Queiroz K. de. 2007. Species concepts and species delimitation. Syst. Biol. 56:879–886.

1454 Rauhut O.W.M., Foth C., Tischlinger H., Norell M.A. 2012. Exceptionally preserved juvenile
1455 megalosauroid theropod dinosaur with filamentous integument from the Late Jurassic of
1456 Germany. Proc. Natl. Acad. Sci. 109:11746–11751.

1457 Raup D.M. 1972. Taxonomic diversity during the Phanerozoic. Science. 177:1065–1071.

1458 R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for
1459 Statistical Computing, Vienna, Austria. URL https://www.R-project.org/

1460 Reichenbacher B., Sienknecht U., Küchenhoff H., Fenske N. 2007. Combined otolith
1461 morphology and morphometry for assessing taxonomy and diversity in fossil and extant
1462 killifish (*Aphanius*, †*Prolebias*). J. Morphol. 268:898–915.

1463 Rieppel O. 2009. Species as a process. Acta Biotheor. 57:33–49.

1464 Rightmire G.P., Margvelashvili A., Lordkipanidze D. 2019. Variation among the Dmanisi
1465 hominins: Multiple taxa or one species? Am. J. Phys. Anthropol. 168:481–495.

1466 Roth V.L. 1992. Quantitative variation in elephant dentitions: Implications for the delimitation of
1467 fossil species. Paleobiology. 18:184–202.

1468 Sahney S., Benton M.J. 2017. The impact of the Pull of the Recent on the fossil record of
1469 tetrapods. Evol. Ecol. Res. 18:7–23.

1470 Saunders I.W., Tavaré S., Watterson G.A. 1984. On the genealogy of nested subsamples from a
1471 haploid population. Adv. Appl. Probab. 16:471–491.

1472 Scanlon J.D. 1996. Studies in the palaeontology and systematics of Australian snakes. .

1473 Schwartz J.H., Tattersall I., Chi Z. 2014. Comment on "A complete skull from Dmanisi,
1474 Georgia, and the evolutionary biology of early *Homo*." Science. 344:360–360.

1475 Simpson G.G. 1937. Patterns of phyletic evolution. GSA Bull. 48:303–314.

1476 Simpson G.G. 1940. Types in modern taxonomy. Am. J. Sci. 238:413–431.

1477 Simpson G.G. 1941. The species of *Hoplophoneus*. Am. Mus. Novit. 1123.

1478 Simpson G.G. 1951. The species concept. Evolution. 5:285–298.

1479 Sites J.W., Marshall J.C. 2003. Delimiting species: a Renaissance issue in systematic biology.
1480 Trends Ecol. Evol. 18:462–470.

1481 Sites J.W., Marshall J.C. 2004. Operational criteria for delimiting species. Annu. Rev. Ecol.
1482 Evol. Syst. 35:199–227.

1483 Smith A.J., Rosario M.V., Eiting T.P., Dumont E.R. 2014. Joined at the hip: Linked characters
1484 and the problem of missing data in studies of disparity. Evolution. 68:2386–2400.

1485 Sneath P.H., Sokal R.R. 1973. Numerical taxonomy. The principles and practice of numerical
1486 classification. San Francisco, USA: W. H. Freeman and Company.

1487 Starrfelt J., Liow L.H. 2016. How many dinosaur species were there? Fossil bias and true
1488 richness estimated using a Poisson sampling model. Phil Trans R Soc B. 371:20150219.

1489 Tayhan Y., Yakin B.Y., Tok C.V. 2016. Shape variability of the head of *Ophisops elegans*
1490 Ménétriés, 1832 (Reptilia: Lacertidae) from Konya, Turkey. Ital. J. Zool. 83:208–212.

1491 Tennant J.P., Mannion P.D., Upchurch P. 2016. Sea level regulated tetrapod diversity dynamics
1492 through the Jurassic/Cretaceous interval. Nat. Commun. 7:12737.

1493 Thorpe J.P. 1982. The molecular clock hypothesis: Biochemical evolution, genetic
1494 differentiation and systematics. Annu. Rev. Ecol. Syst. 13:139–168.

1495 Trueman A.E. 1924. The species-concept in palaeontology. Geol. Mag. 61:355–360.

1496 Tschopp E., Mateus O., Benson R.B.J. 2015. A specimen-level phylogenetic analysis and
1497 taxonomic revision of Diplodocidae (Dinosauria, Sauropoda). PeerJ. 3:e857.

1498 Tschopp E., Tschopp F.A., Mateus O. 2018a. Overlap Indices: Tools to quantify the amount of
1499 anatomical overlap among groups of incomplete terminal taxa in phylogenetic analyses.
1500 Acta Zool. 99:169–176.

1501 Tschopp E., Upchurch P. 2019. The challenges and potential utility of phenotypic specimen-level
1502 phylogeny based on maximum parsimony. Earth Environ. Sci. Trans. R. Soc. Edinb.
1503 109:301–323.

1504 Tschopp E., Villa A., Camaiti M., Ferro L., Tuveri C., Rook L., Arca M., Delfino M. 2018b. The
1505 first fossils of *Timon* (Squamata: Lacertinae) from Sardinia (Italy) and potential causes
1506 for its local extinction in the Pleistocene. Zool. J. Linn. Soc. 184:825–856.

1507 Van Valen L. 1976. Ecological Species, Multispecies, and Oaks. Taxon. 25:233–239.

1508 Villa A., Daza J.D., Bauer A.M., Delfino M. 2018. Comparative cranial osteology of European
1509 gekkotans (Reptilia, Squamata). Zool. J. Linn. Soc. 184:857–895.

1510 Villa A., Delfino M. 2019. A comparative atlas of the skull osteology of European lizards
1511 (Reptilia: Squamata). Zool. J. Linn. Soc. 187:829–928.

1512 Villa A., Kirchner M., Alba D.M., Bernardini F., Bolet A., Luján À.H., Fortuny J., Hipsley C.A.,
1513 Müller J., Sindaco R., Tuniz C., Delfino M. 2019. Comparative cranial osteology of
1514 *Blanus*(Squamata: Amphisbaenia). Zool. J. Linn. Soc. 185:693–716.

1515 Villa A., Tschopp E., Georgalis G.L., Delfino M. 2017. Osteology, fossil record and
1516 palaeodiversity of the European lizards. Amphib.-Reptil. 38:79–88.

1517 Watanabe A. 2016. The impact of poor sampling of polymorphism on cladistic analysis.
1518 Cladistics. 32:317–334.

1519 Watanabe A., Fabre A.-C., Felice R.N., Maisano J.A., Müller J., Herrel A., Goswami A. 2019.
1520 Ecomorphological diversification in squamates from conserved pattern of cranial
1521 integration. Proc. Natl. Acad. Sci. 116:14688–14697.

1522 Wencker L.C.M, Tschopp E., Villa A., Augé M.L., Delfino M. 2021. Phylogenetic value of jaw
1523 elements of lacertid lizards (Squamata: Lacertoidea): A case study with Oligocene
1524 material from France. Cladistics. Early View.

1525 Wiens J.J. 1995. Polymorphic characters in phylogenetic systematics. Syst. Biol. 44:482–500.

1526 Wiens J.J. 1998. Testing phylogenetic methods with tree congruence: phylogenetic analysis of
1527 polymorphic morphological characters in phrynosomatid lizards. Syst. Biol. 47:427–444.

1528 Wiens J.J. 2000. Coding morphological variation within species and higher taxa for phylogenetic
1529 analysis. In: Wiens J.J., editor. Phylogenetic analysis of morphological data. Washington,
1530 DC, USA: Smithsonian Institution Press. p. 115–145.

1531 Wiens J.J. 2007. Species delimitation: New approaches for discovering diversity. Syst. Biol.
1532 56:875–878.

1533 Wiens J.J., Penkrot T.A. 2002. Delimiting species using DNA and morphological variation and
1534 discordant species limits in spiny lizards (*Sceloporus*). Syst. Biol. 51:69–91.

1535 Wiens J.J., Servedio M.R. 2000. Species delimitation in systematics: inferring diagnostic
1536 differences between species. Proc. R. Soc. Lond. B Biol. Sci. 267:631–636.

1537 Wiley E.O. 1978. The evolutionary species concept reconsidered. Syst. Biol. 27:17–26.

1538 Wilkinson M. 1997. Characters, congruence and quality: A study of neuroanatomical and
1539 traditional data in caecilian phylogeny. Biol. Rev. 72:423–470.

1540 Wood B.A., Li Y., Willoughby C. 1991. Intraspecific variation and sexual dimorphism in cranial
1541 and dental variables among higher primates and their bearing on the hominid fossil
1542 record. J. Anat. 174:185–205.

1543 Wood H.E. 2nd. 1931. Lower Oligocene rhinoceroses of the genus *Trigonias*. J. Mammal.
1544 12:414–428.

1545 Xu X., Upchurch P., Mannion P.D., Barrett P.M., Regalado-Fernandez O.R., Mo J., Ma J., Liu
1546 H. 2018. A new Middle Jurassic diplodocoid suggests an earlier dispersal and
1547 diversification of sauropod dinosaurs. Nat. Commun. 9:2700.

1548 Zachos F.E. 2016. Tree thinking and species delimitation: Guidelines for taxonomy and
1549 phylogenetic terminology. Mamm. Biol. - Z. Für Säugetierkd. 81:185–188.

1550 Zachos F.E. 2018. (New) Species concepts, species delimitation and the inherent limitations of
1551 taxonomy. J. Genet. 97:811–815.

1552 Zarapkin S.R. 1939. Das Divergenzprinzip in der Bestimmung kleiner systematischer
1553 Kategorien. Verhandl VII Intern Kong Entomol. 1:494–518.

1554 Zheng X., O'Connor J.K., Wang X., Pan Y., Wang Y., Wang M., Zhou Z. 2017. Exceptional
1555 preservation of soft tissue in a new specimen of *Eoconfuciusornis* and its biological
1556 implications. Natl. Sci. Rev. 4:441–452.

1557 Zollikofer C.P.E., León M.S.P. de, Margvelashvili A., Rightmire G.P., Lordkipanidze D. 2014.
1558 Response to comment on "A complete skull from Dmanisi, Georgia, and the evolutionary
1559 biology of early *Homo*." Science. 344:360–360.

1560 CAPTIONS

1561 Figures and table

1562 Figure 1. Intergeneric (between two specimens of two different genera), intrageneric
1563 (between two specimens of a single genus but different species), and interspecific dissimilarity
1564 for extant lacertid taxa (left, middle, and right columns, respectively). The horizontal black line
1565 in the boxplots represents the median. NS indicates statistically non-significant differences.
1566 Increasing number of stars refers to decreasing significance cutoff (***=0.001, **=0.01,
1567 *=0.05). Generally, intraspecific dissimilarity is significantly lower than intrageneric
1568 dissimilarity, which is significantly lower than intergeneric dissimilarity. The exceptions are
1569 species with low sample size (1 specimen per species of *Gallotia*; 2 specimens of *Iberolacerta*
1570 *monticola*; 4 of *Lacerta media*; 3 of *Podarcis siculus*).

1571 Figure 2: Intraspecific dissimilarities for all extant (a) and extinct (b) lacertid species in
1572 our dataset. Horizontal black lines in the box plots represent the median. Dark boxes represent
1573 "outlier taxa" that were statistically distinguished from more than two other taxa in the dataset.
1574 (b) Extinct lacertid species are compared to overall mean weighted pairwise dissimilarity of
1575 extant species, which is 0.2076 ± 0.0579 character state differences per character scored. Extinct
1576 species have much more variable intraspecific dissimilarity than extant species.

1577 Figure 3: Simulation of missing data in extant species, following patterns observed in
1578 extinct species. Intraspecific, weighted pairwise dissimilarity scores (y-axis) are given for the
1579 whole dataset, the simulated dataset with intermediate values of missing data, the simulated
1580 dataset with the same characters missing from the comparison as in the extinct partner species,
1581 and the extinct partner species. The extinct partner species are *Plesiolacerta lydekkeri* (for
1582 *Lacerta agilis*), *Dracaenosaurus croizeti* (for *L. bilineata*), *Pseudeumeces cadurcensis* (for *L.*

1583 *trilineata*), *Mediolacerta roceki* (for *Podarcis muralis*), "*L.*" *filholi* (for *Psammodromus algirus*),
1584 "*L.*" *siculimelitensis* (for *Timon lepidus*). NS indicates non-significant differences. Increasing
1585 number of stars refers to decreasing significance cutoff (***=0.001, **=0.01, *=0.05). The black
1586 line in the box plots represents the median, diamonds represent the mean.

1587 Figure 4. Principal Coordinate Analysis based on dissimilarity scores highlighting the
1588 different genera (a), and species within *Lacerta* (b), *Timon* (c), and *Podarcis* (d). Genera can be
1589 more easily distinguished in this way than species.

1590 Figure 5: Correlation of average completeness score, AOI, and COI within a species (y-
1591 axis) and number of specimens per extant (squares) and extinct (circles) lacertid species (x-axis).
1592 Trendlines are indicated with solid lines for extant and dashed lines for extinct species
1593 (completeness, long dashes; AOI, intermediate length of dashes; COI, short dashes). AOI seems
1594 most correlated with sample size in extant species, but extinct species show different patterns.

1595 Figure 6: Distribution of missing characters from the pairwise comparisons relative to
1596 percent of maximum dissimilarity observed in extant and extinct lacertid species. There does not
1597 seem to be a general trend of higher dissimilarity or ranges of dissimilarity with more missing
1598 characters.

1599 Figure 7: Observed dissimilarity values relative to sample size subsampled in the four
1600 best-represented lacertid species in our dataset. Variability in the average values of mean
1601 pairwise dissimilarity (triangles) does not overlap with observed maximum (dots) and minimum
1602 values (squares) once sampling includes seven or more specimens.

1603 Table 1: Completeness (C), All Characters Overlap Index (AOI), and Comparable
1604 Characters Overlap Index (COI) within extant and extinct lacertid species in the complete dataset
1605 and partitions.

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)

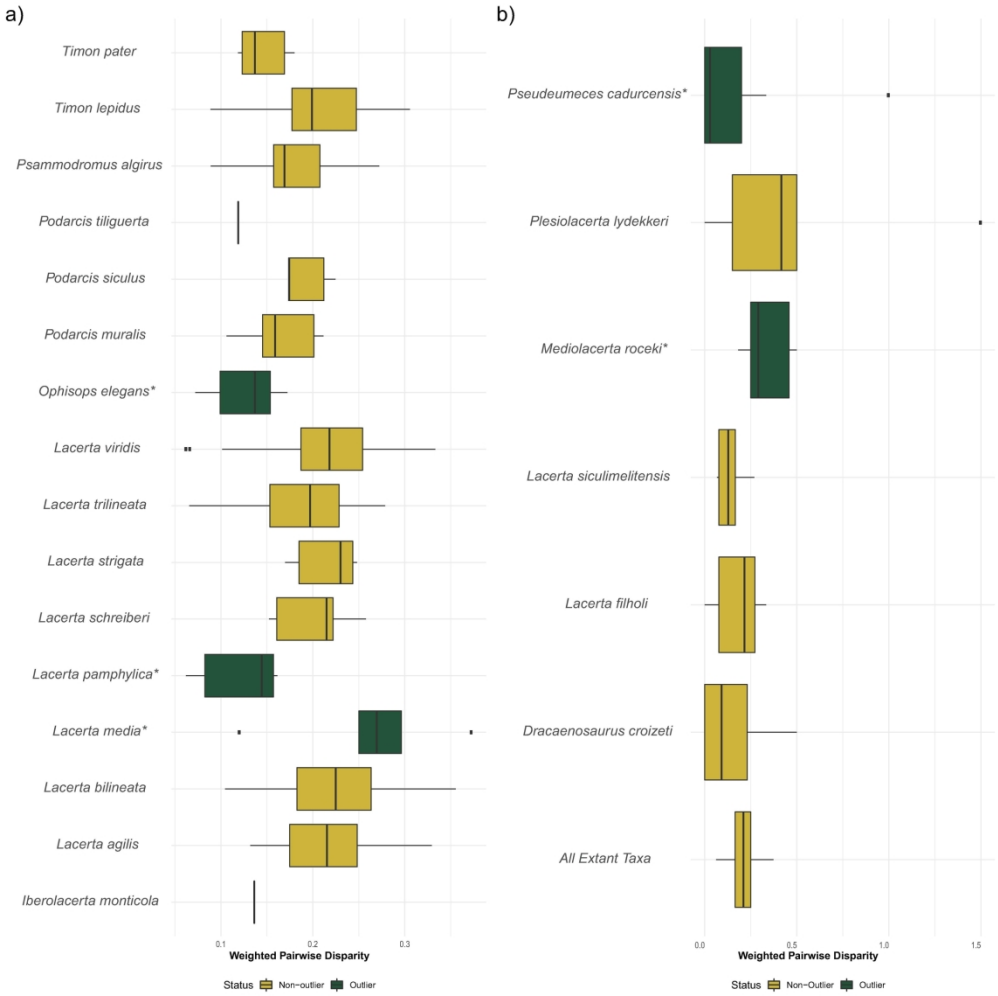

Figure 2: Intraspecific dissimilarities for all extant (a) and extinct (b) lacertid species in our dataset. Horizontal black lines in the box plots represent the median. Dark boxes represent "outlier taxa" that were statistically distinguished from more than two other taxa in the dataset. (b) Extinct 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)

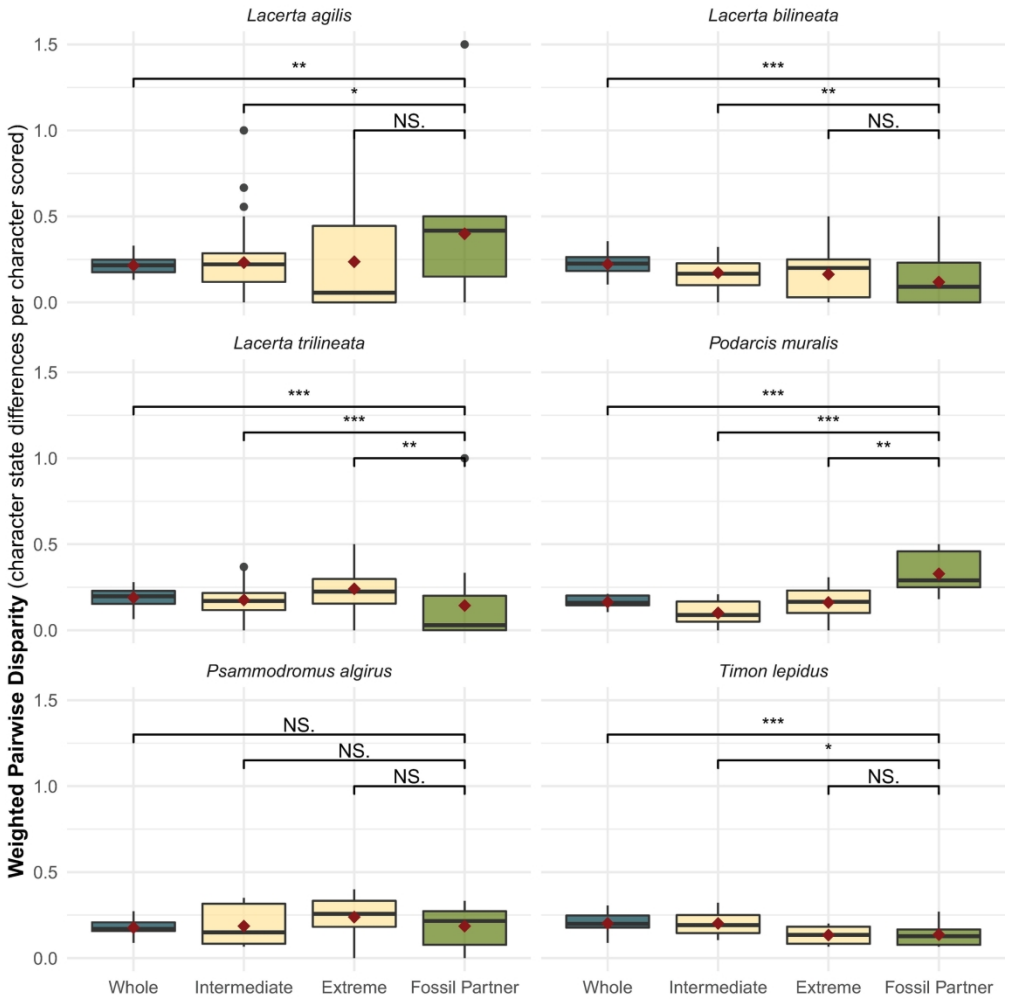

Figure 3: Simulation of missing data in extant species, following patterns observed in extinct species. Intraspecific, weighted pairwise dissimilarity scores (y-axis) are given for the whole dataset, the simulated dataset with intermediate values of missing data, the simulated dataset with the same characters 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)

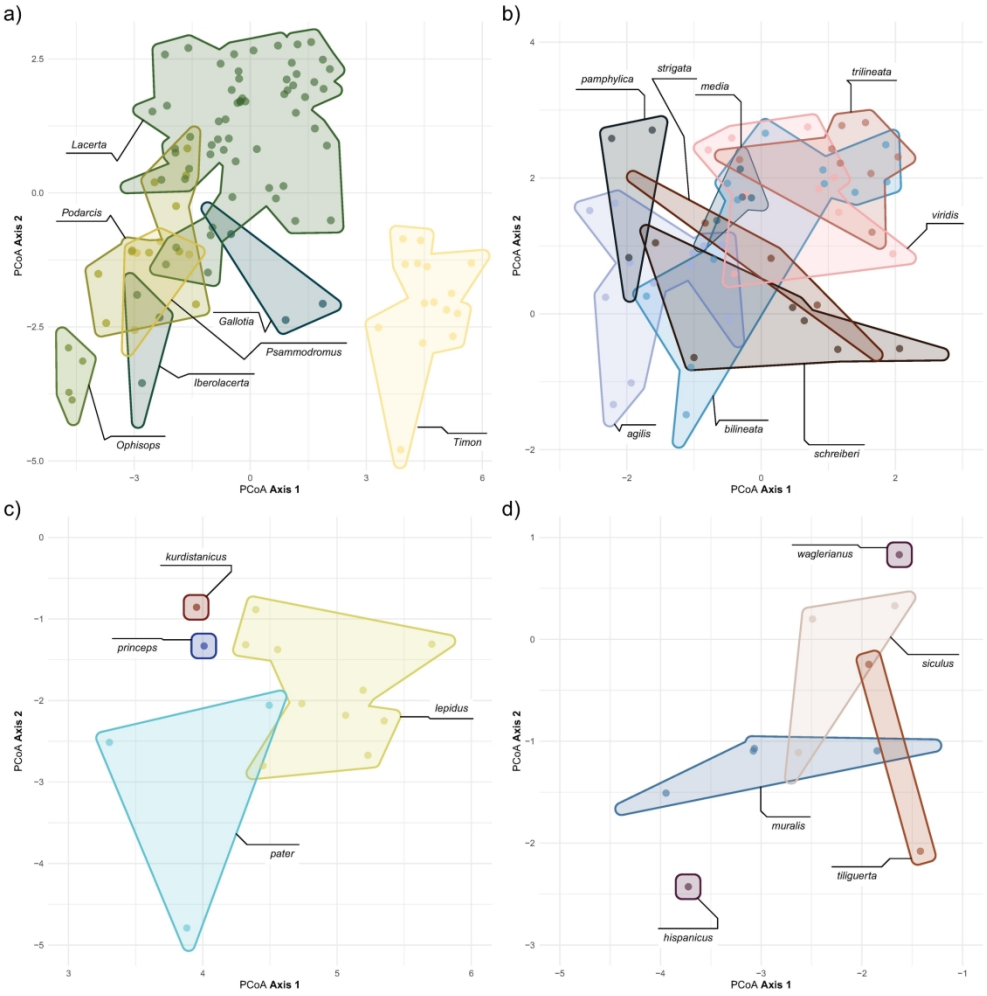

Figure 4. Principal Coordinate Analysis based on dissimilarity scores highlighting the different genera (a), and species within *Lacerta* (b), *Timon* (c), and *Podarcis* (d). Genera can be more easily distinguished in this way than species.

203x203mm (300 x 300 DPI)

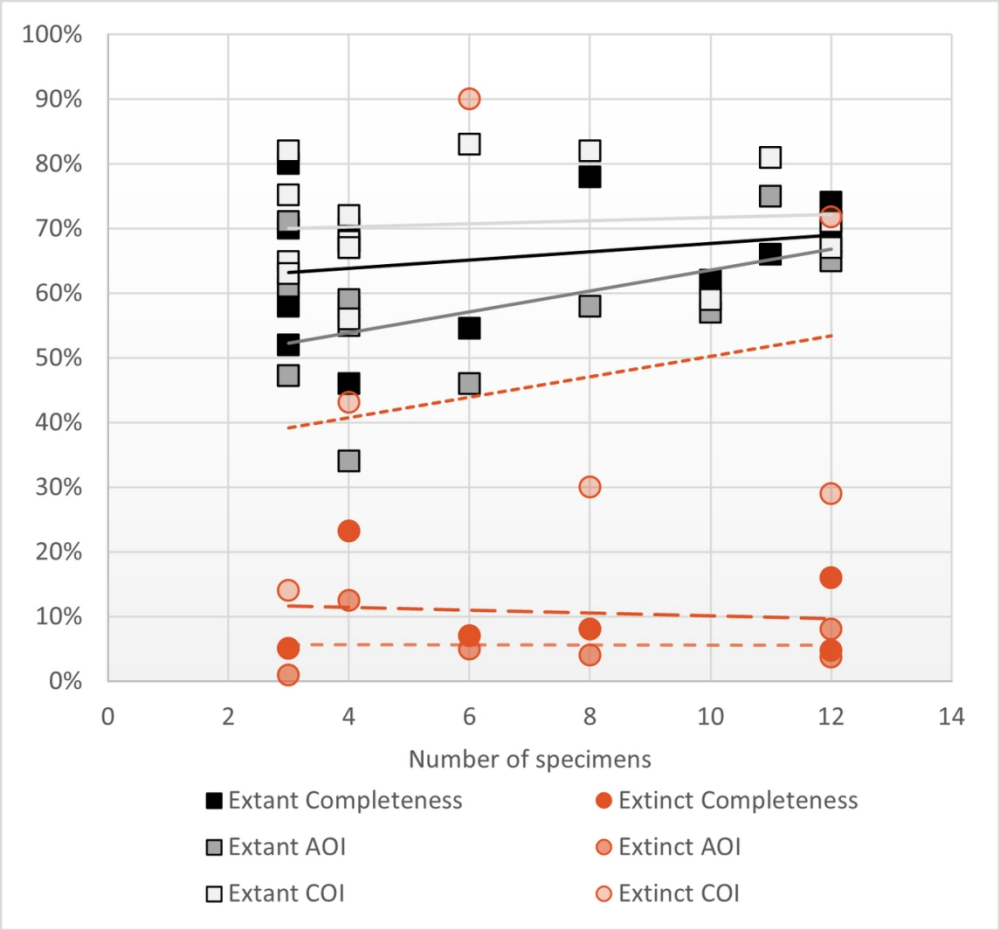

Figure 5: Correlation of average completeness score, AOI, and COI within a species (y-axis) and number of specimens per extant (squares) and extinct (circles) lacertid species (x-axis). Trendlines are indicated with solid lines for extant and dashed lines for extinct species (completeness, long dashes; 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)

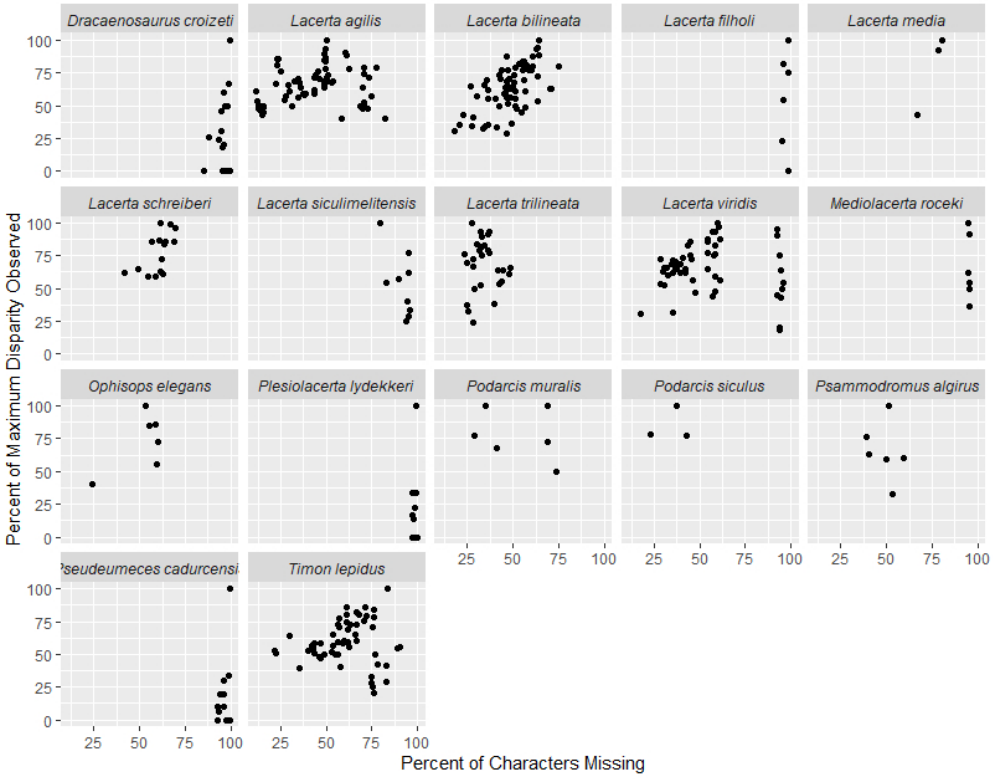

Figure 6: Distribution of missing characters from the pairwise comparisons relative to percent of maximum dissimilarity observed in extant and extinct lacertid species. There does not seem to be a general trend of higher dissimilarity or ranges of dissimilarity with more missing characters.

538x423mm (38 x 38 DPI)

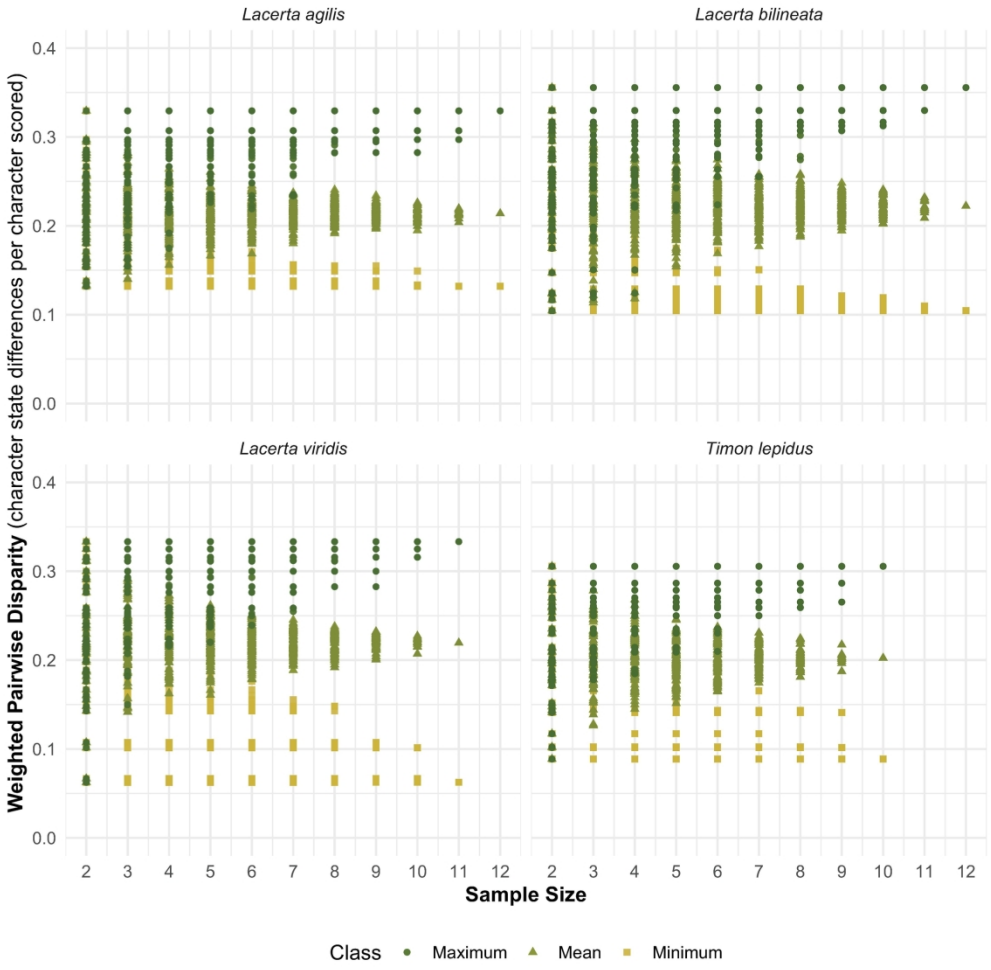

Figure 7: Observed dissimilarity values relative to sample size subsampled in the four best-represented lacertid species in our dataset. Variability in the average values of mean pairwise dissimilarity (triangles) does not overlap with observed maximum (dots) and minimum values (squares) once sampling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